

SCUOLA di DOTTORATO

Dottorato in Diritto ed Economia

Direttore della Scuola di Dottorato
prof. Felice Arena

Collegio dei docenti
Dottorato di Ricerca in Diritto ed Economia
XXXVIII ciclo

Giuseppe Tropea (coordinatore)
Stefania Romeo (Coordinatrice Vicaria)
Alessio Rauti
Federica Maria Tescione
Amalia Chiara Di Landro
Francesco Manganaro
Andrea Filocamo
Giovanni D'Amico
Angelo Viglianisi Ferraro
Graziella Bonanno
Antonino Mazza La Boccetta
Marco Schirripa
Antonio Bruno Pansera
Marina Mancini
Antonio Spadaro
Massimiliano Ferrara
Arturo Capone
Massimo Finocchiaro Castro
Attilio Gorassini
Roberto Siclari
Carmela Salazar
Sabina De Rosis
Claudio Panzera
Tiziana Rumi
Daniele Cananzi
Vincenza Caracciolo
Domenico Nicolò
Ettore Michele Squillaci
Faustino De Gregorio



SCUOLA di DOTTORATO

SCUOLA DI DOTTORATO
Università Mediterranea di Reggio Calabria

DIPARTIMENTO
DIGIES

DOTTORATO DI RICERCA
DIRITTO ED ECONOMIA

S.S.D.
IUS/17
XXXVIII
CICLO

Intelligenza artificiale e diritto penale: nuove soluzioni per nuovi problemi o vecchie soluzioni per vecchi problemi?

Dottorando

Andrea Battaglia

[redacted]

Tutor prof. F. Squillaci

[redacted]

Coordinatore del Dottorato prof. G. Tropea

Firmato digitalmente da Giuseppe Tropea Data: 07.08.2026 22:01:00 CEST
Organizzazione: UNIVERSITAMEDITERRANEA/00163260805

Abstract

This research stems from the observation that artificial intelligence systems are becoming increasingly prevalent in key areas of social life and, consequently, from the need to assess their impact on fundamental aspects of criminal law. In particular, the study examines whether traditional models of attributing liability are still adequate for addressing phenomena characterised by algorithmic decision-making autonomy, or whether a systematic rethinking of these models is required.

The study focuses on the central issue of the causal link, which is called into question by the opacity and unpredictability inherent in machine learning systems (so-called 'black boxes'). In this context, it was deemed appropriate to emphasise the criterion of logical probability as a tool for assessment in situations of uncertainty, whilst proposing a reinterpretation of causality in 'psychic' terms, suitable for describing the relationship between human input and the system's autonomous output. From this perspective, the action of the human agent is not understood as a deterministic cause of the event, but rather as a factor guiding a decision-making process which, whilst not fully predictable, remains anchored to the initial conditions set by the human agent.

As regards culpability, the research highlights the risk of drifting towards forms of liability detached from the facts, suggesting instead a model centred on the management of technological risk. In this light, Regulation (EU) 2024/1689 (AI Act) is considered not only as an instrument of administrative regulation, but also as a possible source of precautionary rules relevant to the assessment of culpability. Of particular importance in this context are the concepts of culpa in istruendo and culpa in vigilando, as well as the need for effective and meaningful human oversight of intelligent systems.

The research also rules out the viability of theories proposing the recognition of autonomous legal personality for artificial intelligence systems, considering it necessary to uphold the centrality of the principle of humanity in punishment. In the business context, it has been considered that the provisions of Legislative Decree No. 231/2001 represent the most appropriate instrument for the attribution of liability, emphasising the category of organisational negligence in contexts characterised by complex technological risks.

The work is structured around three main strands: a first part dedicated to the technical and functional overview of the phenomenon of artificial intelligence; a second part centred on the analysis of attribution models, with particular reference to the profiles of causality and culpability in the light of European legislation; and a third part, finally, aimed at verifying the validity of the categories examined in certain high-risk sectors, such as surgical robotics and autonomous driving. The overall objective is to contribute to the development of a model of criminal law for artificial intelligence capable of combining the effective protection of legal interests with respect for fundamental human rights.

Sommario

La presente ricerca prende le mosse dalla constatazione della crescente incidenza dei sistemi di intelligenza artificiale nei principali ambiti della vita sociale e, conseguentemente, dalla necessità di verificarne l'impatto sulle categorie fondamentali del diritto penale. In particolare, il lavoro si interroga se i modelli tradizionali di imputazione della responsabilità siano ancora idonei a fronteggiare fenomeni caratterizzati da autonomia decisionale algoritmica, ovvero se sia necessario un loro ripensamento in chiave sistematica.

L'indagine si concentra sul nodo centrale del nesso di causalità, messo in crisi dall'opacità e dall'imprevedibilità proprie dei sistemi di machine learning (c.d. *black box*). In tale contesto, si è ritenuto di valorizzare il criterio della probabilità logica quale strumento di accertamento in situazioni di incertezza, proponendo al contempo una rilettura della causalità in termini "psichici", idonea a descrivere il rapporto tra *input* umano e *output* autonomo del sistema. In questa prospettiva, l'azione dell'agente umano non viene intesa come causa deterministica dell'evento, bensì come fattore che orienta un processo decisionale che, pur non essendo pienamente prevedibile, resta ancorato alle condizioni iniziali poste dall'uomo.

Quanto al versante della colpevolezza, la ricerca evidenzia il rischio di derive verso forme di responsabilità sganciate dal fatto, suggerendo invece un modello incentrato sulla gestione del rischio tecnologico. In tale ottica, il Regolamento (UE) 2024/1689 (*AI Act*) viene considerato non soltanto quale strumento di regolazione amministrativa, ma anche come possibile fonte di regole cautelari rilevanti ai fini del giudizio di rimproverabilità. Particolare rilievo assumono, in questa prospettiva, le figure della *culpa in istruendo* e della *culpa in vigilando*, nonché l'esigenza di un controllo umano effettivo e significativo sui sistemi intelligenti.

La ricerca esclude, inoltre, la percorribilità delle teorie che prospettano il riconoscimento di una soggettività giuridica autonoma in capo ai sistemi di intelligenza artificiale, ritenendo necessario mantenere ferma la centralità del principio di umanità della pena. In ambito imprenditoriale, si è ritenuto che la disciplina di cui al d.lgs. n. 231/2001 rappresenti lo strumento più adeguato per l'imputazione della responsabilità, valorizzando la categoria della colpa di organizzazione nei contesti caratterizzati da rischi tecnologici complessi.

Il lavoro si articola lungo tre direttrici principali: una prima parte dedicata alla ricognizione tecnico-funzionale del fenomeno dell'intelligenza artificiale; una seconda incentrata sull'analisi dei modelli imputazionali, con particolare riferimento ai profili di causalità e colpevolezza alla luce della normativa europea; una terza, infine, volta a verificare la tenuta delle categorie esaminate in alcuni settori ad alto rischio, quali la robotica chirurgica e la guida autonoma. L'obiettivo complessivo è quello di contribuire alla costruzione di un modello di diritto penale dell'intelligenza artificiale capace di coniugare l'effettività della tutela dei beni giuridici con il rispetto delle garanzie fondamentali della persona.

INDICE

Introduzione

Capitolo primo

Intelligenza Artificiale: tentativi definitori, casi concreti e delimitazione del tema di ricerca

1. L'intelligenza artificiale come concetto in cerca di definizione.....10;

1.1. Cosa differenzia gli algoritmi di IA dagli algoritmi tradizionali? 13;

1.2. Brevi cenni storici in materia di IA..... 15;

1.3. Il *machine learning*, il *deep learning*, e le principali modalità di apprendimento..... 20;

2. Gli algoritmi a supporto dell'attività giudiziaria ed all'attività di indagine 25;

2.1. I sistemi di IA: autori del reato o mero strumento nelle mani dell'uomo?..... 32;

3. Delimitazione della ricerca: verso una nozione "penalisticamente orientata" di IA 34;

4. Definizioni normative (e giurisprudenziali) dei sistemi di IA 43;

4.1. Breve cronistoria sul cd. *AI Act*. Definizione europea di intelligenza artificiale 43;

4.2. Definizione di IA fornitaci dal Consiglio di Stato italiano.....73;

5. Esempi pratici di sistemi di IA imprevedibili.....75;

5.1. Sistemi in grado di cagionare un'offesa diretta ai beni giuridici.....77;

5.1.2. Gli *embodied systems*..... 77;

5.1.2.1. Il caso degli *autonomous vehicles*..... 77;

5.1.2.3. I diversi livelli di autonomia delle *self-driving cars* 79;

5.1.2.4. Le normative sulle *self-driving cars*..... 80;

5.1.2.5. *Case-studies*: incidenti con la partecipazione di *autonomous vehicles*..... 81;

5.1.2.6. Sistemi di guida autonoma degli aerei 84;

5.1.2.7. Sistemi di armi autonome letali..... 87;

5.1.2.8. *Robot* nel sistema sanitario (cd. *Healthcare robots*) 89;

6. I non-*embodied systems*..... 94;

6.1. I bots..... 97;

6.2. *Algorithmic trading systems*..... 97;

7. Sistemi che cagionano un'offesa indiretta ai beni giuridici..... 100;

7.1. Sistemi sanitari 102

Capitolo secondo

Cenni sul rapporto tra IA e responsabilità civile

1. La responsabilità civile come soluzione immediata..... 108;

1.2. La regolazione dell'IA nel contesto europeo e la responsabilità civile 114;

1.3. Occorre creare una normativa *ad hoc* ovvero è sufficiente adattare la normativa vigente?..... 117;

2. Il codice civile ed i regimi speciali di responsabilità oggettiva in capo all'utilizzatore del sistema di IA 120;

3. La nozione di prodotto e la sua adeguatezza al contesto dell'IA……131;

4. La nozione di produttore e la sua adeguatezza al contesto dell'IA……136;

5. La nozione di difetto e la sua adeguatezza al contesto dell'IA……138;

5.1. La classificazione dei difetti: difetti di fabbricazione, progettazione, informazione nei sistemi di IA..... 144;

5.1.2. Focus sui difetti di progettazione: responsabilità per colpa o assoluta?.....151;

5.1.3. Il difetto di progettazione in relazione ai sistemi di IA 153;

5.2. Verso nuovi criteri di valutazione del difetto: la proposta di revisione della *Product liability directive*.....157;

5.2.1. Le proposte della dottrina 158;

5.3. Un'alternativa: l'agente modello umano.....160;

5.4. Un'altra via: l'agente modello algoritmico 161;

5.5. *Product liability* e *product safety*: verso la preemption *doctrine*.....162;

5.5.1. *Product safety* e tecniche di armonizzazione..... 164;

5.5.2. Sistemi di IA e norme di armonizzazione: brevi cenni sull'*AI Act*..... 169;

5.5.2.1. Le pratiche di IA vietate: approfondimento……171;

5.5.3. I sistemi di IA consentiti..... 175;

5.5.3.1.I sistemi di IA ad alto rischio.....176;

5.5.4. La tesi della non difettosità del prodotto conforme agli *standard* tecnici 179;

5.6. *Product safety* e *product liability*: una separazione concettuale 182;

5.7. Il nesso di causalità ed il suo accertamento 185;

5.7.1. Brevi cenni sui profili sostanziali del nesso di causalità 186;

5.7.2. Aspetti processuali del nesso di causalità 191;

6. Responsabilità per esercizio di attività pericolose ed i sistemi di IA 194;

7. C'è bisogno di una personalità elettronica? 196;

8. Serve un'assicurazione obbligatoria? 199;

9. Responsabilità civile ed IA: spunti di riflessione ..... 202.

Capitolo terzo

Intelligenza artificiale e responsabilità penale

1. Il rapporto tra diritto penale ed IA. Linee generali ed introduttive 208;

2. La macchina intelligente come autore del reato: la tesi di Gabriel Hallevy 210;

2.1. Critiche alla teoria di Hallevy 214;

2.2. Le funzioni della pena e l'erroneo parallelismo tra la responsabilità delle persone giuridiche e la responsabilità delle macchine 218;

3. La responsabilità degli esseri umani a contatto con il sistema di IA: l'utilizzatore 221;

3.1. La responsabilità colposa dell'utilizzatore di un sistema di IA completamente autonomo 223;

3.2. La responsabilità colposa dell'utilizzatore di un sistema di IA parzialmente autonomo 224;

3.2.1. Che tipo di fatto di reato potrebbe essere imputato all'utilizzatore di un sistema di IA parzialmente autonomo? 229;

3.2.2. La responsabilità colposa ed i limiti del principio dell'*human-in-the-loop*..... 234;

3.3. La validità del principio di affidamento: limite alla rimproverabilità? 238;

3.3.1. Prospettive *de jure condendo* e limiti alla rimproverabilità dell'utilizzatore..... 241;

3.4. Le ipotesi abberranti ed i sistemi di IA 250;

4. La responsabilità del produttore per i danni provocati dal sistema di IA 251;

4.1. La rilevanza del principio di precauzione nel diritto penale..... 251;

4.2. La tipicità del reato da IA..... 255;

4.3. Il nesso di causalità ed il suo accertamento 261;

4.3.1. Incompatibilità tra causalità e probabilità del *decision making* algoritmico: è possibile individuare delle leggi scientifiche? 268;

4.4. Chi è responsabile dei danni cagionati dall'IA? Un *too many hands problem* 274;

4.4.1. L'*AI Act* e le aree di competenza nella gestione del rischio 278;

4.5. La difficile dimostrazione della colpa del produttore..... 280;

4.5.1. Problemi di colpa specifica: l'individuazione delle regole cautelari 282;

4.5.2. A proposito di rischio consentito: qual è il rapporto tra regole cautelari scritte e non scritte? 287;

4.5.2.1. L'individuazione delle aree di rischio consentito ed i limiti
dell'istituto 294;

4.5.2.2. L'*AI Act* e gli obblighi di monitoraggio post-market: superamento del rischio consentito? 297;

4.5.2.3. IA e rischio consentito: approccio quantitativo ..... o qualitativo? 301;

4.5.3. La colpa generica nell'IA 307;

4.5.3.1. Un'imprevedibilità prevedibile dell'evento lesivo è rimproverabile? 308;

4.5.3.2. L'incidenza del principio di precauzione sul giudizio di prevedibilità dell'evento 313;

4.5.4. IA e rischio consentito: applicazione di quanto fino a qui scoperto 317;

5. Soluzione forzata allo stato attuale: crisi del diritto penale d'evento e tutela anticipata 321;

5.1. Le fattispecie contravvenzionali di cui all'art. 112 cod. cons 326;

5.1.2. Nuove soluzioni: *de jure condendo* 327;

6. Facciamo un punto: IA e responsabilità penale ..... 329.

Conclusioni

Indice bibliografico

Intelligenza artificiale e diritto penale: nuove soluzioni per nuovi problemi o vecchie soluzioni per vecchi problemi?

Introduzione

L'impatto dei sistemi di intelligenza artificiale nei settori più importanti della vita sociale impone al giurista di ragionare su una sfida sistemica che pone in bilico le stesse fondamenta della responsabilità penale. Il nostro lavoro di ricerca parte dall'esigenza di verificare se le classiche categorie del diritto penale liberale siano sufficientemente idonee per assorbire l'impatto dell'autonomia algoritmica, ovvero se sia necessario un intervento di manutenzione straordinaria dei modelli di imputazione. La domanda che sta alla base dell'intera trattazione concerne, in poche parole, l'alternativa tra la creazione di nuove soluzioni per nuovi problemi ed il recupero di vecchie soluzioni per vecchi problemi. Con questo approccio, l'IA non viene studiata come entità isolata ed a sé stante, ma viene immersa in un contesto generale nonché considerata in una nozione funzionale allo studio. Una nozione, cioè, "penalisticamente orientata". Ciò che interessa per il giurista è l'opacità decisionale derivante dalla c.d. *black box* algoritmica che causa comportamenti emergenti in grado di mettere in crisi le tradizionali regole per l'imputazione causale di un evento e per la rimproverabilità soggettiva. Di fatti, a ben vedere, il nucleo centrale della ricerca è inevitabilmente costituito dal nesso eziologico. Al cospetto di sistemi di *Machine Learning* (ML) capaci di elaborare dati in maniera imprevedibile ed impredicibile dallo stesso programmatore, il tradizionale ricorso alle leggi scientifiche di copertura risulta del tutto infruttuoso. Prendendo come assodati i principi manifestati dalla sentenza Franzese del 2002, si è notato come anche in contesti

di incertezza tecnologica, il criterio della probabilità logica possa fungere da correttivo epistemologico. Nell'analisi del nesso di causalità, fondamentale è risultata la tesi del Viganò secondo cui non si deve appiattire l'accertamento del nesso causale su semplici dati statistici. E ciò vale soprattutto quando l'algoritmo altro non fa che fungere da mediatore tra l'azione umana e l'evento lesivo. In questo senso, nella ricerca si è proposta una analisi teorica del concetto di causalità psichica per spiegare il nesso tra *input* umano e *output* autonomo del sistema intelligente. L'ipotesi considerata è che l'azione umana non causi l'evento in maniera meccanica ma crei le condizioni necessarie per orientare un processo decisionale che, anche se non deterministico, resta ancorato alle premesse date dall'agente umano. Per quanto concerne la colpevolezza, la ricerca si concentra sul rischio di una deriva verso modelli di responsabilità da posizione. Nel corso della ricerca, ci siamo posti in un'ottica di valorizzazione del ruolo del Regolamento UE 2024/1689, il c.d. *AI Act*, non solo come cornice regolatoria amministrativa ma anche come fonte di regole cautelari la cui violazione può cristallizzare il giudizio di rimproverabilità. Riportando alla mente le riflessioni di Donini sulla necessità di un diritto penale che non punisca il semplice evento infausto ma che punisca una gestione negligente del rischio tecnologico, si è sostenuto che la responsabilità penale debba essere ancorata a parametri certi di diligenza. In quest'ottica, le nuove frontiere della tipicità colposa devono incentrarsi sulla *culpa in istruendo* e sulla *culpa in vigilando*. La prima relativa alla selezione dei *dataset* da fornire al sistema intelligente; mentre la seconda relativa al monitoraggio delle *performance* dei sistemi stessi. Seguendo quanto significato anche da G. Ubertis, l'agente umano è investito dell'obbligo di un controllo umano significativo, inteso come obbligo di supervisione critica volta ad evitare l'abdicazione del giudizio penale all'automatismo algoritmico. Altro importante passaggio della ricerca è quello relativo alla confutazione delle teorie che vorrebbero riconoscere una personalità giuridica alle macchine. A nostro giudizio, occorre riaffermare la centralità del principio di

umanità della pena e la funzione rieducativa di questa, anche per infondere una maggiore responsabilizzazione in capo all'agente umano. La soggettività elettronica, ci pare, è una *fictio juris* pericolosa, suscettibile di creare zone d'ombra ingiustificate per i vari centri di potere economico. Con la nostra ricerca abbiamo tentato di dimostrare che il D.Lgs, n. 231/2001 e, dunque, la responsabilità da reato degli enti, sia lo strumento più idoneo ed efficace per intercettare la responsabilità nei contesti imprenditoriali, laddove la colpa di organizzazione appare la categoria più aderente ad una realtà in cui l'evento lesivo è il frutto di carenze strutturali nei protocolli di sicurezza. Concludendo, la trattazione si snoda lungo tre direttrici centrali che fungono da guida per il lettore, conducendolo dalla teoria alla prassi. Una prima direttrice concerne la ricognizione tecnico-funzionale del fenomeno in questione, diretta ad indagare come l'opacità algoritmica sfidi il principio di tassatività proprio del diritto penale. Segue il cuore dommatico della ricerca, incentrato sull'analisi dei modelli imputazionali tra colpevolezza e causalità alla luce della normativa europea. Infine, il lavoro si incentra sull'esame dei settori speciali ad alto rischio come la robotica chirurgica e le auto a guida autonoma, al fine di verificare la tenuta delle posizioni di garanzia nel caso concreto. Lo scopo finale è quello di tracciare le coordinate di un diritto penale dell'intelligenza artificiale che sia, al contempo, efficace nella tutela dei beni giuridici e rigoroso nel rispetto delle garanzie individuali del cittadino.

CAPITOLO PRIMO

INTELLIGENZA ARTIFICIALE: TENTATIVI DEFINITORI, CASI CONCRETI E DELIMITAZIONE DEL TEMA DI RICERCA

1.L'intelligenza artificiale come concetto in cerca di definizione; 1.1. Cosa differenzia gli algoritmi di IA dagli algoritmi tradizionali?; 1.2.Brevi cenni storici in materia di IA; 1.3.Il *machine learning*, il *deep learning*, e le principali modalità di apprendimento.; 2.Gli algoritmi a supporto dell'attività giudiziaria ed all'attività di indagine; 2.1.I sistemi di IA: autori del reato o mero strumento nelle mani dell'uomo?; 3.Delimitazione della ricerca: verso una nozione "penalisticamente orientata" di IA.; 4.Definizioni normative (e giurisprudenziali) dei sistemi di IA.; 4.1.Breve cronistoria sul cd. AI Act. Definizione europea di intelligenza artificiale.; 4.2.Definizione di IA fornitaci dal Consiglio di Stato italiano.; 5. Esempi pratici di sistemi di IA imprevedibili.; 5.1. Sistemi in grado di cagionare un'offesa diretta ai beni giuridici.;5.1.2.Gli *embodied systems*.;5.1.2.1.Il caso degli autonomous vehicles.;5.1.2.3.Idiversi livelli di autonomia delle *self-driving cars*.; 5.1.2.4.Le normative sulle *self-driving cars*.; 5.1.2.5.*Case-studies*: incidenti con la partecipazione di *autonomous vehicles*.; 5.1.2.6.Sistemi di guida autonoma degli aerei.;5.1.2.7.Sistemi di armi autonome letali.;5.1.2.8.*Robot* nel sistema sanitario (cd. *Healthcare robots*).;6.I *non-embodied systems*.;6.1.I *bots*.;6.2.*Algorithmic trading systems*.;7.Sistemi che cagionano un'offesa indiretta ai beni giuridici.;7.1.Sistemi sanitari.

1. L'intelligenza artificiale come concetto in cerca di definizione.

Prima di addentrarci più dettagliatamente nelle questioni giuridiche e, in particolare, di tipo penalistico, occorre procedere ad una ricognizione del fenomeno. Occorre, in altre parole, comprendere e definire il concetto di intelligenza artificiale, facendo riferimento tanto alla scienza giuridica, quanto e soprattutto alle scienze extra-giuridiche.

Solo dopo aver compreso il fenomeno, nei limiti delle nostre capacità da giuristi, potremmo essere in grado di afferrarne la sua dirompenza per il diritto e, in particolare, per il diritto penale.

Non avendo le competenze tecniche necessarie per trattare in maniera approfondita dell'intelligenza artificiale, ci limiteremo ad utilizzare un approccio funzionale, delineando soltanto quelle nozioni che siano a noi, per l'appunto, funzionali per comprendere le conseguenze che lo sviluppo dei sistemi di intelligenza

artificiale può avere sull'ascrizione della responsabilità penale.

Sul concetto di intelligenza artificiale, d'ora in avanti IA, non pare che la scienza sia concorde nell'attribuire una univoca definizione. Infatti, sono numerosi gli studiosi che ritengono più confacente utilizzare la locuzione IA solo come una sorta di metafora per riferirsi a tutti i sistemi tecnologici di ultimissima generazione, per distinguerli dalle ormai acquisite e passate tecnologie.

La difficoltà definitoria è acuita dal fatto che non esiste, in pedagogia ed in sociologia, una definizione unitaria del concetto stesso di intelligenza[1]. Infatti, vi è chi ritiene che la nozione di IA debba variare al variare della concezione di intelligenza cui si aderisce e dalla risposta alle seguenti domande: per poter essere considerato come intelligente, si deve essere in grado di pensare ed agire come un umano ovvero è sufficiente un comportamento razionale, inteso come mera capacità di raggiungere il miglior risultato possibile? L'intelligenza è un attributo del pensiero ovvero si ricava dal comportamento esterno? Dalla combinazione delle varie risposte a dette domande, derivano quattro differenti concezioni di intelligenza e, dunque, di IA.

La prima concezione ritiene che l'IA sia corrispondente al fatto che la macchina agisca come un essere umano; la seconda che l'IA pensi come un essere umano; la terza che pensi razionalmente; la quarta che agisca razionalmente.

Attualmente, la maggioranza degli studiosi tende a ricondurre il concetto di IA a quest'ultimo modello, in forza del quale un agente è considerato intelligente nel momento in cui riesce a conseguire determinati obiettivi in

[1] La teoria delle intelligenze multiple di *Howard Gardner* identifica nove tipi di intelligenza: linguistica, logico-matematica, musicale, spaziale, corporeo-cinestetica, interpersonale, intrapersonale, naturalistica ed esistenziale. Gardner sostiene che queste intelligenze sono indipendenti l'una dall'altra, ma spesso si manifestano in combinazione.

un ambiente incerto e complesso[2]. Ciò detto, sembrerebbe che il discrimine tra l'IA e tutte le precedenti tecnologie sia da rinvenire proprio nella capacità di agire adattandosi ad una serie indefinita di circostanze, prendendo decisioni che massimizzano il risultato atteso.

Invero, il concetto di IA non sembra afferrabile e definibile una volta per tutte. Ciò in quanto trattasi di un concetto che muta al mutare della percezione che la società ha di una determinata tecnologia in un determinato contesto storico e culturale: basti pensare al fatto che oramai siamo abituati ad interagire con strumenti come *Google Maps* e nessuno di noi direbbe, ad oggi, che trattasi di strumenti di IA, nonostante in un certo qual modo esso lo sia. Vi è anche chi sostiene che il riferimento all'intelligenza sia solo uno slogan da operazione di *marketing*, così che algoritmi che funzionano grazie ad oramai ben consolidati meccanismi di automazione vengono talvolta spacciati per sistemi dotati di autonomia[3].

---

[2] Vd. S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 4 e 36 ss. I due elementi citati – quello del perseguimento degli obiettivi e quello dell'ambiente incerto – ricorrono nella maggior parte delle definizioni della nozione di "intelligenza": si vd., a titolo di esempio, J.S. ALBUS, Outline for a theory of intelligence, in *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 21, Iss. 3, May-June 1991, p. 474: «*the ability of a system to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioral subgoals that support the system's ultimate goal*»; D.B. FOGEL*, Review of "Computational intelligence: Imitating life*", in Proc. of the IEEE, vol. 83, iss. 11, 1995, p. 1590: «*Any system [...] that generates adaptive behaviour to meet goals in a range of environments can be said to be intelligent»; A. BERTOLINI, Study on Artificial Intelligence and Civil Liability – Study requested by the JURI Committee, July 2020, p. 19: «a machine or system is deemed intelligent when it can act in, adapt to, and learn from specific environments so as to achieve predefined goals*»; J. HORST, *A Native Intelligence Metric for Artificial Systems, Performance Metrics for Intelligent Systems, Workshop*, in *Proceedings of the Performance Metrics for Intelligent Systems* (PerMIS) Workshop, 2002, disponibile a questo link: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=824478: «*the successful (i.e., goal-achieving) performance of the system in a complicated environment*»; D. POOLE, *Computational Intelligence: A logical approach*, Oxford University Press: «[*An intelligent agent does what] is appropriate for its circumstances and its goal*, *it is flexible to changing environments and changing goals, it learns from experience, and it makes appropriate choices given perceptual limitations and finite computation*»

[3] Lo rileva A. BERTOLINI, *Artificial Intelligence Does Not Exist! Defying the Technology-Neutrality Narrative in the Regulation of Civil Liability for Advanced Technologies*, in Europa e dir. priv., 2022, n. 2, p. 373. Per un'analisi dell'evoluzione dei sistemi tecnologici,

Certamente, quando parliamo di IA non intendiamo riferirci agli androidi che hanno sempre popolato l'immaginario fantascientifico, con sembianze fattive umane e capacità intellettuali simili. Anzi, quando parliamo di IA parliamo di sofisticati *software* algoritmici in grado di svolgere determinati compiti, ma del tutto privi di buon senso, coscienza e, dunque, di una intelligenza ad ampio spettro come quella dell'essere umano. Gli attuali sistemi di IA sono *artificial narrow intelligence*, cioè sistemi deboli. Chiaramente, l'obiettivo della scienza tecnica è quello di sviluppare nel lungo termine sistemi di IA forte, cioè *artificial general intelligence*[4].

Siamo, dunque, ben lontani dall'essere tecnologico senziente ed autocosciente.

1.1. Cosa differenzia gli algoritmi di IA dagli algoritmi tradizionali?

Con il termine algoritmo, di regola, si intende, una sequenza di istruzioni che prescrive di trasformare un

dall'automazione all'autonomia si vd. B. PANATTONI*, AI and Criminal Law: The Myth of 'Control' in a Data-Driven Society*, in *International Review of Penal Law*, 2021, p. 125 ss.

[4] La distinzione tra *weak AI e strong AI* risale a J.R. SEARLE, *Minds, brains, and programs*, in *Behavioral and Brain Sciences*, 1980, n. 3 p. 417. Il dibattito sulla fattibilità tecnica della creazione di sistemi di i.a. con capacità intellettive paragonabili a quelle dell'essere umano è in realtà ancora acceso; si va dagli studiosi che preconizzano un imminente sviluppo di forme di AGI (es. R. KURZWEIL, *The Singularity Is Near: When Humans Transcend Biology*, Viking Penguin, 2005) a coloro che, al contrario, ritengono che tale obiettivo sia ancora molto lontano, o addirittura irraggiungibile (es. gli informatici Craig Silverstein e Mark Andreesen, le cui dichiarazioni sono contenute in K. AULETTA*, Googled: The End of the World as We Know It*, Penguin Press, 2009, p. 327). La molteplicità dei punti di vista in merito è ben testimoniata da un sondaggio condotto tra alcuni esperti del settore, vd. S.D. BAUM E AL., *How long until human-level AI? Results from an expert assessment*, in *Technological Forecasting and Social Change*, vol. 78, iss. 1, p. 185. L'attuale dibattito pare essere andato persino oltre, concentrandosi sul possibile sviluppo di sistemi di i.a. che superino le capacità intellettive umane (c.d. superintelligenza); sul tema vd. N. BOSTROM, *Superintelligenza, Bollati Boringhieri*, 2014, che in tale rivoluzione tecnologica vede una potenziale catastrofe per la società, dalla quale è necessario tutelarsi; vd. anche il sondaggio condotto su 550 esperti nel campo dell'i.a., ai quali è stato chiesto quando, a loro parere, si avranno le capacità tecnologiche per sviluppare una "superintelligenza", vd. V.C. MÜLLER – N. BOSTROM, *Future Progress in Artificial Intelligence: A Survey of Expert Opinion*, in V.C. Müller (ed*.), Fundamental Issues of Artificial Intelligence*, Springer, 2016, p. 555 ss. Sulla possibile futura regolazione dei sistemi AGI vd. T. MAHLER, *Regulating Artificial General Intelligence* (AGI), in B. Custers – E. Fosch-Villaronga (eds.), *Law and Artificial Intelligence*, Asser Press – Springer, p. 521 ss.

determinato *input* in un altrettanto *output*, sulla base di un procedimento predefinito[5]. Lo schema tradizionale è quello dell'*if-this-then-that*. Non è obbligatorio che un algoritmo, per essere tale, sia incorporato su un supporto informatico: sono algoritmi anche le stesse norme giuridiche che prescrivono al giudice di applicare una determinata pena come conseguenza dell'accertamento della commissione di un certo fatto di reato[6]. Una prima differenza tra algoritmi tradizionali ed algoritmi intelligenti è proprio questa. Inoltre, le istruzioni che vengono impartite agli esseri umani possono anche dare per sottintese alcune nozioni, posto che l'essere umano è dotato del buon senso grazie al quale è in grado di colmare le eventuali inadeguatezze dell'algoritmo rispetto al caso concreto; mentre, le istruzioni impartite alle macchine devono essere perfettamente dettagliate poiché esse applicano alla lettera il tutto, non essendo in grado di interpretare le diverse situazioni in base al contesto[7].

L'essere umano potrebbe dare esecuzione ad una regola in modo impreciso e difettoso, mentre il sistema informatico dovrebbe eseguire l'istruzione fornitagli in sede di programmazione perfettamente, non residuando

[5] vd. G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione*?, vol. 1, il Mulino, 2022, p. 50: «*Semplificando al massimo, potremmo dire che gli algoritmi sono sequenze di istruzioni, linee di codice, che possiamo scrivere e che possiamo leggere*»; vd. anche T. GILLESPIE, *The relevance of algorithms*, in T. Gillespie – P. Boczkowski – K. Foot (eds), *Media Technologies: Essays on Communication, Materiality, and Society*, MIT Press, p. 167: «*encoded procedures for transforming input data into a desired output, based on specified calculations*»; N. DIAKOPOULOS, *Algorithmic Accountability: Journalistic investigation of computational power structures*, in Digit. Journal, 2015, 3, p. 400: «*a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome*»; queste ultime due definizioni sono state riprese dal Consiglio d'Europa nel *report Algorithms and Human Rights* (*Council of Europe* DGI(2017)12), 2018.

[6] Vd. D.R. DESAI – J.A. KROLL, *Trust But Verify: A Guide to Algorithms and the Law*, *in Harvard journal of law & technology*, 2017, Vol. 31, n. 1, p. 23: «*we can call the steps for brushing teeth an algorithm*».

[7] D.R. DESAI – J.A. KROLL, *Trust But Verify*, cit., p. 23-24; sull'assenza di senso comune dei sistemi di i.a. vd. C.E.A. KARNOW, *The application of traditional tort law to embodied machine intelligence*, in R. Calo e al. (eds.), *Robot Law*, Edward Elgar Publishing, 2016, p. 75; M.B. MAGRO*, Robot, cyborg e intelligenze artificiali*, cit., p. 1194 ss.

alcuna discrezionalità e valutazioni di opportunità nella fase del *decision making*[8].

Gli algoritmi, di tutti i tipi, possono pertanto essere incorporati in un supporto fisico (*robot*) ovvero avere natura esclusivamente virtuale (*bot*). Nella prima eventualità, l'interazione con il mondo avverrà secondo lo *standard* del *sense, think, act*: il *robot* percepisce l'ambiente circostante, processa le informazioni ed agisce. È proprio detta capacità di interagire con il mondo esterno che rende il *robot* un essere foriero di danni di rilevanza penale alle cose ed alle persone: basti pensare ai droni da guerra, alla chirurgia robotica. Nella seconda eventualità, invece, gli algoritmi agiscono sui sistemi informatici.

1.2. Brevi cenni storici in materia di IA.

Il concetto di Intelligenza Artificiale nasce nell'ottobre del 1950 con un contributo scientifico di *Alan Turing* che esordiva così: *<<I propose to consider the question, "Can machine think?>>*[9]. L'illustre matematico studiava la possibilità di introdurre nella realtà una macchina pensante, proponendo un *test* per verificare quando, in termini operativi, una macchina potesse definirsi intelligente.

Quello successivamente noto come *Test* di *Turing* si fondava su un celebre *imitation game*. Ovverosia, una donna ed un *computer*, in due stanze separate, venivano posti nella condizione di comunicare per iscritto con un comunicatore umano, il quale poteva rivolgere ogni tipo di domanda. Alla conclusione dello scambio, l'intervistatore avrebbe dovuto scoprire l'identità dei due conversanti: nel caso in cui non fosse stato in grado di comprendere correttamente, la macchina avrebbe superato

[8] S. QUATTROCOLO*, Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., p. 14.

[9] A. TURING, *Computing Machinery and Intelligence*, in *Mind*, 1950, p. 433.

il *test*, dimostrando di essere interscambiabile con il conversante umano[10].

Nel 1956, si verifica la nascita ufficiale del campo di ricerca della intelligenza artificiale quando nell'estate un gruppo di brillanti studiosi composto da *John McCarhty*, *Marvin Minsky, Claude Shannon* e *Nathaniel Rochester* invitarono ricercatori di teoria degli automi, reti neurali e studio dell'intelligenza artificiale a partecipare ad un *workshop* di due mesi presso la città di *Dartmouth*[11]. Ed è proprio in occasione dell'evento *Dartmouth Summer Research Project on Artificial Intelligence* che viene utilizzata per la prima volta scientificamente la locuzione *Artificial Intelligence*[12].

Negli anni a seguire, lo studio dell'intelligenza artificiale si è alternato tra periodi di grande entusiasmo e periodi di pause riflessive. Certamente, nei primi anni si verificavano alcuni successi. I primi algoritmi venivano progettati secondo tecniche basate sull'empirismo: si esplorava passo dopo passo un determinato spazio di ricerca. Gli algoritmi tentavano i diversi percorsi possibili e, in caso di errore, tornavano indietro per intraprendere nuove strade fino a quando non avessero raggiunto la soluzione[13].

Sulla base di questa impostazione empiristica, vennero creati programmi capaci di risolvere teoremi matematici[14],

---

[10] L. FLORIDI, *Philosophy and computing. An introduction*, Londra, 1999, p. 134 ss.
[11] S. J. RUSSEL, P. NORVIG, *Artificial Intelligence*, cit. p. 17; G.F. ITALIANO, *Intelligenza Artificiale: passato, presente, futuro*, in PIZZETTI F. (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018, p. 208 ss.
[12] J. MCCARTHY, M.L. MINKSY, N. ROCHESTER, C.E. SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, in *AI Magazine*, 2006, p. 12. Qui si può leggere parte del testo della proposta: «*We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer*».
[13] G.F. ITALIANO, *Intelligenza Artificiale: passato, presente, futuro*, cit., p. 210.
[14] Herbert Gelertner, nel 1959, costruì il *Geometry Theorem Prover*, capace di dimostrare teoremi matematici.

di giocare a dama[15] e si approfondì lo studio delle reti neurali artificiali.

Gli entusiasmi in merito alla potenzialità dell'intelligenza artificiale ben presto dovettero fare i conti con l'emergere dei primi ostacoli. Uno dei maggiori ostacoli era l'intrattabilità di molti problemi che i sistemi di intelligenza artificiale cercavano di risolvere. L'approccio empiristico non funzionava bene per le questioni estremamente complesse[16].

Di pari passo con le prime difficoltà, mutò anche l'atteggiamento delle istituzioni governative che nei primi anni avevano contribuito allo sviluppo della nuova disciplina. Il Regno Unito, ad esempio, decise di sospendere il finanziamento economico a tutti i centri di ricerca sulla intelligenza artificiale.

Una ripresa importante nel campo si ebbe con la nascita dei c.d. sistemi esperti[17]. Agli inizi, i sistemi di intelligenza artificiale erano in grado soltanto di ragionare su problemi astratti. Gli studiosi si resero conto che la capacità di ragionamento è solo una parte dell'intelligenza, dentro cui vi è anche la capacità di risolvere problemi reali[18]. Per potere risolvere problemi reali, occorre che il soggetto che si appresta a risolverli abbia alcune necessarie conoscenze. I sistemi esperti vennero creati proprio per venire incontro a questa esigenza: trattasi, infatti, di programmi informatici in grado di rappresentare e ragionare con la

[15] Ci si riferisce al programma scritto da *Arthur Samuel* e dimostrato in televisione nel 1956. All'inizio Samuel era in grado di sconfiggere il suo programma, ma dopo pochi mesi non riuscì più a vincere una partita. Fu uno dei primi esempi di *machine learning*. Cfr. S. FRANKLIN, *History, motivations, and core themes*, in FRANKISH K., RAMSEY W. M. (a cura di*), The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, p. 19.
Frank Rosenblatt creò *il perceptron*.

[16] S. J. RUSSEL, P. NORVIG*, Artificial Intelligence*, cit. pp. 21-22.

[17] Uno dei primissimi esempi di tale approccio è costituito dal programma DENDRAL, sviluppato nel 1965 da Edward Feingenbaum, Joshua Lederberg e Bruce Buchanan. DENDRAL aiutò ad identificare la struttura molecolare di molecole organiche, analizzando i dati di uno spettrometro di massa ed utilizzando le sue conoscenze in chimica. Ad esso seguì Mycin, un altro sistema esperto che aiutò i medici a diagnosticare ed a trattare le malattie infettive del sangue e meningiti. Cfr. J. RUSSEL, P. NORVIG, *Artificial Intelligence*, cit. pp. 22-23.

[18] S. FRANKLIN*, History, motivations, and core themes*, cit. p. 20.

conoscenza di alcuni soggetti specializzati in un determinato settore allo scopo di risolvere problemi o dare consigli. Ad esempio, si può pensare al sistema *Mycin*, creato per aiutare i medici a diagnosticare malattie infettive del sangue e le meningiti. Tali sistemi, sin da subito, si rivelarono utili in particolar modo per le loro applicazioni industriali[19].

In quegli stessi anni, ripresero anche gli studi sulle reti neurali con la pubblicazione di alcuni importanti volumi in materia, tra cui il lavoro di *Rumelhart*. Tali studi diedero vita all'utilizzo di reti neurali artificiali nelle scienze cognitive e trovarono varie applicazioni pratiche, essendo usate per la rivelazione di frodi, valutazioni immobiliari, valutazione del credito, etc.

Anche questo periodo florido è stato succeduto da un altrettanto periodo di crisi di molte imprese che si occupavano di intelligenza artificiale. Dopo questo arresto, lo sviluppo dell'intelligenza artificiale cercò di ripartire da basi più concrete e solide, usando i risultati provenienti da altre discipline come la teoria del controllo e la ricerca operativa.

L'ultimo ventennio è stato segnato da importanti conquiste. Una delle più importanti si ebbe nel 1997, quando un sistema intelligente costruito da IBM e capace di giocare a scacchi (conosciuto come *Deep Blue*) riuscì a sconfiggere il campione mondiale di scacchi *Gary Kasparov* in un match di sei sfide.

Un successo ancora maggiore si verificò quando un sistema intelligente (*Equational Prover*) dimostrò la congettura di *Robins*, un problema matematico fino ad allora irrisolto, che aveva resistito per molti anni agli sforzi dei matematici.

[19] Ad esempio, tramite l'uso di un sistema denominato XCON nella configurazione di ordini per nuovi sistemi informatici, *la Digital Equipment Corporation* risparmiò all'incirca quaranta milioni di dollari l'anno; cfr. G. F. ITALIANO, *Intelligenza Artificiale: passato, presente, futuro*, cit., p. 214.

Sulla stessa scia, nel 2011, un altro sistema costruito da IBM (Watson) ha vinto il *quiz* televisivo americano *Jeopardy*, battendo i campioni in carica, utilizzando algoritmi di elaborazione del linguaggio naturale per interrogare circa 200 milioni di pagine di contenuto e rispondere alle domande in inglese colloquiale.

Ancor più recente, nel 2017, *GoogleAlphaGo* è riuscita a sconfiggere i campioni del complicato gioco *Go* ed un altro sistema *Libratus* a vincere contro i migliori giocatori di *poker Texas Hold'em*.

Al giorno d'oggi, l'intelligenza artificiale è presente nella vita quotidiana ed impatta vari settori della vita sociale. Può essere utilizzata per tradurre discorsi da una lingua all'altra, per riconoscere discorsi, per effettuare pianificazioni logistiche, per combattere *e-mail*, *spam* e così via. Molte delle sue applicazioni sono già presenti nei telefoni cellulari di uso comune.

Alla base di questo incredibile sviluppo della intelligenza artificiale, non risiede tanto un mutamento delle tecniche impiegate, quanto, piuttosto, la disponibilità di una potenza di calcolo notevolmente superiore ed una enorme quantità di dati.

Negli anni che verranno, questa diffusione e crescita sembra destinata a subire un ulteriore sviluppo. Ci sarà un impiego trasversale delle applicazioni fondate sull'intelligenza artificiale, con l'introduzione di veicoli a guida autonoma, l'uso di prodotti robotici per lo svolgimento di delicate operazioni mediche (*Surgical Robots*) o per l'assistenza post-operatoria, l'uso dell'intelligenza artificiale per operazioni di *law enforcement* (ad es. droni di sorveglianza, algoritmi per la *fraud detection*) o come ausilio nell'attività educativa, fino all'impiego di *robot* per lo svolgimento di faccende quotidiane. Ovviamente, sono previsti anche cambiamenti strutturali economici e sociali e, particolarmente, l'ambito del lavoro potrebbe vedere una graduale sostituzione

degli esseri umani nello svolgimento di determinate mansioni.

1.3. Il *machine learning*, il *deep learning*, e le principali modalità di apprendimento.

Il presente ed il futuro dell'IA è il *machine learning*[20]. Trattasi di una tecnica che consente agli algoritmi di apprendere dai dati, attraverso il metodo induttivo: <<*a computer observes some data, builds a model based on the data, and uses the model as both a hypothesis about the world and a piece of software that can solve problems*>>[21].

Durante la fase del *training*, l'addestratore/programmatore sottopone all'algoritmo differenti casi dai quali quest'ultimo inferisce il modo di agire. Infatti, la forza dei sistemi di ML sta nel fatto che essi sono in grado di trarre generalizzazioni statistiche partendo da esempi concreti e specifici[22]. Alla base di questo approccio giace la considerazione secondo cui non tutte le attività dell'essere umano siano formalizzabili tramite la creazione di regole esplicite e ben definite, ma sono il frutto di esperienza ed imitazione[23]. Si è avuto modo di comprendere come avrebbe avuto maggiore applicabilità concreta e minore margine di errore una tecnologia che non si basasse su regole astratte, ma che si basasse sulla capacità di riconoscere *pattern* statistici nei dati acquisiti ed effettuare generalizzazioni. Se, ad esempio, dovessimo fornire ad un sistema di riconoscimento facciale un *database* di immagini di tante persone, il sistema sarà in grado di discernere i tratti del viso, creando distinte categorie riconducibili a ciascuno

[20] D'ora in avanti lo indicheremo come ML.
[21] S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 651.
[22] P. DOMINGOS, *A Few Useful Things to Know about Machine Learning*, in *Communications of the ACM*, vol. 55, n. 10, nov. 2012, p. 79; nella letteratura giuridica vd. per tutti A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence based practices e tutela dei diritti fondamentali*, in *Arch. pen.* (web), 2021, n. 1, p. 6.
[23] T. Ray, *A new experiment: Does AI really know cats or dogs -- or anything?,* in Zdnet, 28 gennaio 2022, disponibile a questo link: https://www.zdnet.com/article/what-does-ai-know-of-cats-and-dogs-maybe-very-little/.

dei tre individui. Maggiore sarà il numero di immagini nel *database*, più accurata sarà la classificazione. Se volessimo, invece, che un sistema riconoscesse una specifica persona in mezzo a tante, potremmo fornire all'algoritmo una serie di immagini determinate, di modo che il sistema possa identificare i tratti del viso di tale persona e riconoscerli all'interno delle immagini fornite come *input*.

Sarebbe impossibile raggiungere un siffatto risultato applicando l'approccio simbolico e causalistico delle vecchie intelligenze artificiali. Lo sviluppo delle tecniche di ML ha comportato un cambio di paradigma nel tema dell'IA, dando il via alla ricerca sui sistemi autonomi in moltissimi settori. Molteplici sono i benefici di questa tecnica: gli sviluppatori non devono produrre miliardi di istruzioni esplicite *ex ante* nemmeno conoscibili perché i sistemi di ML sono in grado di imparare a 'comportarsi' in base alle esperienze 'vissute'. La logica sottesa allo sviluppo del ML è la seguente: è conveniente e più efficiente insegnare alla macchina ad effettuare generalizzazioni, piuttosto che impartire singoli insegnamenti meramente nozionistici.

Inoltre, i sistemi di ML sono in grado di cogliere dei *pattern* molto sottili nei *data* che gli esseri umani non sarebbero in grado di notare[24]. Grazie al cd. *cloud computing*, i *software* connessi tra di loro sono in grado di imparare dalle reciproche esperienze[25].

Al giorno d'oggi si può affermare che siamo in un periodo florido per la ricerca in materia perché vi è un'ampia disponibilità di dati, i cd. *big data*, ma anche perché vi è stato un incremento di capacità computazionale dei *computer*, sia in termini di potenza sia

[24] Vd. A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, in *Fordham Law Review*, 2018, p. 1094; vd. anche S. RUSSELL – P. NORVIG, Artificial Intelligence, cit., p. 651.

[25] A. CAPPELLINI*, Machina delinquere non potest*?, cit., p. 6.

in termini di velocità e memoria[26]. La tecncica maggiormente utilizzata di ML è fondata sulle reti neurali artificiali, ovverosia sistemi di elaborazione dei dati il cui funzionamento trae ispirazione dal funzionamento dell'apparato nervoso dell'essere umano.

Le reti neurali altro non sono che *software* di modellazione statistica installabili direttamente sul *computer* o su *hardware* ad essi dedicati, che consentono un addestramento ed un'esecuzione delle funzioni algoritmiche più rapidi[27].

In breve: le reti neurali ricevono informazioni (*input*) su uno strato di nodi di ingresso ed emettono risposte nell'ambiente esterno (*output*), tramite l'elaborazione di dati compiuta da un differente strato di nodi (*hidden*). Dunque, la maggior parte delle reti neurali è formata da tre strati: quello degli ingressi che riceve i segnali in ingresso (*input*); quello nascosto che procede all'elaborazione vera e propria dei dati (*hidden*); quello di uscita che raccoglie i risultati dello strato *hidden* e li comunica all'esterno (*output*).

Nel momento in cui le reti neurali sono organizzate su più livelli, siamo dinanzi al cd. *deep learning*. Trattasi, in breve, di un campo di studi branca del ML. La potenza computazionale dei sistemi di *deep learning* dipende dalla abilità dei neuroni artificiali di apprendere tutto ciò che i nodi dei livelli precedenti hanno già elaborato e processato[28]. Il sistema è altamente efficiente ed efficace in

[26] Tra i tanti vd. J. KAPLAN, *Artificial Intelligence*, cit., p. 39; J. SINGH E AL., *Responsibility & Machine Learning: Part of a process*, in SSRN, 2016, p. 2, disponibile a questo link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2860048; R. CALO, *Artificial Intelligence Policy: A Primer and Roadmap*, in *University of Bologna Law Review*, 3(2), p. 402; G.F. ITALIANO E AL*., Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni*, cit., p. 49; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1182.

[27] Vd. M. Negretti, *Intelligenza artificiale, quali sono le architetture hardware dedicate*, *in AI4Business*, 20 gennaio 2021, disponibile a questo link: https://www.ai4business.it/intelligenza-artificiale/intelligenza-artificiale-quali-sono-le-architetture-hardware-dedicate/, che fornisce una spiegazione semplificata dei vantaggi dell'hardware nell'addestramento e nel funzionamento delle reti neurali.

[28] J.D. KELLEHER – B. TIERNAY, *Data Science*, MIT Press, 9th ed., 2018, p. 238.

quanto ogni strato della rete si occupa specificamente di un determinato aspetto dell'informazione da elaborare: ad esempio, in un sistema deputato ad analizzare un'immagine, il livello più basso si potrebbe focalizzare sull'analisi del colore; quello successivo sulle linee e quello dopo ancora su altri elementi dell'immagine, fino ad arrivare all'identificazione dell'immagine[29].

Venendo ora alle principali modalità di apprendimento nel ML, si può asserire che, di solito, si distingue tra tre diverse modalità: i) *supervised learning*; ii) *unsupervised learning*; iii) *reinforcement learning*.

i) Nella prima modalità di apprendimento, il sistema analizza la coppia *input-output* assegnata dall'addestratore ed impara a riconoscere l'*output* che deve essere correlato a ciascun input (cd. *classification o pattern recognition*)[30]. Un esempio di *supervised learning* può essere quello dell'addestramento di algoritmi per la guida autonoma tramite l'applicazione *ReCAPTCHA*[31] di *Google*: il *software* chiede agli utenti di classificare alcune immagini inerenti alla circolazione stradale al fine di riconoscere se l'utente è un essere umano o un *bot* che sta provando a violare un account protetto. Anche se la funzione primaria di detto *software* è la sicurezza informatica, moltissimi hanno ipotizzato che i dati acquisiti con *ReCAPTCHA* vengano usati da *Google* per alimentare i propri algoritmi di guida autonoma installati sulle *Google Cars*[32];

[29] L. DORMEHL, *Thinking Machines: The Quest for Artificial Intelligence and Where It's Taking Us Next, Tarcherperigree*, 2017, p. 50-51.
[30] K.P. MURPHY, *Machine Learning*, cit., p. 2
[31] Acronimo di Completely Automated Public Turing test to tell Computers and Humans Apart, vd. la pagina di assistenza Google (Che cos'è un CAPTCHA?), disponibile al link: https://support.google.com/a/answer/1217728?hl=it.
[32] Così M. Lotkowski, *You Are Building a Self Driving AI Without Even Knowing About It*, Hackernoon, 25 febbraio 2017, disponibile a questo link: https://hackernoon.com/you-are-building-a-self-driving-ai-without-even-knowing-about-it-62fadbfa5fdf; R. Maruzani, *Are You Unwittingly Helping to Train Google's AI Models?*, in Towards Data Science, 26 gennaio 2021, disponibile a questo link: https://towardsdatascience.com/are-you-unwittingly-helping-to-

ii) Nei sistemi di *unsupervised learning*, vengono invece forniti gli input ma non gli output e l'obiettivo dell'algoritmo è trovare gli *interesting patterns*[33] nei dati. Questa tecnica dà la possibilità di abbattere i costi derivanti dalla necessità di ottenere *labeled data* e consente di classificare enormi quantità di dati. La tecnica più diffusa è il *clustering,* tecnica grazie al quale si individuano categorie specifiche all'interno di un *dataset*: si prendano come esempio gli algoritmi che classificano gli utenti di un sito *e-commerce* o di un motore di ricerca sulla base delle loro preferenze di acquisto o del loro comportamento online allo scopo di inviare messaggi pubblicitari personalizzati;

iii) Nel *reinforcement learning*, invece, il sistema impara a risolvere problemi che richiedono di elaborare strategie molto sofisticate, tramite il calcolo delle azioni più promettenti per raggiungere un obiettivo specifico. Tale tecnica viene utilizzata quando l'ambiente non è conosciuto ovvero quando la modellazione computativa richiederebbe un impiego eccessivo di risorse. Uno degli aspetti cardine di tale modalità è la predisposizione di incentivi e disincentivi da parte dei programmatori grazie ai quali il sistema impara a riconoscere sia le azioni più promettenti sia quelle che non lo sono, così da poter modificare e migliorare in autonomia le *performance* future[34].

I sistemi di ML consentono di individuare, dunque, *pattern* ricorrenti impercettibili dall'uomo, ma questo tipo di ragionamento basato sull'autoapprendimento porta con

train-googles-ai-models-f318dea53aee; i vertici di Google, d'altra parte, avevano già ammesso di utilizzare i dati ottenuti attraverso ReCAPTCHA per addestrare Google Street View, vd. S. Perez, Google Now Using ReCAPTCHA To Decode Street View Addresses, in TechCrunch, 29 marzo 2012, disponibile a questo link: https://techcrunch.com/2012/03/29/google-now-using-recaptcha-to-decode-street-view-addresses/.

[33] K.P. MURPHY, *Machine Learning*, cit., p. 2.

[34] G.F. ITALIANO E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni*, cit., p. 52.

sé una carenza di comprensione semantica che potrebbe determinare errori talvolta anche grossolani. Ad esempio, un sistema di riconoscimento di immagini potrebbe scambiare l'immagine di una tartaruga con quella di un'arma[35]. Così come un graffito che ingombra un segnale stradale ovvero una scarsa illuminazione delle strade potrebbe comportare un errore di riconoscimento da parte di un sistema di guida autonoma[36].

Nonostante si stia tentando di sviluppare sistemi di DL resistenti a questi eventi, ci sarà sempre una percentuale di errore perché i sistemi di ML si limitano a riconoscere *pattern* ricorrenti in un *dataset*, ma non sono in grado di discernere e ragionare. Il sistema di ML non può imparare ciò che non ha mai visto: l'accuratezza dei risultati è diretta conseguenza dell'affidabilità dell'algoritmo e della precisione dei dati usati per l'addestramento. L'utilizzo di dati imprecisi od incompleti ha già creato problemi di discriminazione: ad esempio, in un caso emblematico, un algoritmo usato per verificare la corrispondenza tra le fototessere ed i criteri per il rilascio dei documenti di identità scartava le persone con tratti asiatici perché era stato addestrato solo con foto di persone di razza caucasica. Sostanzialmente, il sistema riteneva che gli occhi dei cinesi fossero come occhi chiusi[37].

L'algoritmo potrebbe anche discriminare nel procedimento di *decision-making*, prendendo in considerazione elementi irrilevanti, arrivando a trattare situazioni uguali in maniera diversa. A ciò si aggiunga che l'utente spesso non conosce né i dati esaminati dall'algoritmo né i meccanismi decisionali usati per

[35] A. ATHALYE E AL., *Synthesizing Robust Adversarial Examples*, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
[36] K. EYKHOLT E AL., *Robust Physical-World Attacks on Deep Learning Visual Classification*, cit.
[37] R. DEVILLÉ E AL., *Basic Concepts f AI for Legal Scholars*, cit., p. 11, a cui si rinvia per gli opportuni riferimenti bibliografici.

l'elaborazione. Ne consegue che gli utenti non possano neanche contestare le decisioni automatizzate[38].

2. Gli algoritmi a supporto dell'attività giudiziaria ed all'attività di indagine.

Già da qualche tempo si sta sperimentando l'uso degli algoritmi di IA a favore delle attività di polizia e di giustizia "predittiva". Gli strumenti per il cd. *law enforcement* sono di applicazione statistica con l'obiettivo di prevedere dove, quando e ad opera di chi potrà in futuro essere commesso un determinato reato, per poterne prevenire la commissione. Mentre i sistemi usati per l'attività giudiziaria sono utili a valutare la pericolosità criminale di un autore di un crimine, ovverosia il pericolo di recidiva (cd. *risk assesment tools*)[39].

Trattasi di strumenti che hanno un forte rischio discriminatorio, nonostante questi potenzialmente siano in grado di apportare benefici. Infatti, tali algoritmi possono portare con sé pregiudizi esistenti nella mente dell'essere umano che li programma. L'*input* processato dall'algoritmo può essere in *re ipsa* discriminatorio. Ad esempio, un *risk assesment* tool che includa tra i criteri di valutazione della pericolosità anche dati inerenti alle caratteristiche etniche, religiose e sociali[40].

La letteratura penalistica e processual-penalistica sull'uso dei sistemi di IA nel settore della giustizia penale è oramai sterminata e, nonostante non sia il tema della nostra ricerca, per tale ragione occorre brevemente dar conto del loro funzionamento[41].

---

[38] D. Gorski, *IBM's Watson versus cancer: Hype meets reality*, in *Science-based Medicine*, disponibile a questo link: https://sciencebasedmedicine.org/ibm-watson-versus-cancer-hype-meets-reality/.

[39] F. BASILE, *Intelligenza artificiale ediritto penale*, cit, p. 10.

[40] S. QUATTROCOLO, *Intelligenza artificiale e giustizia: nella cornice della Carta Etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in Leg. pen., 18 dicembre 2018, p. 6, nota 14; vd. anche A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 12 ss.

[41] S. QUATTROCOLO, *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, cit., D. KEHL E AL., *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Berkman Klein Center for Internet & Society, Harvard Law*

Con riferimento ai sistemi di polizia predittiva (*predictive policing*), possiamo distinguerli in due tipi di modelli: l'uno riguarda quei modelli che individuano le zone calde, basandosi su dati ambientali, spaziali e temporali ricorrenti nei reati precedentemente commessi (cd. *hot spots*); l'altro tipo, attraverso l'elaborazione di dati di diversa provenienza come fotografie o informazioni inerenti ad analoghi reati precedenti, consente di profilare il possibile autore di una serie criminale, prevedendone la mossa successiva. Gli algoritmi di quest'ultimo tipo si definiscono di *crime linking*. I sistemi del primo tipo rientrano nella cd. *Place-Based Predictive Policing*[42], mentre quelli del secondo nella cd. *Person-Based Predictive Targeting*[43].

La logica alla base di detti algoritmi di *crime linking* è la seguente: determinate forme di criminalità si manifestano in un arco temporale ed in una zona geografica circoscritti (cd. *near repeat crimes*)[44].

School, 2017; D.K. CITRON, *Technological Due Process*, Wash.U. L. Rev., 2008, p. 1249; in particolare, sulla polizia predittiva vd. A.G. FERGUSON, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, NYU Press, 2017; S. BRAYNE, *Predict and Surveil: Data, Discretion, and the Future of Policing*, Oxford University Press, 2020. In lingua italiana si vedano G. UBERTIS, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in Dir. pen. cont. – Riv. Trim., 2020, n. 4; D. PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi*, Giappichelli, 2022; S. SIGNORATO, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in Riv. dir. proc., 2020, p. 604 ss.; G. CANZIO, *Intelligenza artificiale, algoritmi e giustizia penale*, in Sist. pen., 8 gennaio 2021; O. DI GIOVINE, *Il judge-Bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in Cass. pen., 2020, 951 ss.; V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit.; A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit.; P.P. PAULESU, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Arch. pen. (web)*, 2022, n. 2; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Dir. pen. cont*., 29 maggio 2019; L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, in *Dir. pen. cont*., 2019, 2, 359 ss.; R. BICHI, *Intelligenza digitale, giurmetria, giustizia predittiva e algoritmo decisorio. Machina sapiens e il controllo sulla giurisdizione,* in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 423; M. AMISANO, *Prevedere -e non predire- attraverso gli algoritmi e le loro insidie*, in *Arch pen.*, 2022, n. 2; in generale, sul rapporto tra processo e nuove tecnologie vd. A. GARAPON – J. LASSÈGUE, *Giustizia digitale*, Il Mulino, 2021 (ed. franc. 2018).

[42] A.G. FERGUSON, *The Rise of Big Data Policing*, cit., p. 62.

[43] A.G. FERGUSON, *The Rise of Big Data Policing*, cit., p. 34.

[44] Per l'analisi di alcuni sistemi di polizia predittiva utilizzati in Italia (in particolare, *KeyCrime* utilizzato dalla Questura di Milano e X-LAW utilizzato dalla Questura di Napoli) rinviamo a F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 12-13.

È facilmente intuibile la presenza del rischio discriminatorio, soprattutto avuto riguardo ai sistemi che definiscono gli *hotspots*: questi sistemi di *predictive policing* si auto-alimentano poichè aumentano i controlli e le pattuglie della polizia nelle zone ritenute ad alta densità criminale, con inevitabile conseguente crescita del tasso di reati rilevati in quella zona specifica. Quella zona diventerà sempre più *hot*, mentre altre zone, non presidiate dalla polizia, rischiano di diventare zone franche per il crimine[45].

I *risk assesment tools* sono, invece, sistemi utili a vagliare il pericolo di commissione di nuovi reati da parte di chi ne abbia commesso già uno. Trattasi di valutazione spettante al giudice penale in diverse occasioni, tra cui quando si tratta di applicare una misura cautelare, una misura di sicurezza o una misura di prevenzione, ovvero quando egli debba concedere o meno la sospensione condizionale della pena o l'affidamento in prova ai servizi sociali[46]. Nell'ambito di tali decisioni, i giudici hanno fornito negli anni risposte perlopiù intuitive e basate sul senso comune o, se permesso dalla legge, sulla base di perizie cliniche[47]. L'uso degli algoritmi sembrerebbe promettere valutazioni effettuate su elementi imparziali ed oggettivi; proprio tale pretesa oggettività ed imparzialità consentirebbe agli algoritmi predittivi di trovare un terreno fertile nel nostro Paese, ove pare ormai senza rimedio la crisi di legittimazione degli organi giurisdizionali[48]. Tali fattori potrebbero riguardare l'età, il sesso, l'origine etnica, il livello di scolarizzazione, la situazione familiare e lavorativa, i precedenti penali, il consumo di sostanze psicotrope o alcoliche etc. [49]. Raccolti detti dati, l'algoritmo li mette insieme fornendo un punteggio

[45] F. BASILE, *Intelligenza artificiale ediritto penale*, cit, p. 13.
[46] F. BASILE, *Intelligenza artificiale e diritto penale*, cit, p. 16; A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 2 ss., a cui si rinvia per il riferimento ad alcuni dei sistemi di valutazione del rischio di recidiva già utilizzati in Italia (spec. p. 4).
[47] A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 7.
[48] V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 10-11.
[49] F. BASILE, *Intelligenza artificiale ediritto penale*, cit, p. 17.

e predisponendo una scala di pericolosità che potrebbe essere usata dal giudice per addivenire ad una decisione.

Oltreoceano, tali strumenti sono già usati per la concessione della *parole*, della *probation* e di altre misure alternative alla detenzione, così come anche nella fase di *sentencing* per la commisurazione della pena del colpevole. Celebre è l'algoritmo *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS), che prende in esame informazioni ottenute dall'imputato tramite un'intervista ed i dati concernenti il carico pendente ed il casellario giudiziale[50]. Tale sistema è passato alla cronaca per diverse contestazioni che ha ricevuto a causa delle presunte discriminazioni causate dal suo utilizzo, nonché di alcuni fattori criminalmente irrilevanti. In particolare, un'inchiesta pubblicata nel 2016 sul sito *ProPublica* ha fatto emergere come le previsioni di COMPAS favorissero sistematicamente gli imputati bianchi rispetto a quelli neri, sovrastimando il rischio di recidiva dei secondi rispetto ai primi[51].

Ma l'algoritmo COMPAS è passato alla cronaca anche e soprattutto per il cd. caso *Loomis*. In tale circostanza, l'imputato *Loomis* contestava che l'entità di pena che il giudice di primo grado gli aveva inflitto, usando il sistema COMPAS, fosse eccessiva poiché basata su dati discriminatori come il genere e la razza. Inoltre, si contestava l'opacità del suo meccanismo di funzionamento, dettato da motivi di segreto industriale contrari al principio del giusto processo. La Corte Suprema del Wisconsin respinse il ricorso di *Loomis*, asserendo che, nonostante fosse presente il rischio

[50] S. QUATTROCCOLO*, Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale "predittiva*", in *Cass. pen*., 2019, p. 1748 ss.
[51] J. Angwin e al., *Machine Bias*, in *ProPublica,* 23 maggio 2016, disponibile a questo link: *https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing*. In replica a tale studio, l'azienda produttrice di COMPAS ha commissionato un contro-studio ad alcuni ricercatori del proprio team, il quale avrebbe evidenziato una serie di errori nella metodologia utilizzata dai ricercatori di *ProPublica*, vd. A.W. FLORES E AL. *False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*", in *Federal Probation Journal*, vol. 80, n. 2, 2016. In generale, sul rischio di discriminazione nell'utilizzo di algoritmi nell'ambito della giustizia penale vd. A.Z. HUQ*, Racial Equality in Algorithmic Criminal Justice*, in Duke L.J., 2019, 1043 ss.

discriminatorio in COMPAS, l'uso degli algoritmi predittivi in sede processuale era legittimo perché le valutazioni statistico-computazionali erano state vagliate dal giudice umano. Con riferimento all'eccezione fondata sull'opacità, la Corte asseriva che *Loomis* aveva comunque avuto il diritto di supervisionare gli input e gli output della decisione algoritmica.

Quanto asserito da tale decisione ha consentito agli autori che si sono occupati del tema di rilevare molteplici criticità. Innanzitutto, l'uso di detti algoritmi nell'attività giudiziaria penale crea problemi di compatibilità con l'art. 3 della Costituzione italiana in cui si afferma il principio di eguaglianza dei cittadini dinanzi alla legge. Ma è anche in contrasto con il diritto penale liberale del fatto fondato sulla materialità ed offensività del reato e sulla responsabilità penale come personale[52]. Usare generalizzazioni statistiche da applicare al singolo caso sembra poter condurre a forme di criminalizzazione del tipo d'autore, laddove la pericolosità sociale è desunta da particolari caratteristiche o dall'appartenenza ad una determinata comunità. Vi sarebbe un ritorno non consentito al diritto penale Lombrosiano[53]. Inoltre, la circostanza per cui la pena venga misurata in base ai comportamenti di una comunità di appartenenza frusta ogni basilare esigenza di individualizzazione del trattamento

[52] Sul diritto penale del fatto si vedano, nella dottrina italiana, G. DELITALA, *Il "fatto" nella teoria generale del reato*, Cedam, 1930; F. BRICOLA, *Teoria generale del reato*, in Nov. dig. it., XIX, 1973; G. MARINUCCI, *Soggettivismo e oggettivismo nel diritto penale. Uno schizzo dogmatico e politico-criminale*, in *Riv. it. dir. proc. pen*., 2011, p. 1 ss.

[53] La più attenta dottrina si è preoccupata di segnalare tale pericolo: si vedano, tra gli altri, V. MANES, L'oracolo algoritmico e la giustizia penale, cit., p. 13-14; F. SGUBBI, *Il diritto penale totale. Punire senza legge, senza verità, senza colpa*, Il Mulino, 2019, p. 41; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 21; A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 12-13; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, in C. Piergallini e al. (a cura di), Studi in onore di Carlo Enrico Paliero, vol. 1 – Politica criminale e teoria della pena, Giuffrè, 2023, p. 79; L. D'AGOSTINO, *Gli algoritmi predittivi per la commisurazione della pena*, cit., p. 364; per un esplicito riferimento al determinismo di stampo lombrosiano vd. A. GIANNINI, *Lombroso 2.0: On AI and Predictions of Dangerousness in Criminal Justice, in International Review of Penal Law*, 2021, p. 179.

sanzionatorio, richiesta e sancita dall'art. 27 della Costituzione[54].

Un'ulteriore lesione, strettamente processuale, concerne il diritto ad una difesa effettiva ed al giusto processo, sanciti dagli articoli 24 e 111 della Costituzione. Infatti, i *risk assesment tools*, stante la loro opacità, non sono pienamente controllabili: non è possibile spiegare il risultato del procedimento algoritmico e, dunque, non è possibile contestarne l'attendibilità. Ciò conduce alla violazione del principio basilare dell'*égalité des armes* [55].

Inoltre, l'uso di informazioni personali a scopi predittivi pare nettamente in contrasto con il principio del contraddittorio nella formazione della prova e con il diritto al silenzio dell'imputato, in aperto contrasto con il principio del *nemo tenetur se detegere*[56].

Nell'ambito dell'Unione Europea, fortunatamente, sussiste l'art. 22 del Regolamento Europeo in materia di Protezione dei Dati Personali (GDPR) che impedisce l'integrale sostituzione del giudice essere umano con l'algoritmo poiché «L'interessato ha il diritto a non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona»[57]. Pertanto, detti sistemi in Europa non possono far altro che fungere da ausiliari del giudice.

L'utilizzo di tali sistemi nel settore penale deve essere incoraggiato, a condizione che questi siano trasparenti, accessibili, comprensibili e rispettosi dei fondamentali diritti

---

[54] V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 13-14; M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 21; A.M. MAUGERI, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 14.

[55] S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?,* cit., p. 1761; V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 15. Manes paragona la *black box decision* scaturente dal sistema algoritmico ad una decisione priva di motivazione o con motivazione meramente apparente, contrastante con l'art. 24 Cost. e con l'art. 111, co. 6 Cost.; vd. ibidem, p. 14-15.

[56] V. MANES, *L'oracolo algoritmico e la giustizia penale*, cit., p. 18.

[57] Si noti che l'art. 22, comma secondo prevede deroghe. Sul tema vd., con alcuni accenni critici, M. GIALUZ, Quando la giustizia penale incontra l'intelligenza artificiale, cit., p. 16.

europei, primo tra tutti il divieto di discriminazione. Su questa scia si collocano sia la *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*, emessa dalla Commissione europea per l'efficienza della giustizia del Consiglio d'Europa (CEPEJ) nel 2021, sia la *Risoluzione sull'intelligenza artificiale nel diritto penale ed il suo utilizzo da parte delle autorità di polizia e giudiziaria in ambito penale*, approvata dal Parlamento europeo nel 2021.

Nella Carta Etica approvata dalla CEPEJ sono enunciati dei principi importanti, tra cui il rispetto dei diritti fondamentali, la non discriminazione, la qualità e sicurezza dei dati, la trasparenza, l'imparzialità e la correttezza, nonché la garanzia del controllo umano[58].

In Italia, occorre sottolineare un altro ostacolo normativo all'uso degli algoritmi di giustizia predittiva: si tratta del divieto di perizia criminologica di cui all'articolo 220, comma secondo, del codice di procedura penale[59].

2.1. I sistemi di IA: autori del reato o mero strumento nelle mani dell'uomo?

Dopo aver senza pretese di esaustività tentato di spiegare cosa si intenda per IA, occorre adesso delimitare il tema della nostra ricerca. Noi desideriamo occuparci esclusivamente dei casi in cui l'utilizzo di un sistema di IA dotato di autonomia, da solo o con la partecipazione dell'essere umano, causi dei fatti penalmente rilevanti.

Non pare creare particolari difficoltà l'ipotesi in cui il sistema di IA costituisca un mero strumento nelle mani

[58] Per l'analisi di questi principi si rimanda a S. QUATTROCOLO, *Intelligenza artificiale e giustizia*, cit., p. 4 ss.; D. PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi*, cit., p. 87 ss.; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 81 ss.

[59] S. QUATTROCOLO, *Quesiti nuovi e soluzioni antiche?*, cit., p. 203; contra M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 20, che rileva come i *risk assessment tools* non sembrano finalizzati specificamente ad accertare «il carattere e la personalità dell'imputato e in genere le qualità psichiche indipendenti da cause patologiche» (art. 220, co. 2, c.p.p.)

dell'essere umano che si è dolosamente servito di esso[60]: in questi casi, risponderà normalmente a titolo di dolo l'autore del reato. A ben vedere, potrebbero essere due le ipotesi di IA come mero strumento del reato. Una prima vede il sistema come appositamente designato per commettere specifici reati. Basti pensare ai *software* usati per le truffe *online*. In casi come questo, potrebbe essere sanzionata a monte la creazione del sistema, posto che tutti i suoi possibili utilizzi sarebbero delittuosi[61]. L'altra ipotesi è quella degli strumenti di per sé leciti, ma utilizzati per il perseguimento di scopi illeciti. Si pensi alla *self-driving car* usata come corriere della droga o come autobomba per attacchi terroristici[62]. In ipotesi così semplici, sono sufficienti i reati già esistenti, nonché gli strumenti reattivi dell'ordinamento: infatti il sistema di IA si atteggia in questi casi a semplice cosa che servì o fu destinata a commettere il reato e, pertanto, sottoponibile a confisca penale di cui all'articolo 240, comma 1, del codice penale[63].

In altre ipotesi, l'uso illecito di sistemi di IA di per sé leciti sembrerebbe rendere necessario l'intervento penale. Si pensi

---

[60] S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame?*, cit., p. 425; A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 656-657; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, ivi, p. 713-714; L. PICOTTI, *Cybercrimes and Criminal Relevance of Artificial Intelligence*, cit., p. 72; G.P. DEMURO*, Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 69; S. MASSI, *Affidamento sull'intelligenza artificiale e "disimpegno morale" nella definizione dei presupposti della responsabilità penale*, ivi, p. 665 ss.; A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 7; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 100; S. RIONDATO, *Robotica e diritto penale (robots, ibridi, chimere e "animali tecnologici"*, in D. Provolo e al. (a cura di), *Genetics, Robotics, Law, Punishment, Padova University Press*, p. 601-602; M.B. MAGRO*, Robot, cyborg e intelligenze artificiali*, cit., p. 1206-1207; S. AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, in A.M. Rodrigues (coord.), *Inteligência artificial no direito penal, 2020, Almedina, p. 71-72; D. LIMA, Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challanges for Criminal Law*, in *South Carolina Law Review*, 2018, p. 690; F. BASILE, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, cit., p. 13.

[61] A. CAPPELLINI, *Machina delinquere non potest?*, cit., p. 7, nota 16.

[62] I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 100.

[63] Così S. RIONDATO, Robot: talune implicazioni di diritto penale, in P. Moro – C. Sarra (a cura di), *Tecnodiritto. Temi e problemi di informatica e robotica giuridica*, Franco Angeli, 2017, p. 85.

alle immagini manipolate che potrebbero essere diffuse a fini estorsivi ovvero a fini distorsivi del consenso elettorale (cd. *deepfakes*). In questi casi, comunque, non rileverebbe una eventuale divergenza tra la volontà causale dell'agente e l'evento verificatosi in concreto poiché, ugualmente, saremmo di fronte ad una responsabilità dolosa[64]. Ciò in quanto, ai fini della responsabilità a titolo di dolo, è sufficiente che l'autore del reato abbia ritenuto che la sua azione potesse causare in concreto quell'evento, non essendo rilevante che l'autore stesso abbia previsto nei minimi dettagli il decorso causale ovvero che ne avesse previsto un altro differente[65].

Pertanto, se un drone dovesse colpire non la casa della persona che l'autore voleva uccidere, ma un edificio adiacente, l'utilizzatore risponderà ugualmente a titolo di omicidio doloso, se l'incendio provocato dovesse causare la morte della persona designata[66].

I casi realmente complicati sono quelli in cui il sistema di IA cessa di essere strumento in mano all'essere umano poiché realizza un reato diverso da quello programmato dall'utilizzatore oppure, avendo avuto istruzioni generiche dal programmatore, il sistema scelga quando e come commettere il fatto: qui è difficile continuare a ritenere che il sistema di IA sia un mero strumento nelle mani dell'essere umano.

3. Delimitazione della ricerca: verso una nozione "penalisticamente orientata" di IA.

Per come annunciato nell'introduzione, a noi interessa condurre una ricerca limitatamente alle ipotesi in cui il fenomeno dell'IA ha un impatto dirompente rispetto ai

[64] I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 101; A. CAPPELLINI, *Machina delinquere non potest?,* cit., p. 7-8; contra S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 716-717.

[65] G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 405.

[66] L'esempio è di I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 101.

tradizionali metodi di imputazione della responsabilità criminale. Dunque, non sono oggetto del nostro studio tutti quei sistemi di IA che fungono da mero strumento in mano all'essere umano e che possono essere trattati alla stregua di ogni altro strumento.

Lo studio dell'IA per il penalista è affascinante perché vi sono taluni sistemi connotati dal carattere dell'imprevedibilità. Trattasi dei sistemi di ultimissima generazione: essi non sono soltanto automatici, bensì *autonomi*. Di fatti, è proprio il carattere dell'autonomia decisionale in capo a detti sistemi a destare maggiori grattacapi e preoccupazioni per il penalista[67].

Ebbene, per poter comprendere la dirompenza del fenomeno rispetto ai tradizionali modelli di ascrizione della responsabilità penale, occorre spiegare cosa si intenda per autonomia di un sistema intelligente. Quest'ultima, *prima facie*, potrebbe essere descritta come la capacità di operare nell'ambiente del mondo reale senza alcuna forma di controllo esterno per un esteso periodo di tempo[68]. Stando a tale impostazione, dovrebbero essere considerati autonomi anche i *robot* per le pulizie, poiché in grado di svolgere il compito senza l'intervento umano.

A nostro avviso, invece, ciò che realmente distingue l'autonomia dalla mera automazione è la capacità dei sistemi di prendere decisioni non ricavabili direttamente dagli *input* forniti dai programmatori, ma che sono frutto dell'interazione con gli ulteriori dati provenienti dall'ambiente circostante. Gli algoritmi intelligenti di ultima generazione sono in grado di raggiungere gli obiettivi prefissi

---

[67] Vd., per tutti, L. PICOTTI, *I primi vent'anni della Convenzione di Budapest nell'ottica sostanzialista e la mancata ratifica ed esecuzione del Primo Protocollo addizionale contro il razzismo e la xenofobia*, in Dir. pen. proc., 2022, n. 8, p. 1032; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, in Giur. it., luglio 2019, p. 1717.

[68] G.A. BEKEY, *Current Trends in Robotics: Technology and Ethics, in P. Lin e al. (eds.), Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, 2012, p. 18., la definisce come la capacità «*to operate in the real-world environment without any form of external control, once the machine is activated and at least in some areas of operation, for extended periods of time*».

tramite procedimenti razionali imprevisti e non stabiliti *ex ante*. È irrevocabile in dubbio come non siamo di fronte al campo del libero arbitrio umano, ma che siamo pur sempre nell'ambito del determinismo[69]. Ma, nonostante ciò, vi è chi ritiene che i comportamenti autonomi dei sistemi di IA siano idonei a creare una parvenza di autodeterminazione, posto che dipendono da una enorme quantità di fattori e variabili[70]. Ciò in quanto la decisione presa dai sistemi intelligenti è certamente causata dal tipo di codice matematico ma anche dai dati acquisiti durante l'addestramento e da quelli acquisiti durante la sua "vita" algoritmica: questo conduce all'impossibilità di comprendere sulla base di quali informazioni il sistema abbia assunto la decisione finale.

Ed è proprio la capacità di agire in maniera innovativa, trovando il percorso più efficiente per raggiungere il risultato a costituire il *quid novi* rispetto alle altre tecnologie. Detta capacità costituisce anche l'elemento di successo che l'IA sta raggiungendo in molti settori della vita sociale, tra cui la medicina, la circolazione stradale, la consulenza finanziaria etc.

Per comprendere cosa, a nostro avviso, si debba intendere per autonomia, occorre prendere in considerazione l'esempio che segue. Due ingegneri avevano progettato un algoritmo per il volo di un drone, che consentiva alla macchina di usare

---

[69] Lo sviluppo delle neuroscienze ha messo in parte in crisi questo postulato – che, tuttavia, è irrinunciabile nell'ambito del diritto penale, costituendone il presupposto fondamentale. In argomento vd., per tutti, O. Di Giovine, Ripensare il diritto penale attraverso le (neuro-)scienze?, Giappichelli, 2019; F. BASILE – G. VALLAR, *Diritto penale e neuroscienze, in Giornale Italiano di Psicologia*, n. 4, 2016, p. 799 ss.; C. GRANDI, *Neuroscienze e responsabilità penale: nuove soluzioni per problemi antichi?*, Giappichelli, 2017; i contributi raccolti in S.J. Morse – A.L. Roskies (eds.), *A Primer on Criminal Law and Neuroscience, Oxford University Press*, 2013; e, recentemente, E.D. CRESPO, *"Compatibilismo umanista": una proposta per conciliare le neuroscienze e il diritto penale*, in Riv. it. dir. proc. pen., 2022, fasc. 2, p. 645 ss. Recentemente, sui rapporti tra intelligenza artificiale e neuroscienze vd. i contributi pubblicati nel volume collettaneo J.M. Peris Riera – A. Massaro (a cura di*), Derecho penal, inteligencia artificial y neurociencias. Diritto penale, intelligenza artificiale e neuroscienze*, Roma Tre Press, 2023 e, in particolare, il contributo di A. Massaro, *Intelligenza artificiale e neuroscienze: l'eterno ritorno del diritto penale?,* p. 27 ss.

[70] A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 31; M. SIMMLER – N. MARKWALDER*, Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence*, in Crim. Law Forum, 2019, n. 30, p. 8-9; C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., p. 1769.

il percorso più efficiente per arrivare ad un determinato *target;* avevano inserito questo obiettivo nel codice algoritmo ed iniziato ad addestrare il drone, collocandolo in vari luoghi attorno al *target*: in questo modo, il sistema avrebbe appreso come raggiungere in modo efficiente ed affidabile il *target* previsto. Durante l'addestramento, i due ingegneri avevano rilevato che il drone aveva iniziato ad indirizzare il volo nella direzione completamente opposta rispetto al luogo del *target*. Dopo alcune indagini, era emerso che non si trattava di un errore o di un malfunzionamento poiché il drone aveva imparato che, al ricorrere di determinate circostanze, il modo più rapido ed efficiente per raggiungere il target era volare verso il perimetro esterno dell'area di prova, dove gli ingegneri lo avrebbero recuperato e portato al *target*, così da ripristinare il processo: l'algoritmo, insomma, aveva compreso che essere trasportato dagli ingegneri verso il *target* era il modo più efficiente per raggiungere l'obiettivo[71].

Da tale esempio si comprende come, a ben vedere, l'autonomia si distingue dall'automazione perché i sistemi autonomi hanno la capacità di compensare l'incompletezza delle informazioni ricevute in partenza tramite l'apprendimento[72]. Tali sistemi sono capaci di prendere *decisions under uncertainty*[73].

Stando a quest'ultima considerazione, dunque, l'IA sarebbe autonoma in quanto adattiva, cioè in grado di interagire con l'ambiente circostante, imparare da esso e di

[71] Il caso è riportato in M.A. LEMLEY – B. CASEY, *Remedies for Robots*, cit., p. 1313; gli autori dell'articolo, a loro volta, affermano di aver ripreso l'esempio da una presentazione tenutasi presso la 11th *Annual Stanford Ecommerce Best Practices Conference*, tenutasi nel giugno 2014 (vd. nt. 1 dell'articolo).

[72] S. RUSSELL – P. NORVIG, *Artificial Intelligence*, cit., p. 42: «*To the extents that an agent relies on the prior knowledge of its designer rather than on its own percepts and learning process, we say that the agent lacks autonomy. A rational agent should be autonomous – it should learn what it can to compensate for partial or incorrect prior knowledge*».

[73] A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 37, a cui si rimanda per un'analisi più approfondita del concetto di autonomia, che viene considerata come un costrutto sociale, piuttosto che come un derivato della tecnologia (social attribution of autonomy), vd. Ibidem, p. 31 ss.; vd. anche M. SIMMLER – N. MARKWALDER, *Guilty Robots*?, cit., p. 13.

modificare il proprio comportamento sulla base degli *input* ricevuti[74].

Vista così, l'autonomia è l'elemento che ci spinge a svolgere la ricerca sulle implicazioni che essa ha sulla responsabilità (civile e) penale. Abbiamo ristretto il campo della nostra indagine ai soli sistemi di ML in cui l'*output* non è ricavabile solo dal codice algoritmico ma anche dall'interazione con l'ambiente e dai processi di autoapprendimento[75].

Oltre l'autonomia, ciò che contribuisce a rendere altamente imprevedibile il comportamento di una IA è la sua interattività. Per interattività si intende la circostanza per cui gli algoritmi non operano come isole, ma sono connessi tra di loro e con l'ambiente fisico in cui si trovano. Sarebbe il cd. *internet of things* che lo *European Research Cluster on the Internet of Things* (IERC) definisce come «*a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual "things" have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network*»[76]. Più sistemi intelligenti ed autonomi possono dar vita ad *output* collettivi inaspettati, frutto di interazioni non sempre comprensibili dall'esterno. Così come, gli algoritmi connessi con sensori e

[74] C.E.A. KARNOW, *The application of traditional tort law to embodied machine intelligence*, cit, p. 60; G. SARTOR – A. OMICINI, *The autonomy of technological systems and responsibilities for their use*, in N. Bhuta e al. (eds.), *Autonomous Weapons Systems. Law, Ethics, Policy, Cambridge University Press*, 2016, p. 49

[75] In questo senso anche C.E.A. KARNOW, *The application of traditional tort law to embodied machine intelligence*, cit., p. 52; K.A. CHAGAL-FEFERKORN, *The Reasonable Algorithm*, in *University of Illinois Journal of Law, Technology & Policy*, 2018, n. 1, p. 117; U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product: per i fondamenti di un diritto dell'intelligenza artificiale self-learning*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 96; M. SIMMLER – N. MARKWALDER, *Guilty Robots?*, cit., p. 5; contra A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 37, che sostengono che la categoria dei sistemi di i.a. autonomi sia più ampia di quella dei sistemi di i.a. che utilizzano tecniche di *machine learning*.

[76] Vd. http://www.internet-of-things-research.eu/about_iot.htm

infrastrutture fisiche complesse sono esposti ad una varietà infinita di *inputs*, risultando altamente imprevedibili[77].

L'imprevedibilità è conseguenza dell'autonomia e dell'interattività, dunque. Ma un altro elemento che rende i sistemi di IA così imprevedibili è la loro opacità tecnologica. L'opacità è la conseguenza da pagare per avere sistemi intelligenti così performanti: all'aumentare della precisione ed accuratezza diminuisce la capacità per gli stessi produttori del sistema di comprendere come e perché esso agisca in un determinato modo. Gli algoritmi individuano correlazioni invisibili all'occhio umano, senza essere però in grado di ricostruire la spiegazione causale dei risultati raggiunti.

Questo fenomeno è definito *black box*: è possibile individuare *input* ed *output* ma non è possibile ricostruire cosa accade all'interno della scatola nera, ovverosia la catena causale che ha determinato il passaggio dagli *input* agli *output*[78]. Vi è chi ritiene che la *block box* sia voluta proprio dai produttori per evitare che i sistemi di IA vengano aggrediti da *hackers* o da terroristi[79] ovvero per proteggere il segreto industriale[80]. Visto così il fenomeno, sarebbe possibile promuovere schemi di elaborazione delle informazioni algoritmiche basati sulla trasparenza: questo obiettivo è alla base delle teorie dell'*interpretability* e dell'*explainability*, che

---

[77] L. Longhitano, *L'intelligenza artificiale di Facebook parla una lingua incomprensibile, ma niente panico*, in Wired Italia, 1 agosto 2017, disponibile a questo link: https://www.wired.it/gadget/computer/2017/08/01/intelligenza-artificiale-facebook/; A. Griffin, *Facebook's Artificial Intelligence Robots Shut Down After They Start Talking to Each Other in Their Own Language*, in Independent, 31 luglio 2017, disponibile a questo link: https://www.independent.co.uk/life-style/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html.

[78] Y. BATHAEE, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, 2018, n. 2, p. 905; D. CASTELVECCHI, *Can We Open the Black Box of AI?,* in Nature, 5 ottobre 2016, disponibile a questo link: https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731; il problema è ben conosciuto anche nelle sedi europee, vd. COMMISSIONE EUROPEA, *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità*, COM(2020) 64 final, 19 febbraio 2020, p. 10.

[79] J.A. KROLL E AL., *Accountable Algorithms*, in 165 U. Pa. L. Rev., 2017, p. 639

[80] vd. J. BURRELL, *How the machine 'thinks'*, cit., p. 3-4; F. PASQUALE, *The Black Box Society*, cit., passim; A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1091-1092.

tentano di sviluppare modelli di calcolo grazie ai quali si possa comprendere il rapporto causa-effetto[81]. Questa tecnologia si chiama *white-box*. L'obiettivo di quest'ultima è eliminare l'opacità algoritmica, ricostruendo la trasparenza dell'agire algoritmico.

D'altronde, tutti i testi di *soft-law* e di *hard-law* che si sono occupati di IA richiedono che i sistemi siano trasparenti e che il procedimento di *decision making* sia quanto più possibile tracciabile e ricostruibile. A titolo esemplificativo, l'art. 13 dell'*AI Act*, rubricato Trasparenza e fornitura di informazioni ai *deployer*, *<< 1.I sistemi di IA ad alto rischio sono progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire ai deployer di interpretare l'output del sistema e utilizzarlo adeguatamente. Sono garantiti un tipo e un livello di trasparenza adeguati, che consentano di conseguire il rispetto dei pertinenti obblighi del fornitore e del deployer di cui alla sezione 3; 2.I sistemi di IA ad alto rischio sono accompagnati da istruzioni per l'uso, in un formato appropriato digitale o non digitale, che comprendono informazioni concise, complete, corrette e chiare che siano pertinenti, accessibili e comprensibili per i deployer>>.*

Ma anche gli orientamenti etici per un'IA affidabile pubblicati nel 2019 dal Gruppo di esperti nominato dalla Commissione europea richiedono che <<*i set di dati ed i processi che determinano la decisione del sistema di IA, compresi quelli di raccolta ed etichettatura dei dati, come pure gli algoritmi utilizzati, dovrebbero essere documentati secondo i migliori standard per consentire la tracciabilità ed aumentare la trasparenza*>>[82]. Ma, a quanto pare, vi sono sistemi in cui

[81] La letteratura in materia è davvero amplissima. Dal punto di vista tecnico si vd. per tutti R. GUIDOTTI E AL., *A Survey of Methods for Explaining Black Box Models*, in *ACM Computing Surveys*, 2018, p. 1 ss.; A. D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1110 ss.; dal punto di vista normativo, vd. T. WISCHMEYER, *Artificial Intelligence and Transparency: Opening the Black Box*, in *T. Wischmeyer – T. Rademacher (eds.), Regulating Artificial Intelligence*, Cham, 2020, p. 75 ss.; per una ricostruzione sistematica dei principali lavori in materia di *Explainable Artificial Intelligence* vd. F. SOVRANO E AL., *Metrics, Explainability and the European AI Act Proposal*, in J, 2022, n. 5, p. 126 ss.

[82] COMMISSIONE EUROPEA – GRUPPO INDIPENDENTE DI ESPERTI AD ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, Orientamenti etici per un'i.a. affidabile, 8 aprile 2019, § 76 ss. (pag. 20).

l'opacità è una caratteristica intrinseca ed ineliminabile, connaturata al loro funzionamento. Nella specie, quando gli algoritmi utilizzano le reti neurali artificiali sono opaci *in re ipsa*.

Sarebbero due i fattori che causano questa opacità inevitabile: *in primis*, i *pattern* decisionali potrebbero essere troppo complicati poiché basati su una eccessiva mole di dati tra i quali è impossibile individuare i fattori dirimenti per la decisione[83], poiché trattasi di algoritmi altamente complessi[84] ed *inscrutable*[85]; *in secundis*, specie quando si tratta di sistemi di DL, l'opacità deriva proprio dalla discrepanza tra processo logico-computazionale seguito dalla macchina e struttura semantica propria del linguaggio umano[86]. Breve: l'algoritmo non riesce a spiegarsi con un linguaggio comprensibile all'essere umano con la conseguenza che le disposizioni normative che prescrivono obblighi di trasparenza rimangano lettera morta poiché *ad impossibilia nemo tenetur*[87].

Pare, dunque, confermarsi la regola dell'inversa proporzionalità tra la maggiore accuratezza predittiva e la trasparenza: più è accurato il sistema, meno sarà trasparente[88]. L'opacità algoritmica sembrerebbe essere solo

---

[83] C. REED E AL., *Responsibility, Autonomy and Accountability: legal liability for machine learning, Queen Mary School of Law Legal Studies Research Paper No. 243/2016*, p. 13; J. BURRELL, *How the machine 'thinks'*, cit., p. 2; W.N. PRICE II, *Medical Malpractice and Black-Box Medicine*, in I.G. Cohen e al. (eds.), *Big Data, Health Law, and Bioethics, Cambridge University Press*, 2018, p. 298.

[84] Y. BATHAEE, *The Artificial Intelligence Black Box*, cit., p. 901.

[85] A.D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1094; vd. anche J. BURRELL, *How the machine 'thinks'*, cit., p. 4-5.

[86] C. REED E AL., *Responsibility, Autonomy and Accountability*, cit., p. 13, J. BURRELL, *How the machine 'thinks'*, cit., p. 2; W.N. PRICE II, *Medical Malpractice and Black-Box Medicine*, cit., p. 298.

[87] U. RUFFOLO – A. AMIDEI, *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design*, cit., p. 492; A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, cit,, p. 290; E.M. COTUGNO, *Intelligenza artificiale e tutela del consumatore. I nuovi obblighi informativi*, cit., p. 319; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale*, cit., p. 16.

[88] M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, in M. Ebers – S. Navas (eds.), *Algorithms and Law, Cambridge University Press*, 2020, p. 50. Sulla differenza tra *Weak* e *Strong Black Boxes* vd. anche Y. BATHAEE, *The Artificial Intelligence Black Box*, cit., p. 906; sul *trade off* tra accuratezza delle previsioni algoritmi e spiegabilità vd. anche A.J. LONDON, *Artificial intelligence and black-box medical decisions: accuracy versus explainability, in The*

l'ultima tappa della crisi di comprensione e controllo ha colpito l'uomo nel rapporto con la tecnologia.

Moltissimi studiosi si sono concentrati sul rapporto tra l'essere umano e la società tecnologica. Vi è chi ha asserito che il progetto illuministico di dominio della natura tramite la tecnica si fosse risolto nell'esatto opposto, ovverosia nell'assoggettamento dell'umanità alle esigenze della stessa macchina produttiva, divenuta ormai una seconda natura che lo riporta nelle stesse condizioni di smarrimento e paura che vivevano i nostri avi agli inizi del percorso di civilizzazione[89]. Successivamente, la tecnica è stata vista come forza autonoma in grado di raggiungere risultati non preventivabili né controllabili dall'essere umano[90]. Ma la pietra miliare sul tema è rappresentata dalla pubblicazione di La società del rischio ad opera di Ulrich Beck[91]. La tesi di Beck è che le tecniche di produzione moderna creino rischi che la società non è in grado di controllare: i progressi permettono di affrontare pericoli precedentemente impossibili da affrontare, ma, al contempo, creano nuovi rischi, sempre più sfuggenti[92].

La tecnica, dunque, da strumento di comprensione e dominio sulla forza cieca e schiacciante della natura, è

Hastings Center Report, 49(1), 2019, p. 15 ss.; R. CARUANA E AL., *Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-day Readmission*, in *Proceedings* of the 21st ACM SIGKDD *International Conference on Knowledge Discovery and Data Mining*, 2015; D.F. LLORCA E AL., *Liability regimes in the age of AI: a use-case driven analysis of the burden of proof*, cit., p. 8; contra C. RUDIN, *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, in 1 Nat Mach Intell., 2019, p. 208: «It is a myth that there is necessarily a trade-off between accuracy and interpretability»*.

[89]W. PRIVITERA, *Tecnica, individuo e società: cinque lezioni sulla teoria di Ulrich Beck*, Rubbettino, 2004, p. 20. Sulla tecnologia come "natura" del futuro vd. il recente K. VAN MENSVOOR, NextNature, D Editore, 2022.

[90] M. HEIDEGGER*, La questione della tecnica*, in Id., *Saggi e discorsi* (a cura di G. Vattimo), Mursia, 1976, p. 5-27 (ed. ted. 1957).

[91] U. BECK, *La società del rischio. Verso una seconda modernità*, Carocci, 2000 (ed. ted. 1986).

[92] L'opera di Ulrich Beck ha avuto influenza enorme anche in ambito giuridico. Si vd., per restare in ambito penalistico, J.M. SILVA SANCHEZ*, L' espansione del diritto penale. Aspetti della politica criminale nelle società postindustriali*, Giuffrè, 2004 (ed. spagn. 1999); F. HERZOG, *Società del rischio, diritto penale del rischio, regolazione del rischio*, in L. Stortoni – L. Foffani (a cura di*), Critica e giustificazione del diritto penale nel cambio di secolo. L'analisi critica della Scuola di Francoforte*, Giuffrè, 2004, p. 357; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit.

diventata essa stessa elemento misterioso, imperscrutabile, incontrollabile.

In quest'ottica, i sistemi di IA sembrano ripresentare vecchi problemi della società del rischio. L'*aliquid novi*, per come si è avuto modo di capire, è rappresentato dall'autonomia, dall'imprevedibilità, dall'interattività e dall'opacità. Siamo dinanzi ad una nuova fase del rapporto tra evoluzione tecnologica ed attività umana perché qui l'imprevedibilità non è un errore, ma è proprio il risultato desiderato al fine di raggiungere i risultati più efficienti: si dice che il sistema di IA sia *unpredictable by design*[93].

Dunque, nonostante non sappiamo bene perché e come l'IA sia arrivata in modo efficiente a quel risultato, ci affidiamo ad essa come fosse un oracolo[94].

4. Definizioni normative (e giurisprudenziali) dei sistemi di IA.

Dopo aver delimitato il campo di ricerca ai soli sistemi di IA che godono di autonomia e, dunque, di imprevedibilità, occorre adesso prendere in esame taluni testi normativi in cui si è tentato di dare una definizione dell'IA. Anche se siamo consapevoli che le definizioni legislative di un fenomeno in costante evoluzione non potranno mai essere in grado di circoscrivere l'intero fenomeno di cui ci occupiamo, è obbligatorio per noi giuristi prenderle in considerazione e ricavare spunti riflessivi da esse. Inoltre, le definizioni servono per delimitare l'ambito di applicazione di una normativa[95].

---

[93] R. CALO, Robotics and the Lessons of Cyberlaw, cit., p. 542; J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit., p. 107; vd. anche le riflessioni svolte, quando ancora non erano state sviluppate le tecniche di *machine learning*, da C.E.A. KARNOW, *Liability for Distributed Artificial Intelligence*, cit., p. 192.

[94] V. MANES, *L'oracolo algoritmico*, op. cit.

[95] A. BELVEDERE, (voce) *Definizioni*, in *Digesto*, IV ed., vol. V, Torino, 1989, p. 149 ss.; ID., I*l problema delle definizioni nel codice civile*, Giuffrè, 1977; in ambito penalistico vd. A. Cadoppi (a cura di), I*l problema delle definizioni legali nel diritto penale. Ominis definitio in iure periculosa?*, Cedam, 1996; proprio con riferimento ai sistemi di i.a., e, più precisamente, alle macchine a guida autonoma, vd. E. HILGENDORF, *Automated Driving and the Law*, in E. Hilgendorf – U. Seidel (eds.), *Robotics, Autonomics, and the Law*, p. 172.

4.1. Breve cronistoria sul cd. *AI Act*. Definizione europea di intelligenza artificiale.

L'Unione Europea ha voluto avere un ruolo di primo piano nella creazione di definizioni e regole inerenti al tema dell'IA. Di fatti, prima di arrivare alla pubblicazione sulla Gazzetta Ufficiale dell'Unione Europea del testo finale del Regolamento europeo sull'IA, il percorso è stato lungo e tribolato. Daremo, per tale ragione, brevemente conto della cronistoria.

Ad aprile 2018, la Commissione accoglieva l'invito del Consiglio Europeo per la definizione di una linea comune sull'IA, ricevendo l'invito ad avere un approccio che ponesse i cittadini al centro dello sviluppo dell'IA ed incoraggiasse l'uso di questa tecnologia per aiutare a risolvere le maggiori sfide mondiali, come la cura delle malattie, la lotta per il cambiamento climatico, la sicurezza dei mezzi di trasporto, il contrasto alla criminalità ed il miglioramento della sicurezza informatica.

Veniva, in quell'anno, sottoscritta una Dichiarazione di Cooperazione sull'IA da 25 paesi.

Il Piano Coordinato per l'IA fu presentato il 7 dicembre 2018, delineando la cornice all'interno della quale gli Stati membri avrebbero inserito le proprie iniziative nazionali. L'obiettivo era quello di fornire una tecnologia affidabile che rassicurasse e preparasse la società all'impatto con l'IA; stimolasse la ricerca; sviluppasse linee guida etiche ed un ambiente normativo favorevole alla innovazione; facesse meglio comprendere gli aspetti legati alla sicurezza interna ed agli attacchi esterni alle infrastrutture; favorisse la crescita e lo sviluppo di competenze e talenti. Era sin da allora chiaro il dilemma: da un lato evidenziare e combattere i rischi che questa nuova tecnologia poneva anche sul piano etico e della tutela dei diritti; dall'altra la necessità di non comprimere la ricerca, favorendo invece una partnership pubblico-privata e lo sviluppo di *startup* e piccole e medie imprese innovative.

Nell'aprile 2018, fu istituita la *High Level Expert Group on Artificial Intelligence* (*AI Hleg*)[96]. Il compito affidato a detto organismo si collocava nel solco del Piano Coordinato ed avrebbe rappresentato una costante della posizione europea: definire da una parte "Linee Guida Etiche" di condotta, e dall'altra dare indicazioni sulla politica, sulla normativa, sulla ricerca e sugli investimenti per l'IA. Una relazione del Gruppo venne presentata nell'aprile 2019, poi seguita da ulteriori approfondimenti.

Nel giugno 2018, viene promossa la *AI Alliance*, un *forum* di discussione del quale avrebbero potuto fare parte, a richiesta, rappresentanti della società civile, delle aziende, dei sindacati, dell'Accademia ed esperti indipendenti.

Nel dicembre 2018, viene istituito l'*AI Watch*, un sistema di monitoraggio sullo sviluppo, sull'applicazione e sull'impatto dell'IA in Europa.

Tornando all'*AI Hleg*, nell'aprile 2019, viene presentato un primo rapporto che indica quattro principi etici ed avanza sette richieste per i sistemi di IA. I principi etici sono il rispetto per l'autonomia umana, la prevenzione dei danni, la correttezza e la "spiegabilità". Da questi le sette richieste: intervento e controllo umano; affidabilità e robustezza tecnica; controllo dei dati e della *privacy*, trasparenza, rispetto della diversità, non discriminazione e correttezza; benessere sociale ed ambientale; responsabilità[97].

Il documento più significativo della Commissione che ha preceduto la proposta di Regolamento è il c.d. Libro Bianco sull'IA del 19 febbraio 2020: "Un approccio europeo per l'eccellenza e la fiducia". Il documento si colloca all'interno dei quattro principi e delle sette richieste dell'*AI Hleg*. Il tentativo è di assicurare i diritti ed i valori unitamente al progresso scientifico e tecnologico. Uno degli aspetti più

[96] Trattasi di 52 esperti designati dalla Commissione tra rappresentanti dell'Accademia, della società civile e dell'industria per coadiuvare lo sviluppo di una strategia europea per l'IA.
[97] N.A. SMUHA, *"The EU Approach to Ethical Guidelines for Trustworthy AI*", in *Computer Law Review International*, 2019.

significativi del Libro Bianco è la distinzione tra rischi elevati (*high risk*) che reclamano una normativa severa e nuova, ed altri rischi considerati minori per i quali si ipotizzano semplici adeguamenti alla legislazione già esistente.

Accanto alle proposte di tutela viene richiamata la necessità di misure che favoriscano la ricerca, lo sviluppo, la diffusione ed il controllo dei sistemi di IA. Sono due aspetti complementari perché senza una IA affidabile e non lesiva dei diritti, non potranno esserci tecnologie credibili accettate dai consumatori e dal mercato[98].

Di grande interesse è poi un ulteriore documento della Commissione, l'*Inception Impact Assesment* del 23 luglio 2020, che fa il punto sulle iniziative proposte e sui loro prevedibili effetti. Di grande rilievo sono le quattro opzioni indicate nel documento: la prima (opzione 0) non prevede appunto un intervento legislativo, ma una serie di misure di *soft law* per agevolare lo sviluppo delle aziende; la seconda fa invece riferimento ad una legislazione che introduca un sistema di "etichettatura" volontaria da parte dei produttori al fine di rassicurare i consumatori sull'affidabilità dei prodotti; la terza prevede norme vincolanti, in particolare per i prodotti *high risk* che facciano uso di IA, con un grado crescente di prescrittività; l'ultima considera ulteriori, possibili ipotesi intermedie, tenendo però anche qui conto dei livelli di rischio.

Il Presidente della Commissione Europea, Ursula von der Leyen, sia nel discorso inaugurale del novembre 2019 sia nell'ultimo intervento sullo Stato dell'Unione del settembre 2023, ha più volte richiamato la necessità di una forte presenza dell'Unione stessa su tutti gli aspetti che coinvolgono questo tema.

Nell'ottobre 2020, in quello che viene definito "*October Framework on AI*", una serie di risoluzioni sono state approvate dal Parlamento sui principi etici, sulle

[98] V. LITVINETS, "*A Summary of the European Commission White Paper on Artificial Intelligence*" in *medium.com*, 9 giugno 2020.

responsabilità nell'uso delle nuove tecnologie, sulla proprietà intellettuale[99]. Di particolare interesse, il tema dell'uso militare dell'IA: si ribadisce la necessità di un controllo umano sui sistemi militari che la utilizzino ed un bando per le armi letali autonome che facciano uso di IA.

E dell'IA si sono occupati anche il Consiglio Europeo nel 2017 e nel 2020, con riferimento in particolare all'applicazione dell'IA ai sistemi capaci di produrre rischi elevati; ed il Consiglio dell'Unione Europea che, nelle sue Conclusioni sul Piano Coordinato per lo sviluppo e l'uso dell'IA del 21 ottobre 2020, sottolinea l'esigenza di rendere compatibili i sistemi di IA con i diritti fondamentali.

Venendo alla proposta di Regolamento sull'IA della Commissione Europea, il 21 aprile 2021, quest'ultima ha presentato al Parlamento ed al Consiglio una proposta di normativa per l'IA. L'intenzione era, così come per il *General Data Protection Regulation* (GDPR) sul trattamento dei dati personali e sulla *privacy*, non solo di creare una disciplina europea ma di costruire un *benchmark* per tutti i paesi interessati ad accedere al mercato dell'UE e tenuti dunque a rispettarne le regole.

Nelle linee generali, la proposta si rivolge ai fornitori (*providers*) che collocano sul mercato o forniscono nell'UE servizi di IA, indipendentemente dal fatto di essere o meno registrati nell'UE stessa; agli utilizzatori (*users*) di sistemi di IA all'interno dell'UE; ai fornitori ed agli utilizzatori di sistemi di IA posti in un paese terzo, laddove il prodotto venga utilizzato nell'UE. Così come per il GDPR, le norme devono avere un carattere di extraterritorialità. Il principio ispiratore è quello dei rischi: più alti i rischi ritenuti, più severe le norme, ad iniziare da quelle che proibiscono modelli e pratiche che pongono rischi ritenuti inaccettabili. Proibite le pratiche che usano tecniche subliminali (*dark patterns*) per distorcere il comportamento di persone provocando danni fisici e psicologici; che sfruttino condizioni

[99] Risoluzione del Parlamento Europeo del 20 ottobre 2020.

di vulnerabilità in relazione all'età o alla disabilità fisica o mentale; che siano utilizzate da pubbliche autorità per valutare, giudicare e trattare un soggetto in base a caratteristiche personali o sociali (*social scoring*). È poi vietato l'uso ai fini di polizia di pratiche biometriche di identificazione (*facial recognition*) in luoghi accessibili al pubblico, salvo che per ricerche mirate di vittime di un crimine, per prevenire attentati ed attacchi terroristici, per scoprire, localizzare, identificare o fermare un soggetto accusato di gravi crimini. In dette situazioni, andranno comunque preventivamente valutate la situazione da affrontare, le misure ritenute necessarie, le conseguenze delle misure stesse su diritti e libertà, e sarà necessaria una preventiva autorizzazione di una autorità giudiziaria o di una autorità amministrativa indipendente.

Per quanto concerne i rischi elevati (*high risk*) l'articolo 6 del Titolo III identifica i sistemi di IA che possono creare appunto un grave rischio per la salute, la sicurezza ed i diritti naturali delle persone. Ci saranno tali rischi quando concorrono due condizioni: che i sistemi in esame vengano usati in prodotti disciplinati dalle direttive dell'UE indicate nell'allegato II o in una specifica precedente legislazione (ad es. macchinari, giocattoli e mezzi personali di protezione, automobili, etc.) e che si tratti di sistemi autosufficienti che funzionano in maniera indipendente da altre unità (*stand alone systems*), la cui componente di sicurezza sia sottoposta ad una preventiva valutazione di conformità da parte di un soggetto terzo. Saranno considerati sempre a rischio elevato i sistemi capaci di influenzare i diritti fondamentali indicati nell'Allegato III. Tra questi rientrano certamente quelli di identificazione biometrica e di profilazione; di educazione e formazione; di reclutamento e gestione del personale; di produzione e godimento di servizi pubblici essenziali (come acqua, gas, elettricità, controllo del traffico); di ordine pubblico; di emigrazione, asilo e controllo delle frontiere; di amministrazione della giustizia; quelli legati al processo democratico.

Venendo alle prescrizioni che i sistemi ad alto rischio devono rispettare, bisogna intanto ricordare che diversi requisiti sono già soddisfatti su base volontaria dai diversi operatori, anche alla luce delle indicazioni dell'*AI Hleg* prima richiamate. Con riferimento ai *providers*, essi dovranno sottoporre i sistemi in esame ad una validazione di conformità; adottare adeguati modelli di gestione dei rischi; prevedere processi continui di alimentazione e valutazione (*training* e *testing*); rendere possibile comprendere i limiti e le capacità dei modelli; mantenere un'aggiornata documentazione tecnica; disegnare e sviluppare i sistemi per garantire la massima trasparenza ed un effettivo controllo umano anche per prevenire rischi alla salute, alla sicurezza o ai diritti fondamentali; registrare i sistemi stessi in un *database* dell'Unione ed indicare un rappresentante ai fini della *accountability*. Per assicurare la trasparenza, sarà necessario che le persone che interagiscono con sistemi a rischio elevato siano informate di trovarsi di fronte ad un prodotto basato sull'IA. La validazione per i modelli a rischio elevato è già disciplinata dalle norme vigenti sulla sicurezza dei prodotti, ad eccezione dei sistemi per l'identificazione biometrica remota o per i *network* di infrastrutture pubbliche, per i quali è prevista la validazione di un soggetto terzo, il *provider* potrà procedere ad una auto-valutazione ma dovrà registrare il modello in un *database* pubblico dei "Sistemi di IA a rischio elevato".

Gli articoli dal 26 al 29 della proposta di Regolamento disciplinano le obbligazioni per gli importatori, i distributori e gli utilizzatori dei sistemi a rischio elevato[100]. Al fine di facilitare la ricerca, potranno essere creati degli ambienti di prova regolati (*sandboxes*) per favorire lo sviluppo e la validazione di modelli innovativi (artt. 53 e 54).

Per quanto concerne la governance, l'art. 56 della proposta istituisce un Consiglio europeo per l'intelligenza artificiale (*European Artificial Intelligence Board*), per il coordinamento

[100] MORRISON e FORESTER, *Draft EU Regolation for Artificial Intelligence*, in *Lexology*, 11 giugno 2021, 4.

delle politiche nazionali ed una uniforme applicazione della normativa. Tale *Board* sarà composto dai responsabili delle autorità nazionali di vigilanza e dal Garante europeo della protezione dei dati e sarà presieduto dalla Commissione. Gli Stati membri dovranno istituire autorità nazionali per facilitare una conforme applicazione delle norme del Regolamento e la sorveglianza del mercato. Le autorità nazionali di sorveglianza avranno un potere di controllo e dei requisiti per i sistemi a rischio elevato, con diritto all'accesso documentale e con la possibilità, se necessario, di effettuare test ed ispezioni.

È disciplinato un apparato sanzionatorio che prevede, ad esempio, per quanto riguarda la violazione delle prescrizioni relative alle pratiche proibite, multe fino a 30 milioni o, se si tratta di una società, di una somma fino al 60% del totale annuo dei ricavi della stessa.

La Proposta di Regolamento è stata accompagnata da critiche che valgono, in parte, anche per il testo finale. Una prima critica, comune sia alla Proposta sia al Regolamento finale seguito all'accordo dopo il trilogo tra Commissione, Parlamento e Consiglio riguarda l'imprecisione del linguaggio utilizzato. A volte il linguaggio sarebbe ambiguo, con espressioni e riferimenti che rischiano di mettere in difficoltà sia i destinatari delle norme sia coloro cui spetterà il compito di farle rispettare. Indispensabile un complesso lavoro interpretativo da parte degli organismi delle governance, sia dell'UE sia dei singoli stati. Ciò per evitare probabili contenziosi anche nei rapporti tra i soggetti nazionali e quelli europei preposti al controllo, che porterebbero inevitabilmente ad un defatigante processo di chiarimento in sede giurisdizionale. Tale imprecisione è dovuta, oltre alla novità della disciplina, al continuo ed imprevedibile sviluppo del presupposto tecnologico su cui si vanno costruendo le norme ed alla necessità di mediare tra posizione opposte tra i paesi: tra chi, insomma, privilegia lo sviluppo della ricerca e della produzione e chi, invece, sottolinea i rischi etico-giuridici posti dall'IA.

Un altro aspetto critico è stato sottolineato da diverse associazioni di categoria. Particolarmente, le associazioni che rappresentano piccole e medie imprese hanno lamentato il fatto che, per far fronte alla necessità di adeguamento alle nuove norme, nei prossimi anni le società impegnate nel settore dovrebbero spendere più di 30 miliardi di euro. Ciò rallenterebbe i processi di informazione e favorirebbe l'esodo dei ricercatori e degli imprenditori verso paesi che pongono meno limiti normativi alla ricerca ed alla produzione ed usano strumenti di indirizzo più flessibili ed agevolazioni mirate[101].

Un'altra critica mossa è stata circa la sovrapposizione di sanzioni, come nel caso dei *data breach*, tra quelle previste dal GDPR e quelle contenute nella proposta della Commissione. Nella proposta, poi, è stata lamentata l'eccessiva centralità assegnata ai *providers*, in una fase in cui sono maggiormente affermati i modelli *open source*. In merito alla figura del *provider*, occorre ricordare che molto spesso i modelli di IA non sono sviluppati da un solo soggetto, ma sono il risultato di un assemblaggio di parti diverse ognuna delle quali riferita a specifiche funzioni, successivamente combinate tra di esse: ciò comporta la difficoltà nell'identificazione del provider cui fare riferimento per l'accertamento delle eventuali responsabilità[102].

Un'altra critica mossa è stata la difficoltà di muoversi tra la richiesta di trasparenza e la richiesta della tutela della proprietà intellettuale e dei segreti commerciali, che potrebbe indurre imprese al di fuori dell'Unione a rinunciare ad immettere i propri prodotti sul mercato europeo[103].

Preoccupazioni sono state, inoltre, manifestate per la complessità nel distinguere tra i rischi elevati e gli altri; per la

[101] The EU's approach to artificial intelligence, International Institute for Strategic Studies (IISS), vol. 27, Comment, 24 settembre 2021.
[102] J. BUYERS et alii, *Debate continues over the pros and cons of regulating Artificial Intelligence*, in *Lexology*, 27 luglio 2021.
[103] V. KAREN et alii, *European Commission Proposed Regulation on Artificial Intelligence*, in *Lexology*, 21 ottobre 2021.

difficoltà di identificare con chiarezza il responsabile finale nei casi di violazione delle norme; per l'incertezza nella determinazione e graduazione delle sanzioni. Poi a ciò si aggiunge la delicatissima questione della spiegabilità richiesta sia per i processi di informazione e decisione che per gli algoritmi che sono alla base: una spiegabilità che spesso sfugge agli stessi sviluppatori, a volte non in grado di comprendere le capacità evolutive del sistema.

Dunque, l'approccio fortemente normativo della nuova disciplina vede una forte burocratizzazione dei processi e la previsione di eccessivi e non sempre giustificati obblighi in tutte le fasi procedimentali, senza tutelare in modo convincente i diritti stessi[104].

Infine, su un punto specifico, critiche alla proposta erano già state avanzate il 18 giugno 2021 in una *Joint Opinion* dell'*European Data Protection Supervisor* (EDPS) e dello *European Data Protection Board* (EDPB), sul più volte richiamato tema della possibilità di consentire, sia pure in casi limitati, il riconoscimento automatico della fisionomia umana in luoghi accessibili al pubblico; e su quello della categorizzazione, anche qui in casi particolari, in base all'etnicità, al genere, all'orientamento politico, religioso o sessuale. Lamentandosi contestualmente della mancanza di coordinamento tra le certificazioni richieste per i sistemi di IA e quelle sulla protezione dei dati[105].

Come già anticipato, la Commissione ha avanzato la propria proposta di regolamento il 21 aprile 2021. Il Consiglio dell'UE ha definito il 6 dicembre 2022 la sua *General Approach Policy*. Gli emendamenti del Parlamento Europeo sono stati numerosi e sono databili al 14 giugno 2023. Il trilogo tra le istituzioni, iniziato a fine giugno, si è concluso l'8 dicembre 2023 con un accordo politico, esaminato ed accolto dal Coreper, il Comitato dei Rappresentanti Permanenti, organo del Consiglio dell'Unione, il 2 febbraio 2024. La Commissione

[104] V.G. FINOCCHIARO, *Intelligenza Artificiale. Quali regole?* , *Il Mulino*, 2024, p. 122.
[105] C. HORGAN, *Artificial Intelligence Update*, in *Lexology*, 12 luglio 2021

per il Mercato Interno e le Libertà Civili del Parlamento si è pronunciata favorevolmente sul testo il 13 febbraio 2024, mentre il voto finale del Parlamento è del 13 marzo. Dopo una ulteriore fase di controllo normativo e di *drafting* e l'approvazione finale da parte del Consiglio il 21 maggio, il testo è attualmente pubblicato nel *Journal Officiel*.

L'intero regolamento entrerà in vigore nel 2026, ma le norme sulle pratiche vietate e quelle sull'intelligenza artificiale con finalità generali sono già in vigore. Durante il lungo percorso, sono emerse novità sul piano della regolazione dell'IA sia con l'Ordine Esecutivo di Biden del 31 ottobre 2023 su una IA affidabile, sia con la Dichiarazione di *Bletchley Park* del primo novembre dello stesso anno a conclusione del *summit* organizzato dal Regno Unito sul tema della sicurezza a fronte dei modelli più avanzati di IA. Per non parlare delle diverse leggi sull'IA approvate in Cina.

L'intenzione dell'UE di porsi come capofila negli indirizzi normativi dovrà tenere conto sia dell'approccio meno invasivo dei due paesi anglosassoni sia di quello cinese, caratterizzato da una forte severità. L'esame in sede di trilogo non aveva contribuito ad una approvazione pienamente condivisa e rapida. Il 10 novembre 2023, ad esempio, dopo che i rappresentanti di Francia, Germania ed Italia nel Consiglio avevano manifestato la propria opposizione alla disciplina proposta per i *foundation models*, la delegazione del Parlamento si era addirittura ritirata dalla trattativa[106]. La Commissione proponeva, quindi, il 19 novembre una diversa impostazione che induceva il Parlamento a far circolare il 24 novembre un nuovo documento di lavoro che consentiva finalmente alla presidenza spagnola di presentare il 28 dello stesso mese un testo finale su questo tema controverso. Difficoltà che si è riscontrata, ad esempio, nell'approccio al centrale tema dell'identificazione biometrica da remoto in tempo reale, rispetto alla quale il Parlamento ha a lungo sostenuto una posizione di divieto assoluto. Altri contrasti si

[106] C. HAWES, Bristow LLP, *The EU'S AI Act: cliff hanger or cliff edge?,* in *LeXology*, 4 dicembre 2023.

sono segnalati sulla decisione sui poteri da attribuire all'*AI Office* sulle finalità di valutazione di impatto sui diritti umani (*Fundamental Rights Impact Assesment*, FRIA) e sulle eccezioni alla normativa per ragioni di sicurezza nazionale.

Ciò che bisogna sottolineare è la differenza di impostazione tra un Parlamento sempre attento alla tutela dei principi etici ed ai rischi che l'uso dell'IA può comportare e gli organismi più vicini agli esecutivi comunitari e nazionali, preoccupati degli effetti negativi che una disciplina troppo rigorosa potrebbe comportare per la ricerca, lo sviluppo e la commercializzazione di questa nuova tecnologia.

In gran parte, il testo, nella sua impostazione e nelle linee generali, rispetta quello dell'aprile 2021. Tenteremo di evidenziare i principali elementi di difformità tra la proposta iniziale della Commissione ed il testo pubblicato nel *Journal Officiel*. Un grande aiuto per questa analisi lo fornisce lo stesso testo legislativo che, nella parte II della Introduzione, ricorda tutti i diversi punti sui quali è stato necessario raggiungere un compromesso tra le Istituzioni.

I principali argomenti che hanno richiesto la ricerca di un punto di equilibrio sono stati: l'oggetto e la finalità della normativa; la definizione di intelligenza artificiale; le pratiche proibite; i criteri di identificazione remota e le deroghe al divieto al loro uso da parte delle autorità pubbliche; la valutazione d'impatto sui diritti fondamentali (FRIA); i modelli di IA a finalità generale (GPAI); la governance e le sanzioni; il sistema delle deleghe.

Il primo degli oltre 90 considerata dell'*AI Act* precisa le finalità dell'atto, asserendo che: "Scopo del Regolamento è migliorare il funzionamento del mercato interno predisponendo un quadro normativo uniforme, in particolare per lo sviluppo, la commercializzazione, la messa in servizio e l'uso dei sistemi di intelligenza artificiale nell'Unione conformemente ai valori dell'Unione stessa; promuovere lo sviluppo di un'IA umanocentrica ed affidabile, garantendo un elevato livello di protezione per la salute, la sicurezza ed i

diritti fondamentali racchiusi nella Carta, compresa la democrazia, il governo della legge (*rule of law*) e la tutela ambientale contro gli effetti dannosi dei sistemi di intelligenza artificiale nell'Unione; sostenere l'innovazione". Tali finalità sono articolate nei successivi considerata ed espresse, non sempre in modo convincente, nel testo degli articoli del Regolamento.

Questo Regolamento rappresenta ciò che De Gaulle avrebbe definito un "vasto programma", ovverosia il tentativo di conciliare due impostazioni che in quasi tutti i modelli di regolazione, anche al di fuori dell'Europa, si stanno confrontando: da una parte la protezione dei diritti fondamentali davanti ai rischi che pone questa nuova tecnologia; dall'altra l'esigenza di non intralciare ma favorire e tutelare lo sviluppo di strumenti con grandi capacità trasformative in campo scientifico, produttivo, economico e sociale. Il punto di partenza è la definizione di intelligenza artificiale. L'adozione di una terminologia più o meno ampia comporta l'inclusione o l'esclusione dalla normativa di una pluralità di sistemi, con conseguenze non indifferenti non solo nel campo produttivo ed economico, ma anche in quello sociale, politico e della sicurezza collettiva. Su richiesta del Parlamento, la proposta originaria è stata modificata per avvicinarla a quella dell'OCSE. Per intelligenza artificiale, si intende quindi un "sistema basato su una macchina progettata per funzionare con diversi livelli di autonomia, che può mostrare adattività dopo l'implementazione e che, per obiettivi impliciti o espliciti, deduce, dagli input che riceve, come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali"[107]. La definizione, per alcuni tecnologicamente neutrale[108], è stata da alcuni considerata troppo ampia e tale da poter includere anche, ad esempio, semplici sistemi di autocorrezione, algoritmi lineari come i tradizionali programmi per computer, modelli standard di

[107] Si veda l'art. 3, comma 1, del testo del Regolamento.
[108] D. VAN DER EYCKEN e E. RIGHINI, *Latham & Watkins LLP, Year in review: A rtificial Intelligence Law in European Union*, in *Lexology*, 3 gennaio 2024.

credit scoring o alcune funzioni di *Excel* [109]. Di questo rischio ci si è probabilmente resi conto, ed il considerato n. 6 del Regolamento precisa che i sistemi di IA devono avere caratteristiche tali da essere distinti "dai più semplici sistemi di *software* o di programmazione, e non dovranno necessariamente interessare sistemi fondati su regole definite solamente da persone naturali per eseguire operazioni automaticamente"[110]. Molto probabilmente, in futuro, ci sarà bisogno di ulteriori precisazioni: per esempio, sarebbe necessario un chiarimento sulla frase "macchina progettata per funzionare con diversi livelli di autonomia". Sembrerebbe che la distinzione si individui tra i modelli che producono risultati fondati esclusivamente in base a regole stabilite dall'uomo, in base ad esempio ad un criterio deterministico (*if-this-then-that*), e quindi non autonomi ma vincolati a regole prefisse; e quelli che usano invece criteri di valutazione statistica dei dati acquisiti per analizzare i c.d. pattern usati per fornire poi gli *output*. Solo questi ultimi modelli dovrebbero costituire oggetto del regolamento[111].

L'art. 2, comma 1, del Regolamento identifica i destinatari della normativa. Quest'ultima si applica ai fornitori (*providers*), indipendentemente dal fatto che abbiano sede nell'Unione o fuori di essa; agli operatori (*deployers*, termine che sostituisce il precedente users); agli importatori e distributori dei sistemi di IA; ai fabbricanti (*product manufacturers*) che collochino sul mercato sistemi di IA; ai rappresentanti autorizzati di fornitori non stabiliti nell'Unione; alle persone coinvolte (*affected*) situate (*located*) nell'Unione. Il termine che unifica tutte le categorie è quello di "operatori" (*operators*).

Senza definire nel dettaglio tutti i termini, riteniamo che un approfondimento lo meriti il termine *deployer*. Per *deployer* si

---

[109] P. MULLER, Advant Beiden, *The AI Act – The Agreement and What It Means*, in *Lexology*, 9 gennaio 2021.

[110] D. BAMBERG et alii, Hogan Lovells, *The EU AI Act: an impact analysis*, in *Lexology*, 1 febbraio 2024.
V.D. ROSENTHAL, VISCHER AG, *The EU AI Act and what it means in practice for most companies*, in *Lexology*, 1 febbraio 2024.

intende qualsiasi persona fisica o giuridica, una pubblica autorità, un loro rappresentante o un altro soggetto che faccia uso di un sistema di IA sotto il proprio controllo (*authority*), ad eccezione del caso in cui il sistema stesso sia utilizzato nel corso di una attività personale non professionale.

Abbiamo già accennato al fatto che, come nel caso del GDPR, le nuove regole avranno un carattere di extraterritorialità perché si applicheranno ai soggetti fornitori di sistemi di IA nei paesi dell'Unione indipendentemente dal fatto che abbiano sede o meno al suo interno; nonché ai soggetti posti al di fuori dell'Unione quando il risultato dell'attività dei loro sistemi di IA si rifletta comunque all'interno della stessa Unione[112]. Il *consideratum* n. 11, come clausola di chiusura, prevede che "al fine di prevenire l'elusione di questo Regolamento e di assicurare una protezione effettiva ai soggetti presenti nell'Unione, il Regolamento stesso si applicherà anche a fornitori ed utilizzatori di sistemi di IA registrati presso paesi terzi, nella parte in cui il risultato di tali sistemi sia finalizzato all'uso nell'Unione".

Il Regolamento si applicherà nella *European Economic Area* che include Islanda, Liechtenstein e Norvegia. L'ambito di applicazione delle norme è però limitato da diverse eccezioni. In primo luogo, a seguito di un confronto che partiva da posizioni divergenti tra Commissione e Parlamento, il Regolamento non influenzerà le competenze degli Stati Membri relative alla sicurezza nazionale, né si applicherà ai sistemi di IA usati solo per fini militari, di difesa o, per l'appunto, di sicurezza nazionale. Non si applicherà, altresì, a pubbliche autorità di un paese terzo o ad organizzazioni internazionali che agiscano nel quadro di accordi di cooperazione in campo giudiziario o di polizia. Sono esclusi i sistemi usati ai soli fini di ricerca e sviluppo scientifico, quelli usati per un'attività privata, non professionale; quelli che

[112] D. FUTTER et alii, ASHURST, *The EU ACT is here: What you need to know and what to do next*, in *Lexology*, 9 febbraio 2024.

operano con licenze *open source*, salvo che si tratti di sistemi proibiti o classificati come ad alto rischio.

Gli Stati Membri potranno poi prevedere una normativa più favorevole ai lavoratori a maggior tutela dei propri diritti, anche in contrasto con la disciplina adottata dai datori di lavoro in base alle nuove norme, così disapplicando le norme del Regolamento.

Come già ricordato prima, a fondamento della disciplina europea vi è il criterio dei "rischi": più alti sono i rischi più rigorosa la normativa, a cominciare dai sistemi che pongono rischi considerati inaccettabili e che, per tale ragione, saranno proibiti. Le pratiche di intelligenza artificiale proibite le rinveniamo nei *considerata* da 16 in poi e, successivamente, disciplinate dall'articolo 5. Saranno vietati i sistemi che usano tecniche subliminali, volutamente manipolative e ingannatrici, per distorcere o limitare la capacità di un soggetto ad assumere decisioni consapevoli; sfruttano le vulnerabilità di una persona in relazione all'età, alla disabilità, a specifiche condizioni sociali o economiche per distorcerne il comportamento e provocare ad essa un danno; si servono della caratterizzazione biometrica per catalogare un soggetto in base a razza, opinioni politiche, affiliazione sindacale, credi religiosi o filosofici, vita e orientamento sessuali; valutano o classificano le persone sulla base del loro comportamento sociale o delle loro caratteristiche personali (*social scoring*); utilizzano in luoghi pubblici modelli di identificazione biometrica da remoto in tempo reale per ragioni di polizia (*law enforcement*), salvo che li si debba usare per la ricerca mirata di vittime di rapimento, traffico di esseri umani, sfruttamento sessuale e per rintracciare persone scomparse, oppure per la prevenzione di una specifica, sostanziale ed imminente minaccia alla vita ed alla sicurezza delle persone, per un prevedibile attacco terrorista, o per la localizzazione e l'identificazione di un sospetto in relazione a specifiche ipotesi criminali richiamate nell'Annesso IIa alla Proposta (tra cui l'omicidio, le lesioni gravi, il traffico di organi, lo stupro etc.); si servono di valutazioni predittive

fondate esclusivamente sulla profilazione delle caratteristiche e dei tratti personali di un soggetto; creano raccolte di immagini facciali tramite un'estrazione non mirata di elementi tratti dalla rete o da telecamere a circuito chiuso (CCTV); puntano ad identificare le emozioni personali nei luoghi di lavoro o in quelli educativi.

Per quanto concerne l'impiego dei sistemi di identificazione biometrica appena menzionati è anche previsto che, prima della adozione di tali misure, andranno valutate sia le conseguenze negative derivanti da una loro mancata utilizzazione, sia quelle delle ricadute che il loro uso potrà avere nei riguardi delle persone interessate. Andrà effettuata una Valutazione di Impatto sui Diritti Fondamentali per esaminare le conseguenze che l'uso di questo strumento può avere su detti diritti. Per la loro adozione, sarà sempre necessaria l'autorizzazione di una autorità giudiziaria o di una autorità amministrativa indipendente; in caso di giustificata urgenza si potrà ricorrere al sistema, ma l'autorizzazione deve essere richiesta entro le ventiquattro ore.

Venendo ai cd. Rischi elevati (*high risks*), una prima affermazione di principio è contenuta nel *consideratum* n. 27 e risente del contrasto più volte sottolineato tra Parlamento e Commissione. Per trovare un punto di equilibrio, si sostiene che "l'identificazione dei sistemi di IA a rischio elevato deve essere limitata a quelli che producono un significativo impatto dannoso alla salute, alla sicurezza ed ai diritti fondamentali delle persone nell'Unione, e i limiti dovranno minimizzare ogni potenziale restrizione al commercio internazionale". Tale affermazione dovrebbe orientare la applicazione concreta delle nuove norme nel rapporto complicato tra tutele e sviluppo, nonché servire come ausilio interpretativo negli eventuali contenziosi. I sistemi ad elevato rischio saranno, dunque, quelli che, pur non provocando pericoli inaccettabili, causano un effetto negativo sulla salute, sulla sicurezza e sui diritti fondamentali delle persone.

Una indicazione più puntuale la troviamo nell'Annesso III, richiamato dall'articolo 6, nel quale vengono indicate otto aree nelle quali dovranno ricadere tali sistemi per poter essere considerati a rischio. I settori sono: biometrica, comprensiva dei modelli di identificazione remota se consentiti; infrastrutture critiche, come acqua, gas ed elettricità; educazione e formazione; occupazione, gestione del personale ed accesso al lavoro autonomo (annunci di lavoro, analisi delle candidature, criteri di valutazione dei candidati); accesso e godimento di servizi essenziali privati (come quelli bancari ed assicurativi) e pubblici (salute, assistenza e benefici sociali); sicurezza pubblica, nella parte in cui l'uso è consentito da altre norme dell'Unione o nazionali e con una dettagliata descrizione delle finalità che qualificano a "rischio elevato" l'uso dell'IA da parte delle autorità; immigrazione, asilo e controllo delle frontiere, anche qui ove consentito dalla legislazione vigente; amministrazione della giustizia e processo democratico . Sotto quest'ultimo profilo, è interessante la precisazione contenuta nel *consideratum* 40 che prevede in ogni caso la classificazione come "rischi elevati" di quelli che mirano ad influenzare il comportamento degli elettori in occasione di competizioni elettorali o di referendum.

Dentro le aree richiamate, saranno considerati a rischio elevato i modelli usati come componenti di sicurezza di un sistema o siano essi stessi un prodotto compiuto (*stand alone systems*), nei settori indicati dalla legislazione europea di armonizzazione di cui all'Annesso II, come nel campo aeronautico, automobilistico, medico, dei giocattoli, delle comunicazioni, dei macchinari; e che, inoltre, richiedono una valutazione di conformità, prima del loro ingresso nel mercato, da parte di un terzo soggetto.

Con una modifica aggiunta nella parte finale del confronto, non saranno considerati a rischio elevato i sistemi che non presentano un rischio significativo per la salute, per la sicurezza e per i diritti fondamentali: ad esempio quando eseguono un compito strettamente procedurale; quando

servono a migliorare una precedente attività umana; quando servono ad identificare difformità rispetto a precedenti modelli decisionali; quando il sistema di IA svolge compiti essenzialmente preparatori in relazione alle ipotesi di cui al richiamato Annesso III.

Per quanto concerne le prescrizioni per i sistemi a rischio elevato, bisogna ricordare la premessa contenuta nel comma primo dell'articolo 8: nel valutare il rispetto effettivo delle prescrizioni, bisognerà tenere presente lo "stato dell'arte" dell'IA e delle tecnologie ad essa connesse. Ciò significa che le prescrizioni dovranno sempre essere lette alla luce dello sviluppo tecnologico e non potranno mai esser date per definitivamente acquisite.

Venendo agli obblighi più rilevanti previsti per i fornitori, i cd. *Providers*[113]. Essi dovranno: prevedere, porre in essere, documentare ed aggiornare un sistema di gestione rischi e qualità; rispettare i criteri stabiliti per l'uso dei dati, come ad esempio nel controllo e nell'eliminazione dei pregiudizi *biases*); conservare la documentazione tecnica; sviluppare un sistema capace di conservare e registrare le operazioni effettuate (*record keeping*); assicurare che il sistema abbia seguito tutte le necessarie procedure di valutazione di conformità; predisporre tempestivamente la dichiarazione di conformità UE; apporre il marchio di conformità CE; assicurare la massima collaborazione con le autorità. I *providers*, inoltre, saranno tenuti a fornire agli utilizzatori tutte le informazioni necessarie "per interpretare i risultati del sistema ed utilizzarlo in maniera appropriata"; quelle sulle misure poste in essere per assicurare il controllo umano; la durata prevedibile ed i criteri di manutenzione; nonché le risorse computazionali necessarie per un funzionamento corretto[114].

[113] HUNTON, ANDREWS, Kurth LLP, *Privacy and Information Law Blog*, in *Lexology*, 1 febbraio 2024, 3.
[114] K.C. HALM et alii, DAVIS, WRIGHT, *Tremaine LLP, EU unanimously approved Comprehensive Artificial Intelligence Act*, in *Lexology*, 12 febbraio 2024, 5.

Dovranno anche designare nell'Unione un rappresentante autorizzato.

Venendo agli obblighi dei *deployers* (prima chiamati *users*), essi dovranno: adottare le misure tecniche ed organizzative per assicurare un uso appropriato del sistema; conferire il controllo umano a persone con la necessaria competenza, addestramento, capacità, autorità e supporto; ove vengano da essi inseriti dati aggiuntivi (*input data*), assicurarsi che siano rilevanti e ragionevolmente rappresentativi alla luce delle finalità del sistema; informare i destinatari finali, lavoratori e loro rappresentanti, di trovarsi di fronte ad un sistema a rischio elevato; ove richiesto, effettuare una valutazione di impatto sui diritti fondamentali (FRIAA); monitorare l'andamento delle operazioni e dare immediata notizia al *provider*, all'importatore, al distributore ed alle autorità interessate di ogni serio incidente; conservare per un periodo adeguato, almeno mesi sei, le registrazioni sull'attività svolta[115].

Sono, altresì, previsti obblighi per gli importatori e per i distributori: trattasi sostanzialmente di una verifica di conformità alle procedure indicate dalla normativa.

Prima abbiamo citato il fatto che i *deployers*, prima di fare uso di un sistema ad alto rischio, debbano effettuare una valutazione di impatto sui diritti fondamentali (FRIA). Si tratta di una procedura il cui scopo è quello di verificare l'impatto potenziale del sistema su diritti come la *privacy*, la non discriminazione, la libertà di espressione. La si dovrà effettuare, per esempio, quando bisognerà valutare la credibilità finanziaria di un soggetto ovvero in campo assicurativo per comparare i rischi e le tariffe in relazione all'età o alla salute. Nel processo valutativo, dovranno essere individuate le persone o i gruppi che potrebbero essere influenzati dal sistema; i potenziali, specifici rischi; le misure da adottare per evitarli; quelle necessarie per rispondere al

[115] HUNTON, cit., 4; L.E. STEINKER et alii, Wikborg Rein LLP, *Towards safe, reliable and human centered AI*, in *Lexology,* 13 febbraio 2024, 3.

verificarsi della situazione di pericolo; la durata e la frequenza dell'uso degli strumenti che potrebbero determinare i rischi. Il tutto sulla base di informazioni che il provider è tenuto a fornire.

Un aspetto di importanza particolare è quello dei cosiddetti modelli *General Purpose AI* (GPAI). Si tratta, così come da definizione del *consideratum* 60a, di modelli capaci di compiere un numero elevato ed indeterminato di compiti. Essi sono istruiti con una quantità rilevante di dati e con modalità diverse e possono essere usati direttamente o come componenti di un sistema di IA. Il Regolamento stabilisce obblighi per i GPAI di carattere generale, ai quali ne aggiunge di specifici per quelli che presentano rischi sistemici. Un tipico esempio di GPAI è dato dai modelli generativi di grandi dimensioni (LLM), alla base di prodotti come *ChatGPT* e *Google Gemini*[116], capaci di produrre video, testi, immagini, codici per computer e così via.

Sono esclusi dagli obblighi che ora indicheremo i modelli usati per fini di ricerca e sviluppo, nonché i prototipi prima del loro ingresso in mercato. I providers di GPAI dovranno, dunque, predisporre e tenere aggiornata una relazione tecnica a disposizione delle autorità; rendere disponibile la documentazione necessaria agli utilizzatori; ove registrati al di fuori dell'UE, indicare un loro rappresentante; garantire il rispetto della legislazione dell'UE in materia di *copyright*, in particolare di quella relativa al *text and data mining* (TDM), l'estrazione automatizzata cioè della rete di testi, suoni, immagini e dati digitali con lo scopo di "proprietari" dei dati utilizzati a negare (*opt-out*) l'uso dei dati stessi; provvedere altresì a riassumere il contenuto dei dati utilizzati per addestrare il proprio modello. Dovranno, pertanto, informare gli utenti del fatto che stanno interagendo con un sistema di IA, ed adempiere ad obblighi di trasparenza come quello di procedere al cosiddetto *water-marking*, ovverosia

[116] H. RAVIA et alii, PEARL, COHEN, ZEDEL, LATZER, BARATZ, *The AI Act*, in *Lexology*, 6 marzo 2024.

l'imposizione di una filigrana informatica che identifica un autore e contiene informazioni necessarie sul *copyright*.

Norme più severe vengono previste per i GPAI che presentano rischi sistemici. I rischi sono quelli specificati nel *consideratum* 60m: ad esempio, quelli che in caso di gravi incidenti possono arrecare serie conseguenze alla salute ed alla sicurezza pubblica; o avere negativi effetti sul processo democratico ed economico; o disseminare contenuti illegali, falsi o discriminatori. Un modello GPAI creerà rischi sistemici anche in relazione alla sua potenza: ciò accadrà quando la potenza di calcolo totale sarà superiore a 10^25 FLOPS[117], una soglia che rappresenta attualmente i sistemi più avanzati di GPAI, come *ChatGPT*-4 e *Gemini*[118].

In caso di rischi sistemici, i *providers* dovranno effettuare una valutazione anche avversariale del modello in base a protocolli standardizzati che riflettano lo stato dell'arte; valutare e mitigare i possibili rischi sistemici derivanti dall'utilizzo nell'Unione; registrare e comunicare alle autorità tutte le informazioni relative a seri incidenti e alle necessarie misure da adottare; assicurare un'adeguata protezione *cyber* per i modelli e le infrastrutture fisiche ad essi correlate. I fornitori di GPAI dovranno, poi, predisporre ed aggiornare una documentazione tecnica dettagliata comprensiva dei dati relativi al *training*, alle risorse utilizzate ed a quelle ritenute necessarie sotto un profilo energetico. Ciò anche per consentire ai diversi utilizzatori di comprendere i limiti e le capacità dei modelli in questione[119].

I rischi moderati sono quelli che non pongono immediati pericoli ed il problema principale che si può porre in relazione ad essi è forse la mancanza di chiarezza sulla loro interazione con gli esseri umani. Si prevede, quindi, nel Regolamento che gli utilizzatori debbano essere

[117] FLOPS (*Floating Points Operations Per Second*) indica in informatica il numero di operazioni eseguite in un secondo dall'unità di elaborazione centrale (CPU, Central Processing Unit).
[118] V. il dossier della Camera dei Deputati n. 26, cit., 9.
[119] V.T. RAMMOS e R. GLASER, Taylor Wessing, *The EU AI Act and general-purpose AI,* in *Lexology*, 28 marzo 2024.

correttamente informati del trovarsi davanti ad una macchina e che quanto prodotto sarà artificialmente generato o modificato. Continuerà ad applicarsi su questi strumenti la legislazione in vigore. Un tipico esempio di rischio moderato è costituito dai *chatbots*: nel momento in cui gli utilizzatori sono consapevoli di interagire con un sistema di IA, non si ritiene che possano derivare rischi elevati dall'uso del sistema stesso.

Per i sistemi a rischio minimo, come videogiochi o filtri anti *spam*, non sono previste restrizioni. Le società che li producono o li commercializzano potranno volontariamente predisporre codici di condotta a cui attenersi[120]. Entro diciotto mesi dall'entrata in vigore dell'*AI Act* la Commissione predisporrà un elenco esemplificativo dei sistemi a rischio elevato e di quelli da non considerarsi tali. Detto elenco sarà costantemente aggiornato.

L'articolo 53, proprio per supportare l'innovazione, prevede la possibilità di istituire *sandboxes* regolatorie. Trattasi di ambienti controllati, stabiliti da una autorità pubblica, nei quali possono essere testati, per un periodo di tempo limitato, prodotti e servizi tecnologicamente innovativi prima del loro ingresso nel mercato, in stretto contatto con le autorità vigilanti, usando un regime regolamentare semplificato e, come detto, transitorio.

Gli stati membri dovranno stabilire almeno una *sandbox* regolatoria a livello nazionale e, ove lo ritengano, altre a livello regionale o locale; potranno anche collaborare tra di loro per creare sandboxes interstatali per guidare providers e *deployers* nello sviluppo di sistemi di IA. È anche prevista dall'art. 54 la possibilità di testare sistemi ad elevato rischio in condizioni reali, anche al di fuori dell'ambiente delle *sandboxes*. Sono indicate prescrizioni diverse per consentire alle autorità di effettuare attività ispettive, porre limiti

[120] K.J. NAHRA et alii, Wilmer Cutler Pickering Hale e Dorr LLP, *The European Parliament Adopts the AI Act*, in *Lexology*, 14 marzo 2024; e D.S. MARKS et alii, Benesch Friedlander Copland and Aronoff LLP, *European Union Artificial Intelligence Act: An Overview*, in *Lexology,* 13 marzo 2024.

temporali, richiedere informazioni ed imporre, se necessario, ulteriori e specifiche misure di salvaguardia.

Queste prove in condizioni reali di sistemi a rischio elevato potranno essere effettuate per un massimo di mesi sei ed il piano per la loro realizzazione dovrà essere naturalmente approvato dall'autorità di vigilanza. Sono misure, queste appena indicate, che mirano a facilitare la ricerca e lo sviluppo. Sono rivolte soprattutto alle piccole e medie imprese.

Breve accenno lo meritano anche le norme sul *copyright*[121]. Qui si intrecciano le previsioni del nuovo Regolamento e quelle, in materia, della Direttiva 2019/790, in particolare nel suo articolo 4. I considerata 60i e 60j del testo del Regolamento ribadiscono l'obbligo per i providers di modelli di GPAI introdotti nel mercato europeo di rispettare la normativa comunitaria vigente; con particolare riferimento alla riserva all'uso dei propri dati eventualmente espressa dai titolari dei diritti di *copyright* che potrà essere invocata, in base alla ricordata Direttiva EUCD, indipendentemente dal dove l'estrazione dei dati abbia avuto luogo. Ancora, l'articolo 52c, comma 1, lettera c, prevede che la riserva dei diritti debba essere rispettata "allo stato dell'arte delle tecnologie": ciò comporta per i *providers* la necessità di un continuo adeguamento dei propri sistemi per meglio individuare i contenuti protetti.

I providers saranno anche tenuti a redigere e rendere pubblico un sommario dettagliato dei materiali e dei contenuti utilizzati per il training dei GPAI, secondo un modello definito dall'*AI Office*; e a conservare un registro delle attività di "*training and testing*" usate per i modelli di IA.

L'*AI Office* avrà il compito di garantire il rispetto della normativa sui *copyrights* e di pubblicare i sommari

[121] Su questa materia, tra i diversi autori italiani, ricordiamo G. FINOCCHIARO, op. cit., 93. V. anche E. CARLETTI e L. LUCANI, Studio Legale Previti, *AI Act e copyright: l'approccio del nuovo Regolamento europeo*, in *Lexology*, 29 marzo 2024.

dell'utilizzo dei dati[122]. Leggendo l'articolo 52c e la Direttiva Europea 2019/790, disciplinante la cosiddetta "eccezione TDM" (*Test and Data Mining*), si comprende che l'estrazione e la riproduzione di dati per finalità di *training* è consentita per ragioni di ricerca scientifica da parte di istituzioni scientifiche e culturali e per i soggetti richiamati dall'articolo 4 della Direttiva. I titolari dei diritti di copyright potranno rendere nota "in maniera appropriata" la loro riserva all'uso dei propri dati (*opt out*). Questo conferma che, nonostante alcuni dubbi interpretativi, è necessario che gli sviluppatori dei sistemi di IA ottengano una preventiva autorizzazione dai titolari dei diritti, salvo che non si ricada nella richiamata eccezione di TDM.

Inoltre, per superare le preoccupazioni dei *deployers* per l'utilizzo di dati che potrebbero essere stati acquisiti in contrasto con le disposizioni appena richiamate, alcuni sviluppatori si sono assunti la responsabilità verso gli utilizzatori stessi per i rischi legati ad eventuali violazioni di *copyright*[123]. La più celebre causa sulla violazione di *copyright* in materia di TDM è al momento quella intentata nel Regno Unito da *Getty Images* contro *Stability AI*, per avere violato i diritti di proprietà intellettuale attraverso lo *scraping* (estrazione indiscriminata dalla rete) di propri dati. Al momento in cui si scrive, la causa è ancora in corso nel Regno Unito[124].

Venendo, invece, agli aspetti sanzionatori, occorre tenere in considerazione l'articolo 71. Gli Stati Membri dovranno prevedere sanzioni, anche amministrative, pecuniarie e non, che siano "effettive, proporzionate e dissuasive". All'interno del Regolamento, sono stabilite le soglie entro cui gli Stati possono intervenire, per le violazioni relative all'articolo 5,

[122] F. BANTERLE e A. SCHETTINO, Hogan Lovells, *Copyright provisions in the AI Act: generative AI, transparency and data mining,* in *Lexology*, 9 febbraio 2024.
[123] L. GREENS, Bryne Wallace LLP, *AI update: Examining the relationship between Artificial Intelligence and Copyright Protection*, in *Lexology*, 20 marzo 2024.
[124] P. WATERS, Gilbert + Tobin, *Getty Images vs Stability AI*, in *Lexology*, 24 marzo 2024; L. HARROLD, Keystone Law, *Artificial Intelligence and copyright infringement*, in *Lexology*, 24 aprile 2024.

pratiche proibite, la sanzione potrà raggiungere i 35 milioni di euro o, se si tratta di una società, il 7% del fatturato annuo mondiale; fino ai 15 milioni di euro, o al 3% del fatturato, per le violazioni relative ai rischi elevati o per quelle connesse ai modelli di IA per finalità generali (GPAI); per informazioni incorrette fornite agli organismi ed alle autorità nazionali competenti, 7,5 milioni di euro o l'1% del fatturato annuo mondiale.

Per quanto concerne le imprese minori, l'approccio deve tenere conto del principio di proporzionalità, per porre le sanzioni al livello più basso dei criteri stabiliti per le violazioni appena indicate. Sanzioni sono previste, altresì, per i soggetti pubblici, le istituzioni, gli organismi dell'Unione e le agenzie che non rispettino le norme e le finalità del Regolamento: tali sanzioni saranno erogate dal Garante Europeo per la Protezione dei Dati (EDPS). È stato, inoltre, introdotto il diritto di singoli e di persone giuridiche a sollevare reclami alle autorità di vigilanza su potenziali violazioni del Regolamento; nonché quello di richiedere ai *deployers* una spiegazione chiara ed esaustiva sui sistemi usati.

Venendo al modello di governance, l'articolo 55b del Regolamento richiama l'istituzione da parte della Commissione di un Ufficio per l'Intelligenza Artificiale (*AI Office*). L'Ufficio è stato costituito con una Decisione della Commissione stessa il 24 gennaio 2024. Tra i compiti del nuovo Ufficio, si devono ricordare quelli relativi allo sviluppo di strumenti e metodologie per valutare i modelli GPAI a finalità generali; il monitoraggio sull'efficacia delle norme previste per i modelli stessi; il controllo sui loro potenziali rischi, anche in base alle indicazioni dell'istituendo panel scientifico; l'indagine sulla violazione delle norme ad essi relative, anche attraverso la valutazione di reclami e denunce; l'assistenza per l'applicazione delle regole in materia di pratiche proibite e di sistemi a rischio elevato; il contributo ad un approccio; il contributo ad un approccio strategico dell'Unione coerente con le iniziative

internazionali sull'IA; il monitoraggio dell'evoluzione dei mercati e delle tecnologie. E ancora, l'assistere la Commissione nel predisporre ulteriori misure normative, nell'esercitare le deleghe ad essa attribuite, nel supportare la concreta attuazione del Regolamento. L'*AI Office* farà parte del Direttorato Generale della Commissione per le Comunicazioni e la Tecnologia (DG-CNECT)[125].

Il quadro della *governance* va ben oltre la mera istituzione dell'*AI Office*. L'articolo 56, infatti, prevede la costituzione di un Comitato Europeo per l'IA (*European Artificial Intelligence Board-EAIB*), con il compito di assistere la Commissione e gli Stati Membri per la corretta ed efficace applicazione del Regolamento, per contribuire al coordinamento tra le autorità nazionali competenti, condividere le *best practices*, fornire a richiesta della Commissione o di propria iniziativa raccomandazioni o pareri sulle materie relative all'applicazione del Regolamento stesso. Tale comitato potrà coadiuvare la Commissione nel promuovere la alfabetizzazione in materia di IA; facilitare lo sviluppo di criteri comuni di condotta; cooperare con le istituzioni dell'Unione e con i soggetti interessati nel campo della sicurezza dei prodotti, della *cybersecurity*, della protezione dei consumatori e dei diritti fondamentali. Inoltre, dovrà collaborare con le autorità di paesi terzi e con le organizzazioni internazionali nel campo dell'IA.

Tale Comitato è composto da un rappresentante di ciascuno Stato Membro per la durata di tre anni, ed ai suoi lavori partecipa come osservatore il Commissario Europeo per la protezione dei dati (EDPS). L'*AI Office* potrà partecipare alle riunioni, senza diritto di voto. Il Comitato si articolerà in due sottogruppi permanenti, ad uno dei quali verrà devoluta la sorveglianza del mercato. I componenti

[125] P. ALEXE E f. CUNNINGHAM, Bird & Bird LLP, *AI on track to be formally adopted ahead of the European elections*, in Lexology, 22 febbraio 2024; D.P. COOPER et alii, Covington & Burling LLP, *European Commission Announces New Package of AI Measures*, in *Lexology*, 14 febbraio 2024.

designati dai singoli Stati Membri devono essere in grado di rappresentare in modo effettivo la posizione degli Stati stessi.

Un forum chiamato *advisory forum* avrà, infine, il compito di fornire al Comitato ed alla Commissione pareri e consulenze tecniche. Del forum faranno parte rappresentanti dell'industria, delle piccole e medie imprese, della società civile e dell'Accademia. Nella composizione dello stesso, dovranno trovare rappresentanza equilibrata sia gli interessi commerciali sia quelli non commerciali. I suoi membri sono nominati dalla Commissione ma di esso saranno componenti permanenti i rappresentanti di alcune istituzioni comunitarie, tra cui l'Agenzia per i diritti fondamentali, quella per la *cybersecurity* e il Comitato europeo per la standardizzazione. Oltre questo forum consultivo, l'articolo 58b prevede la costituzione di un panel scientifico di esperti indipendenti, scelti sempre dalla Commissione in base alle loro specifiche e riconosciute competenze nel campo dell'intelligenza artificiale. Ad essi spetterà il compito di coadiuvare sotto un profilo tecnico la Commissione stessa con la loro conoscenza ed esperienza scientifica. Anche gli Stati Membri potranno avvalersi delle competenze del *panel*.

A livello nazionale, il Regolamento prevede che ogni Stato Membro stabilisca o designi almeno un'autorità di notificazione ed un'autorità di sorveglianza del mercato, indipendenti ed imparziali. Tra i compiti dell'autorità di notificazione vi sarà quello di definire, guidare e monitorare le procedure necessarie per l'attività degli organi che dovranno esprimere il loro parere di conformità sui sistemi. L'autorità di sorveglianza sarà responsabile dell'applicazione e del controllo del Regolamento 2019/1020 in materia di sicurezza dei prodotti e di tutela dei consumatori. L'Italia, per come risulta dal disegno di legge in materia di IA presentato, sembrerebbe avere individuato nell'AgID (Agenzia per l'Italia digitale) e nell'ACN (Agenzia per la Cybersicurezza Nazionale) le Autorità Nazionali per l'intelligenza artificiale. Spetterà ad esse il coordinamento e la collaborazione con le altre pubbliche amministrazioni e con le

autorità indipendenti nel settore in esame, tramite un apposito Comitato di coordinamento composto dai direttori generali delle due Agenzie e dal Capo del Dipartimento per la trasformazione digitale della Presidenza del Consiglio.

Per quanto concerne l'entrata in vigore del Regolamento, quest'ultimo è entrato in vigore 20 giorni dopo la pubblicazione in Gazzetta Ufficiale. Le singole norme troveranno applicazione in base ad uno specifico scadenzario indicato nell'articolo 85. Precisamente, sei mesi dopo l'entrata in vigore per quanto concerne i sistemi proibiti; dodici mesi per gli obblighi relativi ai sistemi a finalità generali (GPAI); ventiquattro mesi per quelli a rischio minimo, medio ed elevato, comprese le valutazioni di conformità, qualità, valutazione dei rischi e controllo *post-marketing*; trentasei mesi per i sistemi disciplinati da altre norme dell'Unione, come quelle che prevedono sui prodotti un giudizio di conformità da parte di terzi. Per quanto concerne i sistemi già immessi sul mercato, è stabilita una moratoria di quattro anni per consentire l'adeguamento alle nuove norme da parte delle autorità pubbliche che facciano uso di sistemi di IA a rischio elevato[126].

Una delle critiche mosse al Regolamento è stata l'eccessiva lunghezza del processo di normazione, contrapposta alla speditezza del processo tecnologico che, ovviamente, procedeva più rapidamente dei legislatori. Ciò ha costretto ad una non facile rincorsa per normare fenomeni di cui, a volte, sembrava anche difficile comprendere la portata. Le deleghe alla Commissione di cui all'articolo 73, sono quindi relative sia alle misure per l'attuazione della disciplina primaria, sia a quelle che si renderanno necessarie per un aggiornamento costante della normativa in base al prevedibile sviluppo tecnologico. I poteri delegati sono attribuiti per cinque anni, rinnovabili tacitamente per altri

[126] K. HARFORD et alii, BAKER McKENZIE, *EU Act finally adopted: what's next?*, in *Lexology*, 13 marzo 2024; M. CARLE et alii, Kramer Levin Naftalis & Frankel, *Key considerations regarding the recently passed EU Artificial Intelligence Act*, in *Lexology*, 14 marzo 2024

cinque anni a meno che non vi sia opposizione espressa del Parlamento o del Consiglio.

Tra le deleghe attribuite, si ricordi quella per aggiornare i criteri di classificazione dei modelli di IA a finalità generali e la delega per modificare la soglia FLOPS, quella cioè che stabilisce la potenza di calcolo totale necessaria per considerare un modello a rischio sistemico. Poi vi è il problema della stessa definizione di IA che potrebbe mutare con lo sviluppo della tecnologia; quello della valutazione di conformità e delle specifiche della documentazione tecnica; la disciplina delle *sandboxes*. La Commissione ed il Comitato istituendo saranno dunque impegnati in un lavoro intenso, sempre attento allo sviluppo delle nuove tecnologie. Vi sono, inoltre, delle linee guida demandate dal Regolamento alla definizione della Commissione. Queste riguardano, ad esempio, l'applicazione dei requisiti e degli obblighi per i sistemi a rischio elevato; le pratiche di IA vietate; gli obblighi di trasparenza; i rapporti tra il Regolamento ed altre normative in materia dell'Unione; la catena delle responsabilità. La Commissione dovrà anche favorire tra le imprese l'adozione di un "*AI Pact*", per dare attuazione a parti del nuovo Regolamento anche prima delle scadenze sopra ricordate.

In sede di attuazione del Regolamento, si creeranno questioni inerenti alla sovrapposizione tra la nuova normativa sull'IA e quella già in vigore in campi ad essa vicini. Ad esempio, si ricordi il *Digital Service Act (*Regolamento UE 2022/2065), nella parte in cui in relazione ad esso andrebbero chiariti il regime giuridico ed i requisiti per le VLOP (*Very Large Online Platform*) nel processo di adeguamento alla nuova normativa; il *Data Act* (Regolamento UE 2023/2054), il cui obiettivo primario è migliorare l'accesso e l'uso dei dati derivanti dagli strumenti IoT (*Internet of Things*), che richiederà una riflessione sui processi decisionali basati su informazioni acquisite grazie all'IA da detti strumenti; e ancora andranno tenuti presenti sia il *Cyber Resilience Act* sia l'*European Media Freedom Act*, che mira a

proteggere l'indipendenza ed il pluralismo della stampa e richiederà una valutazione sui processi di acquisizione, conservazione ed uso dei dati[127]. La Commissione ha anche proposto un pacchetto di misure ulteriori come aiuti economici e normativi per start-up e piccole e medie imprese nel campo dell'intelligenza artificiale; una proposta di regolamento per facilitare la creazione di imprese interessate all'uso dell'IA, anche attraverso l'utilizzo di capitali a rischio; finanziamenti a favore del programma europeo dedicato all'IA generativa; iniziative per agevolare la formazione e la qualificazione di talenti; lo sviluppo e la diffusione di spazi europei di dati. La nuova disciplina è un punto di partenza che metterà alla prova la capacità delle strutture dell'Unione e di quelle nazionali a confrontarsi, in un processo tecnologico assai rapido ed in settori sempre più ampi, non solo con altri stati, ma forse ancora di più con soggetti privati capaci di influenzare la ricerca ed il mercato all'interno ed all'esterno dell'Unione stessa.

Venendo alla definizione di IA fornitaci dall'*AI Act*, essa è ricavabile dall'art. 3 del predetto Regolamento, secondo cui per IA si intende*<<un sistema automatizzato creato per operare con vari livelli di autonomia a che presenta adattabilità dopo la diffusione e che, per espliciti o impliciti propositi, deduce dall'input che riceve come generare gli outputs come previsioni, contenuti, raccomandazioni o decisioni che possono influenzare gli ambienti fisici o virtuali>>.*

È interessante constatare come l'autonomia sia la principale caratteristica dei sistemi di IA che ricadono nell'applicazione del Regolamento.

4.2. Definizione di IA fornitaci dal Consiglio di Stato italiano.

[127] V.H. MARKUS e M. PONCZA, Heuking, *EU digital law: Formal agreement in the EU Parliament on the AI Regulation*, *the Cyber Resilience Act and the European Media Freedom Act*, in *Lexology*, 15 marzo 2024; C. MONSEE e S. RICHTER, Taylor Wesing, *The interfaces between artificial intelligence and the Data Act*, in *Lexology*, 9 gennaio 2024.

In Italia, la giurisprudenza amministrativa si è dovuta confrontare con il problema della definizione di IA. In particolare, il Consiglio di Stato[128] si è trovato a dover giudicare circa la legittimità delle procedure inerenti ad una gara per la fornitura di un *pacemaker* di alta fascia per gli enti della sanità lombarda. La lettera di invito ed il capitolato tecnico indicavano, tra i criteri di valutazione dell'offerta, il parametro tabellare "Algoritmo di prevenzione+trattamento delle tachiaritmie atriali": in caso di presenza di entrambi gli algoritmi, di prevenzione e di trattamento, sarebbero stati assegnati quindici punti; nel caso di presenza di uno solo dei due, sette.

La società che ha vinto la gara era la *Microriport*, mentre la prima ad essere esclusa era la *Abbot*. La commissione di gara, nel valutare la sussistenza dei parametri di cui sopra, aveva dato rilievo ai soli algoritmi che erano dotati di autonomia. Entrambi i prodotti erano dotati di autonomia nella prevenzione, ma soltanto uno nel trattamento: il sistema di *Microriport* aveva al suo interno algoritmi autonomi di trattamento che consentivano <<in maniera automatica di contrastare il ritmo prefibrillatorio costituito dal riconoscimento di frequenti ectopie atriale e trattato mediante riduzione/omogeneizzazione dei periodi refrattari atriali>>. Mentre l'algoritmo di *Abbot* non veniva considerato autonomo in quanto poteva essere attivato soltanto presso ambulatori cardiologici. *Abbot* impugnava la aggiudicazione dinanzi al TAR Lombardia, lamentando che il parametro preso in considerazione nulla avesse a che vedere con il carattere dell'autonomia di un sistema. Il TAR, nell'accogliere il ricorso, asseriva che <<il dato testuale della lettera di invito non richiede che l'algoritmo di trattamento, al verificarsi dell'episodio aritmico, sia avviato dal dispositivo medesimo in automatico. Tale caratteristica attiene ad una componente ulteriore, non indicata dalla legge di gara, vale a dire a un algoritmo di intelligenza artificiale nella diagnosi delle aritmie e avvio del trattamento. Fondatamente, pertanto,

[128] Cons. di Stato n. 2891/2021.

Abbot ha dedotto l'erroneità della valutazione della commissione di gara che – pur in presenza di un algoritmo di trattamento delle aritmie nel proprio dispositivo (vale a dire l'algoritmo di trattamento NIPS, pacificamente definibile come tale) – ha attribuito soli 7 punti anziché 15 al dispositivo offerto. Infatti, la commissione ha confuso (…) il concetto di algoritmo con quello di avvio automatico del trattamento>>.

Microport proponeva allora ricorso al Consiglio di Stato avverso la decisione del TAR. Il Consiglio di Stato, pur riconoscendo che l'algoritmo altro non sia che <<semplicemente una sequenza finita di istruzioni, ben definite e non ambigue, così da poter essere eseguite meccanicamente e tali da produrre un determinato risultato>>, accoglieva il ricorso di *Microport*, asserendo che tale termine, quando riferito a sistemi tecnologici sia <<ineludibilmente collegato al concetto di automazione ossia a sistemi di azione e controllo idonei a ridurre l'intervento umano>>. Il Consiglio di Stato esclude che lo strumento di *Abbot* potesse essere qualificato come algoritmo atteso che <<non interviene in funzione di automazione delle funzioni di prevenzione e trattamento delle tachiaritmie atriali come richiesto dall'amministrazione, e dunque correttamente l'amministrazione non lo ha considerato ai fini del punteggio>>.

La critica che occorre fare alla decisione è la seguente: non si può presumere che con l'espressione "dispositivi di alta fascia", la PA intendesse algoritmi dotati di autonomia. Ma a noi interessa solo la definizione che il Consiglio di Stato fornisce di IA, comparandola con quella di algoritmo. L'IA è un <<sistema che non si limita solo ad applicare le regole *software* e i parametri preimpostati (come fa invece l'algoritmo "tradizionale"), ma, al contrario, elabora costantemente nuovi criteri di interferenza tra dati e assume decisioni efficienti sulla base di tali elaborazioni, secondo un processo di apprendimento automatico>>. Breve. Il Consiglio assimila la nozione di IA a quella di ML, aderendo

certamente ad una nozione molto più restrittiva di quella adottata dall'Unione europea nell'*AI Act*.

5. Esempi pratici di sistemi di IA imprevedibili.

Dopo aver ristretto l'ambito della nostra indagine ai soli sistemi di IA, imprevedibili, occorre adesso fornire taluni esempi pratici di sistemi di IA che potrebbero cagionare, o che hanno già cagionato danni a beni giuridici protetti dal diritto penale. Si tratta di fatti che, se compiuti da un agente umano, integrerebbero senza alcun dubbio fattispecie penalmente rilevanti. Il nostro campo di indagine è allo stesso tempo più ampio e più ristretto di quello che potrebbe riguardare la responsabilità aquiliana.

Più ampio in quanto determinati incidenti, come nel caso del medico che abbia instaurato un rapporto professionale con il paziente, ricadono nella disciplina del danno da inadempimento contrattuale di cui all'art. 1218 c.c.; mentre più ristretto in quanto, per come è noto, la clausola generale di cui all'art. 2043 c.c. delinea una nozione atipica di fatto illecito, differenziandosi dall'illecito penale che, invece, necessità costituzionalmente di tipicità espressa dalla legge (art. 25, co. 2, Cost.).

Seguendo una classificazione delle possibili fattispecie fornita da B. Fragasso[129], potremmo distinguere i vari esempi pratici in base alle modalità con cui i sistemi di IA interagiscono con l'esterno e, dunque, in base al tipo di bene giuridico sul quale il sistema può incidere tramite la "commissione" di fatti penalmente rilevanti.

Una prima distinzione è tra i sistemi che possono agire sul mondo esterno direttamente ed i sistemi che, invece, possono incidere sui beni giuridici solo attraverso un filtro umano. Rientrano tra i primi i dispositivi caratterizzati da *embodiement*, come i *robot*, in cui il *software* è incorportato in un *hardware*. Questi sono sistemi che, agendo nel mondo

[129] B. Fragasso*, Intelligenza artificiale e responsabilità penale. Principi e categorie del diritto penale alla prova di una tecnologia "imprevedibile"*, Giappichelli, 2025.

fisico, possono incidere sull'integrità delle persone cagionando lesioni e morti. Si pensi ai *robot* come gli *autonomous vehicles*, i sistemi di guida autonoma degli aerei, i sistemi di arma autonomi letali, gli *healthcare robots*. Sempre nei primi rientrano taluni algoritmi privi di *embodiement* che definiamo *bot*. A seconda dell'ambiente virtuale in cui operano, essi sono in grado di compiere diverse tipologie di illeciti: i *social bots* possono pubblicare *post* diffamatori o di istigazione all'odio online; i sistemi di trading possono alterare i prezzi degli strumenti finanziari, integrando potenzialmente il reato di manipolazione del mercato.

Sistemi che, invece, possono ledere indirettamente i beni giuridici protetti dal diritto penale sono i cd. *intelligent decision support systems*, come, ad esempio i sistemi di supporto alla diagnostica ed al trattamento in ambito medico.

Sistemi in grado di cagionare un'offesa diretta ai beni giuridici.

Gli *embodied systems*.

Il caso degli *autonomous vehicles*.

Il settore automobilistico è quello maggiormente avanzato nella ricerca e nell'uso dell'IA. Per tale ragione, la nostra attenzione si concentrerà maggiormente su detto settore piuttosto che su altri, usandolo come paradigma per comprendere i problemi inerenti all'imputazione della responsabilità penale per i danni da algoritmo intelligente. La letteratura che si è occupata dell'imputazione dei danni derivanti da macchine a guida autonoma è già molto ampia[130]. Lo scopo precipuo dei ricercatori e dei produttori

[130]Senza alcuna pretesa di esaustività si vd., quanto alla dottrina italiana, A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit.; A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit.; U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit.; G. CALABRESI – E. AL MUREDEN, *Driverless Cars*, Il Mulino, 2021, p. 156; A. LEPORE, *I.A. e responsabilità civile. Robot, autoveicoli e obblighi di protezione*, in *Tecnologie e diritto*, 2021, p. 198; M. TAMPIERI, *L'intelligenza artificiale: una nuova sfida anche per le automobili*, in *Contr. Impr.*, 2020, n. 2, p. 748; S. PELLEGATTA, *Automazione nel settore*

dei veicoli a guida autonoma è la completa automazione delle funzioni di guida: scomparirebbero volante e pedali ed il conducente non sarebbe altro che un passeggero. Le *self-driving cars* dovranno essere tra di loro connesse così come dovranno essere connesse con le infrastrutture stradali.

I benefici della definitiva affermazione di veicoli completamente autonomi sarebbero in grado di determinare una drastica riduzione del numero di incidenti e delle vittime ad essi collegate, ma anche l'incremento della mobilità degli anziani e dei disabili[131]. Attualmente nessuna azienda automobilistica ha sviluppato una *self-driving car* pronta per essere testata e messa in commercio. Mancando una normativa di riferimento che disciplini l'attribuzione della responsabilità per il caso in cui il sistema completamente autonomo causi danni a cose o persone, il rischio legale è troppo alto.

Ma anche quando esistenti, i testi normativi prevedono la presenza di un *human-in-the-loop*, ovverosia di una persona che condivide la funzione di guida con l'IA. La rinuncia alla

---

automotive: profili di responsabilità civile, in *Contr. impr*., 2019, n. 4, p. 1429-1431; F. BARTOLINI, *Auto a guida autonoma e problemi di responsabilità civile*, in *R. Giordano e al.* (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, 2022, p. 306-307; U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, in Giur. it., luglio 2019, p. 1704 ss.; sui profili penalistici vd. A. CAPPELLINI, *Profili penalistici delle self-driving cars*, in Dir. pen. cont. – Riv. Trim., 2019, n. 2, p. 326 ss.; L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit. Nella letteratura statunitense si vd. DOUMA – S.A. PALODICHUK, *Criminal Liability Issues Created by Autonomous Vehicles*, in *Santa Clara Law Review*, p. 1157 ss.; K.S. ABRAHAM – R.L. RABIN, *Automated vehicles and manufacturer responsibility for accidents: a new legal regime for a new era*, in *Va. L. Rev*., 2019, p. 127 ss.; M.A. GEISTFELD, *A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, in Calif. L. Rev, 2017, p. 1611 ss.; ID., *The Regulatory Sweet Spot for Autonomous Vehicles*, in Wake Forest L. Rev. 2018, p. 337 ss.; J. GURNEY, *Imputing Driverhood: Applying a Reasonable Driver Standard to Accidents Caused by Autonomous Vehicles*, in P. Lin e al. (eds.) Robot Ethics 2.0, cit., p. 51 ss.; M. ROE, *Who's Driving That Car?: An Analysis of Regulatory and Potential Liability Frameworks for Driverless Cars*, in B.C. L. Rev., 2019, p. 315 ss.; D.G. GIFFORD*, Technological Triggers to Tort Revolutions: Steam Locomotives, Autonomous Vehicles, and Accident Compensation*, in *Journal of Tort Law*, 2018, p. 71 ss.; nella letteratura tedesca vd. E. HILGENDORF, *Automated Driving and the Law*, cit; M. EBERS, *Civil liability for autonomous vehicles in Germany,* 2022, in SSRN, disponibile a questo link: https://ssrn.com/abstract=4027594.

[131] Vd. S. Saripalli, *Are Self-Driving Cars the Future of Mobility for Disabled People?,* in *The Conversation*, 5 ottobre 2017, disponibile a questo link: https://theconversation.com/are-self-driving-cars-the-future-of-mobility-for-disabled-people-84037.

totale autonomia dei veicoli discende proprio dal timore del *responsibility gap* e dal fatto che verrebbe dato agli algoritmi il difficile compito di risolvere questioni eticamente sensibili come quali persone sacrificare in caso di impossibilità di un intervento completamente privo di lesioni. Dagli esperti, questo fenomeno viene individuato come *trolley problem*[132].

5.1.2.3. I diversi livelli di autonomia delle *self-driving cars*.

Il livello di autonomia si definisce in base al ruolo che deve essere svolto dall'utilizzatore per garantire la funzionalità del veicolo ed un corretto livello di sicurezza. La classificazione che pare essere quella maggiormente affidabile ed utilizzata è quella elaborata dalla *Society of Automotive Engineers* (SAE), in collaborazione con *l'International Organization for Standardization* (ISO), pubblicata nel 2014 e più volte modificata.

La SAE individua cinque categorie di *Driving Automation System*. In particolare, al livello 1 (cd. *Drive Assistance*), il veicolo può eseguire autonomamente o il movimento longitudinale o quello laterale ed il sistema autonomo si disinserisce immediatamente su richiesta del conducente. Esempi di tale sistema possono essere il controllo adattivo

[132] Vd. M. CUNNEN E AL., *Artificial Driving Intelligence and Moral Agency: Examining the Decision Ontology of Unavoidable Road Traffic Accidents through the Prism of the Trolley Dilemma*, in *Applied Artificial Intelligence - An International Journal*, 2019, Vol. 33, No. 3, p. 267 ss.; W.B. WENDEL, *Economic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles*, in *California Western Law Review* 129, 2018; I. COCA VILA, *Coches autopilotados en situaciones de necesidad. Una aproximación desde la teoría de la justificación penal, in Cuadernos de política criminal*, 2017, 235 ss.; in generale, sul trolley problem, vd. D. EDMONDS, *Uccideresti l'uomo grasso? Il dilemma etico del male minore*, Raffaello Cortina Editore, 2014. Nota correttamente Ruffolo che, in una fase di transizione, in cui le macchine a guida autonoma si affiancheranno ai veicoli tradizionali, l'utilizzatore dell'*autonomous vehicle* dovrà ragionevolmente attendersi di non essere sacrificato in favore dell'incolumità personale di terzi (es. i pedoni che attraversano all'improvviso un incrocio), conformemente alla logica che sta alla base dell'esimente dello stato di necessità (art. 54 c.p.), vd. U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1708. Vd. anche G. CONTISSA – F. LAGIOIA – G. SARTOR*, La manopola etica: i veicoli autonomi eticamente personalizzabili e il diritto*, in *Sistemi Intelligenti*, 2017, n. 3, p. 601 ss., che propongono che i sistemi di guida autonoma assicurino sempre la possibilità, per il conducente, di decidere in via preventiva i criteri da utilizzare in caso di "scelte tragiche".

della velocità di crociera o il parcheggio del veicolo. Al livello 2 (cd. *Partial Driving Automation*), il veicolo può eseguire autonomamente e simultaneamente sia il movimento laterale che quello longitudinale: può dunque svolgere interamente, in determinate circostanze e per un determinato lasso temporale, la funzione di guida nei suoi aspetti operazionali. Ma, visto che le capacità del sistema di effettuare la *object and event detection and response* (OEDR) sono limitate, il conducente dovrà supervisionare l'attività ed intervenire ogni volta sia necessario. Al livello 3 (cd. *Conditional Driving Automation*) il veicolo può svolgere interamente, in determinate circostanze e per un certo lasso temporale, la funzione di guida: a differenza che per i veicoli di cui al secondo livello, qui il sistema è in grado di riconoscere le situazioni in cui non è in grado di operare, avvertendo il conducente affinchè egli prenda il controllo del veicolo. Pertanto, il conducente non è tenuto a monitorare costantemente il sistema, ma ha il dovere di restare attivo e reattivo, garantendo una *minimal risk condition*. In Germania, già dal 2017 possono circolare questi tipi di veicoli. Anche in Francia, ma solo a livello sperimentale. Al livello 4 (cd. *High Driving Automation*), il veicolo può svolgere interamente la funzione di guida se sussistono le circostanze di tempo e luogo. L'utente è un mero passeggero e non deve quindi intervenire per riprendere la funzione di guida se il sistema presenta guasti o difetti nella circolazione, poiché il veicolo è in grado di effettuare da solo il cd. *fallback*. Ad oggi nessun livello 4 è stato approvato, anche se in Germania, dal 2021, si è autorizzato l'uso di "funzioni di guida autonoma in aree operative definite". Al livello 5 (cd. *Full Driving Automation*), il veicolo svolge la funzione di guida su qualsiasi tipo di strada ed in tutte le circostanze in cui un'automobile tradizionale può essere usata.

I livelli 1 e 2, in verità, non vengono considerati come *automated driving systems* poiché sono in grado di svolgere solo una parte della funzione di guida, lasciando tutto il resto nelle mani del conducente.

5.1.2.4. Le normative sulle *self-driving cars*.

Come spesso accade nel campo dell'innovazione, gli Stati Uniti sono stati tra i primissimi paesi ad avviare le sperimentazioni delle auto a guida autonoma. Ciò è stato possibile grazie alla mancata adesione alla Convenzione di Vienna sulla circolazione stradale del 1986 da parte degli USA, che vincola la circolazione di *autonomous vehicles* alla presenza di specifiche normative.

A prescindere da ciò, Stati come il Nevada hanno approvato specifiche normative dirette a regolare la circolazione di *autonomous vehicles* di livello 2[133]. In California, si è consentita la sperimentazione nel traffico di veicoli *driverless*[134]. Tali testi non affrontano il tema della responsabilità, ma si limitano a prevedere e disciplinare il regime delle autorizzazioni e dei requisiti per la circolazione su strada di detti veicoli.

Con riferimento agli stati aderenti alla Convenzione summenzionata, un primo emendamento del 2016 ha consentito la circolazione di sistemi di guida parzialmente autonoma, purchè vi sia la possibilità per il conducente di disattivarli sempre e dal 2022 sono entrate in vigore modifiche che consentono la circolazione di veicoli a guida completamente autonoma[135]. Ma spetta agli Stati aderenti

[133] Per accedere al *database* aggiornato delle legislazioni statali, negli Stati Uniti, in materia di *Autonomous Vehicles* vd. NATIONAL CONFERENCE OF STATE LEGISLATURES, *Autonomous Vehicles State Bill Tracking Database*, disponibile a questo link: https://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx.

[134] California Code of Regulations, tit. 13, div. 1, ch. 1, art. 3.7, accessibile a questo link: https://www.law.cornell.edu/regulations/california/title-13/division-1/chapter-1/article-3.7.

[135] Gli emendamenti, presentati il 14 gennaio 2021 dallo *United Nations Economic Commission for Europe* (UNECE), sono stati tacitamente approvati ai sensi dell'art. 49(2)(a) della Convenzione; vd. United Nations, Convention on Road Traffic Vienna, 8 November 1968 – *Acceptance of Amendment to Article 1 and New Article 34 Bis to the Convention*, 21 gennaio 2022, disponibile sul sito delle Nazioni Unite a questo link: https://treaties.un.org/doc/Publication/CN/2022/CN.26.2022-Eng.pdf. Già nel 2016, tuttavia, era stato consentito l'utilizzo di sistemi di bordo che influiscono sulla guida del veicolo, purché potessero essere "neutralizzati o disattivati dal conducente" (art. 8, co. 5-bis).

comprendere se e come consentire la circolazione di detti veicoli.

La Germania ha avuto ed ha un ruolo di *leader* normativo sul tema. Infatti, ha approvato la prima disciplina organica in materia di ADS, consentendo la circolazione di veicoli di livello 3 e 4[136]. Con riferimento ai primi, l'utilizzatore è ancora considerato conducente del veicolo. Nel momento in cui è attiva la funzione di guida automatica il conducente può distogliere la propria attenzione, restando, però, percettivo di modo da poter intervenire quando lo chieda il sistema ovvero quando il conducente avrebbe dovuto riconoscere sulla base di circostanze evidenti che non esistevano i presupposti per l'uso delle funzioni di guida automatiche[137]. Per gli utilizzatori del sistema di livello 4, la legge tedesca non prevede alcun obbligo di monitoraggio. Scompare la figura del conducente[138] e compare quella del supervisore tecnico[139], che potrebbe anche trovarsi all'esterno del veicolo e che deve intervenire solo quando vi sia una richiesta diretta da parte del sistema, riportando il veicolo ad una situazione di rischio minimo[140].

In Francia, nel 2019, è stata introdotta una disciplina regolante la circolazione dei veicoli a guida autonoma di livello 3[141]. L'*ordonnance* 14 avril 2021 n. 2021-443 ha, inoltre, modificato il Code de la route francese, inserendo un capitolo all'interno del *Tìtre* 2 rubricato "*Responsabilitè pénale applicable en cas de circulation d'un vehicle à delegation de conduite (articles L123-1 à L123-4*)"[142], che detta una disciplina speciale per la

[136] *Bundesgesetzblatt Jahrgang* 2021 Teil I Nr. 48, cit.; sulla legislazione tedesca vd. diffusamente M. EBERS, *Civil liability for autonomous vehicles in Germany*, cit.; vd. anche L. D'AMICO, *Intelligenza artificiale e auto a guida autonoma. tra prevenzione primaria, colpa penale e rischio consentito*, in Riv. it. med. leg., 2022, fasc. 3, p. 612 ss.
[137] Sez. 1b(1) e (2) StVG, cit.
[138] *Fahrzeugführer*.
[139] *Technische Aufsicht eines Kraftfahrzeugs*.
[140] Sezione 1d(3) StVG, cit.
[141] Vd. Legge n. 2019-486 del 22 maggio 2019, cit. Sulla disciplina francese vd. più approfonditamente M. GIUCA, *Disciplinare l'intelligenza artificiale. La riforma francese sulla responsabilità penale da uso di auto a guida autonoma*, in Arch pen., 2022, n. 2.
[142] Il testo della legge è disponibile su *Légifrance* a questo link:https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043370894.

responsabilità del conducente e del produttore di *autonomous vehicle*. L'art. L123-1, co. 1, esclude l'applicabilità degli obblighi di vigilanza previsti per il conducente di veicolo tradizionale nel caso in cui, al momento dell'infrazione, il sistema era attivo con la funzione di guida automatica[143]. Il conducente dovrà comunque mantenersi sempre in uno stato di allerta in modo da poter tempestivamente riprendere la funzione di guida, se richiesto dal sistema. Al comma 3, l'art. L123-1 stabilisce che laddove il conducente non risponda alla richiesta di riprendere il controllo avanzata dal sistema, si tornerà ad applicare la generale disciplina sulla responsabilità penale del conducente. Sembra, dunque, che in capo al conducente si possa configurare una posizione di garanzia: egli ha l'obbligo di restare allerta e di attivarsi se richiesto dal sistema perché se non lo fa l'evento lesivo gli sarà addebitato a titolo omissivo.

L'art. L123-2, co. 1, invece, prevede che, per il lasso di tempo in cui il sistema esercita il controllo dinamico del veicolo, il produttore risponderà degli eventuali delitti di omicidio o lesioni stradali, sempre che ne venga accertata la colpa. Così il legislatore ha voluto circoscrivere la responsabilità colposa del produttore ai soli casi in cui vi è un controllo totale del veicolo da parte del sistema.

In Italia, non sembra vicina una disciplina organica sulle macchine a guida autonoma. Il decreto del Ministero delle Infrastrutture e dei Trasporti del 28 febbraio 2018 recante Modalità attuative e strumenti operativi della sperimentazione su strada delle soluzioni di *Smart Road* e di guida connessa ed automatica (cd. Decreto *smart roads*)[144], dà

[143] Per controllo dinamico si intende, nello specifico, l'«*exécution de toutes les fonctions opérationnelles et tactiques en temps réel nécessaires au déplacement du véhicule. Il s'agit notamment du contrôle du déplacement latéral et longitudinal du véhicule, de la surveillance de l'environnement routier, des réactions aux événements survenant dans la circulation routière et de la préparation et du signalement des manoeuvres*», vd. Décret n° 2022-1034 du 21 juillet 2022 portant publication de l'amendement à la convention internationale sur la circulation routière de Vienne du 8 novembre 1968, adopté à Genève le 14 janvier 2022 (1), disponibile a questo link:
https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000046081258.

[144] Per un esame approfondito della normativa si vd. L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 817 ss.

la possibilità, previa autorizzazione, di sperimentare su strada i veicoli a guida parzialmente autonoma. La guida deve sempre avvenire sotto il controllo di un supervisore e su specifici tratti stradali dotati di adeguate infrastrutture. Ai sensi dell'art. 11, co. 1, lett. A), la domanda di autorizzazione deve contenere anche <<l'indicazione del proprietario del veicolo a guida autonoma, quale soggetto responsabile ai sensi dell'art. 196 del codice della strada e dell'art. 2054, comma 3, del codice civile, ovvero di altro soggetto obbligato in solido, ai sensi del citato art. 196>>. Al di fuori di tali particolari regimi, la circolazione di auto a guida autonoma deve intendersi implicitamente vietata dal Codice della strada che richiede che ogni auto sia condotta da una persona che se ne assuma la responsabilità[145].

5.1.2.5. *Case-studies*: incidenti con la partecipazione di *autonomous vehicles*.

Ci sono stati parecchi incidenti che hanno coinvolto, direttamente o meno, gli *autonomous vehicles*. Ci sono diversi *database online* che raccolgono e classificano gli incidenti che hanno coinvolto tali veicoli[146]. Daremo brevemente conto dei casi più importanti.

Nel 2018, in un caso, *Umeda v Tesla*, un cittadino giapponese, Yoshihiro Umeda, veniva colpito ed ucciso da un veicolo *TeslaMode* X, mentre si trovava sulla corsia d'emergenza dell'autostrada a prestare soccorso ad un amico motociclista. Al momento dell'incidente, era inserito il sistema *Autopilot* nella Tesla ed il veicolo aveva improvvisamente accelerato colpendo i pedoni sulla corsia di emergenza. I parenti di *Umeda* avevano presentato ricorso contro Tesla per ottenere il risarcimento dei danni, sostenendo la presenza di un difetto di progettazione. Anche se la corte statunitense ha dichiarato di non avere

[145] A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 326. Vd. in particolare l'art. 46 Codice della Strada: «si intendono per veicoli tutte le macchine di qualsiasi specie, che circolano sulle strade guidate dall'uomo».

[146] Tra i più aggiornati vd. il sito *Autonomous Vehicle Crashes* (https://www.avcrashes.net), che dà anche la possibilità di visualizzare i dati attraverso una mappa interattiva

giurisdizione sul caso, accogliendo le istanze di *Tesla*[147], la vicenda merita di essere considerata per valutare le argomentazioni del ricorso, fondate sui principi della responsabilità per danno da prodotto.

I querelanti lamentavano che *Tesla Autopilot* avesse taluni difetti di produzione, in quanto non sarebbe stato in grado di riconoscere la presenza di pedoni e neanche di allertate il conducente del malfunzionamento in corso. *Tesla* avrebbe saputo che il prodotto era difettoso, ma <<(…) *Despite such knowledge, Tesla, acting through its officers, directors, and managing agents, for the purpose of enhancing profits, knowingly and deliberately failed to remedy or otherwise take corrective action to fix the known defects in said vehicles, failed to warn the public, including decedent Yoshihiro Umeda, of the extreme risk of injury that could result and be occasioned by said defects*>>. Inoltre, Tesla avrebbe omesso di informare gli acquirenti sui rischi ed i pericoli legati all'uso del sistema di guida autonoma, promuovendo una *overreliance* del conducente nei confronti delle funzionalità del veicolo.

Quanto ai procedimenti penali, negli Stati Uniti, si sono discussi due casi in cui gli incidenti provocati dagli *autonomous vehicles* hanno determinato l'avvio di indagini penali nei confronti dei conducenti, rei, secondo l'accusa, di avere omesso colposamente di vigilare sul funzionamento del veicolo. In un caso del marzo 2018, una *safety driver* di *Uber*, Rafael Stuart Vasquez, è stata accusata del reato di omicidio stradale per non avere evitato la collisione tra il veicolo che stava testando ed un pedone che stava attraversando la strada. La posizione di *Uber* è stata invece archiviata da parte dell'autorità giudiziaria. La richiesta di rinvio a giudizio dell'agosto 2020 non spiega le precise circostanze dell'incidente. Queste, però, possono essere ricavate dai

[147] I giudici si dichiaravano privi di giurisdizione poiché sia il conducente della Tesla, sia la vittima e i parenti della vittima che avevano presentato la querela, erano tutti cittadini giapponesi; inoltre, la Tesla coinvolta nell'incidente è stata venduta in Giappone.

numerosi articoli pubblicati[148], nonché da un report del *National Transportation Safety Board* (NTSB), l'agenzia federale incaricata ad indagare sugli incidenti che coinvolgono trasporti ed infrastrutture[149]. Secondo quest'ultimo report, il sistema di guida autonoma aveva individuato il pedone 5.6 secondi prima dell'impatto, ma non era stato in grado di classificarlo come pedone né di predire il suo movimento. Le specifiche del sistema non prevedevano infatti la frenata di emergenza per attenuare la collisione, che era affidata al guidatore.

La *safety driver* era distratta e questo è dimostrato da un video della telecamera rivolta verso l'interno del veicolo che mostrava che l'operatrice aveva distolto lo sguardo dalla strada per un lungo periodo di tempo mentre il veicolo si avvicinava al pedone, probabilmente per guarda il cellulare. L'operatore aveva guardato la strada solo un secondo prima dell'impatto, iniziando a sterzare 0,02 secondi prima di colpire il pedone, ad una velocità di 39 miglia orarie.

Nel *report* redatto dall'NTSB si legge che il sistema di guida autonoma sviluppato da Uber è progettato per apportare piccole modifiche alla traiettoria dei veicoli per evitare gli ostacoli, ma non è collaudato per effettuare frenata di emergenza. Dunque nei casi in cui siano individuati degli ostacoli da aggirare, le auto avviano un'*action suppression* per un secondo per studiare un percorso alternativo o per permettere all'operatore di prendere il controllo del veicolo. Effettuata questa *action suppression*, la risposta dell'ADS dipende dalla possibilità di evitare o meno di evitare la collisione applicando la massima frenata. Se la massima frenata è possibile, il sistema frena automaticamente; se non lo è, il sistema invia un allarme all'operatore, avviando

[148] In particolare, un recente articolo ricostruisce la vicenda, vd. L. Smiley, *'I'm the Operator': The Aftermath of a Self-Driving Tragedy*, in *Wired US* , 8 marzo 2022, disponibile a questo link: https://www.wired.com/story/uber-self-driving-car-fatal-crash/.

[149] NTSB Accident report, *Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe*, *Arizona March* 18, 2018, pubblicato il 19.11.2019, disponibile a questo link:
https://www.ntsb.gov/investigations/accidentreports/reports/har1903.pdf.

contemporaneamente un rallentamento graduale del veicolo. Alla base di tale congegno, vi è l'idea che le brusche frenate siano maggiormente sicure se effettuate dal conducente. In base al report, a contribuire all'incidente sarebbero state anche l'inadeguatezza delle procedure di valutazione dei rischi per la sicurezza attivate da *Uber* e la mancanza di meccanismi efficaci per affrontare la "*complacency*" degli operatori nei confronti dell'automazione.

In un altro caso, del dicembre 2019, alla guida della Tesla Model S, vi era un cittadino, Kevin George Aziz Riad, che non era intervenuto per fermare la sua automobile con sistema attivo di *autonomous driving* al semaforo rosso. L'auto si è scontrata con un altro veicolo cagionando la morte di due persone. Riad è stato rinviato a giudizio per omicidio stradale commesso con colpa grave[150].

5.1.2.6. Sistemi di guida autonoma degli aerei.

Tra le più antiche tecniche di IA sperimentate nell'ambito civile, ci sono i sistemi di *autopilot* degli aerei[151]. Oggi si è raggiunta la quasi totale autonomia, essendo in grado di decollare, navigare ed atterrare, con un minimo *input* umano: al pilota è affidata la gestione delle emergenze, dei cd*. edge cases*. [152] In via di sviluppo sono anche sistemi completamenti autonomi di volo che non hanno necessità neanche della

[150] La notizia è riportata da CBS Los Angeles, *Driver to stand trial in Tesla autopilot crash which killed two in Gardena*, 20 maggio 2022, disponibile a questo link: https://www.cbsnews.com/losangeles/news/driver-to-stand-trial-in-tesla-autopilot-crash-which-killed-two-in-gardena/.

[151] Per una ricostruzione dei principali problemi che si pongono in ordine all'individuazione della persona responsabile, in caso di incidenti aerei vd. M.C. ELISH – T. HWANG, *Praise the machine! Punish the human! The Contradictory History of Accountability in Automated Aviation*, in *Comparative Studies in Intelligent Systems, Working Paper* #1 V2, 18 maggio 2015; M.C. ELISH, *Moral crumple zones: Cautionary Tales in Human-Robot Interaction*, in *Engaging Science*, *Technology, and Society 5* (2019), p. 40 ss.; W.B. WENDEL, T*echnological Solutions to Human Error and How They Can Kill You: Understanding the Boeing 737 Max Products Liability Litigation*, J. Air L. & Com. 379, 2019.

[152] M.C. ELISH, *Moral crumple zones: Cautionary Tales in Human-Robot Interaction*, in *Engaging Science, Technology, and Society* 5 (2019), p. 40 ss.; W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You: Understanding the Boeing*.

presenza umana nella cabina di pilotaggio[153]. Detti sistemi sono ovviamente nati per incrementare la sicurezza dei voli e ridurre l'incidenza di errori umani. Dagli anni Cinquanta in poi, nella produzione dei voli, sono diventati lo *standard*.

È proprio la longevità di detti sistemi a darci la possibilità di comprendere i problemi che si possono verificare quando il controllo di un veicolo è affidato al lavoro congiunto di essere umano e macchina. Perché è chiaro che questi sistemi nascono per ridurre al minimo l'incidenza dell'errore umano, ma dalla cogestione del volo possono sorgere nuovi problemi.

Il primo pericolo è la totale deresponsabilizzazione morale dei piloti dovuta all'eccesivo affidamento nella tecnologia. Affidamento che diventa dipendenza tecnologica alla quale consegue l'incapacità di gestire le situazioni di emergenza. Questo, per esempio, è quanto accaduto nell'incidente del volo *Air France* 447, occorso nel 2009: i piloti non riuscirono a risolvere una situazione di stallo, un problema semplice da gestire per un pilota di aereo, determinandone il crollo[154]. A ciò si aggiunga che il funzionamento degli attuali sistemi tecnologici alla base dell'*autopilot* sono altamente difficili da comprendere. Una delle vicende più esemplificative ha coinvolto i velivoli modello *Boeing* 737 MAX. Nel 2018, un velivolo di tale modello precipitò in mare poco dopo il decollo da Giacarta cagionando la morte di tutti i passeggeri. La medesima situazione si è verificata nel 2019 su un volo operato da *Ethiopian Airlines*. Nel 2022, un volo operato da

[153] Vd. S. Lekach, *Self-flying planes could transport passengers one day—but first, packages, in Mashable*, 24 luglio 2021, disponibile a questo link: https://mashable.com/article/self-flying-airplanes; T. Pallini, *I flew on a self-flying plane where pilots sat back as the aircraft taxied, took off, and landed on its own and I'm convinced it's the future of aviation, in Business Insider*, 8 maggio 2021, disponibile a questo link: https://www.businessinsider.com/flying-on-self-flying-plane-convinced-its-future-of-aviation-2021-5?r=US&IR=T.

[154] Per una ricostruzione dell'incidente vd. *Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA), Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris, luglio 2012*, disponibile a questo link: https://bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf; vd. anche l'opinione di un esparto aviatore W. Langewiesche, The Human Factor, in Vanity Fair, ottobre 2014, disponibile a questo link: https://www.vanityfair.com/news/business/2014/10/air-france-flight-447-crash.

China Eastern Airlines si è schiantato subito dopo la partenza. I primi due incidenti sono stati causati da una difettosa interazione tra i piloti ed il sistema di controllo di volo intelligente adottato dalla Boeing, ovverosia il *Maneuvering Characteristics Augmentation System* (MCAS)[155]. Si tratta di un sistema anti-stallo che interviene quando alcuni sensori richiedono una stabilizzazione. In ambedue gli incidenti, i sensori avevano percepito erroneamente che l'aereo si trovava con la parte anteriore impennata ed avevano segnalato tale criticità al MCAS che continuava ad abbassare la parte interiore dell'aereo. I piloti tentavano di riportare il velivolo nella corretta situazione di equilibrio ma, non sapendo che il sistema MCAS si fosse attivato, non riuscivano a capire cosa stesse accadendo. Dunque, un sistema che aveva come obiettivo quello di ridurre i rischi derivanti da errori umani imprevedibili si è trasformato in un ulteriore ed inaspettato fattore di rischio[156].

5.1.2.7. Sistemi di armi autonome letali.

Contesto privilegiato per la ricerca e la sperimentazione delle tecnologie più avanzate è sempre stato quello bellico. Ingenti stanziamenti si sono effettuati negli ultimi decenni per lo sviluppo di arme autonome letali, i cd. *Lethal autonomous weapon systems* (LAWS)[157]. Secondo la definizione

[155] La medesima ricostruzione emerge anche in J. Fallows, *Here's What Was on the Record About Problems with the 737 Max, in The Atlantic*, 13 marzo 2019, disponibile a questo link: https://www.theatlantic.com/ideas/archive/2019/03/heres-what-was-on-the-record-about-problems-with-the-737-max/622213/; per un'analisi del caso sotto la lente della responsabilità per danno da prodotto vd. inoltre W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit.

248 In argomento vd. più diffusamente W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit., passim.

[156] In argomento vd. più diffusamente W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit., passim.

[157] In dottrina vd. N. Bhuta e al. (eds.), Autonomous Weapons Systems, cit.; J. Galliott – D. MacIntosh – J.D. Ohlin (eds.), Lethal Autonomous Weapons. Re-Examining the Law and Ethics of Robotic Warfare, Oxford University Press, 2021; P. SCHARRE, Army of None: Autonomous weapons and the future of war, W.W. Norton&Company, 2018; A. LELE, Disruptive Technologies for the Militaries and Security, Springer, 2019; P. ASARO, How just could a robot war be?, in P. Brey – A. Briggle – K. Waelbers (eds.), Current Issues in Computing and Philosophy, Ios Press, p. 50 ss.; E. FUZAYLOVA, War torts, autonomous weapons systems, and liability: why a limited strict liability tort regime should be implemented, in Cardozo Law Review, 40, n. 3, 2019; W.C. MARRA – S.K. MCNEIL, Understanding "The Loop": Regulating the

più accreditata, si tratta di armi che, una volta attivate, sono in grado di identificare, selezionare ed ingaggiare un obiettivo senza ulteriore intervento umano[158]. Anche se l'essere umano può opporre una sorta di veto alla decisione algoritmica, qualora egli non lo facesse, dette armi sarebbero perfettamente in grado di decidere il *target* da colpire. È per questo che dobbiamo considerarle a tutti gli effetti autonome.

Occorre, invece, escludere dal novero dei sistemi autonomi d'arma i sistemi automatizzati manovrabili a distanza[159] come i *loitering munitions*[160]. Altresì, è necessario escludere dal novero delle armi autonome le armi ed i sistemi d'arma specificamente progettati per difendere le proprie piattaforme, le proprie forze e la popolazione contro minacce dinamiche come missili, munizioni ed aeromobili nemici[161].

Quando è attivata, un'arma autonoma è in grado di identificare gli obiettivi militari grazie a sensori e tecniche di riconoscimento delle immagini basate sul ML, nonché di colpirli. A differenza dei tradizionali droni militari che

---

Next Generation of War Machines, in 36 Harvard Journal of Law and Public Policy, 2013, p. 1139; nella dottrina italiana, D. AMOROSO – G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto: quale ruolo per il controllo umano?*, in *BioLaw Journal*, 2019, n. 1, p. 33 ss.; F. CHESINI, *"Terminator Scenario"? Intelligenza artificiale nel conflitto armato: "Lethal Autonomous Weapons Systems" e le risposte del diritto internazionale umanitario,* in *BioLaw Journal*, 2020, fasc. 3, p. 441 ss.

[158] Così la Department of Defense Directive (DODD) 3000.09, *Autonomy in Weapon Systems*, 21 novembre 2012, mod. 2017, disponibile a questo link: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf: «*weapon system[s] that, once activated, can select and engage targets without further intervention by a human operator*»; nello stesso senso, vd. definizione fornita da International Committee of the Red Cross (ICRC) *Position on Autonomous Weapon Systems*, 2021, disponibile a questo link: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems.

[159] Risoluzione del Parlamento Europeo del 12 settembre 2018 sui sistemi d'arma autonomi (2018/2752(RSP)).

[160] Trattasi di piccoli missili pilotati da remoto, che vengono fatti esplodere direttamente sull'obiettivo militare, ostacolando l'utilizzo di sistemi antirazzo. Si veda S. Valesini, *Nella guerra in Ucraina scendono in campo le armi autonome*, in *Wired Italia*, 24 marzo 2022, disponibile a questo link: https://www.wired.it/article/ucraina-guerra-armi-autonome-droni-kamikaze-robot-killer-russia/.

[161] Così la Risoluzione del 12 settembre 2018 sui sistemi d'arma autonomi (2018/2752(RSP)); vd. ad esempio il sistema *Iron Dome* utilizzato da Israele per intercettare razzi a media velocità e proiettili di artiglieria con traiettoria balistica, vd. I. Orpaz, *How Does the Iron Dome Work*?, in *Haaretz*, 19 novembre 2012, disponibile a questo link: https://www.haaretz.com/2012-11-19/ty-article/.premium/how-iron-dome-works/0000017f-e70b-dc7e-adff-f7af61f70000.

possono controllare autonomamente solo limitati aspetti della propria navigazione aerea, qui l'autonomia riguarda proprio la selezione e l'ingaggio degli obiettivi. Trattasi di funzioni molto delicate dal punto di vista etico perché si tratta di incidere su beni come la vita e l'integrità fisica[162]. L'uso di armi autonome, peraltro, rischia di essere in contrasto con il diritto internazionale umanitario che disciplina le regole che devono essere seguite nel corso dei conflitti armati e prevede norme specifiche per la selezione degli obiettivi militari (cd. *Law of targeting*)[163]. Vi è chi ritiene che, usando sistemi di IA in guerra, si possa ridurre drasticamente il numero di vittime civili[164]. Ma, per come è noto, i sistemi di IA non sono infallibili e commettono errori nell'identificazione e nel riconoscimento di oggetti e persone: il malfunzionamento di un'arma in ambito militare potrebbe risolversi in un evento tragico, cagionando la morte di centinaia o anche migliaia di persone.

E, ovviamente, emerge il problema della attribuibilità del fatto illecito: chi ne risponde? Il programmatore, il produttore, il comandante dell'operazione o chi materialmente preme il bottone o il *display*? Per queste ragioni, molti scienziati hanno proceduto a raccogliere firme per manifestare la necessità che i governi prevengano la corsa agli armamenti intelligenti[165]. Nel 2017, 116 fondatori di importanti imprese del settore robotica e IA hanno inoltrato

---

[162] D. AMOROSO – G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto*, cit., p. 37.

[163] C. GRUT, *The Challenge of Autonomous Lethal Robotics to International Humanitarian Law*, in *Journal of Conflict and Security Law*, vol. 18, iss. 1, 2013, p. 5 ss.; M.N. SCHMITT – J.S. THURNER, *Out of the Loop: Autonomous Weapons Systems and the Law of Armed Conflict*, in *Harvard National Security Journal*, 2013, n. 4, p. 231; in generale, sul targeting vd. M.N. SCHMITT – E. WIDMAR, *The Law of Targeting, Oxford University Press*, 2012; G.D. SOLIS, *The Law of Armed Conflicts, Cambridge University Press*, 2010, p. 519 ss.

[164] Vd. il documento proposto dal governo statunitense nell'ambito del *Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons*, intitolato *Humanitarian Benefits of Emerging Technologies in the Area of Lethal Autonomous Weapons*, 28 marzo 2018.

[165] FUTURE OF LIFE INSTITUTE, *Autonomous Weapons: an Open Letter from Ai & Robotics Researchers*, 9 febbraio 2016, disponibile a questo link: https://futureoflife.org/open-letter-autonomous-weapons.

una lettera aperta alle Nazioni Unite[166]. Vi è stata anche una *Campaign to Stop Killer Robots*[167]. Ad oggi, comunque, non sussiste il divieto di usare questo tipo di armi. Nel 2016, è stato istituito un Gruppo di Esperti Governativi, con il mandato di valutare l'opportunità di regolare e/o vietare le armi autonome. Ad oggi, nonostante siano state proposte normative vincolanti sul punto[168], la Riunione delle Alte Parti contraenti della *Convention on Certain Conventional Weapons* (CCW) è riuscita solo ad adottare principi guida non vincolanti[169].

Questi possono essere sinteticamente riassunti nel modo che segue: il diritto internazionale continua ad applicarsi appieno a tutti i sistemi di armi, SALA compresi; l'uomo deve sempre mantenere il controllo sull'uso dei SALA ed è responsabile per il loro utilizzo; deve essere sempre garantita la sicurezza fisica ed informatica dei SALA; devono essere effettuati test periodici per valutare i rischi; i SALA non devono avere sembianze umane; non si deve ostacolare lo sviluppo di tecnologie a scopi pacifici; la CCW si impegna ad offrire supporto per affrontare lo sviluppo dei SALA nel contesto degli obiettivi e delle finalità della Convenzione, al fine di trovare un equilibrio tra necessità militari e considerazioni umanitarie.

Anche l'UE non ha ancora adottato una normativa vincolante nel settore. Il Parlamento ha approvato due risoluzioni nel 2018[170] e nel 2021[171], in cui esortava l'Alto

---

[166] FUTURE OF LIFE INSTITUTE, *Killer robots: World's top AI and robotics companies urge United Nations to ban lethal autonomous weapons,* disponibile a questo link: https://futureoflife.org/2017/08/20/killer-robots-worlds-top-ai-robotics-companies-urge-united-nations-ban-lethal-autonomous-weapons/.

[167] *Campaign to Stop Killer Robots*, vd. link: https://www.stopkillerrobots.org/.

[168] Le proposte pervenute sono disponibili a questo link: https://meetings.unoda.org/section/ccw-gge-2022_documents_18542_proposals_19869/.

[169] CCW, *Meeting of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects – Final Report*, 13 dicembre 2019, Annex III, disponibile a questo link: https://documents-dds-ny.un.org/doc/UNDOC/GEN/G19/343/64/PDF/G1934364.pdf?OpenElement.

[170] Risoluzione del 12 settembre 2018 sui sistemi d'arma autonomi (2018/2752(RSP)).

[171] Risoluzione del Parlamento europeo del 20 gennaio 2021 sull'intelligenza artificiale: questioni relative all'interpretazione e applicazione del diritto internazionale nella misura in

rappresentante dell'Unione per gli affari esteri e la politica di sicurezza, gli Stati membri ed il Consiglio ad adoperarsi per l'avvio dei negoziati internazionali su uno strumento giuridicamente vincolante che disciplini, anche vietandoli, i sistemi d'arma autonomi letali.

Con riferimento agli USA, i sistemi in parola sono regolati da una direttiva del 2012 redatta dal Dipartimento di Difesa[172]. Tale normativa non vieta lo sviluppo o l'utilizzo di detti sistemi, ma ne subordina l'impiego alla presenza di una serie di requisiti. Tra di essi, si stabilisce che i *LAWS* devono rispettare il diritto bellico e le convenzioni internazionali, che devono essere sviluppati in modo da permettere agli operatori di esercitare livelli appropriati di giudizio umano sull'uso della forza letale; che devono essere sottoposti a continui test di funzionamento; che ogni cambiamento di sistema operativo impone un nuovo *test* di valutazione.

Nonostante non sia facile individuare precisamente le tecnologie impiegate in ambito militare, molte fonti affermano che si sarebbero già verificati alcuni casi di utilizzo di sistemi d'arma autonomi in contesti militari. Ad esempio, nel report redatto dalla Commissione di Esperti sulla Libia, nominata dal Consiglio di Sicurezza dell'Onu, si legge che nel 2020 un drone militare turco avrebbe individuato e colpito automaticamente un obiettivo militare libico, in assenza di input o supervisione umana[173]. Così come, stando ad alcune testate, nella guerra in Ucraina, ambedue le truppe

cui l'UE è interessata relativamente agli impieghi civili e militari e all'autorità dello Stato al di fuori dell'ambito della giustizia penale (2020/2013(INI))

[172] DODD 3000.09, *Autonomy in Weapon Systems*, 21 novembre 2012, modificata nel 2017, disponibile a questo link: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf. La Direttiva sarà probabilmente modificata tra il 2022 e il 2023, proprio per adattarsi alle più recenti evoluzioni tecnologiche nell'ambito degli *AI-enabled autonomous systems*, vd. G.C. Allen, *DOD Is Updating Its Decade-Old Autonomous Weapons Policy, but Confusion Remains Widespread*, in *Center for Strategic & International Studies*, 6 giugno 2022, disponibile a questo link: https://www.csis.org/analysis/dod-updating-its-decade-old-autonomous-weapons-policy-confusion-remains-widespread.

[173] Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council (UN S/2021/229), 8 marzo 2021, § 63 ss., disponibile a questo link: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N21/037/72/PDF/N2103772.pdf?OpenElement.

starebbero facendo uso di armi autonome dotata di sistemi di IA[174].

5.1.2.8. *Robot* nel sistema sanitario (cd. *Healthcare robots*).

Negli ultimi decenni, soprattutto nel settore della chirurgia, si è assistito ad un forte incremento dell'uso di dispositivi robotici in grado di compiere movimenti molto più precisi di quelli dell'essere umano, nonché di svolgere interventi molto delicati, consentendo addirittura al medico di operare da remoto[175].

Anche per detto settore, si è tentato di creare una classificazione dei livelli di autonomia dei sistemi robotici[176]. Al livello 0 si collocano i *robot* che si limitano ad eseguire i comandi dell'operatore umano; al livello 1 i *robot* in grado di correggere o condizionare l'agire umano; al livello 2 quelli

---

[174] R.F. Trager – L.M. Luca, *Killer Robots Are Here—and We Need to Regulate Them*, in *Foreign Policy,* 11 maggio 2022, disponibile a questo link: https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation/. W. Knight, Russia's Killer Drone in Ukraine Raises Fears About AI in Warfare, in Wired US, 17 marzo 2022, disponibile a questo link: https://www.wired.com/story/ai-drones-russia-ukraine/. Vd. però G.C. Allen, *Russia Probably Has Not Used AI-Enabled Weapons in Ukraine, but That Could Change*, in *Center for Strategic & International Studies*, 26 maggio 2022, disponibile a questo link: https://www.csis.org/analysis/russia-probably-has-not-used-ai-enabled-weapons-ukraine-could-change, che contesta il fatto che vi siano prove dell'utilizzo effettivo di *autonompus weapons*, pur non escludendo che la Russia possa averne la disponibilità.

[175] Sulla questione della responsabilità nella robotica chirurgica vd. E. FOSCH-VILLARONGA, *Robots, Healthcare, and the Law. Regulating Automation in Personal Care*, Routledge, 2020; F. GRIFFIN, *Artificial Intelligence and Liability in Health Care, in 31 Health Matrix: Journal of Law-Medicine*, 2021, p. 65 ss.; M. ROE*, Who's Driving That Car?*, cit., p. 340 ss.; G. GUERRA, *Evolving artificial intelligence and robotics in medicine, evolving European law. Comparative remarks based on the surgery litigation*, in *Maastricht Journal of European and Comparative Law*, 2021, Vol. 28(6), p. 805 ss. Nella dottrina italiana, U. RUFFOLO, *L'Intelligenza artificiale in sanità: dispositivi medici, responsabilità e "potenziamento",* in Giur. it., febbr. 2021; M. BASSINI – L. LIGUORI – O. POLLICINO, *Sistemi di Intelligenza Artificiale, responsabilità e accountability. Verso nuovi paradigmi?,* in F. Pizzetti (a cura di), *Intelligenza artificiale*, cit., p. 349 ss.; C. IAGNEMMA, *I 'robot medici': profili problematici in tema di alleanza terapeutica e di responsabilità penale*, in Corti supreme e salute, 2020, n. 2, p. 442 ss.; in generale, sul quadro europeo in materia di device medico-sanitari vd. A. PISANI TEDESCO, *Il nuovo quadro normativo europeo dei dispositivi medici*, in Dir. comm. internaz., 2022, fasc. 3, p. 675 ss.

[176] Vd. G.Z. YANG, *Medical robotics-regulatory, ethical and legal considerations for increasing levels of autonomy*, in *Science Robotics*, 2017, 1 ss.; per un commento più approfondito alla classificazione vd. D. AMOROSO – G. TAMBURRINI*, I sistemi robotici ad autonomia crescente tra etica e diritto*, cit., p. 47 ss.

che eseguono compiti designati dall'operatore umano sotto la sua supervisione; al livello 3 quelli che generano strategie di esecuzione dei compiti sotto la supervisione umana; ai livelli 4 e 5 si trovano i sistemi con maggiore sofisticazione in grado di compiere un'intera procedura chirurgica con o senza la supervisione umana.

Questi ultimi due livelli forniscono prospettive future poiché, attualmente, i modelli usati in ambito sanitario necessitano di un controllo completo e costante da parte del chirurgo. Gli attuali sistemi sono tecnicamente telecomandati dal chirurgo. Il *robot* si limita a tradurre in movimenti precisi ed in tempo reale i comandi forniti dal medico tramite la manipolazione di un controller, eliminando il tremore dei movimenti della mano[177]. Uno dei prodotti maggiormente sviluppati è il *robot* Da Vinci[178], di livello 0. Diversi *robot* usati in sala operatoria hanno raggiunto il livello 1: sono quei sistemi con *software* che assistono il chirurgo nel movimento del *controller*, eventualmente correggendolo ed impedendogli di svolgere determinate manovre che potrebbero risultare pericolose per il paziente[179]. Nelle sale operatorie, vengono anche usati dispositivi di livello 2: il sistema ROBODOC, ad esempio, è in grado di eseguire piani preoperatori relativi al taglio di ossa, sotto supervisione del medico, che può sempre disattivare il sistema. Così come sono già in uso dispositivi con autonomia di livello 3: il sistema STAR, ad esempio, raggiunge questo livello di autonomia quando l'attività concerne le strategie di anastomosi[180]. Per i sistemi di livello 0 ED 1, la macchina è un semplice strumento nelle mani del

[177] Vd. la pagina di Ucla Health, *What Is Robotic Surgery?,* disponibile a questo link: https://www.uclahealth.org/robotic-surgery/what-is-robotic-surgery, in cui il controller viene paragonato ad un joystick per videogames.

[178] Sul suo funzionamento vd. più diffusamente G. GUERRA, *Evolving artificial intelligence and robotics in medicine*, cit., passim; per una panoramica sui device medici basati su sistemi di i.a. attualmente in commercio vd. C. IAGNEMMA, *I 'robot medici': profili problematici in tema di alleanza terapeutica e di responsabilità penale*, cit., p. 442 ss., cui si rinvia anche per l'ampia bibliografia scientifica citata.

[179] Così D. AMOROSO – G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto*, cit., p. 48.

[180] Così D. AMOROSO – G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto*, cit., p. 49.

medico non in grado di compiere attività imprevedibili da parte dell'utilizzatore e del produttore. Dunque, in caso di danni all'integrità fisica, ne risponderanno alternativamente o congiuntamente il medico e/o il produttore. Con riferimento al medico, chiaramente, troverà applicazione il regime di favore di cui all'art. 590-*sexies* c.p. per la responsabilità colposa del medico[181]. Nel caso di danno verificatosi a causa di un difetto del device, sarà il produttore ad essere chiamato a rispondere.

Negli USA, la giurisprudenza è già stata chiamata ad occuparsi in diverse occasioni della responsabilità del produttore per danni causati da malfunzionamenti del robot medico[182]. In diversi casi, il dispositivo Da Vinci si è bloccato nel corso di un intervento, costringendo il chirurgo a continuare le operazioni in modo tradizionale: a seguito dell'operazione, i pazienti hanno manifestato problemi di salute, come ustioni, forti dolori e disfunzioni. Nella maggioranza dei casi, i ricorsi presentati contro la struttura ospedaliera e la società produttrice (*Intuitive Surgical Inc*.) sono stati rigettati poiché mancava la prova del nesso causale tra i danni subiti dal paziente ed il presunto malfunzionamento del *robot*. In un solo caso, *Taylor v. Intuitive Surgical Inc.*[183]*,* la Corte Suprema dello Stato di Washington ha stabilito la responsabilità del produttore per il danno arrecato al paziente, sostenendo che la società non aveva adempiuto all'obbligo di avvertire l'acquirente dei pericoli inerenti all'uso del prodotto. A ben vedere, la casistica inerente al dispositivo Da Vinci pare riprodurre i medesimi problemi della responsabilità per danno da prodotto. Mentre, con riferimento ai sistemi di livello 2 e 3, potrebbero sorgere questioni nuove, legate alla capacità di

[181] C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?,* cit., p. 1757.
[182] Per un'analisi più approfondita della casistica rinviamo a G. GUERRA, *Evolving artificial intelligence and robotics in medicine*, cit., p. 815 ss.
[183] Supreme Court of Washington, Taylor v. Intuitive Surgical, Inc., Case: 92210-1, 9 febbraio 2017, disponibile a questo link: https://case-law.vlex.com/vid/Taylor-v-Intuitive-Surgical-Inc-020917-WASC-92210-1-665020489.

svolgere autonomamente – ma sempre sotto la supervisione umana – talune attività.

Il chirurgo che supervisiona la macchina ha competenze specialistiche che dovrebbero consentirgli di comprendere se e quando il robot presenta malfunzionamenti e d intervenire di conseguenza, disattivando il dispositivo e continuando a lavorare in modo tradizionale. Questo rende le problematiche tale settore di meno difficile risoluzione rispetto a quello dell'*autonomous vehicles*. In quest'ultimo, il conducente non sempre è posto nelle condizioni di poter interpretare gli output del veicolo in modo tempestivo: basti pensare ai robot di livello 2, laddove l'utilizzatore deve essere pronto a riprendere il controllo della guida in ogni momento, senza che vi sia nemmeno un avvertimento sonoro da parte del sistema di guida autonomo. Il conducente di un *autonomous vehicle* potrebbe rendersi conto di un malfunzionamento quando ormai è tardi[184]. Inoltre, il contesto in cui operano gli AV è imprevedibile: il traffico stradale è basato su una interazione costante con gli altri utenti della strada; vi sono innumerevoli segnali stradali da interpretare. Mentre gli *health-care robots* si muovono in un contesto abbastanza prevedibile e con un compito specifico senza alcuna interazione con persone e/o macchine.

6. I *non-embodied systems.*

6.1. I *bots.*

I *bots* altro non sono che programmi *software* basati su sistemi di IA normalmente definiti per contrapposizione con i *robots* che sono caratterizzati dall'incorporazione del *software* in un *hardware*[185]. Tali *software* sono oggi diffusamente

[184] Vd. M. ROE*, Who's Driving That Car?,* cit., p. 339.Si legga, ad esempio, la testimonianza di una conducente di un *autonomous vehicle* coinvolto in un incidente, che alla polizia intervenuta sul posto ha riferito: «When I popped up and I looked and saw a black truck – it happened so fast», in N.E. Boudette, *'It Happened So Fast': Inside a Fatal Tesla Autopilot Accident*, in *The New York Times*, 17 agosto 2021, disponibile al seguente link: https://www.nytimes.com/2021/08/17/business/tesla-autopilot-accident.html.
[185] G. SCOPINO, *Algo Bots and the Law: Technology, Automation, and the Regulation of Futures and Other Derivatives, Cambridge University Press*, 2020, p. 37.

utilizzati in ambito societario per il servizio clienti, per rispondere online ai quesiti più semplici posti dagli utenti (*chatbots*). Ma è il settore dei *social networks* quello in cui si è assistito alla creazione su larga scala di *social bots*, cioè di account sui *social* in grado di produrre contenuti in maniera autonoma e di emulare le interazioni umane. Nella versione più rudimentale, detti *social bots* sono in grado di ripubblicare automaticamente i *post* che contengono determinate parole o *hashtags*; di rispondere a post che soddisfano un certo criterio; di seguire qualsiasi utente che abbia seguito il bot o che abbia pubblicato una specifica parola chiave; di effettuare ricerche sui motori di ricerca di articoli/notizie in base a criteri specifici e di eseguire una ripubblicazione automatica[186]. Purtroppo, i *social bots* sono stati resi famosi dall'uso che ne è stato fatto durante le elezioni statunitensi del 2016 a fini di propaganda politica. I *bots*, infatti, sono in grado di amplificare a dismisura la diffusione di notizie false o tendenziose, ripubblicando costantemente post che contengono specifiche parole chiave ed alterando, così, il dibattito politico pubblico. L'uso dei *bots* a fini di destabilizzazione e polarizzazione del dibattito pubblico è oramai vicenda ben studiata e nota dalla letteratura[187].

Nonostante il fenomeno della diffusione di notizie false o tendenziose durante le elezioni sia sempre esistito, il fenomeno sotto la nostra lente presenta fattori di pervasività

---

[186] Anche A. BESSI – E. FERRARA, *Social bots distort the 2016 U.S. Presidential Election Online Discussion*, in *First Monday*, vol. 21, n. 11, 2016.

[187] P.N. HOWARD – S.C. WOOLEY, *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media, Oxford University Press*, 2018; J.D. OHLIN, *Election Interference. International Law and the Future of Democracy, Cambridge University Press*, 2020; ID., *Did Russian Cyber Interference in the 2016 Election Violate International Law?*, in *Texas Law Review*, 2017, p. 1579; vd. anche i *report* pubblicati dal *Programme on Democracy & Technology (Oxford Internet Institute),* disponibili a questo link: https://navigator.oii.ox.ac.uk/what-is-comprop/; per un approccio penalistico vd. la monografia di T. GUERINI, *Fake news e diritto penale*, Giappichelli, 2020, e la sezione monografica di Dir. pen. cont. – Riv. trim., 2021, fasc. 4, dedicata al tema «La disinformazione tra politica e diritto. Dimensione istituzionale, strategie preventive e dinamiche punitive»; sulla possibile qualificazione – dal punto di vista del diritto internazionale – delle presunte influenze cyber straniere sui processi democratici vd. M.N. SCHMITT, *Foreign Cyber Interference in Elections*, Int'l L. Stud., 2021, p. 739; S. WHEATLEY, *Foreign Interference in Elections under the Non-intervention Principle: We Need to Talk about "Coercion"*, in *Duke Journal of Comparative & International Law*, 2020, p. 161.

ed ampiezza del tutto nuovi. Secondo alcuni studiosi, ciò ha inciso sulle scelte degli elettori: dei ricercatori in *data-science* hanno calcolato che, durante le elezioni USA del 2016, circa il 15% degli *account* che twittava su argomenti di politica era costituito da *bots*, responsabili del 19% della totale conversazione su *Twitter*[188]. Fortunatamente, negli ultimi anni si è registrato un uso molto minore dei social *bots* di propaganda elettorale. Da un lato questa diminuzione sembrerebbe da collegare ai numerosi richiami all'attenzione da parte delle istituzioni sul tema[189]; dall'altra, i *social bots* potrebbero essere diventati più difficili da riconoscere, visto l'incremento nelle capacità di simulazione del linguaggio umano, tramite tecniche di *natural language processing*[190]. Per i *social bots* meno sofisticati che si limitano ad effettuare una serie di azioni sulla base delle istruzioni fornite in sede di programmazione non si creano problemi di attribuzione dell'eventuale fatto illecito al programmatore o all'utilizzatore. I *bots* che usano sistemi di *machine learning*, invece, creano maggiori problemi.

Per esempio, il *social bot* creato da Microsoft nel 2016, @*TayTweets*, era programmato per imitare i modelli di linguaggio di una ragazza americana di 19 anni e per imparare ad esprimersi grazie all'interazione con gli utenti

[188] A. BESSI – E. FERRARA*, Social bots distort the 2016 U.S. Presidential Election Online Discussion*, cit.

[189] In particolare, quanto alle iniziative promosse in ambito europeo, la Commissione europea ha incardinato la strategia di contrasto alla disinformazione online sul sistema di autodisciplina previsto dal *Code of Practice on Disinformation*, un codice di buona condotta, adottato a settembre 2018 e rivisto nel 2022, al quale le piattaforme digitali aderiscono spontaneamente e al quale, ad oggi, hanno aderito i maggiori player della rete. Tra le pratiche di disinformazione online che gli aderenti si impegnano a contrastare vi è anche «*The creation and use of fake accounts, account takeovers and bot-driven amplification*», vd. COM, *The Strengthened Code of Practice on Disinformation*, 16 giugno 2022, Commitment 14. Dal momento che non esistono, ad oggi, criteri vincolanti per qualificare un contenuto come "disinformazione", non sembra che un presunto contenuto disinformativo possa rientrare tra quei "contenuti illegali" la cui conoscenza dà luogo ad un obbligo, per il prestatore di servizio online, di agire «immediatamente per rimuovere i contenuti illegali o per disabilitare l'accesso agli stessi» (Proposta di Regolamento del Parlamento europeo e del Consiglio relativo a un mercato unico dei servizi digitali (legge sui servizi digitali) e che modifica la direttiva 2000/31/CE, art. 5, co. 1, lett. b).

[190] Vd. intervista al data scientist Emilio Ferrara (G. Guglielmi, The next-generation bots interfering with the US election, 28 ottobre 2020), disponibile a questo link: https://www.nature.com/articles/d41586-020-03034-5.

umani su *Twitter*[191]. Dopo soli due giorni dall'attivazione, il *bot* era stato sollecitato da interazioni aggressive e provocatorie derivanti da taluni utenti e, per tale ragione, aveva iniziato a pubblicare *tweet* razzisti e discriminatori come <<*I fucking hate negros, I wish we could put them all in a concentracion camp*>>. *Microsoft* era stata costretta, quindi, a ritirare l'account e ad eliminare i contenuti più discriminatori.

Un altro esempio è il *social bot* @jeff_ebooks: progettato da uno sviluppatore di servizi online olandese, generava contenuti basandosi su un "*markov chain generator*", cioè su un *software* che consente di elaborare frasi di senso compiuto a partire da un testo fornito[192]. Tale *bot* usava, come testo di partenza, frammenti di *tweet* pubblicati dallo stesso creatore dell'*account* automatico. Partendo da singole parole, il *bot* elaborava nuove frasi di senso compiuto. La polizia olandese aveva aperto un fascicolo nei confronti dello sviluppatore, dopo che il *bot* aveva tweettato di trovarsi ad un evento di moda ad Amsterdam e di avere l'intenzione di uccidere.

In ambedue i casi, i sistemi di IA hanno agito con comportamenti non prevedibili e non predeterminati a priori dai programmatori[193], così interrompendo il rapporto diretto tra programmatore e comportamento dei *bot*.

Questioni simili si sono riproposte con riferimento a bot di acquisto automatico online, come quelli che acquistano un numero enorme di biglietti per concerti online da utilizzare a fini di bagarinaggio[194].

---

[191] R. Price, Microsoft is deleting its AI chatbot's incredibly racist tweets, in Business Insider, 24 marzo 2016, disponibile a questo link: https://www.businessinsider.in/Microsoft-is-deleting-its-AI-chatbots-incredibly-racist-tweets/articleshow/51539858.cms.

[192] A. Hern, Randomly generated tweet by bot prompts investigation by Dutch police, in The Guardian (online), 12 febbraio 2015, disponibile a questo link: https://www.theguardian.com/technology/2015/feb/12/randomly-generated-tweet-by-bot-investigation-dutch-police.

[193] Secondo alcuni studiosi, tuttavia, quantomeno il comportamento di Tay avrebbe potuto essere previsto, stante la nota presenza di troll su internet. In particolare, Microsoft avrebbe potuto filtrare i dati di input, escludendo, ad esempio, i tweet che contenevano linguaggio offensivo; vd. J. SINGH E AL., Responsibility & Machine Learning, cit., p. 12.

[194] F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 26.

6.2. *Algorithmic trading systems*.

Con tale espressione, si intendono i sistemi informatici in grado di effettuare autonomamente ordini di acquisto sulle piattaforme di *trading*. È l'algoritmo a decidere la tempistica, il prezzo e la quantità dell'ordine; nella maggior parte dei casi, può anche avviare l'ordine in assenza di intervento umano[195]. Una particolare categoria di *automated trading systems* sono gli *high frequency trading* (HFT): si tratta di algoritmi che consentono di effettuare negoziazioni ad altissima velocità, rendendo possibile perfezionare diverse migliaia di ordini al secondo[196]. Questo beneficio è assicurato dall'abbattimento del periodo di latenza, cioè del lasso di tempo che intercorre tra la decisione economica e l'effettiva esecuzione dell'operazione sul mercato, che si attesta oggi intorno a pochi microsecondi[197]. Lo scopo dell'esecuzione di un elevatissimo numero di operazioni in un breve intervallo

[195] Questa la definizione fornita dalla Direttiva 2014/65/Ue del Parlamento Europeo e del Consiglio del 15 maggio 2014 relativa ai mercati degli strumenti finanziari e che modifica la direttiva 2002/92/CE e la direttiva 2011/61/UE (c.d. MIFID II): «negoziazione di strumenti finanziari in cui un algoritmo informatizzato determina automaticamente i parametri individuali degli ordini, come ad esempio se avviare l'ordine, i tempi, il prezzo o la quantità dell'ordine o come gestire l'ordine dopo la sua presentazione, con intervento umano minimo o nullo e non comprende i sistemi utilizzati unicamente per trasmettere ordini a una o più sedi di negoziazione, per trattare ordini che non comportano la determinazione di parametri di trading, per confermare ordini o per eseguire il trattamento post-negoziazione delle operazioni eseguite» (Art. 4(1)(39)). Vd. anche G. SCOPINO, Algo Bots and the Law, p. 17.

[196] Questa la definizione fornita dalla Direttiva MIFID II: «qualsiasi tecnica di negoziazione algoritmica caratterizzata da: a) infrastrutture volte a ridurre al minimo le latenze di rete e di altro genere, compresa almeno una delle strutture per l'inserimento algoritmico dell'ordine: co-ubicazione, hosting di prossimità o accesso elettronico diretto a velocità elevata; b) determinazione da parte del sistema dell'inizializzazione, generazione, trasmissione o esecuzione dell'ordine senza intervento umano per il singolo ordine o negoziazione, e c) elevato traffico infragiornaliero di messaggi consistenti in ordini, quotazioni o cancellazioni» (Art. 4(1)(40)). La velocità di tali algoritmi sarebbe da imputare non solo alla sofisticata struttura del software e alla potenza computazionale degli hardware, ma anche alla peculiare collocazione fisica di questi ultimi, posti in prossimità del server dei mercati finanziari presso i quali gli algoritmi operano, così da ridurre radicalmente lo spazio che l'impulso elettrico deve percorrere per giungere a destinazione (cd. co-location), vd. A. PUORRO, High Frequency Trading: una panoramica, in Questioni di Economia e Finanza (Occasional Papers) – Banca d'Italia, n. 198, 2013, p. 11.

[197] vd. J. ARIFOVIC – X. HE – L. WEI, *Machine learning and speed in high-frequency trading,* in *Journal of Economic Dynamics & Contro*l, 139, 2022, p. 1; vd, A. PUORRO, High Frequency Trading, cit., p. 10-11, che – significativamente – nel 2012 parla di "millisecondi" e non di "microsecondi", a testimonianza della continua evoluzione tecnologica ci questi sistemi; a Puorro si rimanda anche per la descrizione delle diverse fasi che consentono di trasformare la decisione economica nell'esecuzione dell'operazione.

di tempo è quello di fruttare le minime oscillazioni di prezzo dei titoli che sarebbero impercettibili per un tradizionale investitore. Gli esigui margini di profitto relativi alla singola operazione, se considerati in rapporto alla enorme mole di operazioni, assumono una portata considerevole[198]. Si determina così un reale vantaggio dell'operatore algoritmico rispetto all'investitore umano[199].

Dal punto di vista del diritto criminale, ciò che rileva non è tanto la velocità delle operazioni, quanto l'autonomia operativa degli algoritmi stessi rispetto alle persone che interagiscono con essi. I sistemi di *algorithmic trading* effettuano continuamente scelte autonome di investimento, decidendo di volta in volta tra diverse opzioni possibili. I più recenti algoritmi sfruttano anche tecniche di ML e sono, pertanto, in grado di apprendere dall'ambiente in cui operano e modificare le proprie strategie operative adattandosi al nuovo contesto: l'HFT potrà effettuare una decisione di investimento che non era prevedibile dal programmatore al momento dell'immissione nel mercato[200].

Dunque, anche qui, la questione più rilevante per il penalista è comprendere se i tradizionali meccanismi imputativi della responsabilità penale siano adeguati a regolare anche i casi in cui il fatto illecito scaturisca non da una condotta umana, bensì da un "comportamento" algoritmico. È pacifico che fatti di manipolazione del mercato, previsti e puniti dall'art. 185 T.u.f.[201], possano essere commessi tramite l'uso di algoritmi di *trading*. È stato proprio l'emergere degli HFT a far sorgere nuovi schemi di turbativa

---

[198] G. STRAMPELLI, *L'informazione societaria a quindici anni dal T.U.F.: profili evolutivi e problemi*, in *Riv. soc.,* 2014, p. 996; A.F. TRIPODI, Abusi di mercato e trading algoritmic, cit., p. 747-748.

[199] F. CONSULICH*, Il nastro di Möbius*, cit., p. 199-200; vd. anche A.F. TRIPODI, *Abusi di mercato e trading algoritmic*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 745 ss.

[200] J. ARIFOVIC – X. HE – L. WEI, *Machine learning and speed in high-frequency trading*, cit., p. 6.

[201] Art. 185, co. 1, D.lgs. 24 febbraio 1998, n. 58: «Chiunque diffonde notizie false o pone in essere operazioni simulate o altri artifizi concretamente idonei a provocare una sensibile alterazione del prezzo di strumenti finanziari, è punito con la reclusione da uno a sei anni e con la multa da euro ventimila a euro cinque milioni».

quali il *pinging*, lo *smoking*, il *layering*, lo *spoofing*. Comune a tutti questi è l'immissione e la contestuale cancellazione di migliaia di ordini al fine di far apparire le condizioni di mercato differenti da quelle effettive[202]. Si tratta di strategie che sembrano integrare a pieno titolo quegli "*altri artifici concretamente idonei a provocare una sensibile alterazione del prezzo di strumenti finanziari*" punita dall'art. 185 T.u.f.

La capacità offensiva degli HFT si è notata anche nei cd. *flash crash*, cioè in quei fenomeni di crollo improvviso dei prezzi, seguito da un rapido ritorno alla preesistente situazione di equilibrio, senza che tale oscillazione possa trovare una giustificazione nel mutamento del valore del titolo o in altre ragion attinenti alla struttura del mercato di riferimento. <<*Il flash crash sembra, quindi, avere una natura meccanica, vale a dire indipendente da fattori economici reali ed, invece, molto più legata alle procedure tecniche di funzionamento degli scambi dal punto di vista informatico*>>[203]. Il primo *flash crash* è il crollo registrato il 6 maggio 2010, quando l'indice *Dow Jones* scese di oltre l'8% nell'arco di pochi minuti, recuperando tutte le perdite nei dieci minuti successivi: un evento senza precedenti.

Sono proprio gli HFT ad alimentate queste fluttuazioni dei titoli perché gli HFT, di fronte alla picchiata dei prezzi, sono i primi a vendere e, subito dopo, a ricomprare, anticipando gli investitori fisici[204]. Enigmatico, al riguardo, è il *flash crash* verificatosi il 23 aprile 2013 sui mercati americani quando la notizia, poi scoperta falsa, di un attentato ad Obama aveva indotto gli HFT a vendere in massa titoli finanziari, intuendo un ribasso sui mercati; il rialzo che si verificò nei minuti successivi fu causato dagli stessi HFT che ricominciarono gli

[202] Così F. CONSULICH, *Il nastro di Möbius*, cit., p. 210; per la distinzione tra le varie tecniche si rinvia a G. RUTA, *I.a. nei reati economici e finanziari*, in *Intelligenza artificiale e giurisdizione penale*, cit., p. 65 ss.
[203] F. CONSULICH, *Il nastro di Möbius*, cit., p. 201.
[204] F. CONSULICH, *Il nastro di Möbius*, cit., p. 202; A.F. TRIPODI, *Abusi di mercato e trading algoritmic*, cit., p. 748.

acquisti, maturando complessivamente 600 milioni di dollari di profitti[205].

Pare chiaro che siffatti operatori algoritmici non possano essere considerati meri strumenti nelle mani dell'uomo, posta la loro capacità di *decision-making* e stante l'adattività e l'interconnettività con l'ambiente in cui si trovano con gli altri agenti virtuali.

7. Sistemi che cagionano un'offesa indiretta ai beni giuridici.

7.1. Sistemi sanitari.

Si è già detto che uno dei settori dove si può notare maggiormente lo sviluppo e l'uso dei sistemi di IA è quello medico. Infatti, in molti ospedali sono usati diversi sistemi di supporto alla diagnostica ed al trattamento in grado di analizzare in modo veloce ed accurato un'immensa mole di dati concernenti la *medical knowledge* e la storia clinica del paziente, in modo tale da fornire indicazioni diagnostiche e terapeutiche ai medici competenti. Si chiamano *Clinical Decision Support Systems* (CDSS) e sono molto accurati soprattutto nel campo della radiologia, dell'oncologia e della dermatologia[206].

IBM Watson viene visto come uno dei più promettenti sistemi nel settore medico. Alcuni esperimenti condotti presso lo *Sloan-Kettering Hospital* negli USA suggeriscono che le diagnosi compiute da *Watson* sono molto più accurate di quelle effettuate dai medici. La media di diagnosi di cancro corrette sarebbe del 90% per *Watson* rispetto al 50% dei

[205] F. CONSULICH, *Il nastro di Möbius*, cit., p. 202 e bibliografia ivi citata.

[206] La letteratura medica è ormai molto ampia. Per limitarci al solo settore oncologico, vd. D. WANG E AL., *Deep Learning for Identifying Metastatic Breast Cancer*, *arXiv preprint*, 2016, disponibile a questo link: https://arxiv.org/abs/1606.05718; L. FERRANTE DI RUFFANO E AL., *Computer-assisted diagnosis techniques (dermoscopy and spectroscopy-based) for diagnosing skin cancer in adults*, in *Cochrane Database of Systematic Reviews*, 12, 2018; per ulteriori riferimenti bibliografici rinviamo al report pubblicato dal Consiglio Superiore di Sanità, I sistemi di i.a. come strumento di supporto alla diagnostica, 2021, p. 18, disponibile a questo link: https://www.salute.gov.it/imgs/C_17_pubblicazioni_3218_allegato.pdf.

medici[207]. Il *database* a disposizione di IBM *Watson* è illimitato perché costituito da tutto quanto risulta accessibile su internet e dalle banche dati sanitarie aggiornate da sistemi sanitari nazionali ed ospedalieri. *IBM Watson* non analizza solo i dati disponibili sui sistemi sanitari nazionali ma anche l'intera letteratura statunitense sul tema che interessa. Dunque, la possibilità di *Watson* di setacciare in breve tempo questo insieme di dati e di ricavarne informazioni rilevanti a fini diagnostico-terapeutici è un enorme vantaggio per i medici[208].

In particolare, i CDSS si basano su tecniche di *supervised learning* perché gli algoritmi imparano ad identificare ed a correlare domande e risposte sulla base degli esempi presenti nell'insieme di addestramento, applicando le conoscenze acquisite nell'analisi delle nuove questioni che vengono loro sottoposte. Poi, gli algoritmi generano diverse soluzioni, ciascuna di esse con un punteggio di confidenza associato. In pratica, il medico potrà inserire nel sistema la sequenza genetica del tumore di un paziente e la sua storia clinica: il sistema integrerà queste informazioni con i vastissimi database che ha a disposizione e genererà una serie di ipotetiche diagnosi o piani terapeutici, assegnando loro specifici punteggi di fiducia e classificando le risposte in base alla probabilità che siano corrette. Così, il sistema parrebbe adottare l'approccio tipico della *evidence-based medicine* che impone di usare le migliori risultanze scientifiche.

Visto che le capacità del CDSS sembrano destinate a migliorare ed a crescere superando quelle dell'essere umano, ci si chiede che ruolo debba avere il medico nel processo decisionale: deve essere un mero esecutore delle istruzioni che riceve dall'algoritmo o avrà l'ultima parola su diagnosi e

[207] F. LAGIOIA – G. CONTISSA, *The Strange Case of Dr. Watson: Liability Implications of AI Evidence-Based Decision Support Systems in Health Care*, in *European Journal of Legal Studies*, vol. 12, n. 2, 2020, p. 255; per ulteriori studi sull'accuratezza dei CDSS rinviamo ai dati raccolti nel report del Consiglio Superiore di Sanità, *I sistemi di i.a. come strumento di supporto alla diagnostica*, cit., p. 18-19.

[208] J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit., p. 108.

trattamento da applicare al caso concreto?[209]. la maggioranza dei testi ritiene che il medico debba continuare ad essere il decisore finale, basandosi sul principio dell'*human in the loop*[210].

Ciò in quanto, i CDSS usano tecniche statistico-predittive basate su enormi quantità di dati, la cui qualità ed accuratezza non è sempre vagliabile dal sistema. Dunque, le competenze del medico ed il buon senso da essere umano dovrebbero portare a non accettare quelle raccomandazioni totalmente implausibili. Il ruolo del medico è fondamentale anche nel considerare fattori come un odore particolare o la consistenza al tatto della pelle del paziente o di un nevo[211]. La definitiva scomparsa del medico si troverebbe in aperto contrasto con la necessità che la relazione medico-paziente sia basata sulla fiducia e sulla trasparenza, nonché sul consenso informato <<*nel quale si incontrano l'autonomia decisionale del paziente e la competenza, l'autonomia professionale e la responsabilità del medico*>>[212].

---

[209] Il tema è già oggetto di ampie riflessioni. Ci limitiamo, in questa sede, a rinviare ai principali contributi: W. N. PRICE II, *Regulating Black-Box Medicine*, 116 Mich. L. Rev., 2017, p. 421; J.S. ALLAIN, *From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artifificial Intelligence Systems*, in La. L. Rev., 2013, p. 1049; F. LAGIOIA – G. CONTISSA, *The Strange Case of Dr. Watson*, cit., p. 245; A.M. FROOMKIN E AL., *When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning*, in 61 Ariz. L. Rev. 33, 2019, p. 33; J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit.; S. GERKE E AL., *Ethical and legal challenges of artificial intelligence-driven healthcare*, in A. Bohr – K. Memarzadeh (eds), *Artificial intelligence in healthcare*, Academic Press, 2020, p. 295 ss.; in lingua italiana vd. A. PERIN, *Standardizzazione, automazione e responsabilità medica. Dalle recenti riforme alla definizione di un modello d'imputazione solidaristico e liberale*, in BioLaw Journal, 2019, n. 1, p. 207; U. RUFFOLO, L'Intelligenza artificiale in sanità, cit., p. 502; L. SCAFFARDI, *La medicina alla prova dell'Intelligenza Artificiale*, in *DPCE Online*, Vol. 51, N. 1, apr. 2022, p. 349.

[210] In questo senso A.M. FROOMKIN E AL., *When AIs Outperform Doctors*, cit.; A. PERIN, *Standardizzazione, automazione e responsabilità medica*, cit., p. 230. Quanto alle fonti istituzionali, vd. lo studio commissionato dal Parlamento europeo, *Artificial intelligence in healthcare. Applications, risks, and ethical and societal impacts*, giugno 2022, p. 49, disponibile a questo link: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf, che richiede che sia sempre previsto «a human-in-the-loop mechanism to enable for human oversight and feedback».

[211] F. LAGIOIA – G. CONTISSA, *The Strange Case of Dr. Watson*, cit., p. 275.

[212] Art. 1, co. 2, legge 22 dicembre 2017, n. 219, recante norme in materia di consenso informato e di disposizioni anticipate di trattamento

Dunque, pare doversi concludere che il medico debba continuare ad avere un ruolo fondamentale da decisore finale. Ma, ciò detto, resta da comprendere che valore abbiano le previsioni del sistema algoritmico; in che casi sia legittimo affidarsi al sistema ed in che casi non lo sia; a che condizioni il medico dovrà rispondere per aver assunto ed eseguito una decisione erronea proposta da un sistema di IA. Si tratta, a ben vedere, di individuare i confini della responsabilità medica, in un contesto in cui l'autonomia professionale è condizionata dalle indicazioni provenienti da un diverso polo decisionale[213]. Le indicazioni dei CDSS potrebbero in futuro essere vere e proprie linee-guida e si potrebbe prevedere un meccanismo simile a quello dell'art. 590-*sexies* c.p., ma specifica per le particolarità di detti sistemi.

[213] A. PERIN, *Standardizzazione, automazione e responsabilità medica*, cit., p. 231.

CAPITOLO SECONDO

CENNI SUL RAPPORTO TRA IA E RESPONSABILITÀ CIVILE

1.La responsabilità civile come soluzione immediata.; 1.2.La regolazione dell'IA nel contesto europeo e la responsabilità civile.1.3.Occorre creare una normativa *ad hoc* ovvero è sufficiente adattare la normativa vigente?2.Il codice civile ed i regimi speciali di responsabilità oggettiva in capo all'utilizzatore del sistema di IA.; 3.La nozione di prodotto e la sua adeguatezza al contesto dell'IA;4.La nozione di produttore e la sua adeguatezza al contesto dell'IA.;5.La nozione di difetto e la sua adeguatezza al contesto dell'IA.;5.1. La classificazione dei difetti: difetti di fabbricazione, progettazione, informazione nei sistemi di IA.;5.1.2. *Focus* sui difetti di progettazione: responsabilità per colpa o assoluta?;5.1.3.Il difetto di progettazione in relazione ai sistemi di IA.;5.2.Verso nuovi criteri di valutazione del difetto: la proposta di revisione della *Product liability directive*.;5.2.1.Le proposte della dottrina.;5.3.Un'alternativa: l'agente modello umano.;5.4.Un'altra via: l'agente modello algoritmico.;5.5.*Product liability* e *product safety*: verso la *preemption doctrine*.;5.5.1.*Product safety* e tecniche di armonizzazione.;5.5.2.Sistemi di IA e norme di armonizzazione: brevi cenni sull'AI Act.;5.5.2.1.Le pratiche di IA vietate: approfondimento.5.5.3.I sistemi di IA consentiti.;5.5.3.1.I sistemi di IA ad alto rischio.;5.5.4.La tesi della non difettosità del prodotto conforme agli standard tecnici.;5.6.*Product safety* e *product liability*: una separazione concettuale.;5.7.Il nesso di causalità ed il suo accertamento.;5.7.1. Brevi cenni sui profili sostanziali del nesso di causalità.;5.7.2.Aspetti processuali del nesso di causalità.;6.Responsabilità per esercizio di attività pericolose ed i sistemi di IA.;7.C'è bisogno di una personalità elettronica?;8.Serve un'assicurazione obbligatoria?;9.Responsabilità civile ed IA: spunti di riflessione.

1. La responsabilità civile come soluzione immediata.

Vista la flessibilità delle categorie concettuali del danno civile, il dibattito sui danni provocati dall'IA si è sviluppato soprattutto in ambito civilistico. Il criterio di giudizio civilistico del "più probabile che non" nell'accertamento della causalità, così come la esistenza di modelli di responsabilità oggettiva, senza colpa, forniscono strumenti molto più

duttili. Tramite questi ultimi, ad esempio, il rischio che l'azione dei dispositivi intelligenti cagionino danni può essere attribuito al produttore o all'operatore del sistema indipendentemente dall'accertamento della colpa.

Se si adotta la responsabilità aquiliana, la mancata individuazione di un soggetto responsabile per la commissione di un fatto lesivo comporta che il costo derivante dal danno resti in capo a chi l'ha subito. Ciò sia perché potrebbe trattarsi di eventi imprevedibili anche se l'utilizzatore o il produttore avessero usato l'ordinaria diligenza, sia perché spesso il danneggiato non si trova nella possibilità di poter provare la colpa del produttore o dell'utilizzatore. Adottare una responsabilità civile basata sulla colpa contraddirebbe sia la funzione deterrente sia la funzione compensativa della responsabilità civile stessa. La prospettiva di una immunità quasi certa condurrebbe all'assunzione in capo ai produttori di rischi spropositati[214]; così come, con riferimento alla funzione compensativa, ricorrere al criterio della colpa comporterebbe la diminuzione delle probabilità di soddisfazione dei danneggiati, che resterebbero sforniti di tutela.

Per tali ragioni, il dibattito sul tema del danno da algoritmo, si è incentrato sull'applicabilità dei regimi di responsabilità oggettiva esistenti, ma anche sulla necessità di adattare questi regimi o di introdurne di totalmente nuovi, come una assicurazione obbligatoria, una personalità elettronica[215] ovvero la *Manufacturer Enterprise Liability*[216].

---

[214] G. WAGNER, Robot liability, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, Baden-Baden, 2019, p. 30 ss.

[215] Per "personalità elettronica" si intende la creazione di un patrimonio separato in capo al sistema di i.a., destinato in via esclusiva al soddisfacimento degli obblighi risarcitori derivanti dagli eventi lesivi algoritmici, senza che sia necessario l'accertamento di una colpa in capo al produttore o all'utilizzatore del sistema. L'ipotesi di introdurre una personalità elettronica era stata prefigurata dalla Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, cit., § 59, lett. f)

[216] La proposta di *Manufacturer Enterprise Liability* prevede la creazione di un fondo indennitario, finanziato dalle società produttrici di sistemi di i.a., al quale il danneggiato potrebbe accedere qualora dimostri di aver subìto il danno, senza che sia necessaria la prova del difetto del prodotto.

La letteratura civilistica in materia appare, pertanto, unitaria nell'asserire che le responsabilità senza colpa siano quelle da applicare al tema in esame, ferma restando la residualità della responsabilità aquiliana[217].

Nel modello della responsabilità aquiliana tradizionale, il criterio di imputazione è unico ed è fondato sulla colpa. Il costo del danno è sempre trasferito su colui che lo ha cagionato con colpa o con dolo e il giudizio di responsabilità non riguarda altri soggetti al di là del danneggiato e dal danneggiante[218]. Dietro questo modello si colloca la funzione prettamente punitiva attribuita dalla dottrina risalente alla responsabilità aquiliana. Originariamente, infatti, vi era una perfetta coincidenza tra responsabilità civile e penale perché il risarcimento era ammesso solo nei casi in cui la *lex Aquilia* prevedeva la condanna ad una pena[219]. A ben vedere, la fissazione dei criteri di imputazione della responsabilità dipende dal sistema economico della società e dalle strutture del mercato di riferimento[220]. Il principio della colpa come solo criterio imputativo della responsabilità civile traduceva l'approccio del *laissez faire* dello Stato liberale ottocentesco, basato sulla libertà contrattuale e l'assolutezza dei poteri del proprietario[221]. Il principio per cui non potesse sussistere una responsabilità senza colpa sembrava essere richiesto dalla rivoluzione industriale, perché si credeva fosse più vantaggioso trascurare gli interessi degli individui piuttosto che rinunciare agli interessi dell'impresa, addebitandogli i

[217] Vd. per tutti U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 95 ss.; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema*?, in *Oss. dir. civ. comm*., 2020, fasc. 2, p. 445 ss.; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1719; vd., inoltre, gli autori citati nel prosieguo, ai §§ 2. ss.

[218] Per un'ampia trattazione sul tema del "nessuna responsabilità senza colpa", si veda G. ALPA, *La responsabilità civile. Parte generale*, Utet, 2010, al capitolo secondo, p. 25 ss.

[219] vd. G. PONZANELLI, *La responsabilità civile: profili di diritto comparato*, Il Mulino, 1992, p. 61; G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 159-160 e autori ivi citati; C. PIERGALLINI, *"Civile" e "penale" a perenne confronto: l'appuntamento di inizio millennio*, in *Riv. it. dir. proc. pen*., 2012, n. 4, p. 1300.

[220] G. ALPA, *La responsabilità civile. Parte generale*, cit. p. 159-160.

[221] G. ALPA*, La responsabilità civile. Parte generale*, cit., p. 105.

costi degli incidenti inevitabili[222]. Dunque, dietro la scelta della colpa come basa della responsabilità, stava una scelta politica, volta a sottrarre le imprese dal costo degli incidenti arrecati incolpevolmente ai terzi, nella convinzione che costi eccessivamente gravosi avrebbero minato la solidità economica della nascente industria[223].

Però, da ciò derivava l'impossibilità di risarcire la grandissima parte dei danni perché la maggior parte degli eventi lesivi derivanti dallo sviluppo industriale sono danni anomali ed incolpevoli. Per tale motivo, a partire dagli anni Sessanta del Novecento, si sviluppano le prime riflessioni organiche in materia, soprattutto in relazione al danno da prodotto. Tali riflessioni diventarono un punto di riferimento in materia di responsabilità oggettiva[224]. Anche questo cambio di paradigma è stato determinato dall'evolversi delle strutture economico-sociali dominanti, nonché dal definitivo sorgere del *welfare state*. La produzione ed il consumo di massa determinano la fuoriuscita del danno dalle fabbriche, coinvolgendo tutte le fasce della popolazione. Nel frattempo, vi era stata la promulgazione di leggi che obbligavano l'assicurazione per gli infortuni sul lavoro: detto obbligo aveva fatto venire meno le esigenze di tutela che stavano alla base delle teorie oggettivistiche del rischio d'impresa. La vittimizzazione di massa determinata dai danni da prodotto e

[222] F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, *Edizioni Scientifiche Italiane*, 2003, p. 127; G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 106.
[223] G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 107; F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 126.
[224] P. TRIMARCHI, *Rischio e responsabilità oggettiva*, Giuffrè, 1961; S. RODOTÀ, *Il problema della responsabilità civile, Giuffrè*, 1964; M. COMPORTI, *Esposizione al pericolo e responsabilità civile*, Morano Editore, 1965; P. FORCHIELLI, *Intorno alla responsabilità senza colpa*, in Riv. trim. dir. proc. civ., 1967, p. 1394; successivamente vd. anche, limitandoci ai testi pubblicati nel nostro Paese, G. ALPA, *Colpa e responsabilità oggettiva nella prospettiva dell'analisi economica del diritto*, in *Politica del diritto*, 1976, p. 431 ss.; M. CONFORTINI, *Posizioni della dottrina e della giurisprudenza italiana sul tema della responsabilità civile del produttore per la messa in commercio di prodotti (difettosi e) dannosi,* in *Resp. civ*., 1977, p. 544; C. CASTRONOVO, *Sistema e problema del danno da prodotto*, Giuffrè, 1979. Per una ricostruzione delle fasi che hanno accompagnato lo storico passaggio dal dogma della colpa alla responsabilità oggettiva si rinvia a F. DI CIOMMO*, Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 114 ss.

dalla circolazione stradale è stata ciò che ha fornito linfa vitale alla responsabilità oggettiva[225].

Lo Stato mutò ed assunse un ruolo di regolatore e, per tale ragione, a partire dal secondo dopoguerra, il sistema della responsabilità civile, si fece carico dell'obiettivo di istituire criteri di ripartizione dei rischi che consentissero di assicurare la più ampia tutela ai danneggiati e di distribuire le perdite in maniera razionale ed economica[226]. Dunque, il problema dei costi della allocazione del danno da problema individuale diventa un problema sociale ed il *focus* della responsabilità aquiliana si sposta dalla figura del danneggiante all'istituto del danno. Nel nostro Paese, le riflessioni sul superamento della tradizione monolitica della responsabilità civile partono proprio dalla presa d'atto che già nel codice esistevano modelli di ascrizione della responsabilità non colposi. Ci stiamo riferendo alla responsabilità per danno causato da dipendenti di cui all'art. 2049 c.c., da cose in custodia di cui all'art. 2051 c.c., da animali di cui all'art. 2052 c.c., dalla rovina di edifici di cui all'art. 2053 c.c., da autoveicoli di cui all'art. 2054 c.c., nonché del danno cagionato dallo svolgimento di attività pericolose di cui all'art. 2050 c.c.

Già nell'applicazione giurisprudenziale di dette norme, si faceva ricorso a presunzioni di colpa invincibili[227], nascondendosi dietro il dito del sistema di responsabilità colpacentrico. La dottrina individuò negli artt. 2049-2054 c.c. delle ipotesi di responsabilità senza colpa, il cui criterio unificatore veniva identificato nel rischio di impresa[228], nella creazione del pericolo[229] o nel criterio del *cheapest cost avoider*[230].

---

[225] F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 131; P. TRIMARCHI, *Istituzioni di diritto privato*, Giuffrè, XIX ed., 2011, p. 128.
[226] G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 162-163.
[227] F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 133.
[228] P. TRIMARCHI, *Istituzioni di diritto privato*, cit., p. 128.
[229] M. COMPORTI, *Esposizione al pericolo e responsabilità civile*, cit.
[230] G. CALABRESI – J.T. HIRSCHOFF, *Toward a Test for Strict Liability in Torts*, in 81 Yale Law Journal, 1972, p. 1060 ss.; G. CALABRESI., *The Costs of Accidents. A Legal and Economic Analysis, Yale University Press*, 1970, trad. it. Costo degli incidenti e responsabilità civile. Analisi economico-giuridica, Giuffrè, 1975. Secondo l'A., in realtà, il principio del *cheapest cost*

Tale diverso approccio determina una netta recisione con la funzione punitiva che veniva attribuita alla responsabilità aquiliana fino ad allora[231]. Si arriva ad asserire che il compito precipuo della responsabilità aquiliana non è scoprire l'autore vero del fatto dannoso con l'annessa sua punizione[232]. Quest'ultima è una funzione propria del diritto penale, che, ultimamente non pare assolvere. Il compito della responsabilità civile è creare i criteri giuridici per l'attribuzione del danno, determinando artificialmente le condizioni in presenza delle quali il costo di un danno subìto da un soggetto debba essere trasferito ad un altro soggetto[233].

---

avoider non è limitato al campo della responsabilità oggettiva, ma coinvolge anche quello della responsabilità colposa; esso, infatti, ha proprio l'obiettivo di indirizzare il legislatore nella scelta di quale sia il modello di responsabilità (colposa o oggettiva) maggiormente idoneo (ossia, più efficiente) a regolare i diversi settori economici. In particolare, in base a tale criterio, il costo del danno deve essere posto in capo al soggetto che, tenuto conto delle informazioni di cui disponeva al momento in cui agiva, si trovava nella posizione migliore per valutare, attraverso un'analisi costi-benefici, la scelta più conveniente tra prevenire il danno e risarcirlo.

[231] La funzione "punitiva" della responsabilità civile, in realtà, ha trovato una rinnovata âge d'or negli Stati Uniti, con l'incremento delle condanne al pagamento di *punitive damages*, sanzioni di diritto civile con funzione deterrente e retributiva, applicate dai giudici in aggiunta ai *compensatory damages* in casi di condotte dolose (o gravemente colpose) che abbiano arrecato danni di particolare entità; sul tema vd. L.L. SCHLUETER, *Punitive Damages*, VII ed., vol. I, LexisNexis, 2015; D.G. OWEN, *A Punitive Damages Overview: Functions, Problems and Reform*, in 39 Vill. L. Rev., 1994, p. 363 ss. In Italia, c'è stata una prima apertura all'ammissibilità dei *punitive damages* con la sentenza delle Sezioni Unite Cass. civ, sez. un., 5 luglio 2017, n. 16601, che ha ammesso l'efficacia in Italia di una sentenza straniera recante la condanna al pagamento dei cosiddetti *punitive damages*, affermando che non si tratta di sanzioni contrarie all'ordine pubblico. Sulla sentenza vd. i commenti contenuti in Danno resp., 2017, n. 4, di M. LA TORRE, Un punto fermo sul problema dei "danni punitivi", p. 419; G. CORSI, Le Sezioni Unite: via libera al riconoscimento di sentenze comminatorie di *punitive damages*, p. 429; G. PONZANELLI, *Polifunzionalità tra diritto internazionale privato e diritto privato; Le Sezioni Unite e le funzioni della responsabilità civile*, p. 435; P.G. MONATERI, *Le Sezioni Unite e le funzioni della responsabilità civile*, p.437; vd. anche C. CONSOLO, *Riconoscimento di sentenze, specie USA e di giurie popolari, aggiudicanti risarcimenti punitivi o comunque sopracompensativi, se in regola con il nostro principio di legalità* (che postula tipicità e financo prevedibilità e non coincide pertanto con il, di norma presente, due process of law), in Corr. giur., 2017, p. 1050.

[232] P. BARCELLONA*, Diritto privato e società moderna*, Jovene, 1996, p. 550: «Il problema della responsabilità civile non consiste nella scoperta del "vero" autore del fatto dannoso, bensì nella fissazione di un criterio grazie al quale si può sostituire l'attribuzione automatica del danno con un criterio giuridico». Sulla continua opera di flessibilizzazione delle categorie del diritto penale ad opera della giurisprudenza torneremo nel prossimo capitolo; vd. in part. § 4. ss.

[233] F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 131; P. BARCELLONA, *Diritto privato e società moderna*, cit., p. 150; G. ALPA, *La responsabilità civile. Parte generale*, cit., p. 98. In realtà, a parere di alcuni autori, anche la responsabilità penale si sta, inopportunamente, evolvendo in un diritto "artificiale" vd. F. SGUBBI, *Il reato come*

Nella responsabilità civile, al contrario del diritto penale in cui la responsabilità è personale, può capitare che il soggetto tenuto a risarcire il danno venga individuato in una persona che non abbia commesso materialmente il fatto lesivo[234]. La responsabilità oggettiva, infatti, non assolve solo ad una funzione economico-allocativa ma anche ad una funzione distributiva e pragmatica, nella misura in cui consente al danneggiato di ottenere il risarcimento anche nei casi in cui sarebbero da scovare tra i meccanismi produttivi e decisionali delle grandi imprese[235].

1.2. La regolazione dell'IA nel contesto europeo e la responsabilità civile.

Le motivazioni che hanno spinto per la creazione della responsabilità oggettiva sembrano ripresentarsi nell'attuale dibattito sull'allocazione dei costi derivanti dai danni algoritmici. Infatti, così come all'epoca, anche oggi sussiste l'esigenza di contemperare innovazione industriale e sicurezza dei consumatori[236]. E così come allora, il rischio da scongiurare è che si creino delle sacche di immunità giustificate dall'obiettivo di incentivare il progresso tecnologico e di evitare *chilling effects*[237].

Però, ad oggi, il modello della responsabilità oggettiva non sembra godere di ottima salute in quanto la disciplina della responsabilità per danno da prodotto, che ha rappresentato la forma più compiuto di responsabilità oggettiva per tanto

rischio sociale. Ricerche sulle scelte di allocazione dell'illegalità penale, Il Mulino, 1990, passim e p. 90: «La responsabilità penale è distribuita socialmente, è allocata artificialmente secondo scelte di opportunità politiche».

[234] U. CARNEVALI, *Appunti di diritto privato*, Raffaello Cortina Editore, IX ed., 2010, p. 247.

[235] F. DI CIOMMO, *Evoluzione tecnologica e regole di responsabilità civile*, cit., p. 94; ID., *Valori e funzioni della responsabilità civile nell'epoca del post-turbocapitalismo*, in *Danno e resp*., 2021, n. 2, p. 147.

[236] Le due istanze emergono in maniera plastica nella Relazione di accompagnamento all'AI Act (p. 3), che, tra gli obiettivi della normativa indica sia quello di «assicurare che i sistemi di IA immessi sul mercato dell'Unione e utilizzati siano sicuri e rispettino la normativa vigente in materia di diritti fondamentali e i valori dell'Unione», sia quello di «assicurare la certezza del diritto per facilitare gli investimenti e l'innovazione nell'intelligenza artificiale».

[237] Per un simile paragone tra i primordi dell'attività industriale e l'attuale contesto produttivo vd. A. DAVOLA – R. PARDOLESI, *In viaggio col robot: verso nuovi orizzonti della r.c. auto ("driverless")?*, in Danno e resp., 2017, n. 5, p. 626.

tempo, ha subìto un graduale avvicinamento al paradigma colposo. Ad esempio, gli incerti confini della nozione di "difetto" e l'eccessiva interpretazione dell'esimente secondo cui la responsabilità del produttore è esclusa se <<lo stato delle conoscenze tecniche e scientifiche, al momento in cui il produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso>> (art. 7, lett. e) della direttiva 85/374/CEE[238][239].

Secondo alcuni autori, il graduale abbandono della responsabilità oggettiva, in Italia, sta impattando anche le norme codicistiche speciali in tema di responsabilità aquiliana. Si tratta di un riavvicinamento alle origini, cioè ad ipotesi ritenute di colpa presunta[240].

Nel settore dei danni provocati dall'IA, le proposte avanzate dalla dottrina di introdurre sistemi di responsabilità oggettiva non sono state accettate[241]. Lo stesso Parlamento Europeo aveva proposto nella Risoluzione del 2020 di introdurre un regime di responsabilità oggettiva con obbligo di assicurazione per gli operatori di sistemi di IA ad alto rischio, escludendo che questi ultimi potessero liberarsi dimostrando la propria diligenza, il comportamento

[238] Direttiva 85/374/CEE del Consiglio del 25 luglio 1985 relativa al ravvicinamento delle disposizioni legislative, regolamentari ed amministrative degli Stati Membri in materia di responsabilità per danno da prodotti difettosi.

[239] Nella giurisprudenza della Corte di Cassazione è ormai invalsa l'affermazione in base alla quale la responsabilità per il danno da prodotto difettoso avrebbe natura presunta e non oggettiva, vd. Cass. civ., sez. III, 28 settembre 2018, n. 23447; Cass. civ., sez. III, 19 febbraio 2016, n. 3258; Cass., sez. III, 29 maggio 2013, n. 13458.

[240] F.P. PATTI, *Il declino della responsabilità oggettiva* (a margine dell'art. 2051 c.c.), in Riv. dir. civ., 2019, n. 4, p. 977 ss.; V. ROPPO, *La responsabilità oggettiva*, in P. Sirena (a cura di*), La funzione deterrente della responsabilità civile alla luce delle riforme straniere e dei Principles of European Tort Law*, Giuffrè, 2011, p. 185 ss; ID*., La responsabilità civile di Pietro Trimarchi*, in *Jus Civile*, 2017, n. 6, p. 696 ss.

[241] C. WENDEHORST*, Strict Liability for AI and other Emerging Technologies*, in *Journal of European Tort Law*, 2011, vol. 11, iss. 2, p. 150 ss.; H. ZECH*, Liability for AI: public policy considerations*, in *ERA Forum*, 2021, n. 22, p. 147 ss.; S. LI – M. FAURE – K. HAVU*, Liability Rules for AI-Related Harm: Law and Economics Lessons for a European Approach*, in *European Journal of Risk Regulation*, 2022, p. 8 ss.; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability – Adapting Liability Rules to the Digital Age and Artificial Intelligence*, in *Journal of European Tort Law*, 2021, vol. 13, iss. 1.

autonomo dell'IA o la forza maggiore[242]. Detta proposta, però, non è confluita nella bozza di direttiva sulla responsabilità da IA che ha, invece, optato per l'introduzione di una presunzione del nesso di causalità in caso di colpa e per la possibilità per l'autorità giudiziaria di ordinare la divulgazione degli elementi di prova, qualora emergano indici di fondatezza della domanda risarcitoria[243]. La motivazione principale manifestata per tale esclusione consiste nel fatto che <<il tipo di prodotti e servizi dotati di sistemi di IA che potrebbero incidere sul pubblico in generale e mettere a repentaglio importanti diritti giuridici, come il diritto alla vita, alla salute ed alla proprietà, e che potrebbero pertanto essere soggetti a un regime di responsabilità oggettiva, non sono ancora ampiamente disponibili sul mercato>>. Trattasi, invero, di una motivazione che non ci convince: basti pensare ai danni all'integrità fisica che possono essere causati, già al giorno d'oggi, da sistemi di guida autonoma.

Contestualmente alla proposta di direttiva, la Commissione ha pubblicato una bozza di revisione della direttiva sulla responsabilità per danno da prodotti difettosi, finalizzata ad adeguare la disciplina vigente <<all'era digitale, all'economia circolare e all'impatto delle catene globali del valore>>[244]. Oltre all'allineamento delle nozioni di prodotto e produttore al nuovo costo industriale, la proposta ridefinisce le circostanze che devono essere valutate al fine di verificare la difettosità del prodotto, includendovi l'eventualità che il prodotto abbia la <<capacità di continuare

[242] Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, 2020/2014(INL), art. 4, § 3: «Gli operatori di sistemi di IA ad alto rischio non possono eludere la propria responsabilità sostenendo di avere agito con la dovuta diligenza o che il danno o il pregiudizio sia stato cagionato da un'attività, dispositivo o processo autonomo guidato dal loro sistema di IA. Gli operatori non sono considerati responsabili se il danno o il pregiudizio è dovuto a cause di forza maggiore». La Risoluzione prevedeva inoltre, per gli operatori di sistemi di i.a. non qualificati come ad alto rischio, un regime di responsabilità per colpa, ma con un onere della prova agevolato per l'attore.

[243] Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale, cit.

[244] Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi, COM (2022) 495 final, 28 settembre 2022.

ad imparare dopo la sua diffusione>>. La proposta avanzata dalla dottrina[245] di sopprimere l'esimente del rischio da sviluppo non è stata accolta dalla Commissione Europea che ha mantenuto la causa di esclusione della responsabilità.

Una responsabilità oggettiva pura ad oggi non sembra essere la soluzione perseguita a livello europeo per disciplinare le ipotesi di danno da sistemi di IA, nonostante la Commissione si sia riservato la possibilità di valutare, sulla base degli incidenti di introdurre un regime di responsabilità oggettiva e/o una copertura assicurativa obbligatoria.

1.3. Occorre creare una normativa *ad hoc* ovvero è sufficiente adattare la normativa vigente?

Due sono le posizioni con riferimento alla responsabilità aquiliana per i danni provocati da sistemi di IA. Da un lato si trovano coloro i quali credono che occorra una interpretazione evolutiva delle norme codicistiche attualmente esistenti, posto che esse sono flessibili ed aperte ai mutamenti socio-tecnologici. Dall'altro lato, si trovano coloro che ritengono che gli agenti intelligenti abbiano caratteristiche così disruptive rispetto ai tradizionali sistemi tecnologici da rendere obsolete le ordinarie categorie di responsabilità aquiliana[246]. Tale dibattito sembra riprendere quanto già accaduto con la contrapposizione tra coloro i quali ritenevano necessario

[245] F.P. PATTI, *The European Road to Autonomous Vehicles*, in *Fordham International Law Journal*, 2019, p. 142 ss.; H. JACQUEMIN – J. HUBIN, *Aspects contractuels et de responsabilité civile en matière d'intelligence artificielle*, in H. Jacquemin – A. De Streel (eds.), *Intelligence artificielle et droit, Larcier*, 2017, p. 137; U. SALANITRO*, Intelligenza artificiale e responsabilità: la strategia della Commissione europea*, in *Riv. dir. civ*., 2020, fasc. 6, p. 1246; vd. anche EUROPEAN COMMISSION – EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 42-44.

[246] G. TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi, cit.; M. SCHERER, Regulating Artificial Intelligence Systems*, cit., p. 354 ss.; J. TURNER, *Robot Rules. Regulating Artificial Intelligence, Palgrave Macmillan Cham*, 2019; in realtà, questa posizione va ben oltre la sola questione della responsabilità aquiliana per i danni da algoritmo, e – proponendo una presa d'atto del presunto superamento della dicotomia "uomo-macchina" – auspica un complessivo ripensamento delle tradizionali categorie "umano-centriche" del diritto, vd. ex multis R. CALO*, Robotics and the Lessons of Cyberlaw*, cit.

introdurre un nuovo campo di studi a causa della diffusione del *cyberspace* e coloro i quali, invece, ritenevano che le norme esistenti fossero sufficienti a regolare il nuovo fenomeno. Famosissima, a tal proposito, la provocazione del giudice statunitense F.H. Easterbrook, che sostenne che pensare di introdurre un nuovo corso di studi sulla *cyberlaw* sarebbe stato come creare un ciclo di lezioni sul diritto dei cavalli (*Law of Horses*)[247].

Come anticipato, parte della dottrina sostiene che le norme esistenti abbiano una elevata capacità di mediazione giuridica in relazione ai nuovi fenomeni tecnologici. Vi è chi, infatti, sostiene che le norme sulla responsabilità aquiliana possano essere applicate anche ai danni cagionati da sistemi di IA[248]. Alla base di detto orientamento sta il fatto che i codici dell'Europa continentale sono stati ideati proprio per essere pronti alle evoluzioni ed ai progressi sociali e tecnologici. Infatti, hanno già dimostrato di essere in grado di adeguarsi alle novità della società e delle tecniche produttive, sia pure con minimi interventi normativi[249].

Il sistema della responsabilità aquiliana da sempre si trova ad interfacciarsi con danni provocati da forme di intelligenza naturale non imputabili. Ci riferiamo agli animali, agli incapaci di intendere e di volere. Di fronte a tali comportamenti intelligenti ed imprevedibili non ascrivibili alle persone (o agli animali) che li hanno

[247] F.H. EASTERBROOK, *Cyberspace and the Law of the Horse*, in *University of Chicago Legal Forum*, 1996, p. 207 ss.

[248] Vd. ex multis U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 95 ss.; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema*?, cit., p. 445 ss.; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1719; A. PROCIDA MIRABELLI DI LAURO, *Le intelligenze artificiali tra responsabilità civile e sicurezza sociale*, in P. Perlingieri e al. (a cura di*), Rapporti civilistici e intelligenze artificiali: attività e responsabilità*, Edizioni Scientifiche Italiane, p. 297 ss. In ambito di *common law*, per l'applicabilità dei tradizionali principi di *tort law*, vd. R.A. HEVERLY, *Are Existing Tort Theories Ready for AI? An American Perspective*, in L.A. DiMatteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, 2022, p. 91 ss.

[249] U. RUFFOLO*, La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 100.

compiuti, le norme attribuiscono la responsabilità per fatto altrui ai soggetti che beneficiano di tali forme di intelligenza ed ai soggetti che si trovano nella migliore posizione per impedire il verificarsi degli eventi lesivi.

Anche nei sistemi di IA, gli algoritmi possono essere forieri di eventi dannosi ed imprevedibili. Applicando per analogia i principi della responsabilità aquiliana, si potrebbe attribuirla ai soggetti che beneficiano delle funzionalità dei sistemi di IA ed ai soggetti che si trovano nella posizione migliore per impedire il verificarsi degli eventi lesivi. Tra i motivi per cui non occorre creare nuove norme ma sarebbe preferibile affidarsi ad un'attività ermeneutica evolutiva della normativa esistente, vi è anche la considerazione per cui ogni intervento normativo rischierebbe di avere effetti sistematici anche imprevedibili sulla tenuta dell'intero sistema[250].

E, pertanto, dopo un primo momento in cui le istituzioni europee si dicevano pronte ad introdurre una sorta di personalità elettronica per i sistemi di IA, più di recente Parlamento e Commissione hanno adottato un approccio diverso. Queste ultime si ritengono convinte che non sia necessaria una revisione completa dei regimi di responsabilità correttamente funzionanti, pur ritenendo opportuni <<adeguamenti specifici e coordinati dei regimi di responsabilità per evitare una situazione in cui le persone che subiscono pregiudizi o danni al patrimonio non ottengano un risarcimento>>[251].

Accanto a chi la pensa in questo modo, ovviamente, vi è chi invece ritiene che la peculiarità dei sistemi di IA siano tali da scardinare i tradizionali meccanismi imputativi della responsabilità aquiliana, rendendo obbligato un

[250] A. AMIDEI*, Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1719.

[251] Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, cit., considerando 6. Vd. anche, nello stesso senso le Proposte per una strategia italiana per l'intelligenza artificiale del luglio 2020, elaborate dal Gruppo di esperti MISE.

intervento legislativo di ampio respiro, che traduca la disruption costituita dai sistemi di IA autonomi in paradigmi legislativi *ad hoc*[252]. Vi è chi auspica l'introduzione della personalità elettronica, chi della responsabilità oggettiva degli operatori dei sistemi di IA e la *Manufacturer Enterprise Liability*.

2. Il codice civile ed i regimi speciali di responsabilità oggettiva in capo all'utilizzatore del sistema di IA.

Per come è noto, gli artt. 2049-2054 c.c. prevedono specifiche deroghe alla responsabilità colposa. La loro applicazione al fenomeno dei danni da dispositivi intelligenti potrebbe essere molto più favorevole per il danneggiato rispetto al regime colposo di cui all'art. 2043 c.c.

Le disposizioni degli articoli appena menzionati delineano tutte ipotesi in cui c'è una persona che ha creato il rischio di un evento di danno oppure che beneficia della creazione del rischio stesso, ovvero che si trova nella miglior posizione per controllare il rischio. L'art. 2050 c.c., in particolare, sembra essere la disposizione che meglio si attanaglia alla posizione del produttore.

Nonostante le prime riflessioni in tema di responsabilità oggettiva si siano manifestate in relazione al rischio d'impresa, non tutte le norme codicistiche che disciplinano la responsabilità senza colpa possono essere ricondotte all'attività industriale. L'art. 2049[253] e l'art. 2051[254] sono disposizioni che vengono applicate all'attività produttiva. Ma ve ne sono altre che vengono applicate con riferimento alla sfera personale dei consociati. Basti pensare alla responsabilità del custode di animale di cui all'art. 2052 o

[252] C.E.A. KARNOW, *The application of traditional tort law*, cit., p. 74; P. ASARO, *The liability problem for autonomous artificial agents*, in *AAAI Symposium on Ethical and Moral Considerations in Non-Human Agents*, 2016, p. 190 ss.; A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., passim.

[253] responsabilità dei padroni e dei committenti.

[254] danno cagionato da cosa in custodia.

a quella del proprietario di immobili di cui all'art. 2053 o di autoveicoli di cui all'art. 2054 del codice[255]. Il padrone di un cane, *ex* art. 2052 c.c., dovrà risarcire i danni causati dal suo animale anche qualora dimostri di avere adottato tutte le misure necessarie per evitare il danno, in quanto chi decide di beneficiare della compagnia di un cane deve anche sostenere gli eventuali costi derivanti da questa scelta, anche qualora non abbia colpe[256].

Analogamente si può asserire che chi sceglie di beneficiare dei sistemi più sofisticati di IA dovrà anche risponderne degli effetti collaterali che possono capitare. Dunque, la *ratio* di ritenere responsabile oggettivamente l'operatore del sistema di IA pare in linea con le norme del codice in materia di imputazione oggettiva. Ma, in assenza di un intervento legislativo che definisca il bilanciamento politico tra esigenze produttive e tutela del danneggiato, l'operazione estensiva delle norme deve essere svolta con estrema cautela.

Sono molti gli autori che consigliano di applicare i regimi speciali di responsabilità in caso di danno cagionato da un sistema di IA: l'applicazione di una o dell'altra fattispecie cambia in base alla concezione che si ha dei sistemi di IA, nonché in base alla *ratio* delle varie disposizioni del codice.

Per esempio, se si ritiene che il sistema di IA sia una cosa inanimata si può fare ricorso alla norma sulla responsabilità da cosa in custodia. Nello specifico campo dell'*automotive*, si applicherebbe l'art. 2054 c.c. se si ritiene di considerare i sistemi di IA come una sorta di intelligenza minorata, si può fare riferimento all'art. 2053 c.c. Se, invece, si ritiene di considerare l'algoritmo come una forma di intelligenza paragonabile a quella umana, la

[255] Le considera invece ipotesi di colpa presunta, e non di responsabilità oggettiva, C.M. BIANCA, Diritto civile: Vol. V – La responsabilità, Giuffrè, 2020, p. 718.
[256] È il principio del *cuius commoda, eius et incommoda esse debent*: vd. U. CARNEVALI, *Appunti di diritto privato,* cit., p. 255-256.

disciplina da applicare sarà quella della responsabilità vicaria di cui all'art. 2049 c.c.

Una delle soluzioni più immediate sembrerebbe essere quella di applicare il regime di responsabilità oggettiva prevista per il custode di cose o di animali di cui agli articoli 2051 e 2052 c.c. La *ratio* di entrambe è basata sul principio del *cuius commoda, eius et incommoda esse debent*[257]. Le due norme si differenziano perché diversi sono i soggetti responsabili. Ciò in quanto l'art. 2051 c.c. fa riferimento alla figura del custode, mentre l'art. 2052 c.c. indica il proprietario dell'animale e "colui che se ne serve per il tempo in cui lo ha in uso" come i destinatari della norma. Dunque, ambedue le disposizioni prevedono una responsabilità oggettiva per la persona che si trova nella posizione di esercitare un controllo sulla cosa o sull'animale, dando una forte importanza all'interesse alla custodia. Di fatti, la disposizione non si applica nei confronti del dipendente che detiene la cosa, l'animale nell'interesse altrui o nell'esercizio delle proprie incombenze. Vista così, sarebbe una forma di responsabilità collegata al profitto ed al controllo del rischio, piuttosto che di una responsabilità collegata ad una vigilanza immediata[258]. Anche il riferimento al caso fortuito come prova liberatoria, in entrambi i casi, potrebbe adattarsi perfettamente al nuovo contesto dell'IA. Infatti, il caso fortuito potrebbe esonerare l'utilizzatore di un sistema di IA da responsabilità nei casi in cui l'autoapprendimento abbia causato dei comportamenti assolutamente imprevedibili[259].

[257] P. TRIMARCHI, *Istituzioni di diritto privato*, cit., p. 135.

[258] P. TRIMARCHI, *La responsabilità civile: atti illeciti, rischio, danno*, cit., p. 364.

[259] U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 100. Non si può non rilevare, tuttavia, come l'applicazione dell'esimente del "caso fortuito" – tanto in relazione all'art. 2051 c.c., quanto in relazione all'art. 2052 c.c. – sia, per giurisprudenza costante, particolarmente restrittiva, limitata a «ciò che non è prevedibile oggettivamente, ovvero tutto ciò che rappresenta un'eccezione alla normale sequenza causale, integra il caso fortuito, quale fattore estraneo alla sequenza originaria, avente idoneità causale assorbente e tale da interrompere il nesso con quella precedente, sovrapponendosi ad essa ed elidendone l'efficacia condizionante» (Cass. civ., sez. VI, 12 luglio 2022, n. 31106; vd. anche Cass. civ., sez. III, ord. 1 febbraio 2018, n. 2477). Sulla

Si è già accennato che il richiamo all'una o all'altra disposizione varia a seconda della concezione che si ha del sistema di IA: chi considera l'agente intelligente come una cosa inanimata crede sia applicabile l'art. 2051 c.c.[260] Secondo l'interpretazione maggioritaria, l'art. 2051 c.c. si applica in quei casi in cui il danno scaturisce direttamente dal dinamismo della cosa, indipendentemente dall'azione dell'essere umano[261]. Pertanto, il discrimine tra l'applicazione dell'art. 2043 c.c. o 2050 c.c. in caso di attività pericolose ed il ricorso all'art. 2051 c.c. starebbe nel diverso ruolo della persona fisica nel verificarsi dell'evento lesivo. Nel primo caso, la cosa è sotto la direzione dell'uomo, mentre nel secondo la cosa non è azionata direttamente dall'operatore o, comunque, pur essendo azionata da questo, si sottrae al suo controllo[262].

Chi, invece, ritiene che i sistemi di IA non siano semplici cose inanimate e che, quindi, siano dotate di una certa *agency*, credono sia più opportuna un'applicazione analogica dell'art. 2052 c.c.[263]. Analogia fondata su talune

problematicità di tale orientamento, che riduce il caso fortuito all'assenza di un elemento costitutivo della fattispecie (il nesso causale), vd. F.P. PATTI, *Il declino della responsabilità oggettiva* (a margine dell'art. 2051 c.c.), cit., p. 989.

[260] E. PALMERINI, *Soggettività e agenti artificiali*, cit., p. 469; U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 97; contra l'equiparazione tra "cosa" e sistema di i.a. vd. G. SARTOR, *Gli agenti software e la disciplina giuridica degli strumenti cognitivi*, in *Dir. inf*., 2003, pp. 55 ss.; M. COSTANZA, *Impresa robotizzata e responsabilità*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 108 ss.

[261] P.G. MONATERI, *La responsabilità civile*, in R. Sacco (diretto da), *Tratt. dir. civ*., Utet, 1998, p. 1038. In giurisprudenza vd. Cass., 22 maggio 2000, n. 6616, in Mass. Foro it., 2000; Cass., 7 aprile 2010, n. 8229, in Mass. Giust. civ., 2010. Recentemente sul campo applicativo dell'art. 2051 c.c. vd. in dottrina V. VILLANOVA, *La responsabilità per il danno cagionato da cose in custodia*, in Resp. civ. prev., 2022, p. 760 ss.; D. SANTARPIA, *Un'indagine subiettiva sulla responsabilità da custodia,* in *Resp. civ. prev*., 2018, fasc. 4, p. 1224C ss.; F.P. PATTI, I*l declino della responsabilità oggettiva (a margine dell'art. 2051 c.c.)*, cit., che riflette sull'incertezza applicativa della norma.

[262] P. TRIMARCHI, *La responsabilità civile: atti illeciti, rischio, danno*, cit., p. 367-368.

[263] Vd. C. BOSCARATO, *Who is Responsible for a Robot's Actions? An initial examination of Italian Law within a European Perspective*, in B. Van den Berg – L. Klaming (eds*.), Technologies on the Stand: Legal and Ethical Questions in Neuroscience and Robotics*, Wolf Legal Pubns, 2011, p. 393; M. RATTI, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente alla luce dell'attuale scenario normativo*, in Contr. impr., 2020, n. 3, p. 1181; G. TADDEI ELMI – F. ROMANO, *Il robot tra ius condendum e ius conditum*, in *Informatica e dir*., 2016, n. 1, p. 126 ss ; contrario all'analogia tra sistemi di i.a. e animali è A. BERTOLINI, *Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules*, in *Law Innovation & Tech*., 2013, p. 227 ss., che osserva come gli animali – esseri

caratteristiche comuni di animali e sistemi di IA. Da un lato, l'imprevedibilità ed incontrollabilità dei comportamenti riducibili tramite l'addestramento; dall'altro, l'assenza di personalità giuridica, che non consente loro di essere titolari di un patrimonio.

Le alternative qui prospettate, fondate sull'art. 2051 c.c. e sull'art. 2052 c.c., scontano il limite di essere valide solo se il custode mantiene una qualche forma di governo sul comportamento dei sistemi di IA. Se in futuro si dovesse arrivare ad una crescente autonomia dei sistemi di IA rispetto agli utilizzatori, l'allocazione dei costi a questi ultimi rischierebbe di porsi in contrasto con la finalità preventiva della responsabilità civile, dal momento che questi si troverebbero nell'impossibilità di diminuire il rischio di danno[264].

Discorso a parte merita il settore delle auto a guida autonoma, laddove immediato è il riferimento codicistico all'art. 2054 c.c. Detta disposizione prevede una responsabilità aggravata a carico del conducente, in base alla quale egli è tenuto a risarcire il danno se non prova di aver fatto tutto il possibile per evitare il danno, nonché una responsabilità oggettiva a carico del proprietario, dell'usufruttuario e dell'acquirente con patto di riservato dominio, in solido con quella del conducente di cui ai commi 1 e 2, a meno che questi non provino che la circolazione del veicolo è avvenuta contro la loro volontà. Vi è, altresì, una responsabilità oggettiva a carico di tutte le persone sopra indicate in caso di danni derivanti da vizi di costruzione o da difetti di manutenzione del veicolo.

Una responsabilità ai sensi dei commi 1 e 2 potrebbe certamente configurarsi con riferimento agli incidenti

irrazionali e imprevedibili – non possono essere paragonati ai sistemi di i.a., che seguono pur sempre le istruzioni fornite in sede di programmazione; vd. anche l'accurata disanima di A. DRIGO, *L'"Evocativa" integrazione analogica dell'art. 2052 c.c. per i sistemi emergenti di intelligenza artificiale: un'analisi critica*, in *GiustiziaCivile.com*, 2020.

[264] A DRIGO, *L'"Evocativa" integrazione analogica dell'art. 2052 c.c.*, cit., p. 14; H. ZECH, *Liability for AI: public policy considerations*, p. 154.

provocati da veicoli a guida parzialmente autonoma di cui ai livelli 1,2 e 3. Rispetto a questi, per come si è visto, il conducente mantiene un obbligo di sorveglianza, dovendo attivarsi per riprendere il controllo del veicolo ogni volta che si verifichi una situazione di emergenza o vi sia una richiesta da parte del sistema[265]. Per andare esente da responsabilità, il conducente dovrò dimostrare di <<aver fatto tutto il possibile per evitare un danno>>, non essendo da sola sufficiente la prova di un funzionamento difettoso del veicolo. Nell'ultima ipotesi, infatti, il conducente potrebbe rispondere in concorso con il produttore. Potrà poi essere chiamato a rispondere in solido con il conducente anche il proprietario del veicolo, ai sensi dell'art. 2054, co. 3 c.c. Quest'ultima soluzione trova anche conferma nel cd. decreto *Smart Roads* che stabilisce che la domanda di autorizzazione della sperimentazione su strada di veicoli a guida autonoma deve contenere anche l'indicazione del proprietario del veicolo a guida automatica, quale soggetto responsabile ai sensi dell'art. 196 codice della strada e dell'art. 2054, comma 3, del codice civile, ovvero di altro soggetto obbligato in solido, ai sensi del citato art. 196>> (art. 11, co. 1, lett. a).[266]

Senza un soggetto qualificabile come "conducente" non pare si possano applicare i commi 1 e 2 dell'art. 2054 c.c. in relazione ai veicoli di livello 4 e 5[267]. Ciò in quanto pare

[265] A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, in *Europa e dir. priv.*, 2019, fasc. 4, p. 1002; G. CALABRESI – E. AL MUREDEN, *Driverless Cars, Il Mulino*, 2021, p. 156; A. LEPORE*, I.A. e responsabilità civile. Robot, autoveicoli e obblighi di protezione*, in *Tecnologie e diritto*, 2021, p. 198; M. TAMPIERI, *L'intelligenza artificiale: una nuova sfida anche per le automobili*, in *Contr. Impr.,* 2020, n. 2, p. 748; S. PELLEGATTA, *Automazione nel settore automotive: profili di responsabilità civile*, in *Contr. impr.*, 2019, n. 4, p. 1429-1431; F. BARTOLINI, *Auto a guida autonoma e problemi di responsabilità civile*, in *R. Giordano e al*. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia,* Giuffrè, 2022, p. 306-307.

[266] Decreto del Ministero delle Infrastrutture e dei Trasporti del 28 febbraio 2018, cit.

[267] A. PISANI TEDESCO*, Smart mobility e rischi satellitari e informatici: i possibili scenari di allocazione della responsabilità civile,* in *Dir. comm. internaz*., 2019, fasc. 4, § 4; A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit., p. 1004; S. PELLEGATTA, *Automazione nel settore automotive*, cit., p. 1433; F. BARTOLINI, *Auto a guida autonoma e problemi di responsabilità civile*, cit., p. 307; contra A. DAVOLA – R.

complicato che la persona trasportata in un veicolo che si muove in modo autonomo possa essere considerato "conducente"[268]. D'altronde, il regime dell'art. 2054, commi 1 e 2, è basato sulla capacità di controllo che non sussiste ove l'utilizzatore non sia posto nella posizione di poter governare il veicolo[269]. Attribuire la responsabilità all'operatore di una *self-driving car* colliderebbe anche con logiche di economia del diritto: porre il costo di un incidente in capo ad un soggetto privo dei mezzi per ridurre la probabilità del verificarsi dello stesso non ha senso[270]. Infatti, l'utilizzatore di un veicolo di livello 4 e 5 non può incidere sul movimento del veicolo e non ha neanche le competenze per migliorare il software preposto a regolare la guida automatizzata. Egli è estraneo alla dinamica della quale sarebbe chiamato a sopportare il rischio, così che la regola di responsabilità civile non può sortire nei suoi confronti l'effetto virtuoso di promuovere il costante miglioramento della tecnologia in questione, al fine di raggiungere il più alto livello di sicurezza possibile[271]. Vista l'inapplicabilità del regime di cui ai commi 1 e 2 dell'art. 2054 c.c. per i veicoli di livello 4 e 5, non si può configurare una responsabilità del proprietario del veicolo ai sensi del comma 3 dello stesso articolo, che presuppone una responsabilità per colpa presunta del conducente. Residua spazio applicativo per la responsabilità oggettiva di cui al comma 4 dell'art. 2054 c.c. per i vizi di costruzione del veicolo, da coordinarsi con la disciplina della responsabilità per danno da prodotto difettoso.

Altra dottrina ha proposto di regolare il danno da algoritmo applicando la disciplina della responsabilità

PARDOLESI, *In viaggio col robot*, cit., p. 619, che ritengono applicabile tale norma, contestandone però l'efficienza sul piano della prevenzione.

[268] A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, p. 1004.

[269] A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 619; A. PISANI TEDESCO, *Smart mobility e rischi satellitari e informatic*i, cit., § 4.

[270] A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 619.

[271] A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 619.

vicaria in via analogica[272] ovvero tramite l'introduzione di una clausola di equivalenza tra dipendente e sistema di IA[273]. Questa forma d'imputazione disciplinata dall'art. 2049 c.c. prevede la responsabilità del datore di lavoro per il danno illecito cagionato a terzi dai dipendenti, nell'esercizio delle incombenze alle quali sono adibiti. L'analogia tra il rapporto di preposizione ed il rapporto tra utilizzatore e sistema di IA sembrerebbe giustificato dal fatto che oramai imprenditori e professionisti si affidano a detti sistemi per lo svolgimento di mansioni che prima erano affidate al lavoro umano. Pertanto, estendere la responsabilità vicaria ai danni da IA comporterebbe che il datore non possa più sottrarsi alla responsabilità per i danni provocati da un'attività svolta a suo vantaggio[274].

La *ratio* sottesa a tale responsabilità oggettiva del datore si rinviene nel principio del *cuius commoda eius et incommoda* secondo cui chi trae vantaggio dal rapporto di preposizione deve anche sopportare i costi derivanti dal danno che esso cagiona. Il datore di lavoro, anche nell'ipotesi di dipendente algoritmico si trova nella posizione più adatta per controllare lo svolgimento dell'attività lavorativa e per minimizzare il rischio di eventi lesivi. Il maggiore ostacolo parrebbe essere rappresentato dall'asimmetria tra non imputabilità del

[272] M. COSTANZA*, Impresa robotizzata e responsabilità*, cit., p. 112 ss.; E. KARNER, *Liability for robotics: current rules, challenges and the need*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 120; S. NYHOLM, *Attributing Agency to Automated Systems: Reflections on Human-Robot Collaborations and Responsibility-Loci*, in *Science Engineering Ethics*, 2018, p. 1201 ss.; S. CHOPRA – L.F. WHITE, *A Legal Theory for Autonomous Artificial Agents*, University of Michigan Press, 2011, p. 128 ss.; M. DIAMANTIS, *Vicarious Liability for AI*, Iowa Legal Studies Research Paper No. 2021-27; ID., *The Extended Corporate Mind: When Corporations Use AI to Break the Law,* in 98 N.C. L. Rev., 2020, p. 926 ss.; ID., *Algorithms Acting Badly: a Solution From Corporate Law*, in 89 Geo. Wash. L. Rev., p. 818. Nel sistema di *common law* la responsabilità vicaria è conosciuta anche come *respondeat superior*.

[273] GRUPPO INDIPENDENTE DI ESPERTI AD ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, Orientamenti etici per un'i.a. affidabile, cit., p. 45 ss.; A. BECKERS – G. TEUBNER*, Three Liability Regimes for Artificial Intelligence*, cit., p. 87.

[274] C. WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, cit., p. 159; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema*?, cit., p. 468; M. DIAMANTIS*, Algorithms Acting Badly*, cit., passim.

sistema di IA e responsabilità solidale del dipendente, prevista dalla disposizione codicistica[275].

La soluzione più immediata e diretta sembra essere quella della responsabilità per danno da prodotto[276]. Poste la progressiva autonomia dei dispositivi intelligenti e la conseguente perdita di controllo sugli stessi da parte dell'essere umano, si apre la strada ad una maggiore responsabilizzazione dei soggetti che immettono sul mercato i prodotti e che si trovano nella posizione più adeguata alla valutazione dei rischi della commercializzazione e ad incidere sulla sicurezza dei dispositivi intelligenti[277]. Per come è noto, la differenza tra responsabilità aquiliana e responsabilità per danno da prodotto è il criterio di imputazione: al ricorrere delle circostanze descritte dalla normativa, al danneggiato basterà dimostrare il difetto del prodotto, per ottenere il risarcimento. Il criterio della *strict liability* pare essere indispensabile in relazione ai danni provocati dai sistemi di IA, la cui capacità di autonomia decisionale rende quasi impossibile il riconoscimento di una colpa in capo al produttore per i comportamenti emergenti e lesivi posti in essere dallo *smart product*. Si è già accennato in precedenza come non sia del tutto vero che la disciplina della responsabilità da prodotto difettoso escluda un giudizio sulla rimproverabilità in senso lato del produttore: i confini della responsabilità sono delimitati dalla nozione stessa di difetto che inevitabilmente introduce un elemento di valutazione soggettivo della

[275] M. RATTI, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente*, cit., p. 1181.

[276] Sull'applicabilità della disciplina europea sulla responsabilità per danno da prodotto difettoso ai sistemi di i.a. vd. *ex multis* i contributi pubblicati nel libro collettaneo S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit.; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability*, cit.; C. TWIGG-FLESNER, *Guiding Principles for Updating the Product Liability Directive for the Digital Age*, ELI *Innovation Paper Series – European Law Institute*, 2021; T.S. CABRAL, *Liability and artificial intelligence in the EU: Assessing the adequacy of the current Product Liability Directive*, in *Maastricht Journal of European and Comparative Law,* 2020, vol. 27(5), p. 615 ss.; A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit.; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 125.

[277] G. WAGNER, *Robot liability*, cit., p. 38 ss.

posizione del produttore. La Corte di Cassazione ha più volte asserito che la responsabilità di tal fatta non ha natura oggettiva ma presunta poiché prescinde dall'accertamento della colpevolezza del produttore ma anche dalla dimostrazione dell'esistenza di un difetto del prodotto[278].

Provando a tracciare le linee fondamentali della disciplina della responsabilità per danno da prodotto, occorre citare la direttiva n. 85/374/CEE[279], recepita nell'ordinamento italiano dal D.P.R. 24 maggio 1988, n. 224 e poi confluita nel Codice del consumo agli artt. 114-127. In particolare, l'art. 114 cod. cons. stabilisce che <<*Il produttore è responsabile del danno cagionato da difetti del suo prodotto*>>, sempre che il difetto sia esistente al momento della messa in commercio del prodotto[280]. La responsabilità del produttore è esclusa se non ha messo il prodotto in circolazione; se non ha fabbricato il prodotto per la vendita o per qualsiasi altra forma di distribuzione a titolo oneroso, né lo ha fabbricato o distribuito nell'esercizio della sua attività professionale, se – in caso di produttore o fornitore di una parte componente o di una materia prima – il difetto è interamente dovuto alla concezione del prodotto in cui è stata incorporata la parte o materia prima o alla conformità di questa alle istruzioni date dal produttore che l'ha utilizzata.

L'art. 120 cod. cons. pone a carico del danneggiato l'onere di fornire la prova, oltre che del danno subito, dell'esistenza del difetto del prodotto e del nesso di

[278] Cass. civ., sez. III, 28 settembre 2018, n. 23447; Cass. civ., sez. III, 28 luglio 2015, n. 15851; Cass. civ., sez. III, 19 febbraio 2016, n. 3258; Cass., sez. III, 29 maggio 2013, n. 13458.
[279] Direttiva 85/374/CEE del Consiglio del 25 luglio 1985, cit. In generale sull'applicazione della Direttiva sul danno da prodotto nel panorama europeo vd. P. Machnikowski (ed.), *European Product Liability. An Analysis of the State of the Art in the Era of New Technologies*, Intersentia, 2016; D. WUYTS, *The product liability directive – More than two decades of defective products in Europe*, in *Journal of European Tort Law*, 2014, n. 5, p. 1; D. Fairgrieve (ed.), *Product Liability in Comparative Perspective*, Cambridge University Press, 2005; S. WHITTAKER, *Liability for Products. English Law, French Law, and European Harmonisation*, Oxford University Press, 2005.
[280] vd. art. 118, lett. b), che esclude la responsabilità del produttore se «il difetto che ha cagionato il danno non esisteva quando il produttore ha messo il prodotto in circolazione»

causalità sussistente tra difetto e danno. È una ripartizione dell'onere probatorio che corrisponde a quello dell'ordinaria responsabilità aquiliana, ai sensi degli artt. 2043 e 2697 c.c.

I soggetti che possono far valere la responsabilità del produttore sono l'utilizzatore, l'acquirente del bene ed il terzo danneggiato dal prodotto. Quest'ultima ipotesi non è frequente nella prassi giurisprudenziale, forse perché il terzo ha a disposizione altre modalità per ottenere ristoro. La tendenza a considerare la responsabilità del produttore come rimedio a tutela del compratore-utilizzatore è evidente dalla stessa collocazione della disciplina nel codice del consumo.

Un *Report* finale sull'applicazione della direttiva sul danno da prodotto in ambito europeo del 2018 aveva evidenziato diverse falle nel meccanismo compensativo dei soggetti danneggiati, specialmente con riferimento ai nuovi prodotti ad alta tecnologia. L'introduzione della figura del programmatore di algoritmi spesso venduti senza supporto fisico rende obsolete le categorie di "prodotto" e "produttore" di cui alla direttiva comunitaria. Inoltre, la capacità di auto-apprendimento dei sistemi di IA e la loro opacità tecnologica incidono sul concetto tipico di "difetto" e "nesso di causalità". Tutte le carenze evidenziate nella sede del *Report* e le riflessioni della dottrina e dei gruppi di studio istituiti a livello europeo[281], sono state alla base della proposta di revisione della direttiva, presentata nel 2022 con l'obiettivo di agevolare il danneggiato nell'ottenimento del risarcimento, adeguando la disciplina vigente <<*all'era*

[281] EUROPEAN COMMISSION, Evaluation of Council Directive 85/374/EEC on the approximation of laws, regulations and administrative provisions of the Member States concerning liability for defective products, SWD(2018) 157 final, 7 maggio 2018, disponibile a questo link: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0157&from=EN.

digitale, all'economia circolare e all'impatto delle catene globali del valore>>[282].

Nonostante alcune modifiche contenute nella proposta revisione rispondano ad esigenze di aggiornamento della direttiva, sussiste sempre il problema di fondo, trascurato dalla Commissione Europea, di comprendere in che modo la capacità di *machine learning* incida sulla configurabilità del difetto del prodotto. Non è semplice comprendere se e quando l'auto-apprendimento possa essere considerato alla stregua di un difetto e se il produttore possa allegare, come causa esimente, la circostanza che <<lo stato oggettivo delle conoscenze scientifiche e tecniche al momento dell'immissione del prodotto sul mercato o della sua messa in servizio oppure durante il periodo in cui il prodotto era sotto il controllo del fabbricante non permetteva di scoprire l'esistenza del difetto>>. Si tratta della c.d. esimente da rischio di sviluppo.

3. La nozione di prodotto e la sua adeguatezza al contesto dell'IA.

Venendo alla nozione di "prodotto", occorre partire dal dato testuale. L'art. 115 cod. cons., che recepisce l'art. 2 della direttiva n. 85/374/CEE, intende per prodotto, ai fini della attribuzione della responsabilità[283], <<*ogni bene immobile, anche se incorporato in altro bene mobile o immobile*>> (co. 1), inclusa l'elettricità (co. 2).

[282] Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi, COM (2022) 495 final, 28 settembre 2022.

[283] La precisazione è d'obbligo, poiché la definizione fornita all'art. 115 cod. cons. si applica al solo titolo II della parte IV del Codice del consumo ; la nozione generale di "prodotto" è invece rinvenibile all'art. 3, co., lett. e) cod. cons., che così lo definisce: «qualsiasi prodotto destinato al consumatore, anche nel quadro di una prestazione di servizi, o suscettibile, in condizioni ragionevolmente prevedibili, di essere utilizzato dal consumatore, anche se non a lui destinato, fornito o reso disponibile a titolo oneroso o gratuito nell'ambito di un'attività commerciale, indipendentemente dal fatto che sia nuovo, usato o rimesso a nuovo; tale definizione non si applica ai prodotti usati, forniti come pezzi d'antiquariato, o come prodotti da riparare o da rimettere a nuovo prima dell'utilizzazione, purché il fornitore ne informi per iscritto la persona cui fornisce il prodotto».

Sin dalla prima messa in commercio di prodotti dotati di software, la dottrina maggioritaria[284] ha ritenuto che il codice immateriale dovesse essere considerato come componente del prodotto, quando esso sia un elemento inseparabile rispetto all'hardware e necessario per il suo funzionamento. Così, i difetti del *software* sono tali da rendere difettoso l'intero prodotto[285]. È stata, d'altronde, la stessa Commissione Europea, nel 1988, a confermate che la direttiva si applica ai software allo stesso modo in cui si applica ai prodotti artigianali ed artistici[286]. Seguendo tale prospettiva, dunque, non vi sarebbero particolari ragioni per escludere dalla nozione di prodotto di cui all'art. 2 della direttiva, gli artefatti che incorporino sistemi di IA inseparabili rispetto all'*hardware*.

Il problema si pone rispetto a quei software che non posseggono un corpo fisico, cioè ai *software non embodied*: algoritmi venduti anche separatamente rispetto al prodotto e che sono direttamente scaricabili online e che possono essere sia destinati ad essere incorporati in un *hardware* sia progettati per operare su un sistema operativo virtuale. In ambedue i casi, tali *software* possono essere difettosi e determinare un'offesa all'incolumità personale o al patrimonio. Ciò che non è chiaro è se questi possano essere qualificati come prodotti o come componenti del prodotto, in base alla normativa vigente.

L'evoluzione tecnologica degli ultimi anni ha comportato una crescente fluidità dei sistemi operativi e,

[284] A. ZACCARIA*, La responsabilità del "produttore" di software*, in *Contr. impr.*, 1993, p. 302 e, più recentemente, D. WUYTS*, The product liability directive*, cit., 2014, n. 5, p. 6; D. FAIRGRIEVE E AL., *Product liability directive*, in P. Machnikowski (a cura di), European Product Liability, cit., p. 46-47; A. PENZA, *La responsabilità per danno da prodotti difettosi*, in P.G. Demarchi (a cura di), *I diritti del consumatore e la nuova class action*, Zanichelli, 2010, p. 218; D. RESTUCCIA, *La tutela del consumatore*, in B. Inzitari (a cura di), *Valutazione del danno e rimedi risarcitori*, Giappichelli, 2016, p. 521.

[285] T.G. WOLPERT, *Product Liability and Software Implicated in Personal Injury*, in Def. Couns. J., 1993, p. 521 ss.

[286] Risposta fornita da Lord Cockfield, per conto della Commissione, alla domanda scritta No 706/88 presentata da Mr Gijs de Vries (LDR-NL) il 5 luglio 1988,pubblicata sull'Official Journal of the European Communities, C 114, vol. 32, 8 maggio 1989, p. 42, disponibile a questo link: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C:1989:114:FULL&from=IT.

pertanto, dei prodotti che li incorporano. Il consumatore può decidere di installare su un *hardware* un *software* sviluppato da una società diversa da quella che ha messo in commercio il prodotto o addirittura un *software open source*. Il sistema operativo può essere oggetto di frequenti *updates* che incidono sul funzionamento concreto del prodotto. In entrambi i casi, la conseguenza è che il prodotto che causa il danno può aver subito delle sostanziali modifiche rispetto alla progettazione iniziale. Circostanza, questa, che esclude la responsabilità del produttore, posta l'esistenza dell'esimente nel caso in cui <<*il difetto che ha cagionato il danno non esisteva quando il produttore ha messo il prodotto in circolazione*>>.

Ad esempio, comprata una stampante 3D, un consumatore può decidere di acquistare *online* il *software* per farla funzionare da una società diversa da quella che ha prodotto la stampante, eventualmente sostituendo questo al sistema operativo installato in origine. Oppure, può mantenere il software originario, ma aggiornarlo successivamente. Il *software* e l'aggiornamento, in questi casi, non possono considerarsi come prodotti a sé stanti, non essendo qualificabili come beni mobili ai sensi della direttiva. Essi non sembrano neanche costituire una componente del prodotto, poiché, accedendo al bene fisico dopo la messa in commercio del prodotto, la società produttrice non ha alcun controllo sul loro funzionamento e non potrà, dunque, essere chiamata a risponderne.

Dunque, qualora si verifichi un danno a causa di un *software* non *ab origine* installato nel prodotto, il danneggiato non potrà ottenere risarcimento attraverso il regime della responsabilità per danno da prodotto. Non sembrerebbero rientrare nella definizione di cui all'art. 2 della direttiva nemmeno i prodotti digitali, cioè i *software* che agiscono in ambienti esclusivamente virtuali e che non sono necessari al funzionamento di un *hardware*: basti pensare alle *app* installate su *tablet* o *smartphone*, il cui

funzionamento potrebbe determinare danni all'incolumità personale o al patrimonio.

La caratteristica della separazione del *software* di IA rispetto al supporto fisico ha indotto una parte della dottrina a ritenere che il software potesse essere qualificato come servizio piuttosto che come componente di un prodotto[287]. Il servizio viene definito in negativo dall'art. 57 TFUE, in forza del quale per servizi si intendono <<*le prestazioni fornite normalmente dietro retribuzione, in quanto non siano regolate dalle disposizioni relative alla libera circolazione delle merci, dei capitali e delle persone*>>. I servizi, a differenza dei prodotti, non possono essere facilmente suddivisi in classi, poiché tra di loro eccessivamente eterogenei. Vengono infatti definiti servizi le attività professionali e tutte le altre attività che non si esprimono nella produzione di un bene tangibile[288]. Se il *software* fosse considerato un servizio, la direttiva 85/374/CEE non sarebbe applicabile ai danni da questo cagionati e gli unici rimedi per la persona danneggiata sarebbero contrattuali[289] o fondati sulla disciplina della responsabilità aquiliana.

La parte maggioritaria della dottrina ritiene, tenendo in considerazione il fatto che la *ratio* della direttiva 85/374/CEE sia quella di garantire un livello elevato di tutela nei confronti dei consumatori, che sia possibile una interpretazione estensiva della nozione di prodotto, che comprenda gli algoritmi che siano strumentali rispetto al

[287] Sul punto vd. EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 28; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1726; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 462; K.A. CHAGAL-FEFERKORN, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers*, in *Stanford Law & Policy Review*, 2019, p. 84; T.S. CABRAL, *Liability and artificial intelligence in the EU*, cit., p. 620; A. BERTOLINI, *Robots as Products*, cit., p. 214 e ss.; A. FUSARO, *Quale modello di responsabilità per la robotica avanzata? Riflessioni a margine del percorso*, in *Nuova giur. civ. comm*., 2020, n. 6, p. 1346.
[288] G. ALPA, *Quale modello normativo europeo per l'intelligenza artificiale*?, in *Contr. impr*., 2021, n. 4, p. 1020.
[289] E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema*?, cit., p. 462.

funzionamento di un bene fisico[290]. Resterebbero comunque esclusi dal campo di applicazione della direttiva i prodotti digitali, non riconducibili ai beni mobili di cui all'art. 2 della direttiva 85/374/CEE[291].

È stato auspicato, sia delle istituzioni europee[292] sia da parte della dottrina[293] un intervento normativo che estenda il campo di applicazione della direttiva anche ai danni cagionati da prodotti immateriali. La già citata proposta di revisione della direttiva sulla responsabilità per danno da prodotto accoglie i suggerimenti emersi in sede di dibattito scientifico ed estende la nozione di prodotto ai *software* ed ai *file* per la fabbricazione digitale e la nozione di componente del prodotto a <<qualsiasi articolo, tangibile o intangibile, o qualsiasi servizio correlato, integrato in un prodotto o interconnesso con

[290] J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 64; B.A. KOCH*, Product Liability 2.0 - Mere Update or New Version*?, ivi, p. 105 ss.; ID., *Response of the EU La Institute to the Public Consultation*, cit., p. 31; G. WAGNER, *Robot, Inc.: Personhood for Autonomous Systems*?, in *Fordham Law Review*, 2019, p. 604; ID., *Robot liability*, cit., p. 41; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1723; U. RUFFOLO, *Responsabilità da A.I. e da algoritmo*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 107; J. KNETSCH, *Are Existing Tort Theories Ready for AI? A Continental European Perspective*, in L.A. DiMatteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence,* cit.; M. RATTI, *Riflessioni in materia di responsabilità civile e danno cagionato da dispositivo intelligente*, cit., p. 1186; A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit., p. 1021; G. GUERRA*, Il concetto di difettosità nella realtà che cambia. Un esercizio di microcomparazione,* in *Comparazione e diritto civile*, 2019, fasc. 1, p. 270.

[291] J.S. BORGHETTI, *How can Artificial Intelligence be Defective*?, cit., p. 64; contra M. RATTI, *Riflessioni in materia di responsabilità civile,* cit., p. 1186; A. AMIDEI*, Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, cit, p. 286, a parere dei quali anche a normativa invariata l'ideatore del software non embodied può considerarsi il fabbricante del prodotto e dunque rispondere ai sensi dell'art. 114 cons. e ss.

[292] L'estensione è auspicata dalla Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale, cit., considerando 8; vd. anche EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies,* 2019, cit.

[293] L. COPPINI, *Robotica e intelligenza artificiale: questioni di responsabilità civile*, in *Pol. dir.*, 2018, fasc. 4, p. 729; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability*, cit., p. 31 ss.; C. TWIGG-FLESNER, *Guiding Principles for Updating the Product Liability Directive for the Digital Age,* cit., p. 5; M. BUITEN E AL., *EU Liability Rules for the Age of Artificial Intelligence, Centre on Regulation in Europe*, 9 aprile 2021, p. 51.

esso dal fabbricante di tale prodotto o sotto il controllo di tale fabbricante>> (art. 4, paragrafo 1, n. 3). Per non ostacolare la ricerca e l'innovazione nel campo delle tecnologie più avanzate, comunque, il considerando 13 stabilisce che la direttiva non si applica a quei *software* che sono liberamente accessibili, utilizzabili e modificabili e che sono sviluppati o forniti al di fuori di un'attività commerciale.

4. La nozione di produttore e la sua adeguatezza al contesto dell'IA.

Per quanto concerne, invece, la nozione di "produttore", occorre prendere in esame il dettato dell'art. 115, comma 2-*bis*, cod. cons., che lo definisce come <<*il fabbricante del prodotto finito o di una sua componente, il produttore della materia prima, nonché, per i prodotti agricoli del suolo e per quelli dell'allevamento, della pesca e della caccia, rispettivamente l'agricoltore, l'allevatore, il pescatore ed il cacciatore*>>[294]. Pur ammettendo che la nozione di algoritmo possa essere inclusa in quella di prodotto, resta dubbio se i soggetti coinvolti nei processi di programmazione, sviluppo e aggiornamento degli algoritmi di IA rientrino nella definizione di produttore.

Ad esempio, lo sviluppatore di un *software* che non rientra nella struttura organizzativa del produttore e che non vende il progetto applicativo a quest'ultimo o aggiorna le funzionalità del *software* dopo l'entrata in commercio del prodotto non pare possa rientrare in alcuna delle categorie normative anzidette. Ne consegue che anche qualora si ritenesse di poter estendere la nozione di prodotto al *software*, ugualmente la responsabilità attivabile ai sensi della direttiva n. 85/374/CEE potrebbe configurarsi solo a carico del fabbricante del prodotto immesso nel commercio e non invece in capo all'ideatore

[294] L'art. 3, co. 1, dir. aggiunge «ogni persona che, apponendo il proprio nome, marchi marchio o altro segno distintivo sul prodotto, si presenta come produttore dello stesso».

dell'algoritmo[295]. I rapporti tra questi dovranno dunque essere regolati dalle regole della responsabilità contrattuale. È quanto accade di norma tra fabbricante e *designer* del prodotto: quest'ultimo si occupa soltanto della concezione astratta del prodotto e non delle componenti; mentre il fabbricante dà corpo a tale idea astratta e solo egli risponde dei difetti del prodotto. Al massimo il *designer* può essere chiamato a rispondere per inadempimento contrattuale nei confronti del fabbricante[296]. La non riconducibilità del programmatore alla nozione di produttore non sembra incoerente con le finalità della disciplina sulla responsabilità per il danno da prodotto difettoso[297]. È la casa produttrice a fungere da garante dell'affidabilità del prodotto finito nei confronti del consumatore sia quando sia impegnata direttamente nella progettazione sia quando, come assemblatore, si limiti a selezionare i partner più adeguati[298]. Il fabbricante del prodotto finito è il soggetto che si trova nella posizione più adeguata a svolgere una costante attività di monitoraggio del funzionamento dei propri prodotti e ad intervenire qualora essi rivelino imperfezioni o malfunzionamenti al fine di migliorarne la qualità e l'affidabilità[299].

L'estensione della definizione di produttore potrebbe invece risultare opportuna al fine di includervi i *backend operators*, cioè quei soggetti che monitorano in maniera continuativa le condizioni di sicurezza dei software e forniscono assistenza ed aggiornamenti agli utenti. Le

[295] U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 97; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability*, cit., p. 32

[296] U. RUFFOLO*, La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 97.

[297] A. DAVOLA – R. PARDOLESI*, In viaggio col robot*, cit., p. 627; contra U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 98, che sostiene che lo sviluppatore di un algoritmo di i.a. non può essere assimilabile al tradizionale designer, poiché questi – attribuendo alla macchina la capacità di modificarsi nel tempo – plasma concretamente la configurazione del prodotto finale.

[298] Così A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 627.

[299] A. DAVOLA – R. PARDOLESI*, In viaggio col robot*, cit., p. 627.

operazioni di *backend* hanno un importantissimo ruolo nel plasmare le funzionalità ed il *design* dei *software* ed intervengono dopo la messa in commercio del prodotto. Per tale ragione non sembra corretto imputare al fabbricante eventuali difetti derivanti da tali attività, stante il controllo limitato o nullo che questi può svolgere in materia.

L'estensione della nozione di produttore ai programmatori è stata auspicata dalla maggioranza degli studiosi[300]. Queste sollecitazioni sono state accolte nella proposta di revisione della direttiva sul danno da prodotto che estende radicalmente il novero dei soggetti che possono essere chiamati a rispondere per il danno derivante dal prodotto difettoso. L'art. 7 della proposta di revisione, infatti, include sia i fabbricanti del prodotto o di una componente di esso sia i soggetti che apportano modifiche sostanziali al prodotto.

5. La nozione di difetto e la sua adeguatezza al contesto dell'IA.

Per quanto concerne l'adeguatezza della nozione di difetto, essa rileva non tanto per la questione dell'applicabilità della direttiva sulla product liability. Piuttosto, la delimitazione della nozione di difetto occupa uno spazio più significativo all'interno della nostra trattazione perché interroga il problema specifico dei sistemi di IA, ovverosia l'autonomia e l'imprevedibilità.

Il concetto di difettosità del prodotto rispetto al quale circoscrivere la responsabilità del produttore costituisce il perno intorno a cui ruota l'intero sistema della responsabilità per danno da prodotto. Per far sì che il consumatore ottenga ristoro per il danno subito non basta vi sia un nesso di causalità tra prodotto e danno, ma

[300] Ex multis S. LI – M. FAURE – K. HAVU, *Liability Rules for AI-Related Harm*, cit., p. 8-9; U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 97-98; B.A. KOCH E AL., *Response of the European Law Institute to the Public Consultation on Civil Liability*, cit., p. 39.

occorre che, al momento della messa in commercio, sussista un difetto, ossia una <<condizione di anormale pericolosità del prodotto>>[301]. Non tutti i prodotti che provocano un danno sono difettosi *in re ipsa*: basti pensare alle sigarette o agli alcolici o ai farmaci innovativi e ai vaccini autorizzati dagli organi competenti. Ciò sta a significare che il giudizio inerente alla difettosità del prodotto è autonomo rispetto a quello sull'accertamento del danno concretamente verificatosi: il difetto non può essere dedotto in via indiretta né dal verificarsi dell'evento dannoso né dall'utilizzo conforme del prodotto[302]. Per difetto non si intende un connotato fisico del bene ma si fa riferimento ad un concetto normativo[303] e relazionale[304] legato alle legittime aspettative dei consumatori. Il difetto consiste pertanto in un disallineamento del prodotto rispetto alla sicurezza che la platea degli utenti ha ragionevolmente diritto di attendersi, in ottemperanza al *consumer expectations test*.

L'art. 117, co. 1, cod. cons. detta i criteri per la valutazione della difettosità del prodotto, precisando che <<un prodotto è difettoso quando non offre la sicurezza che ci si può legittimamente attendere tenuto conto di tutte le circostanze, tra cui: a) il modo in cui il prodotto è stato messo in circolazione, la sua presentazione, le sue caratteristiche palesi, le istruzioni e le avvertenze fornite; b) l'uso al quale il prodotto può essere ragionevolmente

[301] U. CARNEVALI, *La responsabilità del produttore. Problemi generali*, in AA.VV., *La responsabilità dell'impresa per i danni all'ambiente e ai consumatori*, Giuffrè, 1978, p. 149; vd. anche Corte di Giustizia UE, 5 marzo 2015, Boston Scientific Medizintechnik GmbH, cause riunite C-503/13 e C-504/13, § 30, che definisce i difetti come «anomale potenzialità di danno che il prodotto può causare alla persona o ai beni del suo utilizzatore».

[302] G. PONZANELLI, *Causa probabile del danno e responsabilità del produttore*, in Danno e resp., 2004, p. 532; Cass. civ., sez. III, 13 dicembre 2010, n. 25116.

[303] C. CASTRONOVO, *La nuova responsabilità civile*, Giuffrè, 2006, p. 692-693; vd. anche P.G. MONATERI, *La responsabilità civile*, cit., p. 716, che parla di "concetto legislativo" (corsivi dell'A.); E. RAJNERI, *The projects of European regulation on liability for damage caused by artificial intelligence. Striking a balance between the interests of consumers, multinationals and SMEs*, in *Riv. dir. banc.*, 2022, fasc. 2, p. 108;

[304] F. CAFAGGI, *Verso un approccio sistematico ai rimedi contro la difettosità dei prodotti,* in S. Patti (a cura di), Il danno da prodotti in Italia, Austria, Repubblica Federale di Germania, Svizzera, Cedam, 1990, p. 85.

destinato e i comportamenti che, in relazione ad esso, si possono ragionevolmente prevedere, c) il tempo in cui il prodotto è stato messo in circolazione>>.

Detti parametri costituiscono un elenco puramente indicativo poiché il giudice, nel valutare il livello di sicurezza che ci si può legittimamente attendere potrà considerare ulteriori elementi tra cui il rispetto della normativa sulla sicurezza del prodotto. A questo proposito, sottolineiamo che il mancato rispetto degli *standard* obbligatori di sicurezza comporta, di per sé, la difettosità del prodotto[305]. Ma non è vera l'affermazione contraria, in quanto l'osservanza di detti *standard* non preclude al giudice di indagare sulla possibile esistenza di difetti nel prodotto[306].

Fin dall'entrata in vigore della direttiva n. 85/374, la dottrina ha riconosciuto proprio nella definizione del difetto il profilo più problematico della disciplina, capace di aumentare a dismisura la discrezionalità giudiziaria e di rendere imprevedibili i giudizi in materia[307]. Nel dettaglio, il tentativo di mediare tra la necessità di fornire una tutela adeguata ai consumatori e l'esigenza di incentivare l'innovazione delle imprese europee[308] si risolve in una formulazione compromissoria ed incerta della norma che risulta ambigua sulla scelta del criterio di imputazione della responsabilità da utilizzare. La clausola generale della "sicurezza che ci si può legittimamente attendere"

---

[305] A. LUMINOSO, *Certificazione di qualità di prodotti e tutela del consumatore-acquirente*, in *Eur. dir. priv.,* 1999, p. 43 e 54; G. SMORTO, (voce) *Certificazione di qualità e normazione tecnica*, in Dig. disc. priv., Agg. II, Utet, 2003, p. 219; A. ZEI, *Tecnica e diritto tra pubblico e privato*, Giuffrè, 2008, p. 263.

[306] E. BELLISARIO*, Il danno da prodotto conforme tra regole preventive e regole risarcitorie*, cit., § 4.

[307] Ex multis E. RAJNERI, *L'ambigua nozione di prodotto difettoso al vaglio della corte di cassazione italiana e delle altre corti europee*, in *Riv. dir. civ*., 2008, n. 5, p. 624 ss.; G. PONZANELLI, *Causa probabile del danno e responsabilità del produttore,* cit., p. 532; E. BELLISARIO*, Il danno da prodotto conforme tra regole preventive e regole risarcitorie*, cit., § 4; M. DELLACASA, *Sulle definizioni legislative nel diritto privato. Fra codice e nuove leggi civili*, Giappichelli, 2004, p. 326 ss.

[308] E. RAJNERI*, L'ambigua nozione di prodotto difettoso*, cit., p. 624; M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 338.

sembra confermare, da un lato, la scelta di fondo del legislatore comunitario per un regime di responsabilità oggettiva; mentre, dall'altro lato, le circostanze che dovrebbero orientare l'interprete nell'applicazione di tale clausola riguardano la condotta del produttore e del danneggiato, piuttosto che le caratteristiche oggettive del prodotto.

Secondo la dottrina italiana, questa valutazione non deve apparire in contrasto con il carattere oggettivo della responsabilità[309]. Per far sì che venga escluso il risarcimento, al produttore non basterà dimostrare di non essere in colpa. Egli dovrà provare l'esistenza di una delle cause tassative di cui all'articolo 118 del codice che non assorbono l'intera casistica della mancanza di colpa. La formula francese della *responsabilité sans faute* sembra essere quella che più emblematicamente esprime il carattere oggettivo della responsabilità: trattasi di un modello di imputazione della responsabilità indipendente dall'accertamento della colpa, ma che non per questo deve rinunciare a qualsiasi considerazione nei confronti delle condotte di attori e convenuti[310]. Dal modello di responsabilità oggettiva inteso in questo modo, dovrebbero distinguersi i sistemi di *absolute responsibility*[311]

[309] G. PONZANELLI, *La responsabilità oggettiva. La responsabilità del produttore. La responsabilità assoluta*, in M. Bessone (a cura di*), Casi e questioni di diritto privato*, Giuffrè, IV ed., 2002, p. 510; F. GALGANO, *Responsabilità del produttore*, in *Contr. impr*., 1986, p. 996; G. VISINTINI, *Trattato breve della responsabilità civile. Fatti illeciti. Inadempimento. Danno Risarcibile*, Cedam, 2005, p. 748.

[310] G. PONZANELLI, *La responsabilità oggettiva. La responsabilità del produttore. La responsabilità assoluta*, cit., p. 510.

[311] D.G. OWEN, *Rethinking the Policies of Scrict Products Liability*, in Vand. L. Rev., 1980, p. 682, che contrappone, proprio in materia di responsabilità per danno da prodotto, la *strict liability* e la *absolute liability*; sul concetto di "responsabilità assoluta" vd., nella dottrina italiana, G. PONZANELLI, *La responsabilità oggettiva. La responsabilità del produttore. La responsabilità assoluta*, cit., p. 509 e 515; G. PONZANELLI – R. PARDOLESI, *La responsabilità per danno da prodotti difettosi*, in *Nuove leggi civ.*, 1989, p. 501. Un'ipotesi, nel nostro ordinamento, di *absolute liability* può essere considerata la responsabilità per fatto del dipendente (art. 2049), che non tollera prove liberatorie fondate sulla diligenza del datore di lavoro; alcuni autori, muovendo da tale considerazione, giungono addirittura a ritenere che l'art. 2049 c.c. non riguardi la responsabilità civile, ma dia luogo esclusivamente ad un'obbligazione di garanzia, vd. D. DE MARTINI, *I fatti produttivi di danno risarcibile*, Cedam, 1983, p. 277.

che, se applicati alla disciplina della responsabilità per danno da prodotto, obbligherebbero il produttore a tenere indenne il consumatore da qualsiasi danno derivante dall'uso del prodotto, trasformandolo dunque in una sorta di assicuratore[312].

Occorre comunque rilevare che nella prassi la distinzione tra regimi di responsabilità per colpa e regimi di responsabilità oggettiva non assoluta tende a scemare. Ne consegue che l'accertamento del difetto di *design* si trasformi spesso in una valutazione della diligenza professionale del produttore[313].

Tale avvicinamento della disciplina della responsabilità per danno da prodotto difettoso ad un modello di responsabilità per colpa incontra problemi peculiari in relazione all'imputazione dei danni provocati da sistemi di IA. Ciò in quanto la capacità di autoapprendimento e di adattamento degli agenti intelligenti non si concilia bene con la necessità di accertare una negligenza in campo al produttore, posto che l'evento lesivo può derivare anche da un comportamento imprevedibile dell'algoritmo.

Quando si tratta della nozione di difetto, occorre anche esaminare le esimenti. Infatti, la nozione di difetto è precisata, per sottrazione, da queste ultime. Il difetto del prodotto non può essere imputato al produttore ai sensi della disciplina comunitaria se <<*il difetto che ha cagionato il danno non esisteva quando il produttore ha messo il prodotto in*

[312] U. CARNEVALI, *La responsabilità del produttore. Problemi generali*, cit., p. 148 e 155; G. PONZANELLI, *La responsabilità oggettiva. La responsabilità del produttore. La responsabilità assoluta*, cit., p. 515.

[313] E. VISENTINI, *L'esimente del rischio di sviluppo come criterio di responsabilità del produttore*. (l'esperienza italiana e tedesca e la direttiva comunitaria), in *Resp. civ. prev*., 2004, p.1267. vd. anche D. CARUSO – R. PARDOLESI, *Per una storia della Direttiva 1985/374/CEE,* in R. Pardolesi (a cura di) *I 25 anni di products liability*, *quaderno speciale di Danno e resp.*, 2012, p. 15, che rilevano come, al di là della qualificazione formale, l'impostazione della direttiva europea come «la seduttiva semplicità del paradigma di responsabilità oggettiva (...) andava evaporando». La dottrina maggioritaria, in Europa, intravede, infatti, nella disciplina della responsabilità per danno da prodotto difettoso un regime de facto colposo, vd. S WHITTAKER, *Liability for Products: English Law, French Law and European Harmonisation,* cit., p. 494 ss.; H. ZECH, *Liability for AI: public policy considerations*, cit., p. 150.

circolazione>> (lett. b, c.d. *later defect defence*), se <<*il difetto è dovuto alla conformità del prodotto a una norma giuridica imperativa o a un provvedimento vincolante*>> (lett. d), o se <<*lo stato delle conoscenze scientifiche e tecniche, al momento in cui il produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso*>> (lett. e, c.d. esimente per rischio da sviluppo o *development risk defence*[314]).

Le lett. b) ed e) devono essere interpretate come disposizioni che contribuiscono a diluire la rigorosità del regime fondato sulla responsabilità oggettiva, poiché introducono una forma di valutazione della diligenza del produttore[315]. Infatti, quest'ultimo non risponde in forza del semplice rapporto di causalità tra creazione del rischio ed evento lesivo che concretizza il rischio creato, ma soltanto in relazione a danni che erano prevedibili sulla base dello stato dell'arte esistente al momento della messa in commercio. Secondo parte della dottrina, il regime delle esimenti scoraggia il danneggiato dal fare ricorso alla disciplina prevista dal codice del consumo perché eccessivamente favorevole al produttore. Ciò induce ad avvalersi della tutela offerta dall'articolo 2050 c.c. che si occupa della responsabilità per l'esercizio di attività pericolose[316]. L'applicabilità della *later defect defence* e della *development risk defence* anche nel caso di danni cagionati da sistemi che utilizzano tecniche di *machine learning* suscita diverse perplessità in dottrina. L'evento lesivo dipende da un comportamento emergente del sistema di IA difficilmente ascrivibile al produttore sulla base delle norme vigenti. La dottrina tende a considerare il comportamento deviante del dispositivo intelligente come un vizio esistente ma non conoscibile, più che come un difetto sopravvenuto, evidenziando il rischio che

[314] Sulla nozione di "rischio da sviluppo" adottata dalla Direttiva vd. D. CERINI, Responsabilità del produttore e rischio da sviluppo: oltre la lettera della dir. 85/374/CE, in Dir. ed econ. assic., 1996, p. 33.

[315] A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit., p. 58.

[316] A. FUSARO, *I prodotti difettosi e pericolosi: le responsabilità*, in G. Alpa (a cura di), *La responsabilità del produttore*, Giuffrè, 2019, p. 380 ss.

l'esimente da rischio di sviluppo possa essere considerata applicabile di *default* in relazione ai sistemi di IA, dal momento che il comportamento di questi è caratterizzato da una limitata prevedibilità[317]. Fosse così, i costi dei sinistri finirebbero per gravare sempre sulla vittima o su altri soggetti chiamati a rispondere oggettivamente a vario titolo ma mai sul produttore, nonostante l'imprevedibilità degli eventi lesivi sia in qualche modo prevedibile da quest'ultimo e, quindi, calcolabile dall'impresa.

A tal proposito, una delle opzioni individuate nell'Impact Assesment di accompagnamento alla proposta di revisione della Direttiva era proprio quella di sopprimere l'esimente del rischio da sviluppo[318]. Anche il Gruppo di Esperti sull'IA nominato dalla Commissione Europea[319] e l'Organizzazione europea dei consumatori[320] auspicava un intervento in tal senso. Tuttavia, detta opzione non è stata ricevuta dalla proposta di revisione della direttiva sul danno da prodotto, probabilmente a causa della forte opposizione da parte delle società produttrici che temevano che una soppressione della *development risk defence* potesse determinare un aumento delle condanne al risarcimento di danni.

---

[317] M. EBERS*, Liability For Artificial Intelligence And EU Consumer Law*, in *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 2021, n. 3, p. 215; G. SPINDLER, *User liability and strict liability in the Internet of Things and for Robots*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 143; sembra invece considerare il "difetto da self learning" un difetto sopravvenuto U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., 1264, che giudica dunque dubbia l'applicabilità della *development risk defense*.

[318] EUROPEAN COMMISSION, *Impact Assessment Report Accompanying the Document Proposal for a Directive of the European Parliament and of The Council on Liability for Defective Products*, SWD(2022) 316 final, 28 settembre 2022, p. 50, disponibile a questo link: https://single-market-economy.ec.europa.eu/document/348b3e35-7d1a-43df-8e9d-296fc09e2c3c_en.

[319] EUROPEAN COMMISSION – EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 42-44.

[320] BUREAU EUROPÉEN DES UNIONS DE CONSOMMATEURS, PRODUCT LIABILITY 2.0, cit., in dottrina vd. anche H. JACQUEMIN – J. HUBIN, *Aspects contractuels et de responsabilité civile en matière d'intelligence artificielle*, cit., p. 137; F.P. PATTI, *The European Road to Autonomous Vehicles*, in *Fordham International Law Journal*, 2019, vol. 43, p. 142 ss. (in relazione alle di macchine a guida autonoma).

5.1. La classificazione dei difetti: difetti di fabbricazione, progettazione, informazione nei sistemi di IA.

Prima ancora che in Europa fosse introdotta una disciplina specifica per la responsabilità per danno da prodotto difettoso, la dottrina[321] ha tracciato una classificazione basata sull'eziologia del difetto in cui si distingue tra vizi di fabbricazione, vizi di costruzione e vizi di informazione.

Questa distinzione non è richiamata formalmente dalla disciplina italiana ed europea che si riferiscono genericamente alla "sicurezza". Ma, nonostante ciò, essa si rinviene nelle circostanze elencate dall'art. 117 del codice del consumo e dall'art. 6 dir. n. 85/374.

In particolare, i difetti di fabbricazione emergono durante il processo di fabbricazione e concernono solo singoli esemplari di una serie di prodotti generalmente non difettosi. Sono vizi tipici dei metodi moderni di produzione in serie che possono essere dovuti ad una disfunzione occasionale del macchinario ovvero a sviste del personale addetto alla fabbricazione o al controllo. Detti vizi si caratterizzano per l'inevitabilità perché, nonostante l'uso delle migliori tecniche produttive essi non possono essere eliminati del tutto. L'altra caratteristica è la prevedibilità perché la loro frequenza è statisticamente calcolata. L'incidenza dei vizi di fabbricazione può essere accertata facilmente poiché basta dimostrare che il singolo prodotto differisca dagli altri prodotti della serie, nelle sue caratteristiche materiali. Con riferimento agli *smart products*, questo può accadere in relazione alle componenti dell'*hardware* o, ad esempio, nel caso dell'installazione incompleta di un *software* di guida autonoma di un

[321] Il merito è soprattutto della dottrina tedesca. Si vedano i riferimenti contenuti in U. CARNEVALI, *La responsabilità del produttore*, cit., p. 30, nota 82; vd. anche G. HOHLOCH, *Prospettive di evoluzione della responsabilità del produttore nel diritto tedesco*, in S. Patti (a cura di), *Il danno da prodotti in Italia, Austria, Repubblica Federale di Germania, Svizzera,* cit., p. 268 ss.

veicolo. Tutti gli errori che coinvolgono il codice dell'algoritmo possono costituire esclusivamente difetti di *design* perché riguardano i criteri di progettazione del prodotto[322].

Trattasi, a ben vedere, dell'unica tipologia di difetto regolata da un regime di responsabilità oggettiva puro, in quanto nel caso in cui il danno sia causato dalla difformità del prodotto rispetto agli altri esemplari della serie, il produttore è ritenuto responsabile ai sensi dell'art. 114 cod. cons. sulla base del solo accertamento del nesso di causalità, indipendentemente da valutazioni circa la diligenza del fabbricante. Nell'individuazione della "*sicurezza che ci si può legittimamente attendere*", l'interprete assumerà come naturale punto di riferimento le caratteristiche degli altri prodotti della stessa serie. L'art. 117, co. 3, conferma questa ricostruzione stabilendo che <<*Un prodotto è difettoso se non offre la sicurezza offerta normalmente dagli altri esemplari della medesima serie*>>.

I difetti di costruzione, invece, derivano da un errore nella progettazione degli artefatti. A differenza dei vizi di fabbricazione, i vizi di costruzione condizionano l'intera serie di prodotti in commercio, aumentando la potenzialità lesiva. In questa categoria rientrano anche quei difetti che non attengono direttamente al prodotto, bensì alla sua confezione, quando essa è tale da non garantire un'efficace conservazione del prodotto[323]. L'art. 117 cod. cons. contempla i difetti di progettazione alle lettere b) e c) ove vengono specificati i criteri ai quali il giudice deve attenersi al fine di verificare l'eventuale difettosità del prodotto.

La progettazione del prodotto deve essere valutata sulla base dell'<<uso al quale il prodotto può essere

[322] M.A. GEISTFELD*, A Roadmap for Autonomous Vehicles: State Tort Liability, Automobile Insurance, and Federal Safety Regulation*, in Calif. L. Rev, 2017, p. 1633; F.P. HUBBARD, *"Sophisticated Robots": Balancing Liability, Regulation, and Innovation,* in Fla. L. Rev., 2015, p. 1854; M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 339.
[323] U. CARNEVALI, *La responsabilità del produttore. Problemi generali*, cit., p. 144.

ragionevolmente destinato e i comportamenti che, in relazione ad esso, si possono ragionevolmente prevedere>> (art. 117, co. 1, lett. b) cod. cons.). Trattasi di un fatto non privo di rilevanza, soprattutto se si considera che la tendenza della regolamentazione europea è quella di garantire sempre la presenza di una persona che possa intervenire in casi di necessità per riprendere il controllo dell'algoritmo o del *robot*[324]. Già in fase di progettazione, il produttore deve prevedere meccanismi idonei ad assicurare un rapido ed efficace intervento da parte dell'utilizzatore, tenendo conto della naturale propensione delle persone a confidare troppo nell'accuratezza dell'algoritmo, nonché dell'aumento dei tempi di reazione e di gestione delle emergenze nel caso in cui le situazioni ordinarie siano gestite da tecniche di automazione.

Ad esempio, l'uso di un veicolo di livello 2 o 3 può portare l'utilizzatore a perdere la concentrazione per intervenire tempestivamente. Il produttore dovrebbe predisporre meccanismi di allerta compatibili con l'incremento dei tempi di reazione. Dovrebbe, altresì, predisporre meccanismi di allerta compatibili con l'incremento dei tempi di reazione e anche un'interfaccia semplice ed intuitiva che consenta all'utilizzatore di comprendere rapidamente quale problema si sta verificando e come agire.

Sempre l'art. 117, co. 1, cod. cons. prevede che la difettosità del prodotto debba essere valutata sulla base del <<tempo in cui il prodotto è stato messo in circolazione>>. Gli *standard* di sicurezza sono correlati allo stato dell'evoluzione scientifica esistente al momento della messa in circolazione. Il comma 2 precisa che un prodotto non può essere considerato difettoso per il solo fatto che un prodotto più perfezionato sia stato messo in

[324] Vd. il già citato art. 14, co. 1, AI Act: «I sistemi di IA ad alto rischio sono progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso».

commercio. Trattasi di una precisazione molto opportuna poiché preclude di ritenere che le legittime aspettative del consumatore debbano sempre coincidere con l'adozione di tutte le misure di sicurezza che la tecnica offre al momento della messa in commercio del prodotto. Il contenuto delle aspettative, invece, è strettamente legato alle scelte di consumo dell'acquirente: egli potrà scegliere un prodotto meno robusto ma a buon mercato; ovvero potrà preferire pagare un prezzo più alto per un prodotto che presenti caratteristiche di maggiore sicurezza e resistenza[325]. Questo anche perché, qualora il prodotto più perfezionato esistente sul mercato venisse considerato il sono ad essere "non difettoso", si creerebbe uno squilibrio concorrenziale tra il produttore di detto bene e tutti gli altri produttori, poiché solo il primo sarebbe esentato dai costi degli incidenti causati dai prodotti[326].

Direttamente collegata all'art. 117, co. 1, lett. b) e c), è l'esimente per il rischio da sviluppo, prevista dall'art. 118, lett. e), che esclude la responsabilità del produttore <<*se lo stato delle conoscenze scientifiche e tecniche, al momento in cui il produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso*>>.

Si è già detto che l'insieme di dette norme induce a ritenere che la verifica sulla sussistenza dei difetti di progettazione non sia fondata su un regime di responsabilità oggettiva pura perché la valutazione della prevedibilità degli impieghi del prodotto da parte dei consumatori e delle conoscenze scientifiche esistenti al momento della messa in circolazione, introducono degli elementi di valutazione sulla colpevolezza del produttore[327].

Infine, vi sono i difetti legati all'erroneità o carenza delle informazioni offerte ai consumatori. Il prodotto è di per sé

[325] P. TRIMARCHI, *Responsabilità civile: atti illeciti, rischio, danno*, cit., p. 439.
[326] G. WAGNER, Robot, Inc., cit., p. 606.
[327] A. BERTOLINI, *Study on Artificial Intelligence and Civil Liability*, cit., p. 58; C. WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, cit., p. 159.

fabbricato nel modo corretto: il difetto consiste nell'omissione delle informazioni sulla corretta modalità d'impiego e sui pericoli che l'utilizzo del prodotto può comportare. La scarsa informazione crea una aspettativa legittima di sicurezza e rende difettoso il prodotto, nel caso di danni derivanti dal suo impiego scorretto. L'art. 117 cod. cons. fa riferimento ai difetti informativi al co. 1, lett. a), ove stabilisce che il giudice debba tenere conto anche del <<*modo in cui il prodotto è stato messo in circolazione, la sua presentazione, le sue caratteristiche palesi, le istruzioni e le avvertenze fornite*>>. Appare evidente che sussiste un'intensa correlazione tra difetti informativi e difetti di progettazione: le legittime aspettative concernenti il *design* del prodotto dipendono anche dalle informazioni fornite al consumatore.

All'obbligo informativo che ricade sul produttore sembra di poter ricollegare due corollari. Per prima cosa, si deve notare che solo eccezionalmente le istruzioni possono costituire mezzi di prevenzione del rischio idonei a sopperire un difetto di progettazione del prodotto. Si pensi ad un farmaco che può causare allergie su alcune categorie di soggetti: in questo caso, un'adeguata avvertenza circa gli effetti allergenici del farmaco è idonea ad evitare che la sua composizione debba subire modifiche con conseguente aumento dei costi e ritardo nello sviluppo tecnologico per fronteggiare il rischio corso da un ristretto numero di utenti[328]. Nella maggioranza dei casi, il sistema degli avvisi informativi non può essere considerato una legittima alternativa ad una maggiore sicurezza strutturale e progettuale del prodotto, ove ragionevolmente possibile. In caso contrario, i costi che dovrebbero essere sostenuti dal produttore per la messa in sicurezza del prodotto sarebbero scaricati sul consumatore[329]. Le avvertenze informative non possono porre sul consumatore un obbligo di diligenza

[328] C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 48.
[329] U. CARNEVALI, *Responsabilità del produttore*, in Enc. Dir., Agg. II, 1998, p. 944; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 48.

spropositato rispetto alle esigenze di normale utilizzo del prodotto: un abito confezionato con materiale altamente infiammabile deve ritenersi pericoloso anche nel caso in cui il produttore abbia informato l'acquirente di tale caratteristica, poiché nel normale utilizzo di un capo di abbigliamento, non può escludersi il contatto accidentale con sigarette, fiammiferi o altre fonti di calore[330]. In secondo luogo, bisogna considerare che quanto più il produttore sorvoli sui rischi del prodotto, tanto più salgono le aspettative di sicurezza dell'utente.

Questi due corollari rischiano di entrare in gioco nel settore più avanzato nello sviluppo di sistemi di IA, cioè quello dell'*automotive*. Il problema si pone in relazione ai veicoli di livello 2 e 3, che sono in grado di svolgere soltanto parzialmente la funzione di guida. Da un lato, la presentazione pubblicitaria di alcuni veicoli semiautonomi tende a dare ai potenziali acquirenti l'idea che la macchina possa guidarsi completamente da sola, senza alcun controllo da parte del conducente. Basti pensare, ad esempio, al sistema di guida autonoma di livello 2 prodotto da Tesla, il cui nome, *Autopilot*, evoca un'esperienza di guida totalmente automatizzata. Ancora, si pensi al video pubblicato sul canale *Youtube* di Tesla, ove si assiste ad una guida nel traffico gestita totalmente dal *robot*, con le mani del conducente posate sul grembo e non sul volante. L'accusa di pubblicità ingannevole rivolta alla società Tesla non ha lasciato indifferenti gli organi giudiziari e di regolazione amministrativa: nel 2020, infatti, il Tribunale di Monaco ha stabilito che "*Autopilot*" e "*Full Self Driving*" sono espressioni fuorvianti ed ha ordinato a Tesla di cessarne l'utilizzo sul sito *web* tedesco[331]. Nel 2022, il *California Department of Motor Vehicles* ha proposto ricorso contro Tesla presso lo *State Office of Administrative Hearings*, allegando che la società avrebbe diffuso dichiarazioni non veritiere sui veicoli di

[330] C. PIERGALLINI*, Danno da prodotto e responsabilità penale*, cit., p. 48; P.G. MONATERI, *La responsabilità civile,* cit., p. 719.
[331] M. Taylor, *German Court Bans Tesla 'Autopilot' Name For Misleading Customers*, cit.

sua produzione, pubblicizzandoli come dotati di funzioni di guida autonome di cui erano sforniti[332]. Dall'altro lato, alcuni difetti di progettazione che emergono dalle prime ricostruzioni dei sinistri non possono essere compensati nemmeno da un'adeguata informazione fornita dal produttore. Un veicolo di livello 2, in grado cioè di guidare autonomamente a talune condizioni, non pare offrire <<*la sicurezza che ci si può legittimamente attendere*>>, nel caso in cui sia incapace di riconoscere alcune luci nel traffico, nonostante il consumatore venga informato di tale carenza.

5.1.2. *Focus* sui difetti di progettazione: responsabilità per colpa o assoluta?

I *design defects* sono sempre stati considerati i difetti più problematici nell'ambito della responsabilità per danno da prodotto[333]. Non esiste, innanzitutto, un *design* del prodotto assolutamente sicuro: il giudice, quindi, è chiamato a valutare quale sia il livello di sicurezza che ci si può legittimamente attendere da un prodotto. Tale valutazione lascia evidenti spazi alla discrezionalità[334]. Il mercato offre diversi modelli dello stesso prodotto per qualità, prezzo e progettazione e, ovviamente, per sicurezza.

Tra i criteri sui quali i consumatori fondano le proprie legittime aspettative di sicurezza rientra anche il metro comparativo con riferimento alla generalità delle merci offerte sul mercato. Dal punto di vista probatorio, un altro elemento che rende complessa la valutazione del difetto di progettazione consiste nel fatto che l'utente può influire sulla sicurezza offerta dal prodotto. Non sempre è agevole comprendere se l'evento lesivo sia conseguenza del difetto del prodotto o di uno scorretto impiego del prodotto da

[332] R. Mitchell, *California DMV accuses Tesla of falsely advertising Autopilot and Full Self-Driving features*, cit.
[333] U. CARNEVALI, *Responsabilità del produttore*, in *Enc. Dir.*, cit., p. 943; P. TRIMARCHI, *Responsabilità civile: atti illeciti, rischio, danno,* cit., p. 438.
[334] U. CARNEVALI, *Responsabilità del produttore*, in *Enc. Dir.,* cit., p. 943.

parte del consumatore[335]. Nonostante ciò, la più grande difficoltà che s'incontra nell'accertamento dei difetti di *design* consiste proprio nell'ambigua formulazione normativa, stante l'incerto ruolo che il sistema complesso assegna all'eventuale comportamento negligente del produttore. Questa questione si intreccia con i tortuosi profili probatori della disciplina, caratterizzati dall'asimmetria informativa di colui il quale sopporta l'onere della prova rispetto al convenuto produttore, nonché dalla complessità tecnica dell'accertamento degli elementi della fattispecie, tra cui il nesso di causalità. A questo riguardo, sussistono due orientamenti opposti: da un lato quello della giurisprudenza di merito secondo cui l'onere della prova del danneggiato deve essere alleggerito essendo sufficiente la prova del malfunzionamento del prodotto per dimostrare la sussistenza del difetto. Al produttore spetta invece l'onere di fornire un'eventuale prova liberatoria ai sensi dell'art. 118 codice del consumo. Questa posizione determina uno svilimento del requisito della difettosità[336], con conseguente trasformazione della disciplina in oggetto in un regime di responsabilità assoluta di natura indennitaria.

La giurisprudenza di legittimità, dall'altro lato, afferma che il difetto del prodotto non può dedursi né dal verificarsi di un evento dannoso, né dall'utilizzo normale del prodotto[337]. Si asserisce che <<il danno non prova, di per sé, né direttamente né indirettamente il difetto, né la pericolosità del prodotto in condizioni normali di impiego, ma solo una più indefinita pericolosità di per sé insufficiente per istituire la responsabilità del produttore, se non sia anche in concreto accertato che quella specifica condizione di insicurezza del prodotto si pone al di sotto

[335] U. CARNEVALI, *Responsabilità del produttore*, in *Enc. Dir.,* cit., p. 943.
[336] M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 342 ss., che si interroga sull'utilità stessa del concetto di "difetto" ove letto alla luce dei meccanismi presuntivi utilizzati dalla giurisprudenza di merito.
[337] Cass. civ., sez. III, 29 luglio 2015, n. 15851; Cass. civ., sez. III, 26 giugno 2015, n. 13225; Cass. civ., sez. III, 29 maggio 2013, n. 13458; Cass. civ., ord. 23 maggio 2013, n. 12665; Cass. civ., sez. III, 15 marzo 2007, n. 6007.

del livello di garanzia di affidabilità richiesto dall'utenza o dalle leggi in materia>>. Si dice, inoltre, che <<la responsabilità del prodotto difettoso ha natura presunta (e non già oggettiva), poiché prescinde dall'accertamento della colpevolezza del produttore, ma non anche dalla dimostrazione dell'esistenza di un "difetto" del prodotto; sicché grava sul soggetto danneggiato (…) la prova del collegamento causale, non già tra prodotto e danno, bensì tra difetto e danno[338].

Il difetto di progettazione in relazione ai sistemi di IA.

Applicare la disciplina sulla responsabilità per danno da prodotto agli eventi lesivi provocati dai sistemi di IA è problematico in quanto occorre stabilire quando vi sia un difetto[339]. Se l'esigenza di aggiornare la nozione di prodotto e di produttore è comune a molti nuovi strumenti della tecnica e può essere soddisfatta agevolmente, la presunta inadeguatezza del concetto di difettosità è legata all'irrompere di sistemi di IA capaci di *decision making* autonomo. La nozione di difetto, per come anticipato, presuppone un qualche tipo di svista, disattenzione o sottovalutazione da parte del produttore, che non per forza è presente nel caso in cui il danno sia conseguenza di una decisione autonoma da parte dei prodotti, dovuta alla loro capacità di modificarsi nel corso del loro ciclo vitale.

I problemi più rilevanti nascono in relazione ai difetti di progettazione[340]. Per quanto concerne i difetti di fabbricazione emergono questioni che non divergono da

[338] Cass., 29 maggio 2013, n. 13458, cit.

[339] EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 28.; A. AMIDEI*, Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, cit., p. 293.

[340] S. NAVAS, *Robot Machines and Civil Liability*, in M. Ebers – S. Navas (eds.), *Algorithms and Law*, cit., p. 168; A. BERTOLINI, *Robots as products*, cit., p. 239; G. GUERRA, *Il concetto di difettosità nella realtà che cambia*, cit., p. 272; G. WAGNER, *Robot, Inc*., cit. p. 605; P.M. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 170.

quelle già affrontate nell'ambito della tradizionale disciplina della responsabilità per danno da prodotto difettoso. Se, ad esempio, per la distrazione di un operatore o per il malfunzionamento di un macchinario, in un drone viene installato un codice incompleto, il produttore potrà sempre essere chiamato a rispondere dell'eventuale danno cagionato dal drone, perché la mera difformità del singolo prodotto rispetto al design della serie è sufficiente a dimostrare l'esistenza del difetto.

Anche con riferimento ai difetti di progettazione, in realtà, può individuarsi un caso non controverso riguardante difetti genetici, presenti già nel momento in cui il prodotto viene messo in commercio. Basti pensare al caso in cui alla capacità di apprendimento non si accompagni la presenza di misure strutturate di prevenzione degli eventi lesivi o l'adozione di un adeguato sistema di cybersicurezza[341]. In queste ipotesi, che saranno le più frequenti, non vi saranno problemi rilevanti con riferimento alla nozione di difetto: alcune difficoltà potrebbero invece emergere con riferimento all'accertamento del nesso di causalità. Ed è proprio con riferimento al design che possono sorgere ipotesi più complesse che non hanno precedenti nella casistica tradizionale, poiché interamente collegate alle modalità peculiari di funzionamento dei sistemi di IA. Sono tutte quelle ipotesi in cui il danno si sia verificato a causa di un comportamento autonomo ed imprevedibile dell'algoritmo, derivante dall'apprendimento e dall'interazione continua con l'ambiente circostante.

Infatti, per come abbiamo già avuto modo di anticipare, il comportamento dei sistemi di IA non è prevedibile poiché non può essere deterministicamente fatto risalire al

[341] Vd. U. RUFFOLO, *Responsabilità da A.I. e da algoritmo*, cit. p. 107; ID., *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 247. L'Autore esaurisce il campo dei possibili difetti nell'assenza di "blocchi etici", ossia di misure che incorporino nella macchina "pensante" i principi del *neminem laedere*. In realtà, come diremo nella parte che segue, l'incorporazione di questi "vincoli etici" non sempre impedisce il verificarsi di eventi lesivi.

codice algoritmico: esso deriva infatti dall'acquisizione di dati derivanti dall'ambiente esterno, dalla capacità di apprendimento e dall'eventuale connessione con l'attività di altri sistemi di IA. Gli agenti intelligenti utilizzano tecniche di *machine learning* che decidono sulla base di previsioni statistiche che non sono infallibili e che possono creare conseguenze dannose imprevedibili da parte degli stessi sviluppatori, programmatori o produttori[342].

Perciò è complicato stabilire se il comportamento lesivo sia derivante o meno da un difetto di progettazione. A meno di non voler considerare la stessa capacità di apprendimento come un difetto sarà complicato stabilire quando il sistema intelligente sia stato progettato in maniera difettosa[343]. Non è del tutto chiaro quale debba essere lo standard di comparazione rispetto al quale si debba valutare la difettosità del prodotto.

Non è chiaro se lo standard sia la decisione che, in condizioni normali, avrebbe potuto essere presa da una persona fisica o se, invece, quella che avrebbe preso un algoritmo "ragionevole"[344]. Ma, comunque, anche se si dovesse considerare la deviazione dell'algoritmo rispetto al percorso decisionale ordinario come integrante la nozione di difetto, la lettera della norma sembrerebbe

[342] K.A. CHAGAL-FEFERKORN, *Am I an Algorithm or a Product*?, cit., p. 84, in relazione alla *product liability* statunitense. Analogamente, quanto alla disciplina europea, vd. A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea,* cit., p. 1720; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema?*, cit., p. 463; M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, in M. Ebers – S. Navas (eds.), *Algorithms and Law*, cit., p. 58; C. WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, cit., p. 159; J.S. BORGHETTI, *How can Artificial Intelligence be Defective*?, cit., p. 70 ss.; A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, cit., p. 74.

[343] C. WENDEHORST, *Strict Liability for AI and other Emerging Technologies*, cit., p. 159; J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, cit., p. 70 ss.; EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 28; S. BECK, *The problem of ascribing legal responsibility in the case of robotics*, cit., p. 475.

[344] J.S. BORGHETTI, *How can Artificial Intelligence be Defective*?, cit., p. 68-71; G. WAGNER, *Robot Liability*, p. 42-5; U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., p. 1261; parafrasiamo qui il titolo di un testo che ha avuto molto successo in questo ambito, R. ABBOTT, *The Reasonable Robot. Artificial intelligence and the law*, Cambridge University Press, 2020.

escludere che il produttore debba rispondere del danno cagionato dal prodotto intelligente, posto che l'art. 7 della direttiva prevede un'esimente nel caso in cui il vizio emerga successivamente alla messa in circolazione del prodotto, nonché un'esimente per il rischio da sviluppo.

Se la tendenza è quella di approssimare sempre più la nozione di *design defect* ad una nozione di prevedibilità, per il produttore, dell'esito lesivo derivante dal prodotto, si finisce allora per cadere di nuovo nel *responsibility gap*, che il riferimento ad un modello oggettivistico intendeva evitare. Visto che la caratteristica che differenzia l'IA dai tradizionali algoritmi è proprio l'autonomia e l'imprevedibilità del *decision making*, lo spazio per l'individuazione di una responsabilità del produttore si restringe notevolmente. La verificazione in concreto dell'evento algoritmico lesivo fuoriesce dalla sfera di dominabilità e controllo dell'essere umano. La prevedibilità di tale evento è ostacolata anche dalla capacità dei sistemi di IA di interagire con l'ambiente circostante e con gli altri algoritmi intelligenti, nonché dall'opacità dei meccanismi di funzionamento degli algoritmi di *machine learning*.

Sono proprio le caratteristiche intrinseche dei sistemi di IA ad avere un impatto determinante sulla conformazione dei difetti di *design*. Non si tratta più di vizi seriali[345]. Ciò in quanto il comportamento dei sistemi di IA non dipende soltanto dal codice installato in essi durante la fase di progettazione ma anche dai nuovi dati che detti sistemi acquisiscono grazie all'interazione con l'ambiente e con altri dispositivi. Pertanto, pare non si possa applicare, nell'ambito dell'accertamento di presunti *design defects* relativi a sistemi di IA, quella agevolazione probatoria a cui la Corte di Giustizia ha fatto a volte ricorso, in forza

[345] A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, cit., p. 292; U. SALANITRO*, Intelligenza artificiale e responsabilità*, cit., p. 1264, che giunge a chiedersi se questa asimmetria tra tradizionali difetti di design e difetti derivanti dal self learning renda difficilmente ascrivibile la seconda categoria alla prima

della quale <<l'accertamento di un potenziale difetto di prodotti appartenenti al medesimo gruppo o alla medesima serie di produzione consente di qualificare come difettosi tutti i prodotti di tale gruppo o di tale serie, senza che occorra dimostrare il difetto del prodotto interessato>>[346].

Verso nuovi criteri di valutazione del difetto: la proposta di revisione della *Product liability directive*.

La proposta di revisione della direttiva aggiunge nuovi criteri di valutazione della difettosità del prodotto e riformula quelli già esistenti, nel tentativo di definire con maggiore precisione la nozione di difetto e di rispondere alle difficoltà applicative della disciplina con riferimento i nuovi prodotti intelligenti.

L'art. 6 della proposta di revisione prevede, infatti, che un prodotto sia considerato difettoso quando non offre la sicurezza che il grande pubblico può legittimamente attendersi, tenuto conto di tutte le circostanze, tra cui <<*a) la presentazione del prodotto, comprese le istruzioni per l'uso, l'installazione e la manutenzione; b) l'uso e abuso ragionevolmente prevedibili del prodotto; c) gli effetti sul prodotto dell'eventuale capacità di continuare ad integrare dopo la sua diffusione; e) il momento in cui il prodotto è stato immesso sul mercato o messo in servizio oppure, qualora il fabbricante mantenga il controllo sul prodotto dopo tale momento, il momento in cui il prodotto è uscito dal controllo del fabbricante; f) i requisiti di sicurezza del prodotto, compresi i requisiti di cibersicurezza rilevanti per la sicurezza; g) qualunque intervento di un'autorità di regolamentazione o di un operatore economico di cui all'articolo 7 in relazione alla sicurezza del prodotto; h) le specifiche aspettative degli utenti finali cui è destinato il prodotto*>>.

[346] Corte di Giustizia UE, 5 marzo 2015, Boston Scientific Medizintechnik GmbH, cit.; si trattava, in questo caso, di difetto inerente a pacemaker; vd. in questo senso AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1724.

Chiaramente positivo è il tentativo di individuare le circostanze valutative del difetto tarate sugli *smart products*. La proposta, infatti, prende atto di talune caratteristiche tipiche dei prodotti intelligenti, che possono incidere sulla formazione delle aspettative da parte dei consumatori e sulla conseguente conformazione dei difetti, ovvero il fatto che a volte questi sono rivolti a specifiche categorie di consumatori. Il testo, tuttavia, non chiarisce come debba essere valutata la caratteristica più dirompente nell'ottica dell'attribuzione della responsabilità: la capacità di apprendimento. La lett. c) la colloca tra i criteri che debbono essere apprezzati dal giudice ma non si comprende in che modo debba influire nel giudizio di difettosità. Non si comprende se la capacità di *machine learning* debba essere valutata come circostanza che induce a ritenere che il prodotto sia difettoso o se l'auto-apprendimento debba essere considerato come elemento che contribuisce ad escludere la difettosità del prodotto.[347] Pare ovvio che la capacità di *machine learning* non possa essere considerata come un difetto di per sé[348]. Ma, nonostante ciò, si deve comunque considerare che la capacità di autoapprendimento dei sistemi di IA potrebbe sia aumentare le aspettative di sicurezza del prodotto ai sensi dell'art. 117, co. 1, codice del consumo, sia diminuirle. Sarebbe forse stata necessaria una presa di posizione più trasparente e radicale su quale debba essere lo *standard* di valutazione della difettosità in caso di comportamento emergente ed imprevedibile di un sistema di IA.

5.2.1. Le proposte della dottrina.

[347] Lo nota anche O. DHEU E AL., *The European Commission's Approach To Extra-Contractual Liability and AI – A First Analysis and Evaluation of the Two Proposals*, CiTiP Working Paper Series, 6 ottobre 2022, in SSRN, disponibile a questo link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4239792, p. 35-36; vd. anche G. SPINDLER, *Different Approaches for Liability of Artificial Intelligence – Pros and Cons – the New Proposal of the EU Commission on Liability for Defective Products and AI Systems*, 13 febbraio 2023, in SSRN, disponibile a questo link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354468, p. 11-12.

[348] J. DE BRUYNE – J. TANGHE, *Liability for Damage Caused by Autonomous Vehicles: A Belgian Perspective, in 8 Journal of European Tort Law*, iss. 3, p. 370.

In dottrina sono state proposte diverse interpretazioni evolutive del *consumer expectation test*. Ciò che a noi interessa è la questione circa la configurabilità del difetto nell'ipotesi di comportamento emergente ed imprevedibile del prodotto soggettivizzato. Secondo parte della dottrina, il modo migliore per tutelare il consumatore che abbia subito un danno a causa di un comportamento emergente di un sistema di IA consisterebbe nell'eliminazione del requisito del difetto. In forza di ciò, per ottenere il risarcimento, il danneggiato può limitarsi alla dimostrazione del nesso di causalità tra prodotto ed evento lesivo, senza che sia necessaria la prova di una "condizione di normale pericolosità" del prodotto. È lo stesso orientamento che la giurisprudenza di merito ha cercato di seguire alleggerendo l'onere probatorio a carico del danneggiato fino a ritenere che fosse sufficiente la dimostrazione di un malfunzionamento del prodotto per dimostrare la sussistenza del difetto, spettando al produttore l'onere di fornire un'eventuale prova liberatoria ai sensi dell'art. 118 cod. cons.

L'effetto di detto orientamento sarebbe facilitare la posizione del danneggiato, rendendo puro un sistema di responsabilità oggettivo che invece dovrebbe essere incentrato su valutazioni inerenti alla rimproverabilità del produttore. La stessa giurisprudenza afferma, d'altronde, che il difetto del prodotto non possa essere dedotto dalla mera verificazione di un evento dannoso, ma che esso debba essere oggetto di un autonomo giudizio[349].

A questo risultato, in prospettiva *de jure condendo*, si potrebbe arrivare tramite un alleggerimento dell'onere probatorio raggiungibile con l'applicazione dell'*inference doctrine.* Trattasi di corollario del ben noto principio del *res ipsa loquitur* in forza del quale il solo fatto che il prodotto abbia causato un danno faccia presumere l'esistenza del

[349] Cass. civ., sez. III, 29 luglio 2015, n. 15851; Cass. civ., sez. III, 26 giugno 2015, n. 13225; Cass. civ., sez. III, 29 maggio 2013, n. 13458; Cass. civ., ord. 23 maggio 2013, n. 12665; Cass. civ., sez. III, 15 marzo 2007, n. 6007.

difetto[350]. Questa parrebbe essere anche la soluzione prescelta dalla Commissione europea con riferimento alla proposta di revisione della direttiva sulla responsabilità per danno da prodotto. L'art. 9, paragrafo 2, lett. c) sancisce che la difettosità del prodotto possa essere presunta nel caso in cui l'attore provi che <<il danno è stato causato da un malfunzionamento evidente del prodotto durante l'utilizzo normale o in circostanze ordinarie>>. Il paragrafo 4 dello stesso articolo stabilisce che, nel caso in cui l'autorità giudiziaria ritenga che l'attore stia incontrando difficoltà eccessive nel provare il carattere difettoso del prodotto, a causa della sua complessità tecnica o scientifica, la difettosità si può presumere nel caso in cui <<il prodotto ha contribuito a cagionare il danno>>.

Un'alternativa: l'agente modello umano.

Secondo altri, la difettosità del prodotto andrebbe valutata in rapporto all'ipotetico comportamento di un agente modello umano[351]. Ciò al fine di verificare se una persona ordinariamente diligente, nello svolgere le funzioni delegate alla macchina, avrebbe potuto evitare il verificarsi dell'evento lesivo. La *ratio* di detto orientamento è da rinvenire nel fatto che i sistemi di IA, essendo più funzionali rispetto ai sistemi tradizionali, creerebbero una legittima aspettativa di sicurezza nei consumatori e, per questo motivo, dovrebbero essere considerati privi della <<sicurezza che ci si può legittimamente attendere>> quando causano danni che una persona fisica sarebbe stata in grado di evitare.

[350] D.C. VLADECK, *Machines Without Principals: Liability Rules and Artificial Intelligence*, in Wash. L. Rev., 2014, p. 128; S. DORMONT, *Quel régime de responsabilité pour l'intelligence artificielle*, in *Contrats, concurrence, consommation*, 2018, études 19, § 4.

[351] A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, cit. p. 1019; D.C. VLADECK, *Machines Without Principals*, cit., p. 128; R. ABBOTT, *The Reasonable Robot*, cit., p. 59; J.K. GURNEY, *Imputing Driverhood: Applying a Reasonable Driver Standard to Accidents Caused by Autonomous Vehicles*, in P. Lin e al. (eds.), Robot Ethics 2.0, cit., p. 55 ss.

Diverse le critiche che possono essere mosse a siffatto orientamento. Gli errori che può commettere un sistema di IA, di fatti, sono diversi da quelli commettibili dalle persone fisiche. Non sarebbe ragionevole un paragone in questo senso[352]. L'*automotive* è un settore abbastanza esemplificativo. Le cause degli incidenti automobilistici ad oggi sono sempre riconducibili alla distrazione dei conducenti, alla sonnolenza o all'alterazione delle percezioni dovuta all'assunzione di alcol o stupefacenti. Si tratta di eventi che non possono di certo coinvolgere sistemi di guida autonoma che potrebbero, al più, causare incidenti dovuti a difficoltà nel riconoscimento di segnali o suoni. A volte può bastare un adesivo o un graffito applicato sul segnale di *stop* per indurre il sistema in errore.

Qualsiasi conducente umano avrebbe potuto riconoscere il segnale di "*stop*" e, in ottica di diligenza media, si sarebbe fermato. Ecco perché, se intendessimo paragonare l'errore umano a quello algoritmico in un caso come questo, si arriverebbe sempre all'affermazione della presenza di un difetto, giungendo ad una soppressione *de facto* del requisito del difetto poiché l'IA nei casi concreti quando sbaglia lo fa in maniera molto stupida, se comparato con il comportamento umano corrispondente[353].

Un'altra via: l'agente modello algoritmico.

Altra parte della dottrina ritiene invece che l'evitabilità da parte di un conducente umano diligente di un singolo evento lesivo sia da ritenersi irrilevante[354]. Lo *standard* di

[352] G. WAGNER, *Robot, Inc*., cit., p. 605; M.A. GEISTFELD*, A Roadmap for Autonomous Vehicles*, cit., p. 1644–47; U. SALANITRO, *Intelligenza artificiale e responsabilità*, cit., p. 1261; A. FUSARO, *Quale modello di responsabilità per la robotica avanzata*?, cit., p. 1350.

[353] J.M BISHOP, *Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It*, in *Front. Psychol*., 5 gennaio 2021, disponibile a questo link: https://doi.org/10.3389/fpsyg.2020.513474. Vd., a tal proposito, gli esempi di errori dei sistemi di *machine learning* che abbiamo riportato *supra*, cap. I, § 2.2.

[354] G. WAGNER*, Robot*, Inc., cit., p. 605; ID., *Robot Liability*, cit., p. 44-45; J.S. BORGHETTI, *How can Artificial Intelligence be Defective?*, p. 63 ss.; concordi anche G. GUERRA*, Il concetto di difettosità nella realtà che cambia*, cit., p. 278; J. DE BRUYNE E AL., *Tort Law and Damage*

valutazione della difettosità di un prodotto dovrebbe essere piuttosto costituito dalla comparazione con un modello di algoritmo mediamente accurato.

A nostro giudizio, tale soluzione è maggiormente coerente con lo spirito e con la lettera della normativa: è proprio sulla base del funzionamento e sulle condizioni di sicurezza medie dei prodotti che si formano le legittime aspettative dei consumatori. Il rischio insito all'adesione a questa linea è il seguente: che il sistema che ha causato il danno venga ritenuto difettoso ogniqualvolta si dimostri che sul mercato esiste un algoritmo che, nel caso concreto, avrebbe evitato l'incidente[355]. Tale impostazione sarebbe incompatibile con l'art. 117, comma 2, cod. consumo, che esclude che un prodotto sia considerato difettoso per il solo fatto che un altro più perfezionato sia stato messo in commercio in qualunque tempo. Inoltre, non è detto che un altro algoritmo che avrebbe, nel caso specifico, evitato il danno sia per ciò solo più sicuro. Nella comparazione tra la sicurezza del sistema che ha causato il danno e quella media di prodotti analoghi non deve essere valutata la singola *performance* in relazione ad un evento lesivo specifico, quanto piuttosto la prestazione complessiva dell'algoritmo, da valutarsi sulla base di analisi statistiche[356]. A questo fine potrebbe essere opportuna l'individuazione di una soglia di riferimento al di sopra della quale l'algoritmo potrà essere considerato difettoso.

Product liability e *product safety*: verso la *preemption doctrine*.

La disciplina della responsabilità per danno da prodotto deve essere letta unitamente alla normativa inerente alla sicurezza dei prodotti che definisce i requisiti che debbono

Caused by AI Systems, in J., De Bruyne (ed.), *Artificial Intelligence and the Law, Intersentia*, 2021, p. 383.

[355] G. WAGNER, *Robot, Inc*., cit., p. 608; J. DE BRUYNE E AL., *Tort Law and Damage Caused by AI Systems*, cit., p. 383.

[356] J.S. BORGHETTI, *How can Artificial Intelligence be Defective*?, cit., p. 63 ss.; J. DE BRUYNE E AL*., Tort Law and Damage Caused by AI Systems*, cit., p. 383; concorde anche G. GUERRA, *Il concetto di difettosità nella realtà che cambia*, cit., p. 278.

essere soddisfatti affinchè la merce possa essere messa in circolazione.

La disciplina della *product safety* ha come obiettivo quello di <<*garantire che i prodotti immessi sul mercato ovvero in libera pratica siano sicuri*>>, per come dispone l'art. 102, comma 1, codice del consumo. Tale disciplina interviene *ex ante* fornendo le condizioni di accesso al mercato dei prodotti. La normativa sulla *product liability*, invece, bisogna applicarla successivamente al verificarsi dell'evento lesivo ed ha la funzione di assicurare una riparazione economica al soggetto danneggiato dal prodotto: interviene laddove la disciplina sulla sicurezza non sia stata capace di impedire la commercializzazione di prodotti difettosi.

Nonostante gli scopi e le prospettive delle due discipline siano differenti, queste finiscono per influenzarsi a vicenda. Vista la complementarietà tra le due discipline, alcuni autori, prendendo spunto dal modello statunitense della *preemption doctrine*, sono arrivati ad immaginare una integrazione totale tra la normativa in materia di *product liability* e quella in materia di *product safety*, protendendo verso una perfetta assimilazione tra la nozione di "prodotto conforme" agli *standard* in materia di sicurezza e quella di "prodotto non difettoso"[357].

La questione del regime applicabile al danno da prodotto conforme è dibattuta da tempo in dottrina ed in giurisprudenza: fin dalle prime applicazioni della

[357] In particolare, è Enrico Al Mureden che si è fatto esplicito promotore, in Italia, di questo indirizzo; vd. la seconda parte dell'articolo U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1713 ss.; E. AL MUREDEN, *La sicurezza dei prodotti e la responsabilità del produttore*, Torino, 2017; ID., *Il danno da ''prodotto conforme''. Le soluzioni europee e statunitensi nella prospettiva del "Transatlantic Trade and Investment Partnership" (T.T.I.P.*), in Contr. e impr., 2015, p. 388; ID., *Sicurezza "ragionevole" degli autoveicoli e responsabilità del produttore nell'ordinamento giuridico italiano e negli Stati Uniti*, in Contr. e impr., 2012, p. 1506; nello stesso senso, specificamente in materia di sistemi di i.a., C. AMATO, *Product Liability and Product Security: Present and Future, in S. Lohsse e al. (eds.), Liability for Artificial Intelligence and the Internet of Things*, cit., p. 79 ss.; E. MARCHISIO, *In support of "no-fault" civil liability rules for artificial intelligence*, in *SN Social Sciences*, 2021, 1: 54, p. 10 ss.; in chiave problematica vd. G. GUERRA, *Il concetto di difettosità nella realtà che cambia*, cit., p. 273 e 276.

disciplina ci si è posti il problema di valutare se il rispetto degli *standard* tecnici di sicurezza del prodotto che regolano le condizioni e i requisiti di accesso al mercato delle merci sia sufficiente per considerare un prodotto non difettoso e, dunque, per escludere la responsabilità oggettiva del produttore nel caso di eventi lesivi, ovvero se la conformità agli stessi debba essere considerata condizione necessaria ma non sufficiente.

La prima opzione applicata ai danni provocati dai sistemi di IA avrebbe il vantaggio di superare la problematica dell'individuazione del difetto algoritmico. Ciò andrebbe a scapito del danneggiato il quale si troverebbe privo di tutela risarcitoria, posto che un prodotto immesso sul mercato, tramite autorizzazione o autocertificazione, sarà presumibilmente conforme agli standard stabiliti dalla normativa sulla *product safety*.

5.5.1. *Product safety* e tecniche di armonizzazione.

Riteniamo sia opportuno un breve cenno alle norme in materia di sicurezza dei prodotti. A livello generale, abbiamo una legislazione orizzontale che si riferisce ad ogni tipologia di prodotto: la direttiva 01/95/CE *General Product Safety Directive*, c.d. GPSD[358].

Ai sensi dell'art. 1, comma 2, tale direttiva si applica <<nella misura in cui non esistano, nell'ambito della normativa comunitaria, disposizioni specifiche aventi lo stesso obiettivo che disciplinano la sicurezza dei prodotti in questione>>. Trattasi, dunque, di normativa con applicazione residuale. Le disposizioni specifiche cui si riferisce il testo appena citato costituiscono la c.d. legislazione verticale, ovverosia il secondo livello di regolazione della *product safety*, composta da una serie di direttive che disciplinano gli *standard* di sicurezza di determinate e specifiche categorie di prodotti: i farmaci

[358] Direttiva 2001/95/ce del Parlamento Europeo e del Consiglio del 3 dicembre 2001 relativa alla sicurezza generale dei prodotti.

(dir. 2001/83/CE); i presidi medico-chirurgici (dir. 2007/47/CE); i cosmetici (dir. 76/768/CEE); il materiale elettrico (dir. 2006/95/CE); gli autoveicoli (dir. 2007/46/CE); i motoveicoli a due o a tre ruote (dir. 2002/24/CE); gli pneumatici (dir. 1992/23/CE); i giocattoli (dir. 2009/48/CE); le macchine (dir. 2006/42/CE)[359]. La dir. 01/95/CE resta comunque applicabile in questi settori per quanto concerne <<gli aspetti e i rischi o le categorie di rischi non soggetti a tali requisiti>> (art. 1, comma 2, GPSD).

Ovviamente la normativa sulla *product safety* non pretende che il prodotto sia completamente immune da rischi: così come per la definizione di prodotto difettoso, l'art. 2, lett. b), dir. 01/95/CE si accontenta di un concetto relazionale di sicurezza, espresso non soltanto in termini di totale assenza di rischi, ma anche di <<rischi minimi, compatibili con l'impiego del prodotto e considerati accettabili nell'osservanza di un livello elevato di tutela della salute e della sicurezza delle persone>>, in funzione di vari elementi, tra cui le sue <<caratteristiche>> (n. 1), il suo eventuale <<effetto>> su altri prodotti (n. 2), la sua <<presentazione>> (n. 3) e le <<categorie di consumatori>> cui esso è destinato (n. 4), con particolare riguardo a quelle maggiormente vulnerabili (bambini ed anziani). E ancora, parallelamente a quanto stabilito dall'art. 6, comma 2, della direttiva sul danno da prodotto, l'art. 2, lett. b) dir. 01/95/CE stabilisce che <<la possibilità di raggiungere un livello di sicurezza superiore o di procurarsi altri prodotti che presentano un rischio minore non costituisce un motivo sufficiente per considerare un prodotto come non sicuro o pericoloso>>. Il "prodotto pericoloso" è concepito come il negativo del prodotto sicuro: *ex* art. 2, lett.c), dir. 01/95/CE è pericoloso

[359] La c.d. "direttive macchine" – particolarmente rilevante nel settore che qui ci interessa, poiché destinata ad applicarsi quantomeno nel settore della robotica – è attualmente soggetta ad un procedimento di revisione, vd. Proposta di Regolamento del Parlamento Europeo e del Consiglio sui prodotti macchina; Bruxelles, COM (2021) 202 final, 21 aprile 2021, che ha proprio l'obiettivo di adeguare la normativa sulla sicurezza dei macchinari ai nuovi prototipi di "robotica intelligente".

<<qualsiasi prodotto che non risponda alla definizione di prodotto sicuro>>.

La definizione generale di sicurezza viene, poi, dettagliata nelle direttive verticali che stabiliscono i requisiti essenziali che i prodotti debbono soddisfare per poter liberamente circolare nel mercato europeo. Tuttavia, tali direttive indicano solo il traguardo, ma non illustrano le modalità tecniche attraverso cui tali requisiti devono essere soddisfatti[360]. I dettagli sono delineati tramite il rinvio a norme tecniche emanate da organismi privati (c.d. enti di normalizzazione), che presentano il vantaggio di essere aggiornate costantemente in base all'evoluzione delle conoscenze scientifiche e tecniche, nonché di rappresentare lo stato dell'arte di una determinata materia[361].

Tra le norme tecniche, un particolare rilievo è stato assunto, a far data dalla Risoluzione del Consiglio delle Comunità Europee del 1985 sul c.d "Nuovo Approccio", dalle c.d. "norme armonizzate", ovverosia *standard* elaborati dagli organismi europei di normalizzazione (*European Standardisation Organisations*), che hanno l'obiettivo di favorire una maggiore armonizzazione tra le normative interne e, di conseguenza, di agevolare la circolazione dei beni[362].

---

[360] In questi termini U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, in *Resp. civ. prev*., fasc. 1, 2005, contenuto in E. Al Mureden (a cura di), *La sicurezza dei prodotti e la responsabilità del produttore. Casi e materiali*, Giappichelli, 2020, p. 70.

[361] U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 63.

[362] Per un approfondimento sul funzionamento del processo europeo di armonizzazione e standardizzazione, rinviamo a COMMISSIONE EUROPEA, La guida blu all'attuazione della normativa UE sui prodotti 2022, 29 giugno 2022, 2022/C 247/01, nonché ai lavori di M.P. EGAN, *Constructing a European Market: Standards, Regulation, and Governance*, Oxford University Press, 2001 e B. VAN LEEUWEN, *European Standardisation of Services and Its Impact on Private Law: Paradoxes of Convergence*, Hart Publishing, 2017. Sullo specifico problema della delega di poteri regolatori ad enti privati e dello scarso controllo democratico sull'esercizio di tali poteri vd. G. SPINDLER, *Market processes, standardisation and tort law*, in *European Law Journal*, 2002, vol. 4, iss. 3, p. 316 ss. R. VAN GESTEL – H.W. MICKLITZ, *European Integration through Standardization: How judicial review is breaking down the club house of private standardization bodies*, in *Common Market Law Review*, 2013, p. 145 ss.; con

Le norme armonizzate vengono adottate in base ad una particolare procedura che inizia da un mandato della Commissione e si conclude con la pubblicazione dei riferimenti della norma adottata nella Gazzetta Ufficiale dell'UE. Gli organismi europei di normalizzazione ai quali la Commissione può richiedere l'elaborazione degli standard tecnici sono il Comitato europeo di normalizzazione (CEN), il Comitato europeo di normalizzazione elettronica (CENELEC) e l'Istituto Europeo per le norme di Telecomunicazioni (ETSI, *European Telecommunications Standards Intitute*).

Elaborati gli *standards* vengono inviati alla Commissione. Essa effettua un controllo sulla corrispondenza rispetto alla richiesta di armonizzazione, nonché rispetto ai principi ispiratori della disciplina generale in materia di standardizzazione. Occorre comunque rilevare che la Commissione normalmente non esamina nel merito gli *standard* tecnici emanati dagli organi di standardizzazione, ma si limita ad un'operazione di mera comparazione formale.

Le norme armonizzate vengono successivamente trasposte in norme tecniche nazionali a cura degli enti nazionali di normalizzazione[363] e sostituiscono le norme tecniche vigenti nei diversi paesi europei, che spesso contrastano tra di loro. L'adesione alle norme armonizzate da parte delle imprese assicura una presunzione di sicurezza dei prodotti; mentre la conformità agli altri *standard* tecnici elaborati a livello europeo costituisce soltanto uno dei vari elementi che deve essere preso in

riferimento all'AI Act vd. M. EBERS, *Standardizing AI. The Case of the European Commission's Proposal for an 'Artificial Intelligence Act*', in L.A. DiMatteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence*, cit., p. 340 ss.

[363] U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 63. Tra gli enti nazionali di normalizzazione vi sono l'Ente Nazionale Italiano di Unificazione (UNI) e il Comitato Elettrotecnico Italiano (CEI) in Italia; il Deutsches Institut fur Normung (DIN) in Germania; la British Standardization Institution (BSI) nel Regno Unito; l'Association Frangaise de Normalisation (AFNOR) in Francia

considerazione dalle autorità preposte al controllo per valutare l'eventuale insicurezza del prodotto.

È comunque vero che la presunzione debba essere considerata iuris tantum, causando un'inversione dell'onere della prova nella valutazione della sicurezza del prodotto; in caso di osservanza delle norme armonizzate, spetterà all'autorità preposta al controllo dimostrare l'eventuale non conformità del prodotto ai requisiti di sicurezza stabiliti dalla legislazione verticale o orizzontale. Dunque, pur essendo liberi se aderire o meno alle norme armonizzate, i produttori sono incentivati a farlo, in quanto la conformità consente loro di essere esonerati dalla difficile opera interpretativa dei requisiti stabiliti in via generale dalle direttive e, poi, dall'onere della prova circa la sicurezza del prodotto[364].

Per quanto concerne l'attestazione della conformità del prodotto alle norme tecniche, essa può avvenire o per mezzo della certificazione di organi specializzati o attraverso una dichiarazione di conformità redatta e documentata dal fabbricante stesso[365]. Per la maggior parte dei prodotti, il controllo *pre-marketing* è effettuato direttamente dal produttore, che, apponendo il marchio CE, certifica la conformità del prodotto agli standard tecnici. La *ratio* di tale "Nuovo Approccio" si ispira all'idea per cui <<Il produttore, pur possedendo le conoscenze dettagliate relative al processo di progettazione e produzione si trova nella posizione migliore per eseguire la procedura completa di valutazione della conformità. La procedura di valutazione

[364] M. VEALE, *Demystifying the Draft EU Artificial Intelligence Act*, in *Computer Law Review International*, 2021, n. 4, p. 105; va comunque sottolineato che le norme armonizzate possono acquisire carattere vincolante, se richiamate in leggi o regolamenti che stabiliscono i requisiti essenziali di sicurezza.

[365] U. CARNEVALI, *La norma tecnica da regola di esperienza a norma giuridicamente rilevante. Ricognizione storica e sistemazione teorica. Ruolo dell'UNI e del CEI*, in *Resp. civ. e prev*., 1997, contenuto in E. Al Mureden (a cura di) La sicurezza dei prodotti e la responsabilità del produttore. Casi e materiali, cit., p. 52.

della conformità dovrebbe quindi rimanere obbligo del solo fabbricante>>[366].

Per un ridotto numero di prodotti invece, ad esempio i farmaci, è un'autorità pubblica ad effettuare una valutazione di conformità alle norme tecniche ed a concedere l'autorizzazione alla immissione in commercio.

5.5.2. Sistemi di IA e norme di armonizzazione: brevi cenni sull'*AI Act*.

L'*AI Act* detta i requisiti di sicurezza che i sistemi di IA, definiti dall'art. 3 come ogni <<*sistema basato su una macchina che è progettato per operare con vari livelli di autonomia e che può mostrare capacità di adattamento dopo l'implementazione e che, per obiettivi espliciti o impliciti, deduce dagli input ricevuti come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali*>> dovranno soddisfare per essere introdotti nel mercato[367].

---

[366] Decisione n. 768/2008/CE del Parlamento europeo e del Consiglio del 9 luglio 2008 relativa a un quadro comune per la commercializzazione dei prodotti e che abroga la decisione 93/465/CEE, considerando 21.

[367] Sulla proposta di regolamento si vedano i commenti di L. FLORIDI, *The European Legislation on AI: a Brief Analysis of its Philosophical Approach*, in *Philosophy & Technology*, 2021, n. 34, p. 215 ss.; M. EBERS, *Standardizing AI*, cit.; W.G. VOSS, *AI Act: The European Union's Proposed Framework Regulation for Artificial Intelligence Governance*, in 25 Journal of Internet Law, 2021, p. 7 ss.; M. VEALE, *Demystifying the Draft EU Artificial Intelligence Act, cit.; J. SCHUETT, Risk management in the Artificial Intelligence Act*, arXiv preprint, 2022, disponibile a questo link: https://aps.arxiv.org/pdf/2212.03109.pdf; F. SOVRANO E AL., Metrics, *Explainability and the European AI Act Proposal*, in J, 2022, n. 5, p. 126 ss.; nella letteratura italiana vd. G. ALPA, *Quale modello normativo europeo per l'intelligenza artificiale?*, cit., p. 1009 ss.; U. RUFFOLO – A. AMIDEI, *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design, strumenti di certificazione preventive e «autodisciplina» di settore*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?,* vol. 1, cit., p. 489 ss.; C. CASONATO – B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 2021, 3, p. 415 ss.; G. FINOCCHIARO, *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in Dir. inf., 2022, n. 2, p. 303 ss.; F. PIZZETTI, *La proposta di Regolamento sull'IA della Commissione Europea presentata il 21.4.2021* (COM (2021) 206 final) tra Mercato Unico e competizione digitale globale, in Dir. internet, 2021, n. 4, p. 591 ss.; per i profili di interesse penalistico vd. F.C. LA VATTIATA, *Brevi note "a caldo" sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale*, in *Dir. pen. uomo*, 30 giugno 2021; A. LAVORGNA – G. SUFFIA, *La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell'ambito della giustizia penale*, in *Dir. pen. cont. – Riv. Trim*., 2021, n. 2, p. 88 ss.

Il regolamento introduce una disciplina speciale rispetto al quadro normativo già esistente sulla sicurezza dei prodotti. La dir. 01/95/CE, infatti, continuerà ad avere vigore come <<rete di sicurezza>> al fine di garantire che i sistemi collegati a prodotti che non sono ad alto rischio in conformità al presente regolamento e che pertanto non sono tenuti a rispettare i requisiti ivi stabiliti siano comunque sicuri al momento dell'immissione sul mercato o della messa in servizio>>.

L'*AI Act* è entrato in vigore il primo agosto 2024, ma la sua applicazione completa è scaglionata, con diverse scadenze chiave: divieti di pratiche di IA inaccettabili ed obblighi di trasparenza in vigore dal 2 febbraio 2025; obblighi per i modelli di IA ad uso generale e le loro norme di governance in vigore dal 2 agosto 2025; piena applicazione per la maggior parte dei sistemi ad alto rischio, come quelli in ambito medico, assunzioni e sicurezza, e piena operatività per aziende e professionisti dal 2 agosto 2026; ulteriori disposizioni per sistemi ad alto rischio integrati in prodotti regolamentati e sistemi di IA ad alto rischio immessi sul mercato prima del 2026 in vigore dal 2 agosto 2027; termine ultimo per alcuni obblighi specifici per i sistemi ad alto rischio individuato nel 2028. Il cuore della normativa e gli obblighi più stringenti per le imprese iniziano a decorrere tra il 2025 ed il 2026, con una gradualità che copre fino al 2027-2028, per permettere un adeguamento progressivo alle nuove regole.

Pertanto, ciò che commenteremo di seguito è la composizione del regolamento attuale, considerando che tra la proposta di regolamento del 2021 e la sua definitiva approvazione ci sono state ulteriori negoziazioni tra Parlamento, Consiglio e Commissione (c.d. triloghi) che hanno comportato qualche differenza tra i due testi.

L'UE ha scelto, con l'*AI Act* di approcciare un modello *risk-based*, decidendo di vietare l'immissione in commercio

di una serie di pratiche di intelligenza artificiale. Immaginando una sorta di rappresentazione a cerchi concentrici, si può pensare a pratiche di IA vietate poiché comportanti rischi socialmente inaccettabili ed a pratiche consentite ma solo con il rispetto di dettagliati obblighi di trasparenza e di informazione.

Al vertice della classificazione si trovano, per l'appunto, le pratiche espressamente vietate dall'art. 5 del regolamento, in quanto ritenute dannose per i diritti e le libertà fondamentali. Le pratiche vietate includono: manipolazione e inganno, tramite tecniche subliminali o manipolative; sfruttamento della vulnerabilità, rivolto a bambini, anziani o persone in difficoltà; *social scoring*, ossia la valutazione delle persone basata su comportamenti sociali o personali; predizione del rischio criminale individuale, basata esclusivamente su dati biometrici o profili personali; *scraping* indiscriminato di immagini facciali, per creare *database* di riconoscimento senza consenso; riconoscimento delle emozioni, vietato in scuole e luoghi di lavoro salvo talune espresse eccezioni; categorizzazione biometrica, usata per inferire razza, orientamento politico, religione e altro; identificazione biometrica in tempo reale, vietata nei luoghi pubblici per la sorveglianza di massa.

L'adozione di questi divieti non è casuale, ma funzionale a preoccupazioni concrete legate alla sorveglianza di massa, alla discriminazione automatizzata ed alla manipolazione del comportamento umano. Il regolamento introduce sanzioni severe che possono arrivare fino a 35 milioni o al 7% del fatturato globale per chi vìola le disposizioni.

5.5.2.1. Le pratiche di IA vietate: approfondimento.

Per come appena visto, diverse sono le pratiche che l'*AI Act* considera socialmente inaccettabili. Occorre, sia pure

senza pretese di esaustività, cercare di comprendere il perché dietro tale scelta.

Si è visto, innanzitutto, che il regolamento vieta l'uso di tecniche subliminali o manipolative che influenzano il comportamento umano senza che l'individuo ne sia consapevole. Detto divieto ha lo scopo di proteggere la capacità decisionale autonoma e la libertà di autodeterminazione propria di ogni essere umano. Le tecniche subliminali possono includere stimoli visivi o sonori impercettibili, messaggi nascosti in contenuti digitali o il condizionamento inconscio attraverso *pattern* di interazione con un sistema di IA. La manipolazione può anche avvenire tramite sistemi di c.d. *nudgning* avanzato, che inducono le persone a compiere azioni che non avrebbero scelto altrimenti. Ad esempio, un tipo di IA che modifica in modo impercettibile il tono di un *chatbot* per influenzare le risposte dell'utente, spingendolo ad acquistare un prodotto o a fornire dati personali senza neanche accorgersene. Tale divieto si applica, però, solo se la manipolazione conduce ad un danno significativo, come la perdita di danaro, danni psicologici o discriminazioni. È un divieto particolarmente rilevante soprattutto nel settore pubblicitario, dei *social* e dell'*e-commerce*, laddove tecniche persuasive avanzate potrebbero essere implementate per ottenere un vantaggio commerciale ingiusto.

In secondo luogo, il regolamento vieta i sistemi di IA che sfruttano vulnerabilità legate ad età, disabilità o situazioni socio-economiche: i sistemi di IA non possono, dunque, approfittare della fragilità cognitiva dei bambini, della ridotta capacità di giudizio di persone con disabilità mentali o delle difficoltà finanziarie di chi si trova in situazioni di povertà. Si pensi ad un giocattolo interattivo che spinga un bambino ad effettuare acquisti senza la consapevolezza del genitore: questo tipo di IA è considerato inaccettabile poiché mina la dignità umana e può portare a danni economici e psicologici. Il divieto non si applica alle IA che forniscono assistenza inclusiva come

strumenti per persone con disabilità, a patto che non abusino della vulnerabilità dell'utente per fini ingannevoli.

In terzo luogo, sono proibiti i sistemi di *social scoring*, ovverosia di classificazione degli individui in base al loro comportamento sociale o alle loro caratteristiche personali. Tale divieto si ispira alla preoccupazione che detti sistemi possano portare a discriminazioni sistemiche e trattamenti iniqui. Per esempio un sistema di IA che analizza le interazioni sui *social media*, il comportamento finanziario o la partecipazione a eventi politici per decidere chi può accedere ad un prestito o a un impiego sarebbe vietato. In alcuni paesi, il *social scoring* è stato implementato per limitare l'accesso a servizi pubblici in base a criteri di affidabilità sociale, creando disuguaglianze e restrizioni ingiuste. Il regolamento non vieta tutte le forme di valutazione automatizzata, ma soltanto quelle che comportano trattamenti ingiustificati o sproporzionati in contesti non correlati ai dati analizzati.

Ancora, il regolamento vieta di valutare il rischio che una persona commetta un crimine sulla base di caratteristiche personali, profili psicologici o dati biometrici. Trattasi di sistemi pericolosi e discriminatori poiché potrebbero portare a pregiudizi razziali o sociali, violando il principio di presunzione d'innocenza. Un *software* che attribuisce un punteggio di pericolosità basato sul luogo di nascita, sulla situazione economica o sull'etnia di una persona sarebbe vietato. Tuttavia, l'*AI Act* non vieta l'uso di strumenti predittivi che si basano su dati oggettivi e verificabili come la mappatura delle zone ad alto tasso di criminalità. L'obiettivo del divieto è quello di evitare la profilazione ingiustificata e garantire che le valutazioni di rischio siano sempre supportate da elementi fattuali e verificabili.

Inoltre, il regolamento vieta gli strumenti usati per il riconoscimento delle emozioni nei contesti di lavoro ed

istruzione, per evitare violazioni della *privacy* e discriminazioni: sistemi che analizzano espressioni facciali, tono di voce o gesti per valutare lo stato emotivo di una persona possono essere usati in modo invasivo. Un'azienda non può, ad esempio, costantemente monitorare il livello di *stress* dei dipendenti. Sono, tuttavia, previste eccezioni per motivi medici o di sicurezza come il monitoraggio dello stato d'animo del paziente.

Il regolamento vieta, altresì, i sistemi che usano dati biometrici per classificare le persone in base a caratteristiche sensibili come etnia, religione, orientamento sessuale o politico. Tali sistemi, infatti, possono portare a discriminazioni e violazioni della *privacy*: un esempio è una IA che analizza tratti facciali per dedurre l'appartenenza etnica ed influenzare decisioni di assunzione o concessione di prestiti. Il divieto si applica anche alle forze dell'ordine, tranne che per l'etichettatura di dati biometrici già raccolti legalmente.

Infine, l'uso di IA per il riconoscimento facciale in tempo reale nei luoghi pubblici è proibito, tranne in casi eccezionali come la lotta al terrorismo o la ricerca di persone scomparse. Questa misura mira a prevenire la sorveglianza e la violazione della *privacy*. Il divieto si applica alle forze di polizia ma non vieta il riconoscimento facciale post-evento, ad esempio per identificare un sospetto dopo un crimine.

In conclusione, l'*AI Act* segna un punto di svolta per l'UE, ponendola come leader globale nella regolamentazione dell'IA. Le proibizioni stabilite mirano a prevenire abusi evidenti come la manipolazione occulta, la sorveglianza indiscriminata e la discriminazione algoritmica. Ma fino a che punto una regolamentazione così stringente può favorire lo sviluppo di una IA etica, sicura ed affidabile e quando invece rischia di soffocare l'innovazione? L'IA ha il potenziale di trasformare la

società ma anche di ampliare pregiudizi, violare la *privacy* e minacciare i diritti fondamentali. L'*AI Act* risponde proprio a queste preoccupazioni, ponendo limiti chiari ed obblighi ben precisi agli sviluppatori ed agli utilizzatori. Senza una solida normativa, il rischio è quello di un *far west* tecnologico, dove le grandi aziende possono operare senza vincoli etici. Dunque, la regolamentazione è fondamentale per garantire trasparenza, responsabilità ed affidabilità nell'uso di queste tecnologie. Dall'altro lato, però, un eccesso di regolamentazione può portare a rallentare l'innovazione creando ostacoli burocratici per le aziende europee che potrebbero trovarsi in svantaggio rispetto ai competitors d'oltreoceano; scoraggiare investimenti e ricerca poiché i finanziatori potrebbero preferire paesi con meno prescrizioni; creare incertezze interpretative costringendo le aziende a navigare in un panorama normativo complesso e soggetto a continue revisioni.

L'UE deve trovare un punto di equilibrio tra protezione dei diritti e libertà fondamentali ed innovazione. Le linee guida della Commissione cercano di chiarire le norme ma c'è il rischio che l'*AI Act* debba essere costantemente aggiornato per adattarsi a sempre nuove tecnologie. Occorre, dunque, un approccio più dinamico con normative capaci di evolversi senza correre il rischio di essere obsolete già al momento della loro entrata in vigore.

L'*AI Act* è un primo passo molto importante nella regolazione dei sistemi di IA, ma la sua efficacia dipenderà da come verrà applicato ed aggiornato nel tempo.

5.5.3. I sistemi di IA consentiti.

I sistemi di IA che non rientrano nelle categorie vietate possono invece essere commercializzati, a patto che rispettino i requisiti stabiliti dal regolamento, che variano a seconda del rischio che essi pongono per la sicurezza e per il rispetto dei diritti fondamentali dei cittadini.

Sono previsti, in particolare, sistemi ad alto rischio, sistemi a rischio limitato e sistemi a rischio minimo o nullo. L'art. 52 dell'*AI Act* impone ai fornitori di sistemi a rischio limitato alcuni obblighi generici di trasparenza al fine di consentire alle persone di essere a conoscenza della natura artificiale di detti sistemi. Tra i sistemi a rischio limitato, vi sono le *chatbots*, i *deepfakes* e gli *emotion recognition systems*. Per quanto concerne le *chatbots*, sono i fornitori a dover progettare il sistema in modo tale che gli utenti siano informati che stanno interagendo con una macchina piuttosto che con un essere umano. Non sono previsti obblighi di trasparenza in capo agli utenti. Solo chi ha progettato la *chatbot* dovrà garantire il rispetto degli obblighi di trasparenza e non anche gli utenti che usano concretamente il *software*. Con riferimento agli *emotion recognition system*s ed ai *deepfakes*, invece, gli obblighi di trasparenza ricadono soltanto sugli utenti.

Vi sono, inoltre, sistemi di IA a rischio nullo o minimo. Questi potranno essere sviluppati senza obblighi, ma i fornitori potranno scegliere liberamente di applicare i requisiti di sicurezza stabiliti per le altre categorie di sistemi, così come di aderire ai codici di condotta di cui all'art. 69 *AI Act*.

5.5.3.1. I sistemi di IA ad alto rischio.

I sistemi ad alto rischio sono soggetti ad obblighi rigorosi. Essi comprendono i sistemi di IA che fungono da componenti di sicurezza in prodotti disciplinati dal diritto dell'UE (art. 6, par. 1, Allegato I), oppure che presentano rischi significativi per la salute, la sicurezza o i diritti fondamentali in settori sensibili quali l'istruzione, l'occupazione, l'applicazione della legge e i servizi pubblici (art. 6, par. 2, Allegato III). Questi sistemi devono soddisfare requisiti tecnici e organizzativi ben dettagliati, tra cui la gestione dei rischi, la trasparenza, la supervisione umana e le valutazioni di conformità prima dell'immissione sul mercato.

A norma dell'art. 6, paragrafo 5, dell'*AI Act*, la Commissione Europea è incaricata di adottare entro il 2 febbraio 2026 le linee guida che spieghino come attuare l'art. 6 in pratica. Ciò include come interpretare i criteri di classificazione e come applicare le esenzioni di cui all'art. 6, par. 3 la Commissione dovrà, inoltre, fornire esempi pratici di sistemi di IA che debbono essere classificati come ad alto rischio e di quelli che, invece, non devono esserlo.

Per addivenire a ciò, la Commissione ha avviato una consultazione sull'IA ad alto rischio, aperta dal 6 giugno al 18 luglio 2025. A detta consultazione hanno partecipato fornitori ed utilizzatori di sistemi di IA, organismi industriali ed associazioni, autorità pubbliche ed autorità di regolamentazione, il mondo accademico ed esperti indipendenti, nonché diverse organizzazioni della società civile. Sarà molto importante l'esito di questa consultazione che, per come anticipato, si avrà entro il febbraio 2026, poiché i sistemi di IA ad alto rischio dovranno soddisfare norme complete che riguardano la governance e la qualità dei dati, i meccanismi di controllo umano, obblighi di trasparenza, robustezza, accuratezza e sicurezza informatica, nonché valutazione della conformità prima dell'immissione sul mercato.

Ma tralasciando gli esiti, alla data in cui si scrive non noti, della consultazione, occorre riflettere sui requisiti per la messa in commercio dei sistemi ad alto rischio. Deve, innanzitutto, essere predisposto un sistema di gestione dei rischi finalizzato all'individuazione dei rischi prevedibili associati a ciascun sistema di IA ad alto rischio e all'adozione di adeguate misure di gestione dei rischi; i *dataset* usati per addestrare e validare i sistemi devono rispondere a parametri di alta qualità affinchè il loro rendimento sia sicuro ed in linea con la destinazione d'uso, dovendo consistere in dati pertinenti, sufficientemente rappresentativi, immuni da errori e

completi[368]; prima di immettere un sistema nel circuito economico occorre predisporre la relativa documentazione tecnica che dimostri il rispetto dei requisiti previsti dal Regolamento; i sistemi devono essere programmati e sviluppati in modo tale da consentire la registrazione automatica degli eventi che si verificano durante il loro impiego, in modo tale da assicurare un livello di tracciabilità del funzionamento del dispositivo intelligente durante tutto il suo ciclo di vita adeguato alla finalità prevista del sistema; i sistemi devono essere sufficientemente trasparenti da consentire agli utenti di interpretare l'*output* di sistema e di utilizzarlo adeguatamente; ogni sistema deve essere accompagnato da istruzioni adeguate e comprensibili, concernenti: identità del fornitore e del rappresentante, caratteristiche capacità e limiti del sistema, le eventuali modifiche che possono riguardare il sistema, le misure di controllo umano, una stima della durata di funzionamento del sistema ed ogni misura di manutenzione e cure inclusi gli aggiornamenti; i sistemi devono essere progettati e sviluppati in modo tale da garantire il loro effettivo controllo da parte di una persona fisica all'atto dell'utilizzo, al fine di prevenire o minimizzare i rischi per la salute, la sicurezza o i diritti fondamentali, che possono emergere in relazione alla destinazione d'uso ed ai prevedibili usi impropri del device (c.d. principio del *human-in-the-loop*); i sistemi di IA devono garantire un livello appropriato di precisione e robustezza; con particolare riferimento ai sistemi dotati di *machine learning* bisognerà assicurare la possibilità di intervento su eventuali *output* affetti da *bias* attraverso puntuali misure correttive; i sistemi devono essere dotati di soluzioni tecniche volte a garantire la sicurezza informatica.

L'art. 40 dell'*AI Act* rinvia per la definizione dei requisiti tecnici alle norme armonizzate, prevedendo una

[368] Non è di agevole comprensione cosa si intenda per "pertinenza" e "rappresentatività" dei dati. Inoltre, è letteralmente impossibile garantire codici algoritmici privi di errore.

presunzione di conformità ai requisiti stabiliti dal titolo III, capo 2, per il caso in cui il produttore si attenga a detti *standard*. Prima ancora che l'*AI Act* delegasse agli organismi europei di normalizzazione la predisposizione di standard tecnici in materia di sistemi di IA, CEN e CENELEC avevano già creato nel 2019 un *focus group* sull'IA finalizzato ad elaborare *standard* per la messa in commercio di *smart products*[369].

I sistemi ad alto rischio sono soggetti al regime di auto-certificazione CE[370]. Solo le applicazioni ad alto rischio elencate nell'ambito dell'identificazione biometrica e categorizzazione delle persone fisiche devono usare organismi notificati specifici per l'*AI Act*. Trattasi comunque di disposizione transitoria poiché, una volta introdotti *standard* armonizzati o specifiche comuni anche in quest'ambito sarà sufficiente l'autocertificazione di conformità da parte del produttore.

5.5.4. La tesi della non difettosità del prodotto conforme agli *standard* tecnici.

Secondo parte della dottrina, il concetto di prodotto sicuro dovrebbe essere considerato coincidente con quello di prodotto non difettoso: un prodotto in grado di

[369] Vd. la notizia sul sito di CEN-CENELEC, CEN and CENELEC launched a new Joint TC on Artificial Intelligence, 3 marzo 2021, disponibile a questo link: https://www.cencenelec.eu/news-and-events/news/2021/briefnews/2021-03-03-new-joint-tc-on-artificial-intelligence/. Sul processo di standardizzazione dei sistemi di i.a., a livello europeo ed internazionale, vd. M. EBERS, Standardizing AI, cit., p. 8 ss.; G. MOBILIO, *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica*, in *Biolaw Journal*, 2020, n. 2, p. 413 ss. Si noti tuttavia che la legge sull'AI presenta un meccanismo misto, poiché per i sistemi biometrici destinati a essere messi in servizio dalle autorità di polizia, di immigrazione o di asilo o dalle istituzioni dell'UE, la valutazione di conformità è effettuata da un apposito ente pubblico, vd. AI Act artt. 43(1), 63(5); lo notano M. VEALE, *Demystifying the Draft EU Artificial Intelligence Act*, cit., p. 106; C. CASONATO – B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, cit., p. 430 ss.

[370] Si noti tuttavia che la legge sull'AI presenta un meccanismo misto, poiché per i sistemi biometrici destinati a essere messi in servizio dalle autorità di polizia, di immigrazione o di asilo o dalle istituzioni dell'UE, la valutazione di conformità è effettuata da un apposito ente pubblico, vd. AI Act artt. 43(1), 63(5); lo notano M. VEALE, *Demystifying the Draft EU Artificial Intelligence Act*, cit., p. 106; C. CASONATO – B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, cit., p. 430 ss.

soddisfare tutti i requisiti per la messa in commercio non potrebbe essere considerato difettoso e dunque dare luogo a responsabilità ai sensi della disciplina di cui alla dir. n. 85/374/CEE. Non si tratta di un orientamento nuovo nell'ambito del dibattito sulla responsabilità per danno da prodotto[371]. Trattasi, comunque, di un orientamento che è stato oggetto di un rinnovato interesse in relazione alle potenzialità lesive degli *smart products*, in quanto fornisce un criterio stabile per la individuazione delle ipotesi di difettosità del prodotto, offrendo una rassicurazione a tutti coloro i quali immetteranno un sistema di IA sul mercato che sia conforme agli *standard* tecnici delineati dall'*AI Act* e dalle norme armonizzate[372]. Stando a detto orientamento, infatti, i sistemi di IA immessi sul mercato in conformità ai requisiti dell'*AI Act* non potranno essere considerati difettosi e, di conseguenza, il danneggiato non potrà chiedere il risarcimento del danno subito sulla base del regime oggettivo previsto dalla dir. n. 85/374/CEE.

La conformità agli *standard* tecnici concorrerebbe ad integrare la nozione di "sicurezza che ci si può legittimamente attendere" che rappresenta uno degli elementi negativi della fattispecie di prodotto difettoso, ai

[371] Vd., oltre agli autori precedentemente citati, U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 75; ID., *La norma tecnica da regola di esperienza a norma giuridicamente rilevante*, cit., p. 50 ss.; P. TRIMARCHI, *La responsabilità civile: atti illeciti, rischio, danno*, cit., p. 419; M. TRONTI, *La direttiva CEE relativa al riavvicinamento delle disposizioni legislative regolamentari ed amministrative degli stati membri in materia di responsabilità per danno da prodotti difettosi*, in Giur. merito, 1988, n. 4, p. 695.

[372] U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1713 ss.; E. AL MUREDEN, *La sicurezza dei prodotti e la responsabilità del produttore*, Torino, 2017; ID., *Il danno da ''prodotto conforme''. Le soluzioni europee e statunitensi nella prospettiva del "Transatlantic Trade and Investment Partnership"* (T.T.I.P.), in *Contr. e impr.*, 2015, p. 388; ID., *Sicurezza "ragionevole" degli autoveicoli e responsabilità del produttore nell'ordinamento giuridico italiano e negli Stati Uniti*, in *Contr. e impr.*, 2012, p. 1506; nello stesso senso, specificamente in materia di sistemi di i.a., C. AMATO, *Product Liability and Product Security: Present and Future*, in S. Lohsse e al. (eds.), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 79 ss.; E. MARCHISIO, *In support of "no-fault" civil liability rules for artificial intelligence*, in *SN Social Sciences*, 2021, 1: 54, p. 10 ss.

sensi degli artt. 6 dir. e 117 codice del consumo[373]. Le aspettative ragionevoli, dunque, in base a tale impostazione, non potrebbero essere tali da pretendere un prodotto più sicuro rispetto a quello che rispetta le norme tecniche. Queste ultime, di fatti, nascendo dal bilanciamento tra le esigenze di certezza e prevedibilità della normativa ed il rischio accettabile, costituiscono lo stato dell'arte in materia di progettazione e fabbricazione dei prodotti[374].

Nella generalità dei casi, il rispetto degli *standard* stabiliti dalle norme armonizzate deve tradursi in una valutazione di non difettosità dei prodotti, poiché sarebbe <<illogico e contraddittorio>> che un prodotto valutato come sicuro in fase di *pre-marketing* possa essere considerato difettoso in un giudizio di responsabilità[375].

La *ratio* dell'orientamento è quella di mandare un messaggio tranquillizzante al produttore, il quale deve essere consapevole che, attenendosi alla normativa, egli non potrà essere chiamato a rispondere per gli eventuali danni causati da un difetto *design*. Il danneggiato potrebbe, al massimo, rivolgersi all'utilizzatore che in alcuni casi potrebbe essere ritenuto responsabile. Qualora sia lo stesso utilizzatore ad essere danneggiato dal prodotto conforme, i costi non possono essere trasferiti su terzi. In ogni caso, resta ferma la possibilità che il produttore sia chiamato a risarcire i danni derivanti da difetti di fabbricazione o da difetti di informazione.

Tale orientamento si ispira alla c.d. *preemption doctrine* statunitense che trova fondamento nella *Supremacy Clause*, ovverosia nell'art. 6, clausola 2, Costituzione USA. Tale clausola stabilisce la espressa supremazia delle norme

[373] U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 75; C. AMATO, *Product Liability and Product Security: Present and Future*, cit., p. 79.
[374] U. CARNEVALI, *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, cit., p. 73.
[375] Ibidem.

federali rispetto alle norme statali, in caso di contrasto tra di esse. Dunque, nell'ambito della responsabilità per danno da prodotto, tale clausola è usata per escludere la responsabilità del produttore nel caso di danno cagionato da un prodotto conforme agli *standard* tecnici federali ma difforme dalle norme previste dalle singole giurisdizioni[376]. In tal caso, l'applicazione della *preemption doctrine* impone di esimere da responsabilità il produttore a patto che gli *standard* abbiano valore di livello massimo di sicurezza e non minimo.

Larga applicazione ha avuto questa teoria con riferimento ai prodotti medici ed agli autoveicoli. Infatti, se un dispositivo medico risponde agli standard di sicurezza stabiliti dalla *Food and Drug Administration*, il produttore non può essere chiamato a rispondere nel caso in cui il prodotto abbia provocato un evento lesivo. Occorre specificare che, comunque, la *preemption doctrine* non gode di uniforme applicazione sul territorio statunitense, tanto che in diversi casi si è arrivati all'adozione di decisioni *contra preemption*.[377]

Product safety e *product liability*: una separazione concettuale.

La maggioranza degli Autori crede che la *preemption doctrine* sia contraria alla direttiva sul danno da prodotto ed al rapporto generale tra *product safety* e *product liability*[378]. Se la sicurezza che ci si può legittimamente

[376] Sulla preemption doctrine nel contesto della product liability statunitense vd. T.O. MCGARITY, *The Preemption War: When Federal Bureaucracies Trump Local Juries*, Yale University Press, 2008; K.N. HYLTON, *Preemption and Products Liability: A Positive Theory*, in *Supreme Court Economic Review*, 2008, p. 205 ss.; C.M. SHARKEY, *Products Liability Preemption: An Institutional Approach*, in *Geo. Wash. L. Rev*., 2008, p. 449 ss.

[377] Così G. GUERRA*, Il concetto di difettosità nella realtà che cambia*, cit., p. 267, che riporta un caso deciso dalla Corte Suprema degli Stati Uniti nel 2005 (n. 03-388.27 del 27 aprile 2005, in Resp. civ. prev., 2006, p. 392) in cui i giudici stabiliscono– con riferimento ai danni cagionati da un pesticida – che la conformità ai requisiti federali non è idonea ad escludere ex se la risarcibilità del danno.

[378] E. BELLISARIO, *Il danno da prodotto conforme tra regole preventive e regole risarcitorie*, cit., passim; EAD*., La rilevanza del criterio presuntivo della conformità alle norme armonizzate*, in Pers. merc., 2012, p. 156 s.; C. CASTRONOVO*, La responsabilità civile*, Giuffrè, 2018, p. 802; A. FUSARO, *I prodotti difettosi e pericolosi: le responsabilità*, cit., p. 390 ss.; C.M. VERARDI,

attendere altro non fosse che quella sancita dagli *standard* di settore, non si comprenderebbero gli sforzi del legislatore per circoscrivere la nozione di difetto, quando sarebbe stato sufficiente un semplice rinvio alla disciplina sulla sicurezza del prodotto. Un elemento che fa pensare alla ragionevolezza di tale interpretazione è fornito dalla disciplina codicistica in tema di responsabilità per la circolazione di veicoli: il comma 4 dell'art. 2054 c.c., nel riferirsi alla responsabilità del proprietario e del conducente per i vizi di costruzione del veicolo, implica che i veicoli, nonostante omologati, possano avere dei vizi e dare luogo a responsabilità del proprietario, del conducente e del produttore[379].

I due corpi normativi avrebbero una funzione diversa. La disciplina della *product safety* serve a garantire che i prodotti immessi sul mercato ovvero in libera pratica siano sicuri e ad assicurare un'applicazione efficace e coerente del requisito generale di sicurezza (considerando 14 dir. 2001/95). A tale scopo, vengono imposti *standard*, controlli e requisiti che debbono essere soddisfatti affinchè il prodotto ottenga l'omologazione e, quindi, l'autorizzazione alla messa in commercio. Trattasi di criteri statici che riguardano il prodotto astratto e che debbono essere valutati *ex ante*, a differenza di quanto accade per la responsabilità per danno da prodotto. Ma, soprattutto, si tratta di condizioni minime il cui rispetto consente la messa in commercio ma non esenta il

Art. 6 – Esclusione della responsabilità, in G. Alpa e al. (a cura di), La responsabilità per danno da prodotti difettosi, Giuffrè, 1990, p. 79 ss.; M. DELLACASA, *Sulle definizioni legislative nel diritto privato*, cit., p. 352 ss.; vd. per una prospettiva gius-economica G. AFFERNI, *L'analisi economica del diritto*, in G. Alpa (a cura di), *La responsabilità del produttore*, Giuffrè, 2019, p. 349-350; Specificamente con riferimento alla conformità dei sistemi di i.a., vd. A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1721; B.A. KOCH, *Product Liability 2.0*, cit., p. 108; A. PROCIDA MIRABELLI DI LAURO, *Le intelligenze artificiali tra responsabilità civile e sicurezza sociale*, cit., p. 313 ss.; implicitamente, con riferimento allo specifico settore dell'automotive, vd. A. ALBANESE, *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, p. 1020 ss. e F. BARTOLINI, *Auto a guida autonoma e problemi di responsabilità civile*, cit., p. 311; A. FUSARO*, Quale modello di responsabilità per la robotica avanzata?,* cit., p. 1348-1349.

[379] U. RUFFOLO, *Intelligenza artificiale e automotive: le responsabilità da veicoli self-driving e driverless*, in Id. (a cura di), Intelligenza artificiale - Il diritto, i diritti, l'etica, cit., p. 164.

produttore dal risarcimento del danno cagionato da un prodotto difettoso.

Nel complesso, la definizione di *standard* tecnici comporta l'adozione di scelte politiche frutto di un bilanciamento tra rischi industriali e benefici delle innovazioni tecnologiche. Niente a che vedere con finalità di razionalizzazione dei giudizi e riduzione della discrezionalità dei giudici nazionali che sono addotte spesso a giustificazione della *preempition doctrine*. La normativa sulla *product liability*, al contrario, ha a che fare con il profilo privatistico della responsabilità civile, nell'ambito della quale non interessa tanto se un'attività poteva o non poteva essere esercitata quanto se si è verificata in concreto una seria minaccia di danno a terzi[380]. Inoltre, a differenza dell'indagine sulla sicurezza, la difettosità di un prodotto di cui alla Dir. 85/374/CEE è per sua natura da condursi *ex post* in base a valutazioni concernenti non solo l'astratto ma anche le modalità di reazione rispetto all'uso concreto da parte del consumatore[381].

Anche il considerando n. 16 della Dir. 2001/95/CE fa comprendere che le discipline europee di *product safety* e *product liability*, per quanto connesse, debbano rimanere tra loro separate, precisando che la direttiva in parola non debba interferire con i diritti previsti dalla Dir. 85/374/CEE in materia di responsabilità per danno da prodotti difettosi[382]. Nel codice del consumo la sicurezza

[380] Così M. COMPORTI, *Esposizione al pericolo e responsabilità civile*, cit., p. 175-176, con riferimento al rapporto tra disciplina autorizzativa allo svolgimento di attività pericolose e applicazione degli artt. 2049-2054 c.c

[381] A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1722; U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1709.

[382] A. AMIDEI*, Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1722; U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1709. La stessa Risoluzione del Parlamento del 20 ottobre 2020, cit., al considerando 8, sottolinea che «qualsiasi aggiornamento del quadro della responsabilità per danno da prodotti difettosi dovrebbe andare di pari passo con l'aggiornamento della direttiva 2001/95/CE del Parlamento europeo e del Consiglio, del 3 dicembre 2001, relativa alla sicurezza generale dei prodotti, al fine di garantire che i sistemi di

dei prodotti è trattata in maniera separata rispetto alla responsabilità da prodotto difettoso e l'art. 111 codice del consumo stabilisce, in materia di *product safety*, che <<*sono fatte salve le disposizioni…in materia di responsabilità per danno da prodotti difettosi*>>. La disciplina sulla responsabilità per danno da prodotto ha l'obiettivo di porre il costo dei danni derivanti dal prodotto sul produttore, anche qualora questi sia incolpevole. Un'esenzione da responsabilità per il fabbricante di prodotto conforme abbasserebbe l'asticella della tutela del consumatore in maniera non indifferente[383].

Se il mancato rispetto della legislazione verticale o orizzontale sulla *product safety* determina pacificamente la difettosità del prodotto, non è vera l'affermazione contraria in forza di cui un prodotto conforme potrà essere considerato difettoso nel caso in cui presenti un'anomala condizione di pericolosità. Superare tale orientamento in relazione ai sistemi di IA non sembra un'evenienza auspicabile, sia perché sono assenti ragioni specifiche per discostarsi dal regime ordinario sia perché non pare opportuno che, proprio in relazione a prodotti per cui il rischio di eventi dannosi non è del tutto preventivabile, sia sufficiente il rispetto di requisiti minimi di commercializzazione per beneficiare di un'immunità.

Il nesso di causalità ed il suo accertamento.

Non è semplice attribuire al produttore per i danni derivanti dai sistemi di IA soprattutto con riferimento all'accertamento del nesso di causalità, sia sul piano sostanziale, sia sul piano strettamente processuale.

L'art. 1 dir. 85/374/CEE e l'art. 114 codice del consumo stabiliscono che <<il produttore è responsabile del danno causato da un difetto del suo prodotto>>. L'onere della

IA integrino la sicurezza e la protezione fin dalla progettazione», considerando, dunque, implicitamente, ancora valida la distinzione tra le due discipline.

[383] U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, cit., p. 1709.

prova del nesso è posto a carico del danneggiato. Anche ai fini della responsabilità aquiliana, per come è noto, il nesso di causalità è regolato dagli artt. 40 e 41 c.p. e questa dimensione logico-sostanziale della causalità è identica sia nel diritto civile sia nel diritto penale[384]. Diversi, invece, restano gli *standard* probatori richiesti per l'accertamento: nel diritto civile vige la regola del "più probabile che non" ispirato a logiche di normalità causale della preponderanza dell'evidenza; nel penale, invece, la regola è dell'"oltre ogni ragionevole dubbio"[385].

5.7.1. Brevi cenni sui profili sostanziali del nesso di causalità.

Bisogna chiedersi se, innanzitutto, il carattere adattivo ed imprevedibile dei sistemi di IA possa essere considerato idoneo a determinare una interruzione del nesso di causalità tra danno e difetto, ai sensi dell'art. 41, comma 2, c.p.[386]. Ci si chiede, cioè, se tali caratteri siano considerabili come *novi acti intervenientes*, per usare termini latini usati dalla cultura giuridica britannica[387].

---

[384] Cass. civ., sez. un., 11 gennaio 2008, n. 576; Cass. civ., sez. III, 11 maggio 2009, n. 10741; Cass. civ., sez. III, 8 luglio 2010, n. 16123.

[385] Il principio è stato affermato dalle note dieci sentenze "gemelle" emanate dalle Sezioni unite civili l'11 gennaio 2008 (numeri dal 576 al 585), in tema di responsabilità aquiliana del Ministero della salute per danni da trasfusione di sangue infetto; vd. la nota di R. BLAIOTTA, *Casualità e colpa: diritto civile e diritto penale si confrontano*, in Cass. pen., 2009, fasc. 1, p. 78 ss.; più di recente, il principio è stato ribadito in giurisprudenza da Cass., sez. L, 3 gennaio 2017, n. 47; Cass., sez. I, 30 giugno 2021, n. 18584; sull'impatto che la differenza tra i due standard probatori ha sugli esiti dei processi vd. F. STELLA, *Giustizia e modernità*, cit., p. 97-98.

[386] In argomento, nella letteratura straniera, vd. M. BASHAYREH – F.N. SIBAI – A. TABBARA, *Artificial Intelligence and legal liability: towards an international approach of proportional liability based on risk sharing*, in *Information and Communications Technology Law*, 2020, 30 (2), p. 173–174; Z.Y. LEE, *Deep learning artificial intelligence and the law of causation: application, challenges and solutions, in Information and Communications Technology Law*, 2021, vol. 30, iss. 3, p. 255 ss.; in materia penale vd. S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame*?, cit., p. 431-432; C. PIERGALLINI, *Intelligenza artificiale,* cit., p. 1760 ss.; M.B. MAGRO*, Robot, cyborg e intelligenze artificiali*, cit., p. 1211; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati,* cit., p. 716 ss.

[387] A. DU BOIS-PEDAIN, *Novus actus and beyond: attributing causal responsibility in the criminal courts*, in *The Cambridge Law Journal*, 2021, Vol. 80, suppl. S1, p. S61 ss.; S. CHOPRA – L.F. WHITE, *A Legal Theory for Autonomous Artificial Agents,* cit., p. 122.

Trattasi di questione di fondamentale rilevanza sia se si aderisce all'impostazione scientifico-naturalistica basata sulla teoria della *condicio sine qua non*, sia se si ritengono preferibili le teorie della causalità adeguata e della causalità umana. Nel primo caso, saremmo di fronte ad una c.d. causalità sorpassante, in cui il fattore sopravvenuto – la decisione autonoma adottata dal sistema - <<*soppianta, azzerandolo, il rischio iniziale ascrivibile al programmatore*>>[388], introducendone uno nuovo, dipendente direttamente dalla decisione algoritmica intelligente.

Anche basandosi su impostazioni "normative", il comportamento emergente del sistema di IA che ha causato un danno potrebbe essere visto come fattore eccezionale, riferendosi alla causalità umana o non ordinario riferendosi alla causalità adeguata e, quindi, interruttivo del nesso causale tra condotta del produttore e danno.

A ben vedere, queste ricostruzioni non possono risultare completamente soddisfacenti[389]. Infatti, in sede di programmazione i più moderni e sofisticati sistemi di IA sono appositamente dotati di capacità di apprendimento. L'abilità di assumere decisioni autonome è, per così dire, una imprevedibilità prevedibile poiché estrinsecazione dei features attribuiti nella fase di *training* con il risultato che l'evento lesivo derivante dal *decision making* artificiale non potrà essere considerato né frutto di una serie causale autonoma, né conseguenza di fattori anormali o eccezionali. È difficile considerare il comportamento emergente del sistema di IA come <<*un rischio nuovo ed*

[388] C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1761.

[389] C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1761-1762; S. CHOPRA – L.F. WHITE, *A Legal Theory for Autonomous Artificial Agents,* cit. p. 137; C. MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale,* in Dir. pen. cont. – Riv. trim., 2022, n. 2, p. 56; S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame*?, cit., p. 432-433; contra S. PREZIOSI*, La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 716 ss., in cui l'autore implicitamente pare sostenere che "l'interposizione" algoritmica interrompa il nesso di causalità

incommensurabile, del tutto incongruo rispetto al rischio originario attivato dalla prima condotta[390]>>, soprattutto visto il tradizionale rifiuto della giurisprudenza a riconoscere cause sopravvenute interruttive[391].

Messa da parte l'opzione ermeneutica basata sull'art. 41, comma 2, c.p., bisogna capire su quali basi si possa accertare il nesso di causalità tra danno ed effetto, dal momento che, ad oggi, il settore dell'IA non è dotato di un corredo nomologico di leggi scientifiche abbastanza consolidato che siano in grado di collegare l'immissione di un prodotto sul mercato con l'evento infausto.

Primariamente, occorre sottolineare che la dimostrazione del rapporto causale nell'ambito della responsabilità per danno da prodotto è da sempre stata considerata uno degli elementi più problematici della disciplina[392]. Lo stesso art. 1 della dir. 85/374/CEE non precisa se sia sufficiente la dimostrazione che il difetto sia stato condicio sine qua non dell'evento lesivo o se, al contrario, sia necessario che l'evento corrisponda ad una conseguenza normale, prevedibile o non eccezionale del fatto. La giurisprudenza dei diversi Stati ha mostrato un'ampia eterogeneità nell'interpretazione degli elementi necessari a soddisfare l'onere della prova relativa al nesso causale tra difetto del prodotto ed evento dannoso[393].

In linea generale, nel diritto civile, si è assistito al passaggio dalla teoria della condicio sine qua non al

---

[390] Cass. pen., 3 marzo 2022, n. 19882; Cass. pen., sez. IV, 13 maggio 2019, n. 20270; Cass. pen., sez. IV, 3 maggio 2016, n. 25689.

[391] R. BLAIOTTA, *Causalità giuridica*, Giappichelli, 2010, p. 209 ss., che ripercorre le (pochissime) sentenze che hanno riconosciuto un'interruzione del nesso causale; in argomento vd. anche l'approfondita indagine di A. VALLINI, *"Cause sopravvenute da sole sufficienti" e nessi tra condotte. Per una collocazione dell'art.41, comma 2, c.p. nel quadro teorico della causalità "scientifica",* in Studi in onore di Franco Coppi, vol. I, Torino, Giappichelli, 2011, p. 341 ss.

[392] Vd., nella dottrina statunitense, D.G. GIFFORD, *The Challenge to the Individual Causation Requirement in Mass Products Torts*, in Wash. & Lee L. Rev. 873, 900–15 (2005); J. STAPLETON, *Product Liability, Cambridge Univerity Press*, 1994, p. 280 ss.

[393] S. WHITTAKER, *Liability for Products. English Law, French Law, and European Harmonisation*, cit., p. 512 ss.; D. WUYTS*, The product liability directive – More than two decades of defective products in Europe*, cit., p. 25 ss.

criterio molto più generico dell'aumento del rischio[394]. Da questo si è arrivati al criterio del più probabile che non, in cui l'attore deve dimostrare che la condotta del convenuto, con probabilità superiore al 50%, abbia causato il danno. Ora si è giunti ad un criterio ancora più ineffabile e diluito: oggi è sufficiente che la condotta del convenuto sia stata idonea ad aumentare il rischio del danno. Ciò escludendo le conseguenze illecite dannose che si sarebbero verificate con eguale probabilità[395]. Pure in questo settore, il diritto civile è riuscito ad interiorizzare quel modello di rischio tipico della società post-moderna[396], che invece scatena tante preoccupazioni nell'ordinamento penale, orientato naturalmente ai principi di colpevolezza, di legalità, di offensività, nonché alla regola processuale dell'oltre ogni ragionevole dubbio.

Proprio nell'ambito degli studi sulla responsabilità per danno da prodotto, si è assistito all'affermarsi del concetto di *black box*[397]. Concetto che descrive quei sistemi tecnologici dei quali si conoscono *input* ed *output*, senza che sia possibile comprendere scientificamente i processi interni che trasformano i primi nei secondi: non è verificabile con certezza quale *input* abbia provocato quale *output* ed in che modalità.

All'opacità dei prodotti moderni, spesso di per sé dotati di complessità tecnica, l'IA aggiunge la capacità di elaborare statisticamente quantità immense di dati. Quando si è di fronte a sistemi di IA dotati di reti neurali, essi hanno anche una tecnica di computazione e decisione non sempre comprensibile secondo i tradizionali schemi di ragionamento umani. Un esempio emblematico è il caso Toyota. In questo caso, i ricorsi presentati dai danneggiati

[394] Per una concezione probabilistica della causalità vd., nella dottrina statunitense, D. ROSENBERG*, The causal connection in mass exposure cases: a «public law» vision of the tort system*, in *Harvard Law Review*, 1984, p. 851; G.O. ROBINSON, *Probabilistic causation and Compensation for tortious risk*, in *Journal of Legal Studies*, 1985, p. 779; E.D. ELLIOTT, *Why Courts? Comment on Robinson*, in *Journal of Legal Studies*, 1985, p. 799.
[395] G. PONZANELLI, *La responsabilità civile: profili di diritto comparato*, cit., p. 90.
[396] C.E. PALIERO, *L'autunno del patriarca,* cit., p. 1240.
[397] C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., passim.

allegavano che gli incidenti si erano verificati per causa di una accelerazione improvvisa dei veicoli, alla quale i conducenti non potevano rimediare neanche azionando il freno. Nonostante l'intensa attività investigativa condotta dagli organi inquirenti della *National Highway Traffic Safety Administration* (NHTSA), le indagini non poterono individuare la causa precisa del malfunzionamento[398].

Problemi di accertamento maggiori potrebbero porsi con riferimento agli incidenti derivanti da macchine a cui sono incorporati sistemi di IA. Qui, oltre che con la complessità tecnologica degli artefatti, occorrerà fare i conti con la capacità di apprendimento e di *decision-making* autonomo degli algoritmi, nonché con la loro permeabilità ai dati provenienti dall'esterno.

Per alcuni, si potrebbe installare degli *event data recorders* (EDR), funzionali a ricostruire *ex post* la dinamica del fatto lesivo. Il Regolamento UE 2019/2144[399] ha disposto che, dal luglio 2022, tutte le automobili di nuova immatricolazione debbano essere dotate di EDR. Una misura siffatta potrebbe benissimo essere estesa alla disciplina di altri settori: l'art. 12 dell'AI Act prevede, ad esempio, l'installazione obbligatorio sui sistemi ad alto rischio di strumenti di registrazione automatica degli

[398] Va comunque rilevato che, nonostante l'impossibilità di accertare con precisione la causa degli incidenti, la Corte distrettuale dell'Oklahoma, nel 2013, ha condannato Toyota al risarcimento del danno, nonché al pagamento di *punitive damages*, ritenendo che l'incidente fosse più probabilmente causato da un difetto del veicolo che da una condotta del conducente – di fatto, applicando dunque la teoria del *res ipsa loquitur* (United States District Court, Oklahoma County, Jean Bookout and Charles Schwarz v. Toyota Motor Corporation, Case no. CJ-2008-7969, 24 ottobre 2013), a tal proposito vd. M. Cronin Fisk, Toyota Settles Oklahoma Acceleration Case After Jury Verdict, in Bloomberg, 25 ottobre 2013, disponibile a questo link: https://www.bloomberg.com/news/articles/2013-10-25/toyota-settles-oklahoma-acceleration-case-after-jury-verdict. Successivamente a tale condanna, Toyota ha proceduto a chiuso con accordi transattivi (c.d. settlement) tutti i procedimenti che erano aperti a suo carico per gli incidenti derivanti dalle suddette accelerazioni improvvise. Sulla complessa vicenda giudiziaria che ha coinvolto la casa automobilistica vd. T. WINKLE, *Development and Approval of Automated Vehicles: Considerations of Technical*, *Legal, and Economic Risks*, in M. Maurer e al. (eds.), Autonomous Driving, Springer, 2016, p. 598; R.I. STEINZOR, (Still) *"Unsafe at Any Speed": Why Not Jail for Auto Executives?,* in *Harvard Law & Policy Review*, 2015, p. 443 ss.

[399] Regolamento (UE) 2019/2144 del Parlamento europeo e del Consiglio del 27 novembre 2019 relativo ai requisiti di omologazione dei veicoli a motore e dei loro rimorchi.

eventi[400]. Questa disposizione si deve leggere in collegamento con l'art. 8 della proposta di revisione della dir. 85/374/CEE, che prevede che il giudice possa ordinare al convenuto la divulgazione degli elementi di prova a sua disposizione, qualora la domanda risarcitoria appaia fondata: tra gli elementi di prova si ritiene che possano rientrare anche i dati raccolti dai *logs* durante l'incidente e nelle fasi immediatamente precedenti e successive.

L'installazione di questa scatola nera è utile per escludere una responsabilità dell'utilizzatore del sistema di IA. Ma, comunque, potrebbe restare complesso in alcune circostanze risalire con certezza alla causa del danno. La causa potrebbe essere un difetto di fabbricazione o di costruzione. Così come potrebbe essere dipendente da fattori esterni ed ambientali che potrebbero perfino saldarsi tra di loro in una "*web of causation*" destinata a rilasciare <<*effetti di tipo interattivo dovuto alla sommatoria di alcune cause o sinergico capaci di esitare in un fenomeno qualitativamente diverso rispetto alla mera sommatoria*>>[401].

Le semplificazioni probatorie fondate sul *re ipsa loquitur* potrebbero rendere più probabile l'accoglimento del ricorso, per il danneggiato. Già, ad oggi, vengono usate dal giudice. Se la proposta di revisione della direttiva sul danno dovesse essere approvata, esse saranno in grado di alleggerire l'onere probatorio per il ricorrente.

5.7.2. Aspetti processuali del nesso di causalità.

L'accertamento del nesso di causalità dell'evento lesivo algoritmico incontra problemi già visti nell'ambito della responsabilità per danno da prodotto. L'art. 120 codice del consumo pone a carico del danneggiato l'onere di fornire

[400] Tale misura era già stata proposta dall' EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, Liability for Artificial Intelligence and other emerging digital technologies, cit., p. 47;

[401] C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1762; vd. anche A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 9;

la prova sia del danno subito sia dell'esistenza del difetto del prodotto sia del nesso di causalità tra difetto e danno. La dottrina ha sempre segnalato la difficoltà per il ricorrente di provare tutto ciò: è una *probatio* diabolica[402]. Il contesto della produzione industriale comporta una netta simmetria informativa tra produttore e consumatore, in relazione alle specifiche caratteristiche del prodotto. Questa carenza di informazioni tecniche si accentua con riferimento al funzionamento dei sistemi di IA, posto che la capacità di auto-apprendimento ed interazione aggiunge un livello di complessità all'individuazione del fatto causativo del danno[403]. Come detto *supra*, perfino il produttore avrebbe difficoltà ad individuarlo. Nonostante sia vigente l'anzidetto riparto dell'onere della prova, la CGUE ha ritenuto compatibili con la Direttiva le regole nazionali che consentono al giudice di ritenere provato il nesso di causalità in presenza di indizi gravi, precisi e concordanti della connessione eziologica, anche in assenza di dati scientifici univoci al riguardo. I giudici della CGUE riconoscono che, specie se oggetto del giudizio sono prodotti ad alta complessità tecnica o scientifica, precludere questa possibilità vorrebbe dire imporre uno standard probatorio talmente alto da escludere ogni modalità di prova diversa dalla prova certa a livello scientifico, rendendo <<*eccessivamente difficile o impossibile l'affermazione della responsabilità del produttore, in tal modo compromettendo l'effetto utile della dir. 85/374*>>[404].

[402] Vd., quando ancora non esisteva una disciplina speciale in materia di responsabilità per danno da prodotto difettoso, le riflessioni di U. CARNEVALI, La responsabilità del produttore. Problemi generali, cit., p. 154-157 e G. ALPA, Responsabilità dell'impresa e tutela del consumatore, Giuffrè, 1975, p. 430 ss.

[403] P. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 175.; S. CHOPRA – L.F. WHITE, *A Legal Theory for Autonomous Artificial Agents*, cit., p. 144; S. WHITTAM, *Mind the compensation gap*, cit., p. 3 e p. 7; A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1724.

[404] Corte di Giustizia UE, 21 giugno 2017, N. W, L. W, C. W/Sanofi Pasteur MSD SNC, *Caisse primaire d'assurance maladie des Hauts-de-Seine, Carpimko*, causa C-621/15, citata da A. AMIDEI, *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, cit., p. 1724. In quel caso, la Corte di Giustizia aveva ritenuto difettoso un vaccino, ritenendo «pacifico che la ricerca medica non ha permesso né di stabilire né di escludere l'esistenza di un nesso di causalità» tra il danno subito e l'iniezione del vaccino.

Il carattere troppo gravoso dell'onere della prova a carico del danneggiato è stato anche sottolineato dalle istituzioni europee che nella quarta relazione sull'applicazione della direttiva hanno messo in campo l'ipotesi di introdurre un rovesciamento dell'onere della prova a favore del danneggiato. Inoltre, si è prospettato di prevedere meccanismi di *disclosure* obbligatoria in capo al produttore. Tali posizioni sono poi confluite nella proposta di revisione della direttiva sulla *product liability*. L'art. 9 della proposta di revisione della Direttiva sul danno da prodotto prova a rispondere alle problematiche emerse nell'applicazione concreta della disciplina, prevedendo una presunzione di difettosità del prodotto ed una presunzione di causalità tra difetto e danno, entrambe *iuris tantum*.

Con riferimento al rapporto tra danno e difetto, essa prevede che <<*si presume l'esistenza del nesso di causalità tra il carattere difettoso ed il danno nel caso in cui sia stato provato che il prodotto è difettoso e che la natura del danno cagionato è generalmente coerente con il difetto in questione*>>. Anche al di fuori di detta ipotesi, il paragrafo 4 dell'art. 9 prevede che, se la complessità tecnologica del prodotto rende troppo difficile, per il ricorrente, la prova del nesso di causalità, così come la prova del difetto, il giudice può presumere la loro sussistenza nel caso in cui: <<*a) il prodotto ha contribuito a cagionare il danno; e (b) è probabile che il prodotto fosse difettoso, oppure che il carattere difettoso dello stesso è una causa probabile del danno, o entrambi tali elementi*>>. Il riferimento alla probabilità di causazione del danno è talmente vago da apparire privo di portata precettiva, dal momento che l'indagine sul nesso di causalità è necessariamente basata su stime probabilistiche. Per porre rimedio all'asimmetria informativa tra consumatore e produttore, l'art, 8 della proposta prevede che l'autorità giudiziaria, se la domanda di risarcimento pare plausibile, possa ordinare la divulgazione di <<*pertinenti elementi di prova a sua disposizione*>>, nei limiti di quanto sia <<*necessario e*

proporzionato per sostenere la domanda>> e tenendo conto <<*dei legittimi interessi di tutte le parti, compresi i terzi interessati, specialmente per quanto riguarda la protezione delle informazioni riservate e dei segreti commerciali*>>. Qualora al convenuto venga ordinato di divulgare informazioni che costituiscano un segreto commerciale, gli organi giurisdizionali devono adottare <<*le misure specifiche necessarie per tutelarne la riservatezza se tali informazioni vengono utilizzate o se ad esse è fatto riferimento nel corso di un procedimento giudiziario*>>.

Tutte queste disposizioni devono per forza essere messe in correlazione con l'art. 9, paragrafo 2, della proposta, in forza del quale sussiste una presunzione di difettosità del prodotto nel caso in cui il convenuto non abbia rispettato l'obbligo di divulgare i pertinenti elementi di prova a sua disposizione a norma dell'articolo 8, paragrafo 1.

6. Responsabilità per esercizio di attività pericolose ed i sistemi di IA.

Parte della dottrina ritiene che, in capo al produttore di sistemi di IA, il regime di responsabilità oggettiva previsto dall'art. 2050 c.c. per l'esercizio di attività pericolose possa configurarsi[405]. Le ipotesi di responsabilità oggettiva viste *supra* si attanagliano bene alla posizione dell'utilizzatore o fruitore del sistema di IA. Mentre, l'art. 2050 c.c. delinea un'ipotesi di responsabilità oggettiva che ben potrebbe applicarsi nei confronti del produttore, in quanto esercente di un'attività pericolosa[406]. In particolare, stando a questa prospettiva, il ricorso all'IA può rendere pericolose attività che la giurisprudenza normalmente non ritiene tali: ad esempio i sistemi medici o i veicoli a guida autonoma che, nonostante promettano benefici in termini di riduzione

[405] In questo senso, U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 108 ss.; ID., *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 248 ss.; U. RUFFOLO – E. AL MUREDEN, *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, in Giur. it., luglio 2019, p. 1712.
[406] Sulla cumulabilità tra la tutela di cui alla dir. 85/374/CEE e quella prevista dall'art. 2050 c.c. vd. diffusamente U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 251 ss.

degli incidenti, possono essere fonte di nuovi tipi di rischi[407], generando pericolosità settoriali anomale[408]. Il ricorso alla disciplina oggettiva di cui all'art. 2050 c.c. potrebbe rivelarsi particolarmente utile, per il danneggiato, nel caso in cui emerga un rischio da sviluppo, che, per come abbiamo già evidenziato, è suscettibile di escludere la responsabilità da prodotto difettoso.

Visto che, ad oggi, non sembrano ancora delineati chiaramente i rapporti tra l'appena citata esimente e la capacità di autoapprendimento dei sistemi di IA, la disciplina di cui all'art. 2050 c.c. potrebbe costituire un'adeguata soluzione per ristorare la persona che abbia subito un danno provocato da un prodotto difettoso. Il fatto che vi sia un rischio da sviluppo, al contrario, costituisce indice di specifica pericolosità del prodotto[409]. La responsabilità di cui all'art. 2050 c.c. sussiste quando l'esercizio dell'attività pericolosa non venga accompagnata dall'adozione di misure idonee ad evitare il danno. Leggendo l'art. 2050 in base alla tradizionale interpretazione di analisi economica del diritto, esso impone all'esercente i costi delle misure preventive, addossandogli i costi da risarcimento successivi in caso di eventi infausti[410]. La prevedibilità, quindi, non si pone sul piano del difetto, quanto, piuttosto, in relazione alle misure preventive, la cui idoneità deve essere valutata *ex*

[407] U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 249-250; ID., *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 108-109; contra A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 625 e A. ALBANESE*, La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione,* cit., nota 89, che, con particolare riferimento alle macchine a guida autonoma, escludono l'applicabilità dell'art. 2050 c.c., rilevando che, aderendo a questa lettura, «si finirebbe per qualificare come pericolosa un'attività (la guida automatizzata), la quale, se conforme alle regole della tecnica, è paradossalmente più sicura della normale attività di guida 'umana'».

[408] U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 248.

[409] U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 250-251.

[410] U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 109.

ante, alla luce delle tecnologie e conoscenze esistenti al momento della produzione[411].

7. C'è bisogno di una personalità elettronica?

Una delle proposte più suggestive e radicali in campo di allocazione di costi derivanti dall'uso dei sistemi di IA è quella di introdurre una forma di responsabilità diretta patrimoniale dell'agente algoritmico, tramite la creazione di una personalità elettronica[412]. Così ragionando, al sistema di IA verrebbe attribuito un patrimonio separato finalizzato al risarcimento dei danni, indipendentemente dall'accertamento di una colpa.

Si adduce come principale motivazione alla base di questa idea la presunta inadeguatezza degli attuali regimi di responsabilità, tutti incentrati sulla capacità di governo di una persona fisica su una cosa o su una attività. Detta proposta è condivisa solo da chi ritiene non sia applicabile la disciplina della responsabilità extracontrattuale[413].

La personalità elettronica consentirebbe di contemperare le pretese risarcitorie del danneggiato con le esigenze di innovazione e di prevedibilità delle decisioni giudiziarie che fanno capo alle imprese. Vi sarebbero benefici in ordine probatorio per il ricorrente. Alla creazione di siffatta tipologia di personalità, molto probabilmente corrisponderebbe la creazione di un

---

[411] U. RUFFOLO, *La responsabilità da artificial intelligence, algoritmo e smart product*, cit., p. 110.

[412] Tra i maggiori sostenitori vd. G. TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, cit., passim; S. BECK, *The problem of ascribing legal responsibility in the case of robotics*, cit., p. 479 ss., nonostante quest'ultima riconosca come la personalità elettronica possa costituire soltanto una tra le soluzioni che il nostro ordinamento necessita per adattarsi al nuovo fenomeno dell'i.a.; per un'approfondita panoramica sulle diverse posizioni, da quelle più risalenti a quelle più recenti vd. M. HILDEBRANDT, *Criminal Liability and 'Smart' Environments*, in R.A. Duff (ed.), Philosophical Foundations of Criminal Law, Oxford University Press, 2011, p. 507 ss.; EAD, *Law for Computer Scientists and Other Folk, Oxford University Press*, 2020, p. 237 ss.

[413] Vd. ad es. E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema*?, cit., p. 467, che considera la soluzione "ridondante", poiché ritiene che molti dei problemi che il nuovo status giuridico si propone di affrontare possono essere risolti attraverso i classici schemi della responsabilità per danno da prodotto o attraverso l'introduzione di un obbligo assicurativo.

registro degli agenti artificiali idoneo ad identificare ciascun sistema di IA sul mercato. Ciò per consentire alle persone che si interfacciano con i sistemi di IA di informarsi sul patrimonio, sui proprietari e sugli eventi lesivi già verificatisi per valutarne l'affidabilità[414].

L'introduzione di una responsabilità patrimoniale diretta per i sistemi di IA non deve farci pensare che si voglia riconoscere una sorta di capacità morale in capo ad essi, con conseguente attribuzione di uno *status* legale equiparabile a quello dell'uomo[415]. Infatti, questa proposta è data da esigenze pratiche di allocazione dei costi economici. Non già della presa d'atto di una autocoscienza in capo agli agenti intelligenti, idonea da spingere il legislatore a prendere in considerazione la loro dignità tramite l'ampliamento della nozione di soggettività giuridica umana[416]. La ratio è la medesima di quella che ha spinto il legislatore a riconoscere autonomia patrimoniale alle persone giuridiche.

Dunque, dalla personalità elettronica in questo senso considerata, occorre tenere distinto tutto quel dibattito inerente all'eventuale creazione di uno status legale in capo al sistema di IA, paragonabile a quello degli umani. Suggestioni di questo tipo restano tali e dovranno essere rinviate al momento in cui vi saranno sul mercato sistemi in grado di compiere ogni attività intellettiva e non agenti che si limitino a svolgere specifici compiti.

---

[414] S. BECK, *The problem of ascribing legal responsibility in the case of robotics*, cit., p. 475. La proposta era già stata avanzata da C.E.A. KARNOW, *Liability for Distributed Artificial Intelligence*, cit., p. 147 ss.

[415] Distinguono accuratamente i due piani A. BERTOLINI – F. EPISCOPO, *Robots and AI as Legal Subjects? Disentangling the Ontological and Functional Perspective*, in Front. Robot. AI, 2022, vol. 9, p. 3; G. WAGNER, *Robot*, Inc., cit., p. 608; U. PAGALLO, Vital, Sophia, and Co.—*The Quest for the Legal Personhood of Robots, in Information*, 2018, n. 9, p. 231 ss.; U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 264; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in cerca di un problema*?, cit., p. 445 ss.; S. BECK, The problem of ascribing legal responsibility in the case of robotics, cit., p. 479.

[416] U. PAGALLO, Vital, Sophia, and Co., cit., p. 231 ss.; U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 264.

La possibilità di introdurre una personalità elettronica in capo ai *robot* sembrava avere trovato una sponda istituzionale nella Risoluzione del Parlamento europeo sulla regolazione della robotica, che, al paragrafo 59, lett. F), proponeva la creazione di <<*uno status giuridico specifico per i robot nel lungo termine, di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato, nonché eventualmente il riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi*>>[417]. Da questa breve descrizione, tuttavia, non è agevole stabilire se la Risoluzione prefigurasse il riconoscimento di una piena capacità d'agire in capo ai sistemi di IA o se invece fosse propensa all'introduzione di una responsabilità patrimoniale diretta dell'agente algoritmico[418]. Ma, comunque, quale che fosse la prospettiva, tale soluzione pare essere stata abbandonata, visto che tutte le più recenti iniziative europee non accennano più alla proposta[419]. Ciò anche in forza delle molteplici critiche ricevute[420].

La dottrina maggioritaria ha sottolineato gli aspetti critici della proposta di *electronic personhood*[421]. Molti autori

[417] Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, cit., vd. § 59, lett. f).
[418] Su quest'ambiguità vd. U. PAGALLO, Vital, Sophia, and Co., cit., p. 239.
[419] Vd. Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza artificiale, cit; Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL)); Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, L'intelligenza artificiale per l'Europa, COM/2018/237 final.
[420] Vd. COMITATO ECONOMICO E SOCIALE EUROPEO, *L'intelligenza artificiale – Le ricadute dell'intelligenza artificiale sul mercato unico (digitale), sulla produzione, sul consumo, sull'occupazione e sulla società*, INT/806, 2017, § 1.12 e 3.33; vd. anche la lettera firmata da numerosi esperti nel settore dell'intelligenza artificiale, Open Letter to the European Commission Artificial Intelligence And Robotics, 2018, disponibile a questo link: www.robotics-openletter.eu.
[421] Ex multis U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione*?, vol. 2, il Mulino, 2022, p. 234 ss.; G. COMANDÉ, *Intelligenza artificiale e responsabilità tra liability e accountability*, cit., p. 180; E. PALMERINI, *Soggettività e agenti artificiali: una soluzione in*

hanno notato come l'introduzione di una personalità elettronica potrebbe aggravare quella tendenza alla antropomorfizzazione dei sistemi di IA ed alla creazione di aspettative sociali nei confronti del loro funzionamento, nonché all'estensione di diritti e garanzie un tempo riconosciuti solo agli uomini[422]. La responsabilità diretta del sistema di IA potrebbe anche limitare in modo del tutto irragionevole la tutela risarcitoria del danneggiato, posto che circoscriverebbe la responsabilità di produttori e/o di utenti all'ammontare del patrimonio dell'ente robotico, così assurto a limitato capitale di rischio[423]. Così come per gli enti collettivi, la limitata esposizione economica dei finanziatori-azionisti della persona elettronica sarebbe un incentivo ad assumere più rischi a costi inferiori, con aumento prevedibile degli eventi dannosi[424].

8. Serve un'assicurazione obbligatoria?

Poste le difficoltà probatorie che il danneggiato potrebbe incontrare nell'ottenere un risarcimento del danno da IA ai sensi della disciplina del danno da prodotto, alcuni autori hanno proposto di introdurre un fondo indennitario finanziato dalle società produttrici a cui il ricorrente può accedere se dimostra di avere subito un danno[425]. Soprattutto nel settore delle automobili a

---

cerca di un problema?, cit., p. 467; A. BERTOLINI – F. EPISCOPO, *Robots and AI as Legal Subjects?,* cit., passim.

[422] R. ABBOTT – A. SARCH, *Punishing Artificial Intelligence*, cit., p. 377. Sul possibile riconoscimento di diritti in capo ai sistemi di i.a. vd., per tutti, D.J. GUNKEL, Robot Rights, MIT Press, 2018; J.C. GELLERS, *Rights for Robots. Artificial Intelligence, Animal and Environmental Law*, Routledge, 2021.

[423] U. RUFFOLO, *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, cit., p. 263-264; B.A. KOCH, Product Liability 2.0, cit., p. 115; vd. anche EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *Liability for Artificial Intelligence and other emerging digital technologies*, cit., p. 38; G. WAGNER, Robot, Inc., cit., p. 611-2.

[424] G. WAGNER, Robot, Inc., cit., p. 612.

[425] Nella dottrina statunitense, la proposta è stata avanzata, con accenti diversi, da D.C. VLADECK, *Machines without principals: liability rules and artificial intelligence*, cit., p. 146 ss.; K.S. ABRAHAM – R.L. RABIN, *Automated vehicles and manufacturer responsibility for accidents,* cit., p. 105. In Italia, sono favorevoli A. DAVOLA – R. PARDOLESI, *In viaggio col robot,* cit., p. 628; U. SALANITRO*, Intelligenza artificiale e responsabilità*, cit., p. 1274; A. DAVOLA, *A Model for Tort Liability in a World of Driverless Cars: Establishing a Framework for the Upcoming Technology*, in Idaho L. Rev., 2018, p. 591 ss..; G. FINOCCHIARO, *Intelligenza*

guida autonoma con l'appellativo di *Manifacturer Enterprise Responsibility* (MER), ma non c'è motivo di escludere che tale soluzione possa trovare applicazione anche con riferimento ad altri settori.

Sarebbe un vero e proprio regime di responsabilità oggettiva pura posto che il danneggiato verrebbe esonerato dalla necessità di provare il difetto del prodotto. Prova che, per come abbiamo visto infra, sembra essere uno degli aspetti più controversi della disciplina della dir. n. 85/374/CEE. L'istituzione di un fondo indennitario avrebbe anche la funzione di coprire i danni anonimi, tramite una socializzazione dei rischi, ed avrebbe anche il vantaggio di consentire un risparmio delle spese legali per ambedue le parti[426]. Ne risulterebbe anche un incentivo per i consumatori a sostituire i prodotti tradizionali con i sistemi di IA.

Secondo alcuni autori, il fondo dovrebbe essere alimentato pro quota da tutti gli operatori che progettano, programmano ed impiegano sistemi di IA sulla base delle quote di mercato occupate. Così che il contributo di ogni produttore al fondo sarebbe proporzionato all'entità del rischio introdotto nella società, quantificato anche in base al numero di dispositivi intelligenti immessi sul mercato[427].

Altri, invece, ritengono che una simile socializzazione dei costi incentiverebbe il fenomeno dei c.d. *free rider*[428],

---

artificiale e responsabilità, cit.; sui vantaggi e gli svantaggi dell'introduzione di *compensation funds* vd. diffusamente G. BORGES, *New Liability Concepts: the Potential of Insurance and Compensation Funds*, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, cit., p. 153 ss.

[426] U. SALANITRO, Intelligenza artificiale e responsabilità, cit., p. 12

[427] D.C. VLADECK, *Machines without principals: liability rules and artificial intelligence*, cit., p. 146 ss.

[428] Parla di perdita del "steering effect" tipico del sistema di responsabilità (per colpa o limitatamente oggettivo) G. BORGES, New Liability Concepts: the Potential of Insurance and Compensation Funds, cit., p. 162. Simili problemi di free riding erano già emersi in relazione alla Market Share Liability, ossia a quell'orientamento giurisprudenziale che, negli Stati Uniti, ha consentito ai ricorrenti di ottenere il ristoro per i danni subìti, nonostante questi non fossero in grado di identificare la società produttrice responsabile della componente del prodotto che aveva causato il danno, vd. A.M. MARINO, *Market Share Liability and Economic*

posto che il fatto che il costo dei danni venga ripartito tra tutte le aziende disincentiva l'adozione di misure di sicurezza più performanti, e propongono di usare il criterio della frequenza e della gravità degli incidenti in cui il prodotto è coinvolto come parametro per la suddivisione dei costi[429].

Bisogna comunque rilevare che c'è una parte della dottrina che si oppone al modello puro di responsabilità oggettiva delineato dalla MER, asserendo che tale soluzione potrebbe ritardare l'ingresso sul mercato di sistemi di IA già sicuri poiché i produttori potrebbero preferire attendere fin a quando non siano in grado di poter calcolare con precisione i danni derivanti da essi, così da internalizzare il costo degli incidenti, trasferendoli sui consumatori[430]. L'attuale modello di responsabilità oggettiva fondato sulla prova di una difettosità del prodotto, potrebbe essere accompagnato dalla costituzione di un fondo pubblico di garanzia. Tale fondo sarebbe destinato a risarcire il danneggiato solo in via residuale nel caso in cui il prodotto sia privo di difetti e, quindi, il patrimonio del produttore non sia aggredibile: così si raggiungerà l'obiettivo di evitare che il costo del danno resti in capo alla vittima[431].

In alternativa, il Parlamento europeo nella Risoluzione sul diritto civile e la robotica del 2017 ha proposto un regime di assicurazione obbligatorio per tutti i dispositivi intelligenti, prendendo spunto da quanto avviene per i

---

Efficiency, in *Southern Economic Journal*, 1991, p. 667 ss. In generale, sulla Market Share Liability vd. Sindell v. Abbott Labs., 607 P.2d 924 (Cal. 1980), il primo caso in cui una corte ha adottato tale orientamento giurisprudenziale; per una ricostruzione delle criticità di tale orientamento – che, di fatto, costituisce una deroga alla necessità di provare il nesso di causalità – vd. M.A. GEISTFELD, *The Doctrinal Unity of Alternative Liability and Market-Share Liability, in 155 University of Pennsylvania Law Review*, 2006, p. 447 ss.

[429] K.S. ABRAHAM – R.L. RABIN, *Automated vehicles and manufacturer responsibility for accidents*, cit., p. 169; U. SALANITRO*, Intelligenza artificiale e responsabilità*, cit., p. 1274.

[430] A. DAVOLA – R. PARDOLESI*, In viaggio col robot*, cit., p. 628; A. DAVOLA, *A Model for Tort Liability in a World of Driverless Cars*, cit., p. 606-607; una soluzione analoga è proposta da C. SCHROLL, *Splitting the bill: creating a national car insurance fund to pay for accidents in autonomous vehicles*, in *Northwestern University L. Rev.*, 2015, p. 803 ss.

[431] A. DAVOLA – R. PARDOLESI, *In viaggio col robot*, cit., p. 628-629.

veicoli a motore. A ciò si integrerebbe eventualmente un fondo pubblico finalizzato a garantire la possibilità di ristoro nel caso in cui i sistemi di IA siano privi di copertura assicurativa[432].

9. Responsabilità civile ed IA: spunti di riflessione.

Nello studio fino a qui effettuato, si sono evidenziate alcune analogie tra il contesto industriale in cui si sono sviluppate le prime applicazioni dell'imputazione oggettiva del rischio di impresa ed il moderno ed attuale dibattito sulla responsabilità per danni da IA. Così come al tempo, anche oggi emerge la necessità di bilanciare l'innovazione con la sicurezza dei consumatori. Altresì, il rischio da scongiurare è che si creino delle sacche di immunità provocate dall'obiettivo di incentivare il progresso tecnologico e di evitare *chilling effects*.

Quando si è compreso che le responsabilità civile e penale avrebbero dovuto assolvere a scopi e funzioni differenti, si è iniziato a ragionare di imputazione oggettiva. Compito della responsabilità civile non è la ricerca del vero autore del fatto dannoso meritevole di biasimo da parte dell'ordinamento quanto piuttosto l'individuazione di criteri giuridici per l'attribuzione del danno idonei a determinare artificialmente le condizioni in presenza delle quali il costo di un danno subito da un soggetto debba essere trasferito su un altro.

Anche se si sta assistendo ad un ritorno a forme sempre più miste di responsabilità, il criterio di allocazione del danno nel diritto civile resta slegato dalla ricerca del soggetto personalmente rimproverabile per aver provocato il danno. Infatti, accanto alla clausola generale dell'art. 2043 c.c., l'ordinamento civile prevede regimi in grado di imputare il danno da IA all'utilizzatore ed al

[432] Risoluzione del Parlamento europeo del 16 febbraio 2017, cit., § 58 e 59, lett. b); in argomento vd. G. BORGES, *New Liability Concepts: the Potential of Insurance and Compensation Funds*, cit., p. 153 ss.

produttore di un sistema senza che si debba necessariamente provare la colpa. Il vantaggio derivante dall'artificialità del diritto civile moderno consente di poterne plasmare le categorie rispetto alle scelte di gestione del rischio assunte dal decisore pubblico. Dunque, il legislatore potrebbe intervenire per chiarire quale sia la allocazione dei costi che ritiene ottimale a seconda degli scopi che si prefigge.

Noi ci siamo concentrati sull'applicabilità dei regimi di responsabilità oggettivi codicistici per quanto riguarda la posizione dell'utilizzatore e del regime contenuto nella direttiva sul danno da prodotto per la posizione del produttore, tralasciando la disciplina generale della responsabilità per colpa di cui all'art. 2043 c.c.: questi sono i regimi ad essere destinati all'applicazione quasi totalitaria nel caso di danni da IA.

Con riferimento specifico all'utilizzatore, molti autori si sono espressi sulla possibilità di applicare i regimi speciali previsti dagli artt. 2049-2054 c.c., con l'esclusione dell'art. 2050 c.c. applicabile meglio alla posizione del produttore. Dette disposizioni si riferiscono a situazioni in cui vi è una persona che ha creato un rischio di evento dannoso oppure che beneficia della creazione di un rischio o che si trova nella posizione migliore per controllare un rischio. Nonostante l'individuazione di un criterio unificatore sia poco agevole, pare potersi asserire che la ratio generale delle disposizioni in parola sia riassumibile nel principio del *cuius commoda, eius incommoda esse debent*. È, dunque, in tale ottica ragionevole ritenere che i soggetti che beneficiano dei più sofisticati algoritmi di IA debbano anche rispondere civilmente degli eventuali effetti collaterali da essi provocati.

Il riferimento ad una fattispecie in luogo dell'altra dipende dalla concezione che si ha dei sistemi di IA, nonché dalla ratio delle norme codicistiche. Così se si considera il sistema di IA come una cosa inanimata, si

potrà ipotizzare di fare ricorso all'art. 2051 c.c. Se si ritiene che il sistema di IA possa essere considerata una forma di minorata intelligenza, si potrà fare riferimento all'art. 2502 c.c.. Se, infine, si considera l'agente algoritmico come una forma di intelligenza paragonabile a quella dell'essere umano la disciplina di riferimento sarà l'art.2049 c.c., disciplinante la responsabilità vicaria.

Se la *ratio* di ritenere responsabile oggettivamente l'operatore del sistema di IA parrebbe in linea con le disposizioni codicistiche in materia di imputazione oggettiva, in assenza di un intervento legislativo che si preoccupi di definire il bilanciamento politico tra istanze di tutela del danneggiato ed esigenze di produzione si sistemi evoluti, l'applicazione estensiva delle norme esistenti nel codice dovrebbe ispirarsi alla prudenza.

È prevedibile che la progressiva attribuzione di autonomia ai sistemi di IA e la conseguente perdita di controllo da parte dell'essere umano aprirà la strada ad una maggior responsabilizzazione delle società produttrici.

Per questi, si è presa in considerazione la disciplina della responsabilità per danno da prodotto difettoso, dando atto della proposta di revisione della direttiva europea su tale tema. Per come si è detto, la Commissione estende la nozione di prodotto agli algoritmi che possono essere venduti anche separatamente rispetto all'*hardware* e la nozione di produttore ai programmatori e sviluppatori di *software* e codici algoritmici.

Si è constatata, inoltre, la inadeguatezza della nozione di difettosità rispetto alla capacità dei sistemi di IA di prendere decisioni autonomamente. A meno che non si consideri la stessa capacità di apprendimento come un difetto, sarà difficile stabilire quando i sistemi di IA siano difettosi, dato che son progettati proprio per imparare con l'esperienza e comportarsi in base alle nuove acquisizioni. Ed anche se si considerasse la mera deviazione

dell'algoritmo rispetto al percorso decisionale ordinario come difetto, la responsabilità del produttore verrebbe comunque esclusa sulla base della *later defect defense* ovvero della *risk development defense*.

Se, in breve, la tendenza è quella di paragonare sempre di più la nozione di difetto alla valutazione circa la prevedibilità dell'evento lesivo, si finisce per cascare di nuovo nel *responsibility gap* che il modello puro oggettivo voleva proprio evitare. Resta il problema di definire come incida il *machine learning* sulla configurabilità di un difetto del prodotto, anche perché la volontà di introdurre tra i criteri per la valutazione della difettosità del prodotto quello inerente agli <<effetti sul prodotto dell'eventuale capacità di continuare ad imparare dopo la sua diffusione>> non pare essere risolutiva posto che non si spiega in quale modo l'autoapprendimento possa incidere sul giudizio di difettosità. Ci si chiede, dunque, se la capacità di *machine learning* debba essere considerata come indice di difettosità del prodotto o, al contrario, se debba indurre ad escluderla.

Ciò posto, ci siamo soffermati su alcune interpretazioni del *consumer expectations test* proposte in dottrina. L'eliminazione di fatto del requisito del difetto che trasformerebbe la disciplina in una forma di responsabilità oggettiva pura; il riferimento ad una ipotetica condotta di un agente modello umano o algoritmico e l'adozione della *preemption doctrine*, che ci darebbe la possibilità di ritenere sempre non difettoso il prodotto conforme agli standard tecnici di settore. Questo orientamento pare da rifiutare, alla luce della differenza strutturale e funzionale tra la *product safety* e la *product liability*. Mentre altre delle soluzioni potranno essere poste a fondamento di pronunce giurisprudenziali per la risoluzione dei primi casi relativi alla imputazione dei danni da prodotto intelligente.

Molto interessante la proposta di giudicare la difettosità del prodotto in base alle sue *performance* complessive,

paragonate alle prestazioni di un modello di algoritmo mediamente accurato, così da evitare l'impossibile confronto tra due modelli decisionali, quello algoritmico e quello umano, incompatibili tra loro.

Sempre sulla responsabilità del produttore, si è approfondito il tema del nesso di causalità, sia dal punto di vista sostanziale sia processuale. Avendo escluso che l'autonomia decisionale dell'algoritmo sia tale da interrompere il nesso di causalità ai sensi dell'art. 41, comma 2, c.p., abbiamo sottolineato come l'opacità algoritmica possa rendere complicatissima la valutazione del percorso decisionale interno alla c.d. *black box*. Il criterio dell'aumento del rischio dovrebbe a nostro giudizio rendere più semplice la prova del nesso causale e lo stesso vale a dirsi per la presunzione del *res ipsa loquitur*, ovverosia della presunzione dell'esistenza del nesso di causalità tra il difetto del prodotto ed il danno.

Infine, abbiamo notato come l'estrema mobilità della responsabilità aquiliana sia tale da consentire di poter affiancare ad essa nuovi istituti di allocazione del rischio. Stiamo facendo riferimento alle proposte indirizzate all'istituzione di una personalità elettronica, di un fondo di compensazione per i danni da IA.

Tutto ciò ci fa comprendere come, nel diritto civile, ve ne siano di soluzioni percorribili per offrire ristoro economico alla persona danneggiata da sistemi di IA. Il modello civilistico si pone in antitesi con quello penalistico che, per come è noto, non consente semplificazioni probatorie né criteri di imputazione oggettivistici. Ad essi si oppongono rispettivamente i principi di presunzione di innocenza e di personalità della responsabilità penale assieme all'alto *standard* probatorio dell'oltre ogni ragionevole dubbio.

Capitolo terzo

INTELLIGENZA ARTIFICIALE E RESPONSABILITÀ PENALE

1. Il rapporto tra diritto penale ed IA. Linee generali ed introduttive.;2.La macchina intelligente come autore del reato: la tesi di Gabriel Hallevy.;2.1.Critiche alla teoria di Hallevy.;2.2.Le funzioni della pena e l'erroneo parallelismo tra la responsabilità delle persone giuridiche e la responsabilità delle macchine.;3.La responsabilità degli esseri umani a contatto con il sistema di IA: l'utilizzatore.;3.1.La responsabilità colposa dell'utilizzatore di un sistema di IA completamente autonomo.;3.2.La responsabilità colposa dell'utilizzatore di un sistema di IA parzialmente autonomo.;3.2.1.Che tipo di fatto di reato potrebbe essere imputato all'utilizzatore di un sistema di IA parzialmente autonomo?;3.2.2.La responsabilità colposa ed i limiti del principio dell'human-in-*the-loop*.;3.3.La validità del principio di affidamento: limite alla rimproverabilità?;3.3.1.Prospettive *de jure condendo* e limiti alla rimproverabilità dell'utilizzatore.;3.4.Le ipotesi abberranti ed i sistemi di IA.;4.La responsabilità del produttore per i danni provocati dal sistema di IA.;4.1.La rilevanza del principio di precauzione nel diritto penale.;4.2.La tipicità del reato da IA.;4.3.Il nesso di causalità ed il suo accertamento.;4.3.1.Incompatibilità tra causalità e probabilità del *decision making* algoritmico: è possibile individuare delle leggi scientifiche?;4.4.Chi è responsabile dei danni cagionati dall'IA? Un *too many* hands problem.;4.4.1.L'*AI Act* e le aree di competenza nella gestione del rischio.;4.5.La difficile dimostrazione della colpa del produttore.;4.5.1.Problemi di colpa specifica: l'individuazione delle regole cautelari.;4.5.2.A proposito di rischio consentito: qual è il rapporto tra regole cautelari scritte e non scritte?;4.5.2.1.L'individuazione delle aree di rischio consentito ed i limiti dell'istituto.;4.5.2.2.L'*AI Act* e gli obblighi di monitoraggio post-market: superamento del rischio consentito?;4.5.2.3.IA e rischio consentito: approccio quantitativo o qualitativo?;4.5.3.La colpa generica nell'IA.;4.5.3.1.Un'imprevedibilità prevedibile dell'evento lesivo è rimproverabile?;4.5.3.2.L'incidenza del principio di precauzione sul giudizio di prevedibilità dell'evento.;4.5.4.IA e rischio consentito: applicazione di quanto fino a qui scoperto.;5.Soluzione forzata allo stato attuale: crisi del diritto penale d'evento e tutela anticipata.;5.1.Le fattispecie contravvenzionali di cui all'art. 112 cod. cons.;5.1.2. Nuove soluzioni: de jure condendo.;6.Facciamo un punto: IA e responsabilità penale.

1. Il rapporto tra diritto penale ed IA. Linee generali ed introduttive.

Si è visto che la responsabilità extracontrattuale può essere adattabile ai nuovi sistemi intelligenti e può essere capace di allinearsi alle decisioni sulla gestione del rischio assunte a livello politico. Le alte garanzie che invece circondano il sistema costituzionale del diritto penale lo rendono un diritto rigido, per così dire inidoneo a contenere fenomeni incerti e globalizzati che caratterizzano la società post moderna. L'IA non fa altro

che rendere ancora più acuti quei problemi scaturiti dalla crisi del diritto penale d'evento di fronte alla società del rischio.

Conviene, pertanto, chiedersi se bisogna rassegnarsi di fronte all'impotenza ed all'irrilevanza del diritto penale nell'era della complessità tecnologica ovvero se si debba rinunciare all'intervento penale. Posto che, per quanto ci riguarda, non è possibile rinunciare al diritto penale costituzionale, in questa sede ci limitiamo a illustrare le principali frizioni che i sistemi di IA creano con le logiche classiche del diritto penale italiano.

Si è già fatto riferimento al fatto che le caratteristiche intrinseche dell'IA sono tali da rendere molto problematica, se non impossibile, l'accertamento dell'imputazione penale del danno provocato dai dispositivi intelligenti. Dispositivi che, tanto più sono autonomi, interattivi e connotati da opacità tecnologica, tanto più sfuggono al controllo umano. Ciò che si determina è una interposizione dell'imponderabile *decision making* algoritmico tra produttore/utilizzatore ed evento lesivo.

Un primo problema che si pone è quello dell'accertamento del nesso eziologico tra comportamento umano ed evento lesivo. A tal riguardo, se la proposta introduzione nel processo civile di presunzioni di causalità potrà certamente agevolare i ricorrenti, dall'altro canto una strada siffatta non è percorribile in campo di prova nel processo penale. Processo penale in cui la regola sostanziale della presunzione di innocenza si ripercuote come standard probatorio dell'oltre ogni ragionevole dubbio.

La crisi del modello nomologico dell'*if-this-then-that*, incide anche sulla struttura del reato e, in particolare, rende praticamente impossibile l'accertamento della colpa in capo al produttore. La mancanza di leggi scientifiche e di generalizzazioni causali sul funzionamento dei sistemi

di IA comporta la perdita di controllo dell'utilizzatore e del produttore sull'agire algoritmico. A ciò consegue che sia possibile venga meno il rimprovero personale nei confronti dell'agente[433]. Il rispetto cogente del principio di colpevolezza impone che sussista una connessione psicologica, almeno come colpa, tra azione umana ed evento. Connessione che, considerata l'assenza di un agente modello pienamente capace di controllare i rischi derivanti dalla produzione e dall'utilizzo dell'IA, rischia di perdersi.

Si tenterà di comprendere se gli elementi costitutivi del reato ed i principi generali del diritto penale riescano a tenere botta rispetto all'avanzare del nuovissimo contesto tecnologico, escludendo nettamente la possibilità avanzata in dottrina di ritenere responsabile penalmente la stessa macchina.

Infatti, a nostro giudizio, confutare quelle tesi che riconoscono come possibile una responsabilità penale diretta in capo ai sistemi di IA appare come una questione preliminare da trattare prima del prosieguo dell'indagine. Per noi, insomma, per poter parlare di diritto penale, non si può che affermare l'indissolubile legame tra intelligenza umana e diritto penale stesso: non esiste diritto più umano del diritto penale.

2. La macchina intelligente come autore del reato: la tesi di Gabriel Hallevy.

Vista la difficoltà di accertare una responsabilità penale in capo alle persone che si trovano a contatto con l'algoritmo intelligente, alcuni Autori hanno ipotizzato una forma di responsabilità diretta dei sistemi di IA[434].

---

[433] Vd. per tutti L. PICOTTI*, Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 816; C. PIERGALLINI*, Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen*., 2020, n. 4, p. 1769 ss.

[434] Il principale teorizzatore della possibile configurazione di una responsabilità penale diretta in capo ai sistemi di i.a. è il penalista israeliano Gabriel Hallevy, vd. G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015; ID., "*I, Robot – I, Criminal*" — *When Science Fiction Becomes Reality: Legal Liability of AI Robots committing Criminal*

Il principale sostenitore di questa tesi è il penalista israeliano Gabriel Hallevy, il quale ritiene che negare la capacità penale degli agenti artificiali sia solo colpa di pregiudizi antropocentrici. Secondo l'Autore, nei fatti commessi dai sistemi di IA sarebbero agevolmente identificabili in capo agli agenti artificiali l'*actus rea* e la *mens rea*, ovverosia i requisiti che, nei paesi di *common law*, rappresentano il fondamento della responsabilità penale e che sono analoghi nella sostanza all'elemento oggettivo e soggettivo della teoria bipartita del reato. In particolare, l'Autore fa riferimento agli *embodied AI systems*, cioè a quei *robot* dotati di IA. Ma non si vedono motivi per ritenere le sue argomentazioni come precluse agli altri tipi di sistemi di IA.

Hallevy, aderendo ad una concezione propriamente materialistica dell'elemento oggettivo del reato[435] in forza di cui <<*any material performance through factual-external presentation is considered as an act*>>[436], ritiene che i *robot* siano capaci di compiere azioni penalmente rilevanti. L'esempio adoperato dall'Autore è quello del danno provocato da un braccio idraulico di un macchinario dotato di IA, paragonando questo movimento meccanico ad una azione penalmente rilevante. Inoltre, posto che la

Offenses, in *Syracuse Sci. & Tech. L. Rep.* 1, 2010, p. 1 ss.; ID., *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 2010, p. 171 ss.; ID., *Dangerous Robots – Artificial Intelligence vs. Human Intelligence*, 11 febbraio 2018, in SSRN, disponibile a questo link: https://ssrn.com/abstract=3121905; nello stesso senso, ma con accezioni spesso molto diverse tra loro, vd. Y. HU, *Robot Criminals*, in U. Mich. J. L. Reform, 2019, p. 487 ss.; C. MULLIGAN, *Revenge Against Robots*, in *South Carolina Law Review,* 2018, vol. 69, p. 579; M. SIMMLER – N. MARKWALDER, *Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence*, in *Crim. Law Forum*, 2019, n. 30, p. 1 ss.; F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 433 ss. Possibilisti, in una prospettiva di lungo periodo, in cui elemento oggettivo e soggettivo del reato siano configurabili anche in capo ai sistemi di i.a., sono F. BASILE, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 2019, fasc. 10, p. 31 ss.; ID., *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, cit., p. 16-17; M. HILDEBRANDT, *Criminal Liability and 'Smart' Environments*, in R.A. Duff (ed.), *Philosophical Foundations of Criminal Law*, *Oxford University Press*, 2011, p. 526.

[435] In questo senso A. CAPPELLINI, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018, p. 12 e 14.

[436] G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems*, cit., 62.

macchina è in grado di non agire di fronte ad un dovere legittimo di agire, l'Autore ritiene che essa possa anche commettere fatti tipici omissivi. Così ragionando, anche il movimento di un braccio meccanico non dotato di IA può essere considerato un *actus reus*, non necessitando la c.d. *suitas* nella nozione di elemento oggettivo così rappresentata[437].

Molto più complicata è la dimostrazione dell'esistenza dell'elemento soggettivo nell'azione robotica, sia come dolo[438] sia come colpa[439]. Con riferimento al dolo, secondo Hallevy il comportamento robotico può essere caratterizzato sia da cognizione sia da volizione. L'agente algoritmico, infatti, è in grado di acquisire tramite sensori, dati dell'ambiente esterno, nonché di rielabolarli, appprezzando la probabilità di verificazione di un evento lesivo. Il sistema sarebbe anche in grado di confrontare il proprio comportamento agli obiettivi specifici che gli vengono attribuiti, sulla base di un autonomo *decision making*.

Con riferimento, invece, alla colpa algoritmica, l'Autore crede che il sistema possa percepire o comunque rappresentarsi in modo errato delle circostanze di fatto corrispondenti alla fattispecie astratta incriminatrice, eventualmente a causa di un fallace processo di auto-apprendimento. Secondo Hallevy, tale circostanza è definibile come *unawareness* ed in questi casi sarebbe possibile addivenire ad una valutazione di negligenza del sistema di IA, qualora emerga che l'"agente modello algoritmico" avrebbe potuto riconoscere l'errore percettivo ed evitare l'evento lesivo.

[437] Concordano con Hallevy – pur ritenendo necessario il carattere volontaristico del fatto tipico – F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 440-441.
[438] C.d. *general intent* che negli ordinamenti di *common law* ricomprende *intention*, *knowledge* e *recklessness*.
[439] *Negligence*.

Provando a trattare le linee essenziali di una sorta di teoria generale del reato algoritmico, l'Autore si preoccupa anche di verificare la compatibilità di siffatto modello con il classico paradigma della responsabilità penale, arrivando a volte a risposte affermative[440] ed altre volte giungendo alla conclusione per cui servirebbe adattare in via interpretativa gli istituti del diritto penale classico.

Un argomento usato a sostegno della propria tesi è il parallelismo con la responsabilità da reato degli enti. Hallevy, in particolare, osserva che l'esclusione degli agenti intelligenti dai soggetti del diritto penale sia dovuta ad una sorta di pregiudizio antropocentrico paragonabile a quello che per anni ha portato la dottrina ad escludere una responsabilità da reato degli enti. Dunque, Hallevy paragona il superamento del brocardo *societas delinquere non potest* con quello del *machina delinquere non potest*.

Per l'Autore, addirittura, sarebbe più immediato il percorso atto a superare il brocardo citato posto che i sistemi di IA hanno *A Body to Kick and A Soul to Damn*. L'argomento dell'assenza di anima e corpo veniva infatti utilizzato per escludere una responsabilità penale degli enti[441]. Le pene robotiche a cui andrebbero incontro detti sistemi potrebbero essere, secondo Hallevy, il *training* robotico, la privazione della libertà personale o presunta tale e la disattivazione[442]. Dunque, la finalità sarebbe solo special-preventiva: di riabilitazione, tramite un programma di addestramento con lo scopo di re-indirizzare i processi di *machine learning*, eventualmente accompagnato da una detenzione; se il percorso di

[440] Ad esempio l'Autore ritiene applicabile l'istituto della legittima difesa.

[441] L'espressione è notissima nel dibattito angloamericano in tema di corporate liability, vd. J.C. COFFEE, *"No Soul to Damn: No Body to Kick": An Unscandalized Inquiry into the Problem of Corporate Punishment*, in *Michigan Law Review*, 1981, p. 386 ss. L'espressione è stata poi ripresa da vari autori, proprio con riferimento al dibattito sulla responsabilità diretta dei sistemi di i.a.; vd. P.M. ASARO, *A Body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, in Lin e al. (eds.), *Robot Ethics*, cit.

[442] 18G. HALLEVY, ult. op. cit., p. 212 ss. Altre pene sono immaginate da Y. HU, *Robot Criminal Liability Revisited*, cit., p. 529; vd. anche M.A. LEMSLEY – B. CASEY, *Remedies for Robots*, in 86 U. Chi. L. Rev., 2019, p. 1311 ss., p. 1389 ss.; F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 460.

riabilitazione non funzionasse, verrebbe incapacitato il sistema attraverso la completa disattivazione, quasi come fosse una pena capitale.

In detta prospettiva, è evidente l'impossibilità delle pene robotiche di svolgere alcuna funzione di tipo general-preventivo o retributivo, posto che lo stesso autore riconosce che i sistemi di IA non sono capaci di autoconsapevolezza o sofferenza, né sono in grado di orientare il proprio comportamento in base al rispetto di norme imperative[443].

2.1. Critiche alla teoria di Hallevy.

La maggioranza degli Autori in dottrina critica la posizione di Hallevy e la ritiene del tutto incompatibile con i principi cardine del diritto penale. Per prima cosa, si contesta la ricostruzione dogmatica sul piano della presunta esistenza di un *actus reus* dell'agente robotico[444]. Il fatto tipico, infatti, non deve essere soltanto materiale ma deve essere psicologicamente imputabile all'autore. E questo sia negli ordinamenti di *common law* sia negli ordinamenti di *civil law*[445]. Nell'ordinamento italiano, l'art.

[443] In realtà, nei suoi primi lavori, Hallevy giunge ad individuare uno spazio anche per la funzione general-preventiva, poi abbandonata nella sua monografia. Riconoscono invece una possibile efficacia general-preventiva alle pene robotiche F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 460. Secondo altri, invece, la pena nei confronti delle macchine dovrebbe avere finalità di "vendetta", rispondendo così, catarticamente, alle aspettative retributiviste dei consociati vd. C. MULLIGAN, *Revenge Against Robots*, cit., passim; in argomento vd. le acute riflessioni di A. CAPPELLINI, *Reati colposi e tecnologie dell'intelligenza artificiale*, cit., p. 33.

[444] C. PIERGALLINI, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato*?, cit., p. 1769; D. LIMA, *Could AI Agents Be Held Criminally Liable*?, cit., p. 680; M. FREITAS – F. ANDRADE – P. NOVAIS, *Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible*, in P. Casanovas – U. Pagallo – M. Palmirani – G. Sartor (eds.) *AI Approaches to the Complexity of Legal Systems*, Springer, 2014, p. 152; U. PAGALLO, *AI and bad robots. The criminology of automation*, in M. R. McGuire – T. Holt (eds.), *The Routledge Handbook of Technology, Crime and Justice*, Routledge, 2017, p. 645; B. PANATTONI, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale*, in *Dir. inf.*, 2021, fasc. 1, p. 345, nt. 112; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law. Do we need a "new" criminal law for autonomous robots*?, in WWU Münster, disponibile a questo link: https://miami.uni-muenster.de/Record/ff3df8a9-0daf-4dd8-a015-af91cb499f1a, p. 12.

[445] Si vd. ad es., nel contesto statunitense il *voluntary act principle*, espresso dal § 2.01 del *Model Penal Code*: «2.01 Requirement of Voluntary Act; Omission as Basis of Liability; Possession as an Act. (1) A person is not guilty of an offense unless his liability is based on

42, comma 1, c.p. vieta che l'agente possa essere punito per un fatto commesso senza "coscienza e volontà". Tralasciando il dibattito dottrinale tra chi ritiene la c.d. *suitas* come elemento del fatto tipico[446] e chi, invece, la colloca nella colpevolezza, qui serve soltanto sottolineare che l'azione robotica non è espressione di coscienza e volontà. Il sistema di IA si limita a rielaborare le informazioni esterne tramite un algoritmo che è privo della capacità di autodeterminazione che caratterizza l'agire umano.

I comportamenti dei sistemi di IA derivano dal codice e dall'interazione di esso con l'ambiente esterno. Nell'azione algoritmica non c'è spazio per la scelta ed il governo di un corpo. Caratteristiche, queste, fondamentali affinchè si possa parlare di comportamenti coscienti e volontari. Ma pur volendo riconoscere in capo al sistema di IA una capacità di azione, pare impossibile muovere un rimprovero di colpevolezza all'agente intelligente, per come imposto dal principio di personalità della responsabilità penale di cui all'art. 27, comma 1, Cost.[447]. Il

---

conduct which includes a voluntary act or the omission to perform an act of which he is physically capable. (2) The following are not voluntary acts within the meaning of this Section: (a) a reflex or convulsion; (b) a bodily movement during unconsciousness or sleep; (c) conduct during hypnosis or resulting from hypnotic suggestion; (d) a bodily movement that otherwise is not a product of the effort or determination of the actor, either conscious or habitual». Nella dottrina statunitense vd. anche M.S. MOORE, *Act and Crime: The Philosophy of Action and its Implications for Criminal Law*, in *Oxford University Press*, 1993; per una ricostruzione delle teorie tradizionali sul voluntary act requirement vd. M. CORRADO, *Is There an Act Requirement in the Criminal Law*?, in University of Pennsylvania Law Review, 1994, p. 1529 ss.

[446] G. FIANDACA – E. MUSCO, *Diritto penale. Parte Generale*, Zanichelli, VIII ed., 2019, p. 232; T. PADOVANI, *Diritto penale*, Giuffrè, XII ed., 2019, p. 146 ss..; F. PALAZZO, *Corso di diritto penale. Parte generale*, Giappichelli, VI ed., 2016, p. 316-317; F. MANTOVANI, *Diritto penale. Parte generale*, Cedam, XI ed., 2020, p. 327; M. DONINI, *Teoria del reato. Una Introduzione*, Cedam, 1996, p. 114.

[447] In Italia, C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1769 ss.; A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 14; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in Riv. it. dir. proc. pen., 2021, n. 1, p. 96; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 71; M.B. MAGRO, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica*, in *Leg. pen*., 10 maggio 2020, p. 9 ss.; EAD., *Robot, cyborg e intelligenze artificiali*, in A. Cadoppi –S. Canestrari – A. Manna – M. Papa (a cura di), Cybercrime, Utet, 2019, p. 1202 ss.; A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 747; F. MAZZACUVA, *The impact of AI on corporate criminal liability: algorithmic misconduct in the prism of derivative and holistic*

sistema di IA non è capace di discernimento semantico o di ragionamento ma si limita a riconoscere dei patterns ricorrenti di un *dataset*. Trattasi di attività che simula l'intelligenza umana ma è priva di un corrispettivo psicologico e cognitivo[448].

Riflettendo sulla configurabilità o meno di un dolo algoritmico, è vero che il sistema può rappresentarsi la commissione di un fatto; è vero anche che la rappresentazione che il sistema si fa può arrivare a comprendere la contrarietà del fatto rispetto all'ordinamento penale ed è anche vero che il sistema può agire valutando la convenienza della commissione del fatto. ma questo è un dolo esangue, privo di contenuto, che si limita a replicare esteriormente processi psicologici inesistenti. I soggetti artificiali, quindi, possono agire con tipicità dolosa, ma questa nulla ancora dice riguardo alla colpevolezza dolosa[449]. Di regola, la dimostrazione della prima è sufficiente a provare la seconda: non potendo indagare la psiche dell'agente, la prova dell'intenzionalità è affidata a meccanismi presuntivi basati sull'osservazione dell'atteggiamento esteriore del soggetto.

È un'eccezione quella per cui il dolo così individuato non è sufficiente a muovere un rimprovero di

---

theories, in *International Review of Penal Law*, 2021, cit., p. 153-154; R. BORSARI, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, cit., p. 262 ss.; L. PICOTTI, *Cybercrimes and Criminal Relevance of Artificial Intelligence*, in F.J. Castro Toledo (coord.*), La transformación algorítmica del sistema de justicia penal*, Thomson Reuters, 2022, p. 70; nella dottrina straniera, vd. U. PAGALLO, *The Laws of Robots*, Springer, 2015 p. 50-51; ID., *Vital, Sophia, and Co. – The Quest for the Legal Personhood of Robotics*, in *Information*, 2018, p. 7; S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame?*, cit., p. 417; M. FREITAS – F. ANDRADE – P. NOVAIS, *Criminal Liability of Autonomous Agents*, p. 153-154; P. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 176 ss.; T.C. KING E AL., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, cit., p. 108; M. COECKELBERGH, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, in *Science and Engineering Ethics*, 2020, p. 2051 ss.; A. MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, in *International Review of Penal Law*, 2021, p. 118; S. AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, cit., p. 77; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 12 ss.

[448] A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 14; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 12; EAD., *Robot, cyborg e intelligenze artificiali*, cit., p. 1195 ss

[449] A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 14.

colpevolezza all'agente. Ciò avviene quando intervengano nel fatto circostanze anormali che abbiano influito in modo irresistibile sulla volontà o sulle altre capacità psicofisiche del soggetto, quando insiste l'incapacità di intendere e volere o, infine, quando non conosce senza colpa la legge penale.

Lo stesso meccanismo presuntivo non può essere valido per quanto riguarda gli algoritmi: non esiste una massima d'esperienza che ci possa autorizzare a ricollegare l'esteriorizzazione di cognizione e volizione da parte dell'agente di IA ad una capacità di libera autodeterminazione[450]. Pur volendo accettare la tesi funzionalista fatta propria da Hallevy, occorre comunque rilevare che, ad oggi, la percezione sociale dei sistemi di IA è lontana dal poter avere aspettative di conformazione autonoma alle regole della convivenza sociale. Anche considerando i fondamentali concetti di autodeterminazione e libero arbitrio una *fictio iuris*, non pare che tale illusione coinvolga i sistemi di IA. Gli agenti artificiali sono infatti concepiti come oggetti che, nonostante l'imprevedibilità, rispondono in modo deterministico agli *input* ed agli stimoli derivanti dall'ambiente. Attribuire ad essi fatti penalmente rilevanti al posto che agli esseri umani che stanno dietro allo sviluppo ed uso di detti sistemi pare creare un'ingiustificata impunità a favore di questi soggetti[451]. La responsabilità penale diretta dell'IA, dunque, non avrebbe l'effetto di colmare il *retribution gap* determinato dalla impossibilità di identificare una persona responsabile per la commissione dell'illecito algoritmico.

[450] A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 14-15.
[451] S. AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, cit., p. 79-80; F. CONSULICH, *Flash offenders. Le prospettive di accountability penale nel contrasto alle intelligenze artificiali devianti*, in corso di pubblicazione sulla Riv. it. dir. proc. pen., 2022, fasc. 4, p. 1048.

2.2. Le funzioni della pena e l'erroneo parallelismo tra la responsabilità delle persone giuridiche e la responsabilità delle macchine.

Le altre due critiche che si possono muovere alla tesi di Hallevy riguardano le funzioni della pena e l'erroneità del paragone con la responsabilità da reato degli enti. Con riferimento alla prima delle due, sopprimere la funzione retributiva e general-preventiva non pare possibile viste le tradizionali teorie della pena[452]. Non può parlarsi di retribuzione perché le attuali macchine intelligenti continuano ad essere incapaci di percepire sensi di colpa, riprovazione sociale e vergogna o sofferenza[453]. Gli algoritmi non sono soggetti nemmeno al timore della sanzione, posto che né la previsione incriminatrice né l'afflizione in sé della pena nei confronti di un altro agente robotico potrebbero essere in grado di dissuadere rispetto ad un determinato comportamento. La condanna all'esecuzione di una pena non avrebbe alcun significato comunicativo nei confronti degli altri consociati algoritmici, ma sarebbe destinata a rimanere un episodio isolato con effetti circoscritti al solo sistema sottoposto a pena[454].

---

[452] L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 832; ID., *Cybercrimes and Criminal Relevance of Artificial Intelligence*, cit., p. 70; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 9; A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 15; nella letteratura straniera vd. P.M. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 181; U. PAGALLO, *Killers, fridges, and slaves: a legal journey in robotics*, in *AI & Society*, vol. 26, p. 350; D. LIMA, *Could AI Agents Be Held Criminally Liable*?, cit., p. 688-689; M. HILDEBRANDT, *Ambient Intelligence, Criminal Liability and Democracy*, cit., p. 167.

[453] A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 15; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1770; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 10; F. CONSULICH, *Flash offenders*, cit., p. 1026; A. MORAITI, AI *Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, cit., p. 119; R. SPARROW, *Killer robots*, cit., p. 72; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 16.

[454] A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 16; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1770; F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 32; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 99; P.M. ASARO, *A body to Kick, but Still No Soul to Damn*, cit., 181; F. CONSULICH, *Flash offenders*, cit., p. 1022; A. MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, cit., p. 118.

E sul piano special-preventivo? Hallevy trova un possibile spazio per la funzione riabilitativa, riflettendo su un possibile procedimento di *training* che ripristini il normale funzionamento dell'algoritmo. Ma anche con riferimento a questa ipotesi, le similitudini con la tradizionale attività rieducativa svolta in sede di esecuzione penale sembrano ridotte[455]. La risocializzazione, infatti, è possibile soltanto tramite un percorso di impegno personale del condannato che decide di intraprenderlo come messa in discussione del proprio passato. Percorso, questo, non di certo replicabile con un semplice training algoritmico. Il *training* algoritmico e la sua eventuale riprogrammazione non costituiscono uno strumento di pressione indiretta paragonabile alla pena ma, al pari della disattivazione vista come pena di morte algoritmica, corrisponde ad un intervento manipolativo sull'algoritmo, indirizzato a modificarne in modo coercitivo le modalità di agire[456].

Insomma, le pene robotiche di cui si è parlato hanno soltanto fondamento simbolico[457]. Nulla hanno a che vedere con la pena tradizionalmente intesa. Concetto, questo, che rimanda ad una sofferenza, impossibile in assenza di un substrato biologico[458]. Al di là dei profili di responsabilità delle persone intorno alla macchina, ci limitiamo a rilevare che l'ordinamento penale dispone già di strumenti idonei a neutralizzare la pericolosità intrinseca delle cose. Si pensi alla confisca di cui all'art. 240 c.p. ed all'obbligo di richiamo e ritiro di prodotti pericolosi di cui all'art. 104, comma 3 ss. Codice del consumo.

---

[455] C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1770; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 9; A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 16; P.M. ASARO, *A body to Kick, but Still No Soul to Damn*, cit., 181; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 98-99

[456] F. CONSULICH, Flash offenders, cit., p. 1025.

[457] Così S. AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, cit., p. 78; implicitamente anche J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 18.

[458] M.B. MAGRO, *Decisione umana e decisione robotica*, cit., passim; F. CONSULICH, *Flash offenders*, cit, p. 1025 ss.

Anche il parallelismo con la responsabilità da reato degli enti non è fortunato. La secolarizzazione che ha condotto al riconoscimento della capacità penale delle persone giuridiche è stata dettata dalla necessità di attribuire rilievo autonomo ad una colpevolezza propria dell'ente, nella diffusa convinzione che la colpa d'organizzazione non si risolva nell'insieme di colpe riconducibili a quelle persone che di detta organizzazione fanno parte. Gli enti restano comunque costruzioni naturalisticamente fittizie[459]. Essi possono agire direttamente sul mondo, ma operano in modo nettamente diverso da quello delle persone fisiche, tanto che la responsabilità amministrativa degli enti presuppone la commissione di un reato da parte dei soggetti apicali o subordinati.

L'ente non è altro che il destinatario apparente della funzione dissuasiva esercitata dalla previsione di una pena per la società. Il comando della norma è rivolto al *management* della società, affinchè questo adotti le misure necessarie a neutralizzare ovvero a ridurre il rischio di reato[460].

Al contrario, i sistemi intelligenti possono incidere direttamente sul mondo reale senza che sia necessaria una intermediazione da parte delle persone fisiche. La previsione di una sanzione penale nei confronti del sistema di IA non può avere alcuna funzione di pressione nei confronti delle persone che hanno prodotto o messo in

[459] A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 17; vd. anche C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1768.

[460] C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1768; P. SEVERINO, *Intelligenza artificiale e diritto penale*, in U. Ruffolo (a cura di), *Intelligenza artificiale - Il diritto, i diritti, l'etica*, cit., p. 535; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 9; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 70; A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 17; P.M. ASARO, *A Body to Kick, but Still No Soul to Damn*, cit., p. 181; S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame*?, cit., p. 423-4; D. LIMA, *Could AI Agents Be Held Criminally Liable*?, cit., p. 688; S. AIRES DE SOUSA, "*Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, cit., p. 78.

commercio la macchina[461]. Al massimo, è il proprietario della macchina a subire un pregiudizio economico in caso di afflizione della pena all'algoritmo. Nulla comunque di paragonabile all'efficacia dissuasiva che la previsione di una sanzione penale per l'ente può avere sulla gestione del rischio societario.

Il rischio della tesi di Hallevy è di una deresponsabilizzazione delle persone fisiche che producono o usano sistemi di IA che potrebbero beneficiare di schermi penali algoritmici e, così facendo, aggirare le responsabilità derivanti dalle proprie condotte colpose o dolose.

3. La responsabilità degli esseri umani a contatto con il sistema di IA: l'utilizzatore.

Avendo completamente messo da parte la futuristica ipotesi di una diretta responsabilità penale del sistema intelligente, occorre a questo punto valutare se possa configurarsi in capo all'utilizzatore del dispositivo una responsabilità penale. La sempre più presente perdita di controllo dell'operatore sul sistema di IA, insieme all'opacità tipica degli algoritmi di *machine learning*, può costituire un limite alla rimproverabilità dell'utilizzatore, posta l'assenza di controllo sul sistema stesso. Ma, a nostro giudizio, alcuni profili di responsabilità penale per l'utilizzatore di un sistema di IA possono sorgere sia in relazione alla futura ed eventuale messa in commercio di dispositivi completamente autonomi sia in relazione agli attuali dispositivi non del tutto autonomi che prevedono ancora un ruolo di governo dell'uomo nell'agire algoritmico.

3.1. La responsabilità colposa dell'utilizzatore di un sistema di IA completamente autonomo.

[461] A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 17; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1205.

Forse sarebbe corretto, in relazione a sistemi di IA completamente autonomi, parlare di utenti o fruitori del servizio intelligente piuttosto che utilizzatori. Alla assoluta perdita di controllo sull'agire tecnologico corrisponde la impossibilità di muovere un rimprovero penale all'utente per la verificazione di eventi lesivi su cui questi non ha dominio. Sugli utenti, la legge potrebbe creare una posizione di controllo[462] limitata, nel rispetto del principio dell'*ad impossibilia nemo tenetur*[463]. Ad esempio, il proprietario di un veicolo di livello 4 o 5 sarà comunque soggetto all'obbligo di cui all'art. 79 del Codice stradale in forza di cui <<*I veicoli a motore ed i loro rimorchi durante la circolazione devono essere tenuti in condizioni di massima efficienza, comunque tale da garantire la sicurezza*>>. L'utilizzatore di un veicolo di livello 4 o 5 dovrà essere tenuto a curarne la manutenzione, a controllare che non vi siano segnali di malfunzionamento, a curare l'aggiornamento del sistema operativo ed evitare di mettere in moto la vettura quando le circostanze del caso concreto lo sconsiglino. Se, a causa di una di tali imprudenze, dovesse verificarsi un incidente mortale, egli potrà essere chiamato a rispondere di omicidio stradale omissivo ai sensi degli artt. 589-*bis* c.p. e 40, comma 2, c.p.

Ovvio è che una responsabilità potrà emergere anche sul piano commissivo ma sembrano essere più casi di scuola che reali ipotesi: si pensi ad esempio ad un incidente mortale che si verifica perché il proprietario di una auto a guida autonoma appone un adesivo su un sensore del veicolo che lo faccia smettere di funzionare.

Con l'introduzione dell'*AI Act*, l'introduzione di sistemi di IA ad alto rischio completamente autonomi e senza supervisione umana è vietata. Occorre mantenere sempre

[462] Sulla distinzione tra posizioni di controllo e di protezione vd. G. FIANDACA, *Il reato commissivo mediante omissione*, Giuffrè, 1979, p. 189; G. GRASSO, *Il reato omissivo improprio. La struttura della fattispecie*, Giuffrè, 1983, p. 320; F. MANTOVANI, *Causalità, obbligo di garanzia e dolo nei reati omissivi*, in *Riv. it. dir. proc. pen*., fasc. 4, 2004, p. 1007.
[463] A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit, p. 338, con riferimento agli utenti di *self driving cars*.

il controllo di una persona fisica su di essi. La persona che vigila deve essere messa nelle condizioni di <<*decidere, in qualsiasi situazione particolare, di non usare il sistema di IA ad alto rischio o altrimenti di ignorare, annullare o ribaltare l'output del sistema di IA ad alto rischio*>> e di <<*intervenire sul funzionamento del sistema di IA ad alto rischio o di interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga*>>. I sistemi di IA dovrebbero essere progettati e sviluppati in modo da consentire alle persone fisiche di sorvegliarne il funzionamento. Dunque, il fornitore del sistema dovrebbe a tal fine individuare idonee misure di sorveglianza umana adeguate prima dell'immissione del sistema sul mercato o della sua messa in servizio. Tali misure dovrebbero garantire che il sistema sia soggetto a vincoli operativi che il sistema stesso non può annullare e che risponda all'operatore umano. Inoltre, dovrebbe garantire che le persone fisiche a cui è stata affidata la sorveglianza umana dispongano delle competenze, della formazione e dell'autorità necessarie per svolgere tale ruolo.

Non si deve comunque escludere l'introduzione di sistemi di IA completamente autonomi, purchè questi siano costantemente sorvegliati[464].

3.2. La responsabilità colposa dell'utilizzatore di un sistema di IA parzialmente autonomo.

Il principio della necessaria sorveglianza umana (c.d. *human-in-the-loop*) comporta la rinuncia ai benefici che potrebbero derivare dall'attribuzione ai sistemi di IA di una completa autonomia di azione. Il motivo della rinuncia è proprio quello di impedire che gli errori algoritmici provochino offese che potrebbero essere evitate tramite un intervento umano.

[464] In Germania, con l'approvazione della legge sulla circolazione di autoveicoli di livello 4 di autonomia, non si è previsto alcun obbligo di monitoraggio per gli utilizzatori, ma si è stabilito che sia sempre necessaria la vigilanza da parte di un "supervisore tecnico" anche esterno rispetto all'abitacolo del veicolo con l'obbligo di intervenire quando vi sia una richiesta in questo senso da parte del sistema, riportando il veicolo in una situazione di rischio minimo.

Il rischio derivante dall'individuazione di un garante per l'attività algoritmica sconta il rischio che l'utilizzatore si trasformi in una sorta di "capro espiatorio" per l'attribuzione della responsabilità penale anche nei casi in cui non c'era la concreta possibilità di evitare l'evento lesivo[465]. Nonostante le crescenti potenzialità in termini di autonomia dei sistemi di IA sembrino condurre ad uno spostamento della responsabilità da quella dell'utilizzatore a quella del produttore, le ben note difficoltà inerenti all'accertamento della responsabilità del produttore rischiano di scoraggiare le indagini in quel senso, favorendo una ricerca di responsabilità nei confronti del human-in-the-loop. Questo colmerebbe il presunto *retribution gap*[466] e porterebbe a soddisfare il bisogno di pena della società.

3.2.1. Che tipo di fatto di reato potrebbe essere imputato all'utilizzatore di un sistema di IA parzialmente autonomo?

In prima battuta, si può asserire che una responsabilità dell'utilizzatore potrà esistere sia sul piano commissivo sia sul piano omissivo, a seconda che il sistema necessiti della mediazione di una persona fisica per influire sulla realtà ovvero che il dispositivo possa agire direttamente[467].

Con riferimento alla prima ipotesi, facciamo riferimento ai sistemi che forniscono semplici raccomandazioni ed abbisognano di un filtro umano per poter incidere concretamente sui beni giuridici. Essi si chiamano anche

[465] Sul possibile sviluppo di un meccanismo di "capro espiatorio" nei confronti dell'utilizzatore di sistemi di i.a. parzialmente autonomi vd. M.C. ELISH, *Moral crumple zones*, cit., passim; A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 337; B. PANATTONI, *Intelligenza artificiale*, cit., p. 340; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale: prove d'impatto a livello europeo*, cit., p. 16; C. MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 57.

[466] J. DANAHER, *Robots, Lawand the Retribution Gap*, cit., p. 299 ss.

[467] In ogni caso, valgono anche con riferimento ai sistemi parzialmente autonomi le considerazioni già svolte nel precedente paragrafo: l'utilizzatore potrà dunque cagionare un evento lesivo anche manomettendo colposamente il dispositivo intelligente (si ricordi l'esempio dello sticker) o non effettuando la doverosa attività manutentiva.

intelligent decision support systems (c.d. IDSS). Classico esempio sono i sistemi di supporto alla diagnosi in campo medico: in questi casi, è evidente che la condotta penalmente rilevante sarà di tipo commissivo. Se il trattamento sanitario prescritto dal medico utilizzatore di un *Clinical Decision Support System* determina l'insorgere di una patologia, al medico potrà essere imputata la commissione del reato di lesioni o omicidio colposi, al ricorrere di tutti gli elementi della fattispecie.

Con riferimento, invece, all'uso di sistemi di IA che possono recare offese direttamente ai beni giuridici grazie alla loro più ampia autonomia si assisterà ad un sostanziale cambio di paradigma di imputazione penale, da commissiva ad omissiva[468]. Infatti, l'utilizzatore, più che una persona che adopera direttamente il dispositivo intelligente, è il soggetto che beneficia dell'attività algoritmica e che ha l'obbligo di supervisionarla e di intervenire in caso di malfunzionamenti.

Tornando ancora sui veicoli a guida parzialmente autonoma, i veicoli di livello 2 o 3 possono svolgere interamente, a certe circostanze e per un determinato periodo di tempo, la funzione di guida. Resta, comunque, fermo l'obbligo del conducente di supervisionare l'attività di guida e di intervenire ogni volta serva. La differenza tra i due livelli è che i sistemi di livello 3 sono in grado di riconoscere le situazioni in cui non sono in grado di operare avvisando il conducente con un certo preavviso affinchè questo possa acquisire il controllo del veicolo. Meccanismo simile, invece, non è previsto nei veicoli di livello inferiore: il conducente deve quindi costantemente

[468] C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1753; B. PANATTONI, *Intelligenza artificiale*, cit., p. 340; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale*, cit., p. 167; con specifico riferimento agli *autonomous vehicles* vd. L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 826; A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit, p. 334 ss.; ID., *Reati colposi e tecnologie dell'intelligenza artificiale*, cit., p. 28-29; M. GIUCA, *Disciplinare l'intelligenza artificiale*, cit., p. 24.

monitorare il sistema[469]. In ambedue le situazioni, l'utilizzatore è chiamato a supervisionare la funzione di guida e ad intervenire in caso di necessità: nel caso del livello 2 la posizione di garanzia verrà attivata nel momento in cui viene messo in funzione il software di guida autonoma; nel caso del livello 3, la posizione di garanzia verrà attivata solo eventualmente, nel caso in cui l'allarme acustico del veicolo segnali un problema.

Affinchè si possa configurare una posizione di garanzia in capo all'utilizzatore occorre che vengano introdotte norme specifiche. In Francia, ad esempio, che al nuovo L123-1, comma 2, *code de la route*, dispone che <<*Le conducteur doit se tenir constamment en état et en position de répondre à une demande de reprise en main du système de cunduite automatisé*>>. Anche nel codice tedesco si stabilisce che il conducente di un veicolo di livello 3 deve sempre rimanere consapevole delle condizioni di guida dl veicolo a guida autonoma in modo tale da poter riprendere subito il controllo del veicolo quando richiesto dal sistema o quando sia imposto da motivi di sicurezza[470].

Anche il decreto *Smart Road* che disciplina la circolazione sperimentale dei veicoli a guida autonoma nel nostro paese pare già introdurre una posizione di garanzia per il supervisore del veicolo, laddove prevede che il supervisore debba essere in grado di commutare tempestivamente tra operatività del veicolo in modo automatico e operatività dello stesso in modo manuale e viceversa. Ciò considerata la responsabilità del supervisore del veicolo in entrambe le modalità operative.

Con riferimento all'accertamento del nesso di causalità tra omissione dell'azione doverosa da parte dell'utilizzatore ed evento, ovverosia la causalità ipotetica, ai fini dell'accertamento di un reato omissivo improprio

[469]Contra C. PIERGALLINI, *Intelligenza artificiale*, cit., 1753 e A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit, p. 335, che ritengono che il paradigma commissivo resti valido in relazione ai veicoli di livello 2.
[470] M. EBERS*, Civil liability for autonomous vehicles in Germany*, cit., p. 8.

occorre effettuare un giudizio controfattuale. Tale giudizio "predittivo" relativo all'efficacia impeditiva dell'azione doverosa consente di comprendere se, aggiungendo mentalmente l'azione doverosa omessa individuata alla luce della regola cautelare che imponeva quell'azione, si sarebbe reciso il decorso causale sfociato nell'evento[471].

Pertanto, a seguito dell'accertamento del nesso reale, come ad esempio dell'algoritmo di *image recognition* che non ha riconosciuto un segnale di *stop* ed ha causato la morte del conducente, il giudice dovrà accertare se l'azione imposta dalla norma giuridica avrebbe potuto impedire l'evento. Andrà, per esempio, escluso il nesso di causalità se la richiesta di prendere il controllo del veicolo sia giunta troppo tardi al conducente e la pressione sul pedale del freno non avrebbe comunque impedito lo scontro.

Una responsabilità di tipo commissivo potrà residuare nei casi in cui l'utilizzatore, avendo ripreso il controllo del veicolo, commetta un errore nell'effettuare la manovra necessaria. Qui si ristabilisce il normale rapporto di strumentalità tra conducente e veicolo[472]. La stessa circostanza si verifica con riferimento agli *autopilot* usati sugli aeroplani: ripreso il controllo dell'aereo, il pilota che provoca un incidente sarà responsabile a titolo commissivo, anche se nella sua condotta possono

[471] Sulla struttura della causalità nei reati omissivi impropri vd. per tutti G. GRASSO*, Il reato omissivo*
improprio, cit., p. 385 ss.; F. STELLA, *La nozione penalmente rilevante di causa: la condizione necessaria*,
in *Riv. it. dir. proc. pen*., 1988, p. 1249 ss.; ID., *Causalità omissiva, probabilità, giudizi controfattuali.*
L'attività medico-chirurgica, in *Cass. pen*., 2005, 422 ss.; C.E. PALIERO*, La causalità dell'omissione:*
formule concettuali e paradigmi prasseologici, in *Riv. it. med. leg*., 1992, p. 821 ss.; G. MARINUCCI,
Causalità reale e causalità ipotetica nell'omissione impropria, in *Riv. it. dir. proc. pen*., 2009, 523 ss.
[472] A. CAPPELLINI*, Profili penalistici delle self-driving cars*, cit., p. 336.

persistere momenti tipicamente omissivi, tipici di ogni contestazione colposa[473].

L'evento lesivo può anche essere il risultato di più cause, imputabili a più persone. Si pensi ad esempio ad un *healthcare robot* di livello L2 gravemente difettoso che, nel corso di una operazione chirurgica, recida un'arteria provocando il decesso del paziente, mentre il supervisore medico sia inadempiente all'obbligo di sorveglianza poiché si trova al di fuori della sala operatoria. L'evento lesivo, in detta ipotesi, sembrerebbe il risultato di due concause, ovverosia il malfunzionamento del prodotto ed il mancato intervento del medico, sempre che si dimostri la causalità ipotetica. Qui si potrebbe configurare un concorso di condizioni colpose indipendenti ex art. 41, comma 3, c.p.[474].

Questa è una ipotesi che, a nostro giudizio, sarà molto frequente nella prassi, posto che alla maggiore autonomia degli algoritmi consegue la maggiore difficoltà da parte del supervisore di notare il malfunzionamento stesso. Inoltre, la dimostrazione di una responsabilità in capo al

---

[473] F. VIGANÒ, *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, in *Dir. pen. cont. – Riv. Trim.*, 2013, n. 3, p. 391; F. GIUNTA, *Illiceità e colpevolezza nella responsabilità colposa*, *Cedam*, 1993, p. 90; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie" nella prospettiva delle fattispecie colpose causalmente orientate*, *Cedam*, 2003, p. 44.

[474] In questo senso I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 103. Con riferimento ai reati colposi causalmente orientati, di cui ci stiamo occupando, non potrà ovviamente configurarsi una cooperazione colposa ai sensi dell'art. 113 c.p. Sulla differenza tra i due istituti vd. L. RISICATO, *Il concorso colposo tra vecchie e nuove incertezze*, in *Riv. dir. proc. pen.*, 1998, p. 137 ss.; L. CORNACCHIA, *Concorso di colpe e principio di responsabilità per fatto proprio*, *Giappichelli*, 2004, p. 142 ss. Recentemente, sulla funzione incriminatrice svolta dall'art. 113 c.p. vd. F. CONSULICH, *Errare commune est. Il concorrente colposo, il nuovo protagonista del diritto penale d'impresa (e non solo)*, in *Leg. pen.*, 28 marzo 2022 e M. DI FLORIO, *La cooperazione nel delitto colposo: una fattispecie con una (problematica) funzione incriminatrice*, in *Arch pen.*, 2021, n. 1.

produttore incontra problemi rilevanti sia per quanto concerne l'accertamento del nesso di causalità sia per quanto riguarda quello dell'elemento psicologico. Ne consegue che il rischio è che le ipotesi delle concause venga in concreto abbandonata, con conseguente addebito al solo utilizzatore dell'evento lesivo.

3.2.2. La responsabilità colposa ed i limiti del principio dell'*human-in-the-loop*.

Il principio dell'*human-in-the-loop* serve per evitare che l'attribuzione di una completa autonomia agli algoritmi possa comportare rischi eccessivi per la tutela dei beni giuridici costituzionali. Dall'altro lato, vi è il rischio che l'utilizzatore si veda accollati tutti gli eventi lesivi derivanti dall'uso del sistema, indipendentemente dalla individuazione della specifica regola cautelare violata e dalla possibilità, in concreto, di evitare il verificarsi dell'evento lesivo. Verrebbe, in questo modo, creata una responsabilità solo formalmente colposa, ma sostanzialmente oggettiva, da posizione. L'utilizzatore altro non sarebbe che un capro espiatorio.

Sorge la questione del come reperire le regole cautelari positive inerenti al settore del'IA per poter imputare una colpa specifica. Esse, a nostro giudizio, possono essere rintracciate nell'ordinaria regolazione del settore, ove applicabile. Ma sarebbe preferibile che vi sia una apposita normazione, specifica per ogni strumento ed ogni settore di impatto. Nell'elaborazione di dette norme cautelari specifiche e anche nella valutazione della sussistenza della colpa generica, una particolare attenzione deve essere posta con riferimento alla valorizzazione della misura soggettiva della colpa, inerente al potere concreto dell'utilizzatore in quella specifica situazione[475].

[475] Sulla misura soggettiva della colpa vd. il saggio fondamentale di V. DE FRANCESCO, *Sulla misura*
soggettiva della colpa, in *Studi Urbinati*, 1977, p. 273 ss.; più recentemente, D. CASTRONOVO, *La colpa*

Dunque, con riferimento all'accertamento della colpa in capo all'utilizzatore, si pone il problema di stabilire i confini della diligenza esigibile. In questa valutazione, occorre considerare che la perdita di controllo sulla tecnologia ha inevitabilmente un impatto sulla ricostruzione dell'agente modello. Lo standard di diligenza richiesto dovrà per forza di cose tenere conto degli effetti dell'autonomia algoritmica sulle capacità di azione dell'operatore di un sistema di IA. L'agente modello del conducente del veicolo parzialmente autonomo, ad esempio, non potrà essere ricostruito usando gli stessi criteri impiegati per individuare quello della guida tradizionale. Si dovrà tenere conto delle medie conoscenze, con l'aggiunta di eventuali speciali e maggiori conoscenze maturate del soggetto individuale. Non esisterà un unico agente modello utilizzatore di sistemi di IA, ma, ovviamente, tanti modelli quanti sono i settori in cui l'IA promette di trovare (o trova già) applicazione. Il giudice dovrà compiere un gravoso sforzo di astrazione nell'identificazione dell'agente modello, vista la scarsità attuale di una conoscenza estesa e diffusa sull'impatto dell'IA sulle capacità psicofisiche dell'agente. È impensabile pretendere dall'utilizzatore una generica sorveglianza, come quella dell'*human-in-the-loop*, in grado di neutralizzare o minimizzare ogni tipo di rischio da prodotto intelligente.

Nella valutazione della diligenza dell'utilizzatore e della esigibilità della conformazione alle norme cautelari, il giudice dovrà prendere in considerazione le caratteristiche

---

penale, Giuffrè, 2009; ID*., La colpa "penale". Misura soggettiva e colpa grave*, in *Riv. it. dir. proc. pen*.,
2013, n. 4, p. 1723 ss.; A. CANEPA, *L'imputazione soggettiva della colpa. Il reato colposo come punto*
cruciale nel rapporto tra illecito e colpevolezza, *Giappichelli*, 2011; M. GROTTO, *Principio di colpevolezza,*
rimproverabilità soggettiva e colpa specifica, *Giappichelli*, 2012; A.R. DI LANDRO, *Dalle linee guida e dai*
protocolli all'individualizzazione della colpa penale nel settore sanitario. Misura oggettiva e soggettiva
della malpractice, *Giappichelli*, 2012.

dei sistemi di IA. Esse li rendono diversi da ogni tipo di strumento, anche perché sono in grado di ostacolare la capacità di governo dell'utilizzatore stesso. Queste sono, in particolare, l'opacità, la c.d. distorsione dell'automazione e la ridotta capacità di reazione in situazioni di emergenza.

In primo luogo, il funzionamento di tali sistemi è connotato da opacità. Non sempre è possibile ricostruire il *path* causale che ha determinato il passaggio dagli input agli output. La c.d. *black box* di cui si è già parlato *supra* incide chiaramente sull'accertamento della colpa dell'utilizzatore[476]. Gli incidenti aerei che hanno coinvolto i Boeing 737 MAX sono un chiaro esempio. Lì, i sensori avevano erroneamente percepito che l'aereo si trovava in una situazione di stallo ed avevano trasmesso il segnale di attivazione del sistema di IA anti-stallo, nonostante non ve ne fosse il bisogno. Il tentativo dei piloti di riportare l'aereo nella corretta posizione si scontrava con l'opposto tentativo del dispositivo di IA di risolvere una situazione di stallo inesistente. La mancata comprensione da parte dei piloti di cosa stesse succedendo era provocata dall'opacità del sistema, che non dava la possibilità ai piloti di identificare l'errore in cui era incorso il dispositivo e di decidere di disattivarlo[477]. L'opacità è una caratteristica anche dei sistemi di IA che forniscono raccomandazioni agli utenti: se il medico che si serve di un sistema di supporto alla diagnostica ed al trattamento ha accesso alla sola raccomandazione finale fornita dal dispositivo potrebbe risultare complicato per il medico riconoscere l'errore algoritmico e discostarsi dal

[476] In questo senso vd. D. SELBST, *Negligence and AI's Human Users*, cit. p. 1331 ss.; A. CAPPELLINI, *Reati colposi e tecnologie dell'intelligenza artificiale*, cit., p. 30.

[477] Sugli specifici problemi della *black box* e della *technological complacency* nella gestione dell'emergenza in ambito di aviazione vd. M.C. ELISH, *Moral crumple zones*, cit.; M.C. ELISH – T. HWANG_*Praise the machine! Punish the human*!, cit.; W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit.

suggerimento. L'attività diagnostica è finalizzata alla individuazione del trattamento sanitario più appropriato per il caso concreto fondato su ipotesi e valutazioni la cui fondatezza va argomentata, affinchè possa eventualmente essere contestata e messa in discussione.

A poco serve, a nostro giudizio, la specificazione del legislatore europeo all'interno dell'*AI Act* secondo cui i sistemi ad alto rischio devono essere progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l'output del sistema e di utilizzarlo adeguatamente. L'intento del legislatore europeo è lodevole, ma rischia di restare mera dichiarazione di principio[478].

Questo accadrà soprattutto con riferimento a quei sistemi che utilizzano le reti neurali artificiali, la cui opacità è proprio una capacità intrinseca. Infatti, anche nel caso in cui i sistemi venissero dotati di sistemi di interfacce *user-friendly*, non è detto che l'utente sia messo in condizione di ribaltare la decisione dell'IA. Questo accade o perché il tempo di reazione è minimo o perché la spiegazione data dal *machine learning* è descrittiva e probabilistica: non è, dunque, causale.

In secondo luogo, gli effetti collaterali derivanti dalla c.d. *black box* dei sistemi intelligenti si rendono più manifesti se consideriamo uno dei principali problemi inerenti all'automazione: la tendenza delle persone a fidarsi ed affidarsi eccessivamente alla tecnologia ed a non

[478] A. AMIDEI, *Le responsabilità del produttore di intelligenza artificiale «difettosa» tra misure di attenuazione by design e obblighi di trasparenza*, in A. Pajno e al. (a cura di*), Intelligenza artificiale e diritto: una rivoluzione?*, vol. 2, cit,, p. 290; A. GIANNINI, *Intelligenza artificiale, human oversight e responsabilità penale*, cit., p. 16; U. RUFFOLO – A. AMIDEI*, La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design*, cit., p. 492; E.M. COTUGNO, *Intelligenza artificiale e tutela del consumatore. I nuovi obblighi informativi*, cit., p. 319.

metterne in discussione i risultati. Questo fenomeno prende il nome di distorsione della automazione, c.d. *technological complacency*. Tale fenomeno è presente soprattutto in relazione a quei sistemi di IA che si limitano a fornire consigli al decisore umano. Si fa riferimento, ad esempio, alle raccomandazioni date dal sistema di supporto alla diagnosi ed al trattamento sanitario, basate su milioni di dati relativi a patologie e cure e, per questo, capaci di raggiungere livelli di accuratezza superiori a quelli dell'essere umano. Il medico che li usa sarà portato a fidarsi, giungendo finanche a mettere in discussione le sue stesse capacità di giudizio clinico. È quello che viene definito *effetto moutonnier*, cioè effetto gregge, che comporta che l'essere umano aderisca acriticamente alle decisioni algoritmiche[479].

Si può prevedere come, in diversi settori, ci sarà una tendenziale decrescita delle competenze dell'essere umano, a vantaggio dell'affidamento di professionisti ed utenti sul *decision making* algoritmico. L'esempio è quello dei piloti degli aeromobili, i quali, ormai da tanti anni sono abituati a gestire voli con l'aiuto dei sistemi intelligenti[480].

Infine, legato alla questione precedente è il problema del prevedibile incremento degli episodi di distrazione da parte dei soggetti tenuti alla sorveglianza dei sistemi parzialmente autonomi. Il livello di attenzione di un essere umano che si limita a controllare l'attività dell'altro è di

[479] C. CASONATO – B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 2021, 3, p. 429; A. SANTOSUOSSO – G. SARTOR, *La giustizia predittiva: una visione realistica*, in *Giur. it.,* 2022, n. 7, p. 1779 ss.; che riprendono l'espressione di A. GARAPON – J. LASSÈGUE, *Justice digitale. Révolution graphique et rupture anthropologique*, PUF, 2018, p. 279. In argomento vd. anche O. DI GIOVINE*, Il judge-Bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale),* in *Cass. pen.*, 2020, p. 960; D. PERRONE*, La prognosi postuma tra distorsioni cognitive e software predittivi*, cit., p. 117, con specifico riferimento agli algoritmi di supporto alla decisione giudiziaria; C. BURCHARD, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, cit., § 4.1., che parla di *anchoring bias o effect*.

[480] M.C. ELISH, *Moral crumple zones*, cit., p. 12-13: W.B. WENDEL, *Technological Solutions to Human Error and How They Can Kill You*, cit., *passim.*

gran lunga più basso rispetto a quello di colui il quale agisce personalmente[481]. Trattasi, comunque, di un problema che riguarda soltanto i sistemi in grado di ledere direttamente i beni giuridici poiché solo in relazione a questi l'utente è chiamato a reagire tempestivamente ed in condizioni di emergenza. Lo stesso non si può dire in relazione all'utilizzatore di un *intelligent decision support system* che avrà tempo per valutare la raccomandazione fornita dal dispositivo. Nell'accertare la colpa del sorvegliante per non aver impedito il verificarsi di un evento lesivo, sarà necessario prendere in esame il preavviso con cui viene fornito l'allarme da parte del dispositivo intelligente: se esso è minore del tempo di reazione medio, all'utilizzatore non può imputarsi la commissione dell'evento lesivo. La mancata predisposizione di adeguati meccanismi di allerta può risolversi in un difetto di progettazione del prodotto che potrà, eventualmente, essere ascritto al produttore, come responsabilità civile per danno da prodotto e come responsabilità penale.

3.3. La validità del principio di affidamento: limite alla rimproverabilità?

Alcuni Autori, riferendosi ai sistemi di IA, utilizzano la locuzione "soggetti ibridi", con ciò significando lo stretto legame che si crea tra soggetto umano ed agente *software*[482]. Questa ibridazione, tuttavia, non trova alcun riconoscimento in nessuna branca del diritto e non ci sembra possa, a maggior ragione, trovare accoglimento nel

[481] Il problema è stato definito anche "*attentional problem*", vd. B. PANATTONI*, Intelligenza artificiale*, cit., p. 337, che a sua volta cita J. ZERILLI E AL., *Algorithmic decision-making and the Control Problem*, cit., p. 559 ss.; vd. anche S. BECK, *Intelligent agents and criminal law—Negligence, diffusion of liability and electronic personhood*, in *Robotics and Autonomous Systems*, 2016, p. 141; E. HILGENDORF, *Automated Driving and the Law*, cit., p. 187; A. CAPPELLINI, *Profili penalistici delle self-driving cars*, cit., p. 337; P.S. GÜNSBERG, *Automated vehicles – is a dilution of human responsibility the answer?*, in *New Journal of European Criminal Law*, 2022, Vol. 13(4), p. 443.

[482] Questa la tesi di Gunther Teubner. Vd. G. TEUBNER, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, a cura di Pasquale Femia, *Edizioni Scientifiche Italiane*, 2019; A. BECKERS – G. TEUBNER, *Three Liability Regimes for Artificial Intelligence*, *Bloomsbury*, 2022.

diritto penale. Diritto penale che si caratterizza per essere la branca del diritto più umanocentrica, pensata per influenzare il comportamento delle persone fisiche, anche quando sottopone a giudizio ed a sanzione gli enti.

Con riferimento ai sistemi di IA parzialmente autonomi, la condivisione di un ruolo di tipo decisionale tra sistemi di IA ed utilizzatore rischia di complicare la delimitazione delle aree di responsabilità, posto che si crea la necessità di separare con precisione le competenze, le mansioni e le aree di controllo delle differenti forme di intelligenza. Ciò per evitare che il principio dell'*human-in-the-loop* diventi solo una scusa per imputare all'essere umano anche errori propri del dispositivo algoritmico, imputabili al produttore.

Alcuni studiosi hanno invocato i ben noti studi relativi alla attività svolta in *équipe*, posta l'esistenza di uno stretto rapporto di collaborazione tra essere umano e sistema tecnologico. I processi di *decision making* sono, infatti, frazionati e ben distribuiti tra più soggetti. Ma, nelle attività svolte in *équipe*, come quella medica, il processo tradizionale decisionale ed esecutivo è parcellizzato solo tra intelligenze umane[483]. Il problema, al pari delle attività svolte in *équipe*, è proprio quello di capire se e fino a quando l'utilizzatore possa fare affidamento sul funzionamento corretto del sistema e se la sua utilizzazione possa avere un qualche impatto sulla configurabilità della sua responsabilità penale.

Avendo come faro il principio di affidamento, generalmente, si ritiene che l'agente che svolge attività pericolose lecite in collaborazione con altri soggetti, possa fare legittimo affidamento sul fatto che le altre persone svolgano la propria mansione diligentemente,

[483]F. BASILE, *Intelligenza artificiale e diritto penale*, cit., p. 27.

prudentemente ed in maniera perita, rispettando le regole cautelari del settore di riferimento[484].

Occorre, dunque, domandarci se, nell'ipotesi di lesione provocata dal dispositivo intelligente difettoso, si possa configurare un limite alla responsabilità colposa dell'utilizzatore per avere (legittimamente) fatto affidamento nella sicurezza di un prodotto che era stato commercializzato rispettando gli *standard* previsti dall'*AI Act*[485].

A nostro giudizio, la risposta è "dipende". Nel senso che dipende dal fatto che in capo all'agente può esserci una posizione di controllo *ex* art. 40, comma 2, c.p., il cui rispetto comporta la vigilanza dell'operato altrui[486]. L'operato altrui che deve essere vigilato, nel nostro settore, è quello del sistema di IA. Come spesso accade, l'esempio più efficace riguarda il settore delle auto a guida autonoma.

Con riferimento ai veicoli di livello 2, riteniamo non possa operare il principio di affidamento poiché il conducente risulta essere titolare di una posizione di garanzia che si attiva nel momento stesso in cui entra in funzione il sistema di guida autonoma: l'umano non può legittimamente fare affidamento sul corretto funzionamento di un sistema che l'ordinamento giuridico gli chiede di vigilare. Al massimo, la colpevolezza potrà

---

[484] F. MANTOVANI, *Il principio di affidamento nel diritto penale*, in *Riv. it. dir. proc. pen*., 2009, p. 536; M. MANTOVANI, *(voce) Affidamento (principio di),* in *M. Donini (diretto da) Reato colposo*, Enc. dir. – I Tematici, 2022, p. 2; G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 433; D. PULITANÒ, *Diritto penale*, cit., p. 310; F. MANTOVANI, *Diritto penale. Parte generale*, cit., p. 380; G. FIANDACA – E. MUSCO, *Diritto penale. Parte generale*, cit., p. 589 ss.

[485] S. BECK, *Intelligent agents and criminal law*, cit., p. 141; C. MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 58; D'AMICO, *Intelligenza artificiale e auto a guida autonoma. tra prevenzione primaria, colpa penale e rischio consentito*, in *Riv. it. med. leg*., 2022, fasc. 3, p. 608.

[486] È questo uno dei limiti all'operare del principio di affidamento, vd. F. MANTOVANI, Il *principio di affidamento nel diritto penale*, cit., p. 546; C. PIERGALLINI*, (voce) Colpa (diritto penale), in Enc. dir., Annali*, X, 2017, p. 238; G. MARINUCCI – E. DOLCINI – G.L. GATTA*, Manuale di diritto penale*, cit., p. 435.

essere esclusa in forza di una inesigibilità della condotta alternativa lecita se si dimostra che l'agente si trovasse in condizioni tali per cui non si potesse conformare al precetto penale[487].

Invece, se ci si riferisce ai veicoli di livello 3, la posizione di garanzia si attiva solo nel momento in cui vi sia un segnale di allarme emesso dal sistema. Mentre, nelle fasi precedenti, il conducente può legittimamente fare affidamento sul funzionamento del veicolo.

La normazione francese è esemplificativa di quanto fin qui asserito. La *Loi PACTE* del 2019[488] ha distinto chiaramente le differenti aree di responsabilità che coinvolgono produttore e conducente, usando come criterio di demarcazione la circostanza dell'attivazione o meno del *software* di guida autonoma, al momento della verificazione dell'evento lesivo. L'art. L123-2, comma 1, *code de la route* stabilisce che, per il lasso di tempo in cui il sistema esercita il controllo dinamico del veicolo, sarà il produttore a rispondere degli eventuali omicidi o lesioni stradali, sempre che ne ricorrano i presupposti oggettivi e soggettivi. D'altro canto, per lo stesso lasso di tempo, l'art. L123-1, comma 1, esclude la vigenza degli obblighi di vigilanza normalmente previsti in capo al conducente del veicolo tradizionale. L'art. L123-1, comma 2, inoltre, delinea gli obblighi di sorveglianza in capo all'utilizzatore dell'*autonomous vehicle*: nel tempo di attivazione del sistema di guida autonoma, il conducente deve mantenersi sempre in stato di allerta, così da poter riprendere in modo tempestivo la funzione di guida, qualora vi sia una richiesta del sistema. Infine, al comma 3, si stabilisce che quando il conducente non risponde alla richiesta di riprendere il controllo del veicolo avanzata dal sistema, tornerà ad applicarsi la disciplina generale sulla

[487] Sul concetto di inesigibilità vd per tutti, G. FORNASARI*, Il principio di inesigibilità nel diritto penale*, *Cedam*, 1990

[488]Legge n. 486 del 22 maggio 2019, cit. Sulla disciplina francese vd. più approfonditamente M. GIUCA, *Disciplinare l'intelligenza artificiale*, cit.

responsabilità penale del conducente e, quindi, le normali regole cautelari.

La normativa individua una netta divisione delle sfere di competenza tra macchina e utilizzatore, stabilendo che quest'ultimo non è tenuto a monitorare costantemente l'attività del *software*, quando questo è attivo, essendo sufficiente che questi sia abbastanza allerta da poter intervenire in caso di emergenza debitamente comunicata dal sistema.

Uno dei limiti all'affidamento dell'agente si deve individuare nella riconoscibilità del rischio[489], da valutarsi al momento della condotta. Nel caso in cui il prodotto abbia già dimostrato malfunzionamenti, l'utilizzatore non potrà sottrarsi alla responsabilità allegando di avere fatto affidamento sul corretto funzionamento del sistema di IA. Il modello <<a precauzione bilaterale>>[490] tipico della responsabilità per danno da prodotto, in cui entrambe le parti sono tenute a contribuire alla prevenzione del danno, è destinato a caratterizzare l'interpretazione del principio di affidamento nel rapporto tra produttore dei dispositivi intelligenti e l'utilizzatore. In questi casi, potrebbe rinvenirsi un'ipotesi di concorso di cause colpose *ex* art. 41, comma 3, c.p.

Il principio di affidamento, infine, non opera solo a favore dell'utilizzatore, ma anche del produttore. Egli, infatti, non potendo garantire assoluta sicurezza del prodotto fa ragionevole affidamento sul fatto che il consumatore utilizzi il prodotto conformemente alle istruzioni fornite.

3.3.1. Prospettive *de jure condendo* e limiti alla rimproverabilità dell'utilizzatore.

[489] M. MANTOVANI, *Il principio di affidamento nel diritto penale*, cit., p. 545; C. PIERGALLINI, *(voce) Colpa (diritto penale)*, cit., p. 238; G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 434-435; D. CASTRONUOVO, *Responsabilità da prodotto,* cit., p. 334; G. FIANDACA – E. MUSCO, *Diritto penale. Parte generale*, cit., p. 590.
[490] D. CASTRONUOVO*, Responsabilità da prodotto*, cit., p. 334.

L'impatto sulle capacità cognitive degli utilizzatori ha fatto sì che alcuni autori pensassero ad una limitazione della responsabilità in capo a questi ultimi tramite la valorizzazione delle norme cautelari scritte che ancora ad oggi non sono state introdotte.

L'emblema è l'art. 590-*sexies* c.p. che, per le lesioni colpose e l'omicidio colposo commessi nell'esercizio della professione sanitaria, prevede una causa di esclusione della responsabilità, nel caso in cui l'evento si sia verificato per imperizia e siano state <<rispettate le raccomandazioni previste dalle linee guida come definiti e pubblicate ai sensi di legge ovvero, in mancanza di queste, (dal)le buone pratiche clinico-assistenziali, sempre che le raccomandazioni previste dalle predette linee guida risultino adeguate alle specificità del caso concreto>>[491].

Le linee guida, nate all'inizio per innalzare l'accuratezza nelle attività di diagnosi e terapia, si sono rivelate uno strumento prezioso nell'accertamento della colpa del personale sanitario per eventi legati all'esercizio dell'attività professionale. Le linee guida presentano il vantaggio di fornire sia all'operatore medico, *ex ante*, sia al giudice penale, *ex post*, un parametro comune per individuare la condotta doverosa al momento del fatto[492].

Vi è, per tale motivo, chi auspica che il meccanismo dell'art. 590-*sexies* c.p. venga replicato in relazione alle raccomandazioni provenienti dai sistemi di supporto alla diagnosi ed al trattamento clinico. Potrebbe servire

[491] La disposizione deve essere comunque interpretata come se contenesse il limite della colpa lieve: se la colpa da imperizia è grave, la responsabilità penale non potrà essere esclusa, anche nel caso in cui siano state individuate correttamente dal medico le linee-guida adeguate; vd. in questo senso la nota sentenza Mariotti delle Sezioni Unite (Cass., sez. un., 21 dicembre 2017, n. 8770, Mariotti) – una soluzione ermeneutica, com'è stato osservato, difficilmente compatibile con la lettera della norma; vd., per tutti, C. CUPELLI, L'art. *590-sexies c.p. nelle motivazioni delle Sezioni Unite: un'interpretazione 'costituzionalmente conforme' dell'imperizia medica (ancora) punibile*, in *Dir. pen. cont*., 2018, n. 3; R. BARTOLI, *Riforma Gelli-Bianco e Sezioni Unite non placano il tormento: una proposta per limitare la colpa medica*, in *Dir. pen. cont*., 2018, n. 5, p. 245.

[492] Così F. CONSORTE, *Colpa e linee guida, nota a Cass. pen., sez. IV, 23 novembre 2010, n. 8254*, in *Dir. pen. proc.,* 2011, n. 10, p. 1228.

l'introduzione di una tutela dell'affidamento del medico nei confronti della correttezza dei risultati algoritmici che verrebbe meno solo in caso di colpa grave[493].

Fiorella ritiene che la causa di esclusione della responsabilità di cui all'art. 590-*sexies* c.p. possa fungere da modello per una norma che escluda la responsabilità penale dell'utilizzatore del sistema di IA come definiti e pubblicati ai sensi di legge ovvero, in mancanza di questi, secondo le buone pratiche operative, sempre che le raccomandazioni previste dai predetti protocolli o le buone pratiche risultino adeguate alle specificità del caso concreto[494].

Prevedere una norma di questo tipo serve per proteggere l'operatore dalle istanze di colpevolizzazione che sarà destinato a subire, poiché più vicino all'evento algoritmico. Una norma siffatta, però, presuppone l'individuazione di protocolli chiari e rigidi che definiscano in modo preciso gli obblighi di sorveglianza dell'utilizzatore[495]. Ciò non è sempre possibile: non si saprebbe, ad esempio, in che modo poter dettagliare tutti gli obblighi di diligenza che possono emergere nel co-governo di un veicolo parzialmente autonomo? Sempre Fiorella propone di riprendere il contenuto dell'art. 2236 c.c. ed introdurre una disposizione secondo cui <<la responsabilità è esclusa se, tenendo conto della speciale difficoltà dell'attività di controllo dell'IA, l'errore, in cui il garante sia incorso, non sia stato determinato da colpa grave: la colpa è senz'altro grave se il garante trascura di

[493] A. PERIN, *Standardizzazione, automazione e responsabilità medica*, cit., p. 231-232; sostanzialmente nello stesso senso J. MILLAR – I. KERR, *Delegation, relinquishment, and responsibility*, cit., p. 122 ss..

[494] Così A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, in *R. Giordano e al*. (a cura di), *Il diritto nell'era digitale*, cit., p. 662-663.

[495] È lo stesso autore a richiederlo, vd. A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, cit., p. 658-659.

mantenere costantemente monitorati i possibili margini di autonomia non dominabile dell'IA>>[496].

Anche se a noi pare contraddittorio asserire che sia possibile monitorare una autonomia non dominabile, la parte precettiva della disposizione parrebbe tenere in considerazione le ben note difficoltà tecniche che possono emergere nel governo dell'IA. Nel dettaglio, tra i parametri che, per valutare la colpa grave debbono essere presi in esame, importanza specifica deve essere attribuita al grado di violazione della regola cautelare, consistente nel livello di divergenza tra condotta doverosa e comportamento dell'agente[497].

È già da qualche tempo che in dottrina vi è chi desidera limitare l'intervento penale ai soli casi in cui sussista la colpa grave, lasciando alla responsabilità civile la punizione di colpe minori[498]. Il campo dell'IA potrebbe costituire un ulteriore ragione per addivenire a questo cambiamento culturale in tema di colpa penale.

3.4. Le ipotesi abberranti ed i sistemi di IA.

Nulla quaestio nel caso in cui l'essere umano utilizza l'IA come strumento di commissione di illeciti penali: qui a rispondere del reato, sarà, a titolo di dolo, colui che si è

[496] Le disposizioni sarebbero collocate, in ipotesi, all'art. 55-bis c.p., eventualmente rubricato "Limiti all'autonomia ammissibile dell'IA. La colpa punibile del garante"; vd. A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, cit., p. 662-663.
[497] L. CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, in *Arch. pen.*, 2022, n. 2, p. 6-7. Per una ricostruzione sistematica del concetto di colpa grave si rinvia a P.F. POLI, (*voce) Colpa grave*, in *Donini* (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 103. ss.; ID. *La colpa grave. I gradi della colpa tra esigenze di extrema ratio ed effettività della tutela penale*, Giuffrè, 2021
[498] Così M. DONINI, *Teoria del reato. Una introduzione*, *Cedam*, 1996, p. 179; M. RONCO, *Scritti patavini*, *Giappichelli*, I, 2017, p. 406; recentemente vd. L. CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, cit., passim, al quale si rimanda per un'ampia ricostruzione delle ragioni che militano per questa tesi, non nascondendosi, d'altra parte, il rischio che il criterio della gravità si possa risolvere in un incremento di indeterminatezza della fattispecie colposa. Vd. anche D. CASTRONUOVO, *La colpa "penale". Misura soggettiva e colpa grave*, cit., p. 1723 ss. che intravede nella "colpa grave" un grimaldello per introdurre un giudizio "individualizzante" sulla colpa.

consapevolmente servito del dispositivo intelligente[499]. Alla casistica ordinaria, possono aggiungersi ipotesi ben più complesse, laddove l'agente intelligente devia rispetto al percorso che l'utilizzatore aveva in mente, ovvero l'ipotesi in cui il programmatore si limita a fornire istruzioni generali, lasciando al *decision making* algoritmico la scelta di quando e come commettere il fatto.

Partiamo dal caso in cui l'agente devia rispetto al percorso che l'utilizzatore aveva in mente. Visto che una *aberractio causae* non comporterebbe alcun ostacolo per la configurazione del dolo, i casi problematici riguardano la verificazione di eventi diversi da quelli programmati dall'utilizzatore del sistema: l'autonomia del sistema non si presta ad essere ristretta dalla pianificazione criminosa dell'agente[500]. Prendiamo questo esempio: Tizio programma un drone militare, di cui è entrato in possesso illecitamente, per uccidere Caio. Il drone sbaglia mira e colpisce la moglie Sempronia che camminava accanto a lui. Si pensi ancora che Tizio abbia programmato il drone per gambizzare Caio, ma quest'ultimo, per errore nell'esecuzione, colpisca Caio al petto, provocandone la morte.

Nel primo caso, si configura un'*aberractio ictus* di tipo monolesivo[501]. Mentre, nel secondo un omicidio preterintenzionale. Per ambedue i casi, la disciplina codicistica ascrive in maniera oggettiva all'agente la

[499] S. GLESS – E. SILVERMAN – T. WEIGEND, *If robots cause harm, who is to blame*?, cit., p. 425; A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, cit., p. 656-657; S. PREZIOSI*, La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati,* ivi, p. 713-714; S. MASSI, *Affidamento sull'intelligenza artificiale e "disimpegno morale" nella definizione dei presupposti della responsabilità penale*, ivi, p. 665 ss.; A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 7; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 100; S. RIONDATO, *Robotica e diritto penale (robots, ibridi, chimere e "animali tecnologici*", in D. Provolo e al. (a cura di), *Genetics, Robotics, Law, Punishment*, *Padova University Press*, p. 601-602.

[500] Ci riferiamo, qui, a tutte le ipotesi in cui si verifichi un evento diverso da quello voluto dall'agente: oltre alle ipotesi citate nel paragrafo, si pensi altresì *all'aberratio ictus* plurilesiva (art. 82, cpv., c.p.) e alle varie figure di delitti aggravati dall'evento.

[501] I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 110.

verificazione dell'evento diverso da quello voluto. Trattasi di vere ipotesi di responsabilità oggettiva che dovrebbero essere lette alla luce del principio costituzionale di personalità della responsabilità penale e di colpevolezza, aggiungendo quantomeno il limite della colpa. Il principio di colpevolezza, infatti, impone che <<tutti e ciascuno degli elementi che concorrono a contrassegnare il disvalore della fattispecie siano soggettivamente collegati all'agente (siano, cioè, investiti, dal dolo o dalla colpa)>>[502].

Con riferimento all'*aberractio ictus*, l'agente risponderà solo se l'offesa alla persona diversa sia dovuta a colpa: cioè solo se una persona ragionevole si sarebbe accorta in quelle circostanze concrete che l'offesa si sarebbe potuta verificare contro una diversa persona[503]. Al caso del drone che invece di colpire Caio colpisce Sempronia, occorre assimilare il caso in cui il sistema di *image recognition* confonda Caio con il suo gemello, colpendo quest'ultimo. L'errore nell'identificazione della persona, infatti, rileva come errore nell'uso dei mezzi di esecuzione del reato. A ciò consegue che si applicherà l'art. 82, comma 1, c.p., piuttosto che la più rigorosa disciplina dell'*error in persona* di cui all'art. 60 c.p.

Nel caso di preterintenzione, l'agente risponderà di omicidio preterintenzionale o del diverso delitto solo se la verificazione dell'evento diverso avrebbe potuto essere ragionevolmente prevista[504]. Lo stesso meccanismo è indicato dal codice penale per il concorso di persone

[502] C. Cost., sent. 1085 del 30 novembre 1988. Per una completa ricostruzione delle ipotesi di colpa in attività illecite si rinvia, per tutti, a F. BASILE, La colpa in attività illecita: un'indagine di diritto comparato sul superamento della responsabilità oggettiva, Giuffrè, 2005

[503] G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 454; D. PULITANÒ, *Diritto penale*, cit., p. 332-333; G. FIANDACA – E. MUSCO*, Diritto penale. Parte generale*, cit., p. 409. Per una recente e approfondita ricostruzione ermeneutica dell'istituto vd. T. TRINCHERA*, Questioni in tema di aberratio ictus*, in *Riv. it. dir. proc. pen*., 2015, fasc. 3, p. 1412 ss.

[504] D. LIMA, *Could AI Agents Be Held Criminally Liable*?, cit., p. 691. Sull'interpretazione costituzionalmente orientata dell'*aberratio delicti* e della preterintenzione vd. G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 458-459; D. PULITANÒ, *Diritto penale*, cit., p. 331-332; FIANDACA – MUSCO*, Diritto penale. Parte generale*, cit., p. 413.

anomalo di cui all'art. 116 c.p. Detto articolo stabilisce che risponde a titolo di concorso doloso chi, volendo concorrere in un determinato reato, abbia fornito un contributo causale alla commissione di un altro reato che, per decisione di un altro compartecipe, è risultato diverso da quello voluto. Un'interpretazione costituzionalmente orientata, anche qui, impone di ascrivere al concorrente il reato diverso soltanto se una persona ragionevole poteva prevedere che sarebbe stato commesso quel reato diverso[505]. Sia che il sistema di IA sia considerato come un mezzo di esecuzione del reato, sia che esso sia considerato *tertium genus* tra mezzo e concorrente, occorrerà provare che l'agente versi in una "colpa in attività illecita", avendo potuto ragionevolmente prevedere una deviazione rispetto al piano criminoso[506].

È difficile delineare un concetto di prevedibilità in concreto, in relazione al comportamento di sistemi di IA che sono intrinsecamente imprevedibili. O si ritiene che il requisito colposo della prevedibilità del fatto diverso sia integrato di default, nel momento in cui entra in gioco l'agire di dispositivi intelligenti, o si dovrà discernere caso per caso se il sistema intelligente fosse un esecutore affidabile o meno. Quest'ultima verifica sembra complessa se si considera che i sistemi di IA rischiano di essere ancora meno prevedibili di quanto lo siano gli esseri umani.

La giurisprudenza, in tema di art. 116 c.p., si accontenta della sola prevedibilità in astratto, che va accertata tramite l'accostamento delle due fattispecie legali senza tenere conto delle modalità concrete di realizzazione del fatto[507]. Allora, così come la giurisprudenza ritiene sia sempre prevedibile che la rapina sfoci in omicidio, le corti

[505] C. Cost., sent. 42 del 13 maggio 1965.
[506] Così la celebre sentenza Ronci (Cass, sez. un., 22 gennaio 2009, n. 22676), che ha interpretato l'art. 586 c.p. come se già contenesse il limite della colpa.
[507] Ex multis Cass., sez. II, 9 dicembre 1985, n. 3633.

tenderanno ad usare lo stesso criterio quando l'agente intelligente commetterà un reato diverso da quello voluto.

Se il drone uccide Caio, al posto di gambizzarlo, a Tizio potrà essere imputato l'evento morte ex art. 584 c.p. Il caso in cui ad un *robot* intelligente si chieda di compiere un reato di furto in un supermercato e questi commetta invece una strage, i due reati hanno due strutture astratte talmente diverse che potrebbero mettere in discussione la prevedibilità del fatto da parte dell'utilizzatore. Occorre rilevare che un'ipotesi di questo genere ad oggi è solo fantascienza perché non c'è all'orizzonte una prossima introduzione di androidi in grado di poter eseguire fatti di furto.

Infine, se l'IA lede una persona diversa da quella che voleva ledere l'operatore dovrà procedersi ad una valutazione caso per caso, tenendo conto del livello di accuratezza del dispositivo.

Venendo all'ipotesi in cui il programmatore si limita a fornire istruzioni generali, lasciando al *decision making* algoritmico la scelta di quando e come commettere il fatto, in questo caso il programmatore si limita a definire il *genus* del reato senza indicarne *tempus* e *modus*.

L'addebito doloso richiede che l'agente voglia realizzare in maniera consapevole un fatto concreto che corrisponde al fatto descritto dalla fattispecie incriminatrice[508]. L'agente dovrà rappresentarsi quel fatto specifico. Non basta che egli si rappresenti una classe di avvenimenti lesivi perché, così fosse, saremmo nel campo della colpa[509]. Per l'autore del reato algoritmico, potrebbe convenire dare *input* generali al sistema di IA, lasciando che sia quest'ultimo a selezionare il momento, il luogo e le modalità migliori per l'azione. La potenza computazionale elevata dei sistemi di *machine learning* e la loro capacità di adattamento a contesti in continuo e rapido cambiamento

[508] G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 404 ss.
[509] F. CONSULICH, *Flash offenders*, cit., p. 1035.

potrebbero identificare le condizioni più opportune per la commissione del reato.

Su questo tema insiste il *dolus generalis*. Escluderlo, con riferimento ai sempre più sofisticati sistemi di IA, potrebbe rivelarsi una lacuna normativa e potrebbe favorire comportamenti opportunistici da parte degli utilizzatori[510]. L'agente potrebbe sfruttare le sofisticate caratteristiche dei sistemi di IA per commettere fatti di reato e restare indenne. È paradossale che, se l'agente si comportasse con le stesse modalità nei confronti di una persona fisica – istigandola ad esempio a commettere un reato – questi verrebbe punito nel rispetto della disciplina di cui all'art. 110 c.p.

Visto che i sistemi di IA non sono persone fisiche, non può applicarsi la disciplina del concorso di persone nel reato, a meno che di non voler entrare in contrasto con il divieto di analogia ed il connesso principio di tassatività. Ma, pur non essendo persone fisiche, detti sistemi sono dotati di un proprio *decision making* che rende difficile qualificarli come semplici oggetti nelle mani dell'utilizzatore.

Il programmatore che addestra *bots* fornendogli un *dataset* di migliaia di tweet ingiuriosi e diffamatori non potrà rispondere dei post diffamatori pubblicati dal bot sui social. Così come al *trader* che fornisce ai sistemi di *High frequency trading* (HFT) parametri manipolativi del mercato non sarà addebitabile il reato di *market abuse*, difettando la rappresentazione del concreto fatto di

[510] In dottrina c'è addirittura chi propone di estendere l'oggetto del dolo, fino a comprendervi il *dolus generalis* (identificato come "*opaque recklessness*"): «*Under this approach, the use of an AI systems with the awareness that it might engage in criminal activities of a certain kind (e.g. unlawful commercial transactions) suffices for the user's criminal liability for the specific criminal activity performed by the system (e.g. purchase of Ecstasy), even when the user did not foresee that the system would engage in the specific activity*» (F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 459). Una soluzione che ci pare sproporzionata e in potenziale contrasto con i principi di legalità e colpevolezza.

reato[511]. Né tantomeno si potrà individuare una responsabilità da posizione, posta la mancanza di un obbligo giuridico di impedire l'evento[512].

Si viene così a creare un intollerabile *responsibility gap* dal carattere paradossale: <<*the less the human agents know, the more they will be able to deny liability*>>[513]. Agli agenti umani, quindi, converrebbe evitare di scoprire come funziona precisamente e cosa faccia in concreto il sistema di ML. È una soluzione insoddisfacente. Soprattutto se si considera che trattasi di condotte meritevoli di risposta penale poiché generatrici di offese immediatamente percepibili[514].

Tale *responsibility gap* è stato evidenziato specialmente con riferimento al fenomeno degli HFT, sempre più diffusi nel mondo finanziario. Oltre all'ipotesi dell'HFT che reagisce a situazioni che alterano il mercato come la pubblicazione di notizie false vendendo o acquistando titoli rapidamente, sussiste anche la possibilità che sia il *trader* a programmare le pratiche manipolative del mercato. Qui siamo in presenza di un dolo antecedente e generale che, per come è noto, non è rilevante ai fini penalistici.

Vi è chi ha notato un certo collegamento tra queste ipotesi e l'istituto dell'*actio libera in causa*. Quest'ultimo è un meccanismo in forza di cui si colpisce una condotta priva dei necessari elementi soggettivi, a patto che questa sia conseguenza di una precedente libera scelta di agire in quel determinato modo[515]. Anche se è lo stesso Autore a riconoscere i limiti di un simile accostamento. Questo

---

[511] Così F. CONSULICH, *Il nastro di Möbius*, cit., p. 218-219; A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 749-750

[512] D. PIVA, *Machina discere, (deinde) delinquere et puniri potest*, cit., p. 686.

[513] R. WILLIAMS, *Lords select committee, artificial intelligence committee, written evidence* (AIC0206), 2017, disponibile a questo link: http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/70496.html%23_ftn13#_ftnref7.

[514] F. CONSULICH, *Il nastro di Möbius*, cit., p. 219.

[515] F. CONSULICH, *Il nastro di Möbius*, cit., passim; vd. anche F. CONSULICH, *Flash offenders*, cit., passim.

istituto, nel diritto penale italiano, è applicabile solo per le ipotesi di reato commessi da chi si trovi in uno stato preordinato di incapacità di intendere o di volere. Dunque, non si potrebbe applicare *in malam partem* a differenti ipotesi. Occorre anche, ad esempio con riferimento all'istituto dell'art. 87 c.p., che vi sia una piena corrispondenza tra reato programmato e reato commesso. Tale corrispondenza è molto complicata da trovare nei casi di cui trattiamo.

In casi siffatti, la possibilità di muovere un rimprovero al *trader* sembra preclusa[516]. Per questa ragione, occorre spostare il fulcro dalla persona fisica all'ente[517]. Nel nostro ordinamento, pare si possa valorizzare l'art. 8 d.lgs. 231/2011, nella parte in cui stabilisce che la responsabilità dell'ente sussiste anche nel caso in cui <<*l'autore del reato non è stato identificato*>>. La soluzione proposta per il settore delle manipolazioni di mercato pare si possa già trasferire anche in altri settori interessati dall'attività algoritmica. Visto così, l'art. 8 delineerebbe un *tertium genus* imputativo differente da quelli previsti dagli artt. 6 e 7, fondato sulla mancata individuazione dell'autore del reato. La mancata individuazione dell'autore del reato

[516] F. CONSULICH, *Il nastro di Möbius*, cit., p. 224 ss.; F. MAZZACUVA, *The impact of AI on corporate criminal liability: algorithmic misconduct in the prism of derivative and holistic theories*, cit., p. 144; M. DIAMANTIS, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, cit., passim, spec. p. 898; S. AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, cit., p. 69; vd. anche A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, cit., p. 750-751, che sottolinea come, nel nostro ordinamento, una sanzione per il trader sarà comunque sempre applicabile a titolo di colpa, attraverso il ricorso alla disciplina amministrativa in materia di *market abuse* (vd. art. 187-ter T.u.f.).

[517] F. CONSULICH, *Il nastro di Möbius*, cit., p. 224 ss.; M. DIAMANTIS, *The Extended Corporate Mind: When Corporations Use AI to Break the Law*, cit., passim, spec. p. 898; S. AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, cit., p. 73-74; F. MAZZACUVA, *The impact of AI on corporate criminal liability: algorithmic misconduct in the prism of derivative and holistic theories*, cit., p. 144 ss., al quale si rinvia per una panoramica sui regimi applicabili negli ordinamenti di *common law* (in particolare, *respondeat superior* e *identification principle*), basati su regimi di imputazione *lato sensu* oggettivi; A.F. TRIPODI, *Abusi di mercato e trading algoritmico*, cit., p. 750-751.

sarebbe un fatto idoneo di per sé a provare una disorganizzazione dell'ente[518].

L'uso degli HFT in modo abusivo potrebbe rientrare nella casistica anzidetta. Vi possono essere casi di accettazione aziendale o vero e proprio doloso sfruttamento della loro propensione distorsiva del mercato, senza che si possa ricondurre ad una precisa persona fisica la scelta di quella specifica operazione manipolativa[519].

Altra parte della dottrina ha, invece, rilevato come il problema della mancata individuazione dell'autore del reato non sia del tutto assimilabile a quello della difficile attribuzione della responsabilità penale per una attività criminosa commessa da un sistema di IA. Il problema è un altro perché, anche avendo identificato l'operatore dell'HFT, non sarebbe rinvenibile il dolo di quello specifico fatto di *market abuse*. Ciò che manca non è l'autore del reato, ma proprio l'integrazione del reato-presupposto in ogni suo elemento.

Seguendo il principio di precauzione, si potrebbe ragionare sull'opportunità di vietare alcune tecniche algoritmiche o di impedirne l'uso in certi ambiti[520]. Settore per settore, occorrerebbe tenere in considerazione i beni giuridici che vengono in rilievo, nonché i rischi ed i benefici di ogni applicazione dell'attività algoritmica. Restando al settore dei HFT, ad oggi, escludere detti algoritmi di *trading* ad alta frequenza è improbabile, posto che si collocano alla base di numerosissime operazioni di

[518] F. CONSULICH, *Il nastro di Möbius*, cit., p. 225-226. Sul meccanismo imputativo previsto dall'art. 8, d.lgs. 231/2001, vd. diffusamente F. CONSULICH*, Il principio di autonomia della responsabilità dell'ente. Prospettive di riforma dell'art. 8*, in *Rivista 231 – La resp. amm. delle soc. e degli enti*, 2018, n. 4, p. 197 ss.; M.A. BARTOLUCCI*, L'art. 8 d.lgs. 231/2001 nel triangolo di Penrose. Tra minimizzazione del rischio-reato d'impresa e 'nuove forme' di colpevolezza*, in *Dir. pen. cont*., 9 gennaio 2017.

[519] F. CONSULICH, *Il nastro di Möbius*, cit., p. 232.

[520] F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, cit., p. 459.

mercato attuali: eliminarli provocherebbe danni forse maggiori di quelli che causano la loro presenza.

4. La responsabilità del produttore per i danni provocati dal sistema di IA.

L'impresa che produce il sistema intelligente è il soggetto più idoneo a valutare i rischi della commercializzazione dei prodotti intelligenti e ad incidere sulla loro sicurezza. Gli utilizzatori, posto l'aumento di autonomia dei dispositivi intelligenti, saranno sempre più impossibilitati ad esercitare un pieno controllo sulle fonti di rischio. Anche ritenendo possibile il principio dell'*human in the loop*, una responsabilità penale dell'operatore potrà sussistere solo se quest'ultimo sia stato posto nelle condizioni di poter intervenire in concreto sull'attività dell'algoritmo. Così non fosse, parleremmo di responsabilità oggettiva da posizione e, dunque, di una responsabilità in aperto contrasto con la Costituzione.

Ma, vi è da dire che anche le società produttrici perdono il controllo sul comportamento dei sistemi di *machine learning*, poiché esse non ne conoscono appieno tutti i potenziali effetti collaterali.

Ciò che occorre comprendere è se (ed in che modo) si possa configurare una responsabilità del produttore per i reati da intelligenza artificiale. La riconducibilità del fatto algoritmico alla persona del produttore si inserisce in un contesto di base di forzatura o comunque di messa in discussione delle categorie tradizionali penalistiche. Queste ultime risultano incapaci di resistere alla complessità dei fenomeni della post-modernità[521].

[521] Sulla responsabilità penale per danno da prodotto si vd. l'accurata ricostruzione teorica e giurisprudenziale di C. PIERGALLINI*, Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, Giuffrè, 2004, cit.; si vd. inoltre ID., *La responsabilità del produttore: avamposto o Sackgasse del diritto penale*?, in *Riv. it. dir. proc. pen*., 1996, p. 354 ss..; ID., *La responsabilità del produttore: una nuova frontiera del diritto penale*?, in *Dir. pen. proc.*, 2007, n. 9, p. 1125 ss; F. STELLA, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Giuffrè, III ed., 2003; A. BERNARDI, *La responsabilità da prodotto nel*

Occorre avvisare che, da ora in avanti, per comodità espositiva utilizzeremo il termine "produttore" per significare tutte quelle persone che contribuiscono ai processi di sviluppo e produzione dei dispositivi intelligenti come i progettisti, i *data trainers*, i programmatori, i fabbricanti, gli installatori, i distributori ed altri.

4.1. La rilevanza del principio di precauzione nel diritto penale.

Pare necessario, prima di valutare se si possa configurare una responsabilità del produttore, comprendere che ruolo possa avere il principio di precauzione nell'incriminazione delle condotte di sviluppo ed immissione sul mercato di dispositivi intelligenti[522]. L'alone di incertezza ed incalcolabilità circa

sistema italiano: profili sanzionatori, in *Riv. trim. dir. pen. ec.*, 2003, p. 1 ss.; A. DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, in *R. Bartoli* (a cura di), *Responsabilità penale e rischio nelle attività mediche e d'impresa: (un dialogo con la giurisprudenza), Firenze University press*, 2010, p. 437 ss.; C.E. PALIERO, *L'autunno del patriarca, Rinnovamento o trasmutazione del diritto penale dei codici*, in *Riv. it. dir. proc. pen*., 1994, p. 1220 ss.; C. PERINI, *Il concetto di rischio nel diritto penale moderno*, Giuffrè, 2010; D. CASTRONUOVO*, Responsabilità da prodotto e struttura del fatto colposo*, in *Riv. it. dir. proc. pen*., 2005, p. 301 ss.; J.M. SILVA SANCHEZ*, L' espansione del diritto penale*, cit.; G. AZZALI, *La responsabilità penale del produttore per danni alla salute*, in. *Riv. trim. dir. pen. ec*., 1991, p. 849 ss.

[522] Sul rapporto tra principio di precauzione e diritto penale si vedano gli studi monografici di D. CASTRONUOVO, Principio *di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, *Aracne*, 2012; E. CORN, *Il principio di precauzione nel diritto penale. Studio sui limiti all'anticipazione della tutela penale*, Giappichelli, 2013; F. CONSORTE, *Tutela penale e principio di precauzione*, Giappichelli, 2013; vd. anche C. RUGA RIVA, *Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica*, in E. Dolcini – C.E. Paliero (a cura di) *Studi in onore di Giorgio Marinucci*, II, Giuffrè, 2006, p. 1743 ss.; F. GIUNTA*, Il diritto penale e le suggestioni del principio di precauzione*, in *Criminalia*, 2006, p. 227 ss.; A. MASSARO, *Principio di precauzione e diritto penale: nihil novi sub sole?*, in *Dir. pen. cont*., 2011; G. FORTI, *La "chiara luce della verità" e l'"ignoranza del pericolo". Riflessioni penalistiche sul principio di precauzione*, in *M. Bertolino – G. Forti (a cura di), Scritti per Federico Stella*, Jovene, 2007, p. 573 ss.; D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, in *Dir. pen. e proc.*, 2008, p. 647 ss.; S. ZIRULIA*, Esposizione a sostanze tossiche e responsabilità penale*, Giuffrè, 2018, p. 324 ss.; C. PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto "a distanza" tra disvalore d'azione e disvalore d'evento*, in *Riv. trim. dir. pen. econ.,* 2010, p. 225 ss.; C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo,* in G. De Francesco – G. Morgante (a cura di), *Il diritto penale di fronte alle sfide della «società del rischio»*, Giappichelli, 2017, p. 293; da ultimo, vd. F. D'ALESSANDRO*, (voce) Precauzione (principio di*), in Donini (diretto da), *Reato colposo, Enc. dir. – I Tematici*, cit., p. 1001 ss.

il funzionamento dei sistemi di *machine learning* potrebbe portare il legislatore a vietare alcune forme di IA che possano comportare un rischio per la salute e l'incolumità dei consumatori.

Il principio di precauzione impone l'astensione dall'esercizio di un'attività o il limite al suo esercizio al di sotto di una soglia di sicurezza, nel caso in cui vi sia una incertezza scientifica sulla possibilità che l'impiego di una determinata tecnologia o lo svolgimento di una data attività possano provocare danni alla salute ed all'integrità fisica delle persone[523]. Tale concetto si distingue da quello della prevenzione. Quest'ultimo concetto, infatti, mira a ridurre l'incidenza di rischi conosciuti e scientificamente dimostrati; mentre, la precauzione riguarda rischi ignoti che, allo stato delle conoscenze nomologiche, non possono essere interamente dimostrati, né possono essere ragionevolmente esclusi[524].

Detto principio prende piega nel campo della tutela ambientale ed è proprio in questo specifico settore che trova riconoscimento espresso nell'art. 191 TFUE. In detto articolo, si stabilisce che la politica dell'Unione in materia ambientale mira ad un elevato livello di tutela, tenendo conto della diversità delle situazioni nelle varie regioni dell'Unione. Essa è fondata sui principi di precauzione e dell'azione preventiva, sul principio della correzione, in via prioritaria alla fonte, dei danni causati all'ambiente,

[523] C. PIERGALLINI, *Il paradigma della colpa nell'età del rischio: prove di resistenza del tipo*, cit., in *Riv. it. dir. proc. pen.,* 2005, p. 1695. In generale, sul principio di precauzione come criterio di gestione (spesso irrazionale) del rischio vd. C. SUNSTEIN, *Laws of Fear. Beyond the precautionary principle*, *Cambridge University Press*, 2005. M.A. GEISTFELD, *Reconciling Cost-Benefit Analysis with the Principle That Safety Matters More Than Money*, in *New York University Law Review*, April 2001, p. 173 ss.; con riferimento al tema del riscaldamento globale vd. D.A. FARBER, *Coping with Uncertainty: Cost-Benefit Analysis, the Precautionary Principle, and Climate Change*, in *Washington Law Review*, 2015, p. 1659 ss.

[524] D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., pp. 26-27; A. MASSARO, *Principio di precauzione e diritto penale: nihil novi sub sole*?, cit., p. 4 ss.; F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 237-238.

nonché sul principio di "chi inquina paga"[525]. Il campo di applicazione del principio è stato progressivamente esteso anche ad altri settori. Tanto è vero che, ad oggi, vede generale riconoscimento nella Comunicazione della Commissione sul principio di precauzione del 2 febbraio 2000. Qui, la Commissione stabilì che il principio <<trova applicazione in tutti i casi in cui una preliminare valutazione scientifica obiettiva indica che vi sono ragionevoli motivi di temere che i possibili effetti nocivi sull'ambiente e sulla salute degli esseri umani, degli animali e delle piante possano essere incompatibili con l'elevato livello di protezione prescelto dalla Comunità>>.

In molti casi, il principio di precauzione ha ispirato il legislatore penale nella definizione del bilanciamento di interessi tra esercizio di attività pericolose utili e protezione della salute e dell'integrità dei cittadini. Basti pensare al settore della tutela del consumatore, alla disciplina degli organismi geneticamente modificati, alla tutela dell'ambiente e alla prevenzione del contagio da covid-19[526]. L'*AI Act* deve molto del suo complessivo impianto valoriale al principio di precauzione[527]. Il divieto di commercializzare una serie di pratiche di intelligenza artificiale come l'uso di tecniche di identificazione biometrica è una chiara applicazione del principio di precauzione. Lo stesso è a dirsi con riferimento ai sistemi ad alto rischio che possono essere introdotti sul mercato solo qualora siano rispettate cautele volte a minimizzare i rischi per i diritti fondamentali e per la salute e l'integrità fisica delle persone.

[525] Per un panorama completo dei richiami al principio di precauzione contenuti nella normativa internazionale e comunitaria v. G. FORTI, *"Accesso" alle informazioni sul rischio e responsabilità*, cit., 160 ss., e, più recentemente, F. D'ALESSANDRO, *(voce) Precauzione (principio di)*, cit., p. 977 ss.

[526] Per una ricostruzione della normativa italiana ispirata al principio di precauzione vd. F. D'ALESSANDRO, (voce) *Precauzione (principio di)*, cit., p. 980 ss.

[527] B. PANATTONI*, Intelligenza artificiale*, cit., p. 356-357; C. MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 52.

Generalmente, autorevole dottrina maggioritaria tende ad escludere che il principio di precauzione possa giustificare l'introduzione di reati di pericolo concreto[528]. Si crede invece sia possibile ricorrere alla formulazione di fattispecie di pericolo astratto anche senza una base scientifica circa l'effettiva pericolosità di determinate condotte. Ma, sarebbe un pericolo, per così dire, ipotetico[529]. Ciò servirebbe per tutelare beni di rango primario nei casi in cui la rischiosità di una determinata condotta, anche se incerta, pare seriamente possibile, sulla base dei dati scientifici disponibili. Deve comunque rilevarsi che l'orientamento preponderante vuole che la previsione dei reati di pericolo astratto possa giustificarsi solo sulla base di dati eziologici nomologicamente consolidati[530].

Il principio di precauzione può essere usato in sede di definizione delle norme cautelari[531]. Ciò accade con la definizione di limiti-soglia quantitativi all'esercizio di un'attività, che, in contesti di incertezza scientifica, spesso si ispira a "logiche iperprudenziali"[532]. Con riferimento alle cautele positive elastiche, è quanto accade nella

[528]D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 652; C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1748; D. CASTRONUOVO, Principio di precauzione e diritto penale. Paradigmi *dell'incertezza nella struttura del reato*, cit., p. 46; F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 240.

[529] D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 652; G. FORTI, *"Accesso alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione*, in *Criminalia*, 2006, p. 155 ss.; p. 204 ss.; R. BARTOLI, (voce) *Fonti della colpa*, in *M. Donini (diretto da) Reato colposo*, *Enc. dir. – I Tematici*, 2022, p. 526.

[530] M. DONINI, *Il volto attuale dell'illecito penale*, Giuffrè, 2004, p. 120; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 504 ss.; C. RUGA RIVA, *Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica*, cit., p. 1749; A. MASSARO, *Principio di precauzione e diritto penale: nihil novi sub sole?*, cit., p. 5; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 47 ss.; C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, cit., p. 291-292.

[531] D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 652; R. BARTOLI, *(voce) Fonti della colpa*, cit., p. 526.

[532] F. CONSULICH, *(voce) Rischio consentito*, in *M. Donini (diretto da) Reato colposo, Enc. dir. – I Tematici*, cit., p. 1114; F. D'ALESSANDRO, *Pericolo astratto e limiti soglia. Le promesse non mantenute del diritto penale*, Giuffrè, 2012, p. 266. Sul rapporto tra logiche precauzionali e tecnica dei limiti-soglia vd. diffusamente E. PENCO*, Limiti-soglia e responsabilità colposa. Il ruolo incerto delle soglie quantitative, dalla colpa specifica al rischio consentito*, in *Riv. it. dir. proc. pen*., 2019, fasc. 1, p. 195 ss.

materia di nostro interesse, nella definizione dei requisiti di messa in commercio dei sistemi di IA, ai sensi dell'*AI Act*. La minimizzazione di attività potenzialmente pericolose potrebbe rientrare in quel bilanciamento tra istanze contrastanti che rappresenta il fulcro della definizione positiva delle cautele. Il legislatore, così, nel rappresentare gli interessi dei consociati, potrà prevedere che la prudenza imponga di non agire in determinate circostanze, anche se ciò significhi rinunciare a svolgere attività utili per la società. Inoltre, occorre ricordare che la violazione di regole cautelari rigide ispirate alla logica precauzionale potrà dare luogo ad un addebito basato sulla colpa specifica solo se, durante il giudizio, si accerti l'effettiva idoneità preventiva tipica della regola precauzionale, nonché l'efficacia impeditiva dell'evento della condotta alternativa lecita[533].

Il principio di precauzione, in casi siffatti, è visto dal legislatore come direttrice dell'intervento politico. In base ad un bilanciamento di costi e benefici, il decisore rinuncia ai vantaggi che potrebbero derivare dal compimento di una attività perché non vuole rischiare che si verifichino eventi lesivi di cui non è prevedibile la portata[534].

Meno semplici sono i rapporti tra colpa generica e precauzione. Occorre valutare se, in un contesto di incertezza scientifica, la diligenza imponga di astenersi dall'esercitare l'attività potenzialmente pericolosa o comunque di ridurne l'impatto.

4.2. La tipicità del reato da IA.

Stando all'IA descritta dal quadro normativo dell'*AI Act*, la responsabilità del produttore può configurarsi sia per il tipo di produzione, sia per il modo di produzione[535]. La responsabilità per il tipo di produzione si configura in

[533] D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 652.
[534] R. BARTOLI, *(voce) Fonti della colpa*, cit., p. 526
[535] F. BRICOLA, *Responsabilità penale per il tipo e per il modo di produzione* in AA.VV., *La responsabilità dell'impresa per i danni all'ambiente e ai consumatori*, Milano, 1978, p. 75 ss.

base a quei settori della produzione che presentano i livelli di pericolosità altissimi. Livelli che sono difficili da affrontare con le cautele inerenti al processo produttivo. Questo è il caso delle pratiche vietate di cui all'art. 5 dell'*AI Act*. Se, invece, la produzione di dispositivi intelligenti è consentita, la responsabilità del produttore sarà individuabile nel modo di produzione, riferendosi alle cautele usate dal produttore nell'iter di sviluppo e commercializzazione del prodotto, al fine di eliminare o minimizzare i rischi[536].

La qualificazione della condotta del produttore come omissiva o commissiva è da sempre discussa in dottrina. Il produttore, immettendo nel commercio il prodotto, compie una azione qualificabile come commissiva, ma è anche vero che, una volta immesso il prodotto, il suo comportamento è solo omissivo. Un'omissione che si discute se sia semplice momento omissivo della colpa ovvero omissione penalmente rilevante.

Comprendere la demarcazione tra condotte omissive e commissive è una questione atavica del diritto penale tradizionale[537]. Ma, è chiaro che tale questione è aggravata dal contesto incerto della post-modernità in cui gli eventi lesivi sono frutto di cause complesse dotate sia di componenti omissive sia commissive. Il criterio che la dottrina propone di assumere è il seguente: il fatto è commissivo se il soggetto ha introdotto nella situazione concreta un fattore di rischio prima assente, poi sfociato nell'evento. Invece, il fatto è omissivo se il soggetto non ha

[536] Riprendono la classificazione di Bricola, nel settore della produzione di sistemi di i.a., C. PIERGALLINI, *Intelligenza artificiale*, cit., 1753-1754; C. MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 61-62; F.C. LA VATTIATA, *La responsabilità penale per danni da intelligenza artificiale alla prova del processo*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale*, cit., p. 672.

[537] Vd. in questo senso R. BLAIOTTA*, Causalità giuridica*, Giappichelli, 2010, p. 317 ss., che ricostruisce i principali orientamenti dottrinali e giurisprudenziali sul tema; vd. anche G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., p. 103; G. FORTI, *Colpa ed evento nel diritto penale*, Giuffrè, 1990, p. 179 ss.; C.E. PALIERO*, La causalità dell'omissione: formule concettuali e paradigmi prasseologici*, cit., p. 828.

contrastato fattori di rischio già esistenti nella situazione concreta che poi sono sfociati nell'evento[538].

Nonostante questa distinzione valga in astratto, nel concreto essa appare molto più labile[539]. Con riferimento alla responsabilità penale per danno da prodotto, si è notata una tendenza alla fungibilità tra qualificazione omissiva e commissiva del fatto. essa non si afferma tanto i base a posizioni di garanzia ben individuate ma in base alle circostanze del caso concreto[540].

La questione non sussiste con riferimento agli eventi lesivi conseguenza della messa in circolazione di prodotti la cui pericolosità ulteriore rispetto ai confini del rischio consentito fosse già conosciuta nel momento in cui il prodotto è stato immesso nel mercato. Qui, la responsabilità del produttore deve essere ricostruita sul piano commissivo, posto che l'atto di immissione in commercio corrisponde alla colposa introduzione nella società di una nuova fonte di pericolo[541]. La distinzione tra omissione e commissione rischia di diventare meno netta nei casi in cui il prodotto riveli di avere rischi maggiori o diversi rispetto a quelli che le norme cautelari miravano a minimizzare o neutralizzare.

Con riferimento ai tradizionali prodotti, questa è la casistica *standard*. Il produttore immette un prodotto sul mercato in buona fede, all'oscuro della lesività dello stesso[542]. La pericolosità del prodotto potrà emergere successivamente, nella concreta interazione con i

---

[538] F. VIGANÒ, *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, cit., p. 384, al quale si rinvia per gli opportuni riferimenti.
[539] V. MONGILLO*, Imputazione oggettiva e colpa tra ''essere'' e normativismo: il disastro di Viareggio*, in *Giur. it*., aprile 2022, p. 955
[540] C.E. PALIERO*, L'autunno del patriarca*, cit., p. 1241; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 232 ss.; ID., *La responsabilità del produttore: una nuova frontiera del diritto penale*?, cit., p. 1126; ID., *La responsabilità del produttore: avamposto o Sackgasse del diritto penale*?, cit., p. 360; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse: problemi e prospettive*, in *Riv. trim. dir. pen. econ*., 2017, n. 3-4, passim e spec. p. 538.
[541] C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 236.
[542] C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 261.

consumatori e, eventualmente, a causa di una inadeguatezza sopravvenuta delle istruzioni fornite in sede di commercializzazione.

Qui, la colpa eventualmente rimproverabile all'agente difficilmente potrà avere ad oggetto il fatto commissivo della messa in commercio. Il mancato ritiro del prodotto pericoloso, così come la mancata informazione ai consumatori ed alle autorità competenti sulla pericolosità del prodotto, sono fatti che appaiano già a primo impatto omissivi.

La giurisprudenza internazionale ci dà esempi di come, per eludere le difficoltà legate all'accertamento dell'illecito omissivo improprio, si sono qualificate come commissive condotte naturalisticamente omissive. Si pensi al caso dello *spray* per pellami che conteneva sostanze tossiche per il sistema polmonare: in quel caso, noto come *Lederspray case*, la decisione assunta dai vertici della società di non ritirare il prodotto è stata considerata attiva poiché consistente in un facere, ovverosia la decisione in quanto tale[543].

Trattasi di un ragionamento evidentemente fallace poiché eliminando mentalmente la decisione dei vertici della società, la situazione di rischio rimane inalterata: la decisione di non intervenire costituisce una omissione. Il problema sarà quello di verificare se sussistano, per il produttore, obblighi di impedimento dell'evento.

Nell'ordinamento tedesco, si riconosce la responsabilità del produttore per il fatto di avere omesso quelle cautele doverose la cui fonte deriverebbe de facto dall'aver posto in essere una pregressa azione pericolosa, ovverosia

[543] In argomento vd. diffusamente C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 232 ss. Un approccio simile è riscontrabile nella giurisprudenza italiana in materia di esposizione dei lavoratori a sostanze tossiche, in cui la condotta del datore di lavoro è considerata commissiva; per alcuni rilievi critici sul punto si vd. M. ROMANO – F. D'ALESSANDRO, *Nesso causale ed esposizione ad amianto. Dall'incertezza scientifica a quella giudiziaria: per un auspicabile chiarimento delle Sezioni Unite*, in *Riv. it. dir. proc. pen*., 2016, fasc. 3, p. 1147 ss.

dall'aver messo in commercio un prodotto pericoloso[544]. Nel caso appena citato, i giudici hanno ritenuto i vertici della società responsabili del reato di lesioni colpose per omissione con riferimento ai fatti verificatisi prima della decisione di mantenere i prodotti sul mercato, per avere omesso di ritirare il prodotto dannoso dopo che si erano già verificati danni ai consumatori.

Il fare pericoloso precedente che ha dato vita alla posizione di garanzia è stato ricollegato all'immissione nel commercio del prodotto, da cui sarebbe derivato in capo al produttore l'obbligo di impedire che da tale atto derivassero eventi dannosi. In base a questa prospettiva, la posizione di garanzia si attiva solo a seguito della percepibilità del rischio[545].

Non possiamo, per ragioni di economia, trattare il dibattito sulle fonti dell'obbligo di impedimento, che vede contrapposti i sostenitori della teoria formale contro quelli della teoria sostanzialistico-funzionale. In posizione intermedia, si collocano i sostenitori della teoria mista[546]. La giurisprudenza ha spesso dimostrato di condividere l'idea secondo cui una posizione di garanzia possa essere generata dalla creazione di una preesistente situazione di pericolo[547].

---

[544] C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 261.

[545] Per la ricostruzione del caso giudiziario, con particolare attenzione ai profili attinenti alla qualificazione della condotta materiale, vd. C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 233 ss.; vd. anche F. STELLA, *Giustizia e modernità*, cit., pp. 232-233; C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 261.

[546] Sul dibattito vd. per tutti G. FIANDACA, *Il reato commissivo mediante omissione*, cit., p. 129 ss.; G. GRASSO, *Il reato omissivo improprio*, cit., p. 184 ss.; F. SGUBBI, *Responsabilità penale per omesso impedimento dell'evento*, *Cedam*, 1975, p. 165; I. LEONCINI, *Obbligo di attivarsi, obbligo di garanzia e obbligo di sorveglianza*, Giappichelli, 1999, p. 180 ss.

[547] Si è affermata ad esempio la responsabilità del proprietario di un vivaio per il decesso del conducente di un autocarro che, a causa dell'innalzamento del piano stradale realizzato senza rispettare le distanze, rimaneva folgorato per il contatto delle piante trasportate con la linea elettrica sovrastante (Cass., Sez. IV, 10 gennaio 2013, n. 27591); nello stesso senso è stato risolto il caso di un bambino morto dopo essere stato travolto da una traversa per uso ferroviario, mentre si arrampicava per gioco su una catasta non messa in sicurezza (Cass., Sez. IV, 21 maggio 1990, n. 11112).

A nostro giudizio, convince maggiormente la tesi della dottrina maggioritaria secondo cui, nel rispetto del principio di legalità, non possa sussistere una responsabilità per reato omissivo improprio senza norme giuridiche che impongano di attivarsi per impedire uno specifico evento. Possono integrare il disposto dell'art. 40. Comma 2, c.p. norme giuridiche extrapenali ed anche norme di diritto privato. Ma, di certo, non semplici rapporti *de facto*, né precedenti attività pericolose[548].

Anche la dottrina che si è occupata della responsabilità del produttore per il danno derivante da prodotto ha sollevato le stesse perplessità rispetto alla possibilità di configurare una posizione di garanzia *de facto*[549]. La deformalizzazione comporterebbe che la posizione di garanzia diventi un generico "obbligo di prevenzione del danno"[550], tramite cui tutti i partecipanti al processo di produzione e commercializzazione del bene rispondono dell'evento solo per avere aumentato il rischio generico della produzione.

A nostro parere, uno spazio per la tipicità omissiva sarà individuabile in capo al produttore, nel nostro ordinamento, in forza di specifici obblighi di monitoraggio ed intervento post market, collocati nel codice del consumo. Questi hanno l'obiettivo di individuare difetti del prodotto emersi successivamente alla messa in commercio.

---

[548]G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 292; F. MANTOVANI, *Causalità, obbligo di garanzia e dolo nei reati omissivi*, cit., p. 984; G. FIANDACA, *Il reato commissivo mediante omissione*, cit., pp. 204 ss.; G. GRASSO, *Il reato omissivo improprio*, cit., p. 277 ss.; I. LEONCINI, *Obbligo di attivarsi, obbligo di garanzia e obbligo di sorveglianza*, cit., p. 103 ss.

[549] C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 260 ss.; A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, cit., p. 26; A. DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, cit., p. 460; C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale*?, cit., p. 360-36

[550] C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale*?, cit., p. 361.

Nel dettaglio, il codice del consumo prescrive obblighi di controllo, di monitoraggio, scambio di informazioni con le autorità amministrative ed intervento, nel caso in cui emergano rischi ulteriori rispetto a quelli che la cristallizzazione delle norme cautelari voleva prevenire. Dunque, se il prodotto, dopo essere stato introdotto nel mercato sulla base di cautele previste dalla legge o dai regolamenti, rivela una attitudine lesiva diversa o maggiore rispetto a quella che le norme cautelari accettavano come rischio residuo, il produttore dovrà provvedere a fornire le informazioni dovute al consumatore o a ritirare il prodotto. Se non lo fa, l'evento lesivo gli potrà essere contestato a titolo omissivo, sempre che si dimostri che l'azione doverosa avrebbe impedito l'evento[551].

Gli stessi principi saranno validi anche per quanto riguarda la produzione di sistemi di IA, con l'avvertenza che l'abilità di questi ultimi di auto-apprendere e di modificarsi nel tempo potrebbe rendere opportuna la messa a punto di obblighi di controllo rafforzati.

Sistemi sicuri e conformi agli standard tecnici potrebbero rivelarsi più pericolosi di quanto previsto in sede di elaborazione delle norme cautelari. I difetti di design potrebbero essere frutto della singola esperienza del sistema di IA. Dunque, essi potranno coinvolgere anche solo alcuni dei prototipi della serie produttiva: il monitoraggio potrebbe richiedere un'attività di vigilanza capillare da parte del produttore.

4.3. Il nesso di causalità ed il suo accertamento.

Dato che il fatto imputabile al produttore potrà essere di tipo commissivo o omissivo, a seconda del momento in cui

[551]Va tuttavia rilevato che questa è una tipica situazione in cui vi è incertezza sulla qualificazione del fatto. Il mancato rispetto degli obblighi di informazione, di ritiro e di richiamo, così come il giudizio sull'evitabilità dell'evento, potrebbero infatti essere interpretati come elementi attinenti alla colpa, e non al nesso condizionalistico omissivo. Sul punto vd. R. BLAIOTTA, *Causalità giuridica*, cit., p. 318 ss.; C.E. PALIERO, *La causalità dell'omissione*, cit

sorge il rischio residuo derivante dal prodotto, ora occorre occuparci del nesso di causalità reale[552] che intercorre tra l'immissione sul mercato del dispositivo intelligente e l'evento lesivo. Ci pare che sia proprio con riferimento a questo momento che si pongono problemi specifici relativi ai sistemi di IA. Si è compreso, infatti, che l'accertamento della causalità ipotetica altro non fa che riprodurre questioni già ben note in dottrina[553].

L'assenza di leggi scientifiche in grado di spiegare il dispiegarsi della catena causale nel funzionamento degli algoritmi di *machine learning* comporta diversi ostacoli nell'accertamento del nesso eziologico tra l'immissione del prodotto sul mercato e l'evento lesivo[554].

---

[552] In generale, sulla causalità, nella dottrina italiana, è doveroso il richiamo ai primi scritti di Federico
Stella, che è stato il più illustre sostenitore della teoria nomologico-deduttiva, vd. F. STELLA, *Leggi*
scientifiche e spiegazione causale nel diritto penale, Giuffrè, II ed., 2000; ID., *La nozione penalmente*
rilevante di causa: la condizione necessaria, in *Riv. it. dir. proc. pen*., 1988, p. 1217 ss.; ID., (voce)
Rapporto di causalità, in Enc. Giur. Treccani, XXV, 1991; sul rapporto di causalità vd. anch*, ex multis*, e
i lavori di G. FIANDACA*, (voce) Causalità*, in *Dig. Disc. Pen*., vol. II, 1988, Utet, p. 119 ss.; O. DI GIOVINE,
Lo statuto epistemologico della causalità penale tra cause sufficienti e condizioni necessarie, in *Riv. it. dir.*
proc. pen., 2002, fasc. 2, p. 634 ss.; M. DONINI, *Il garantismo della condicio sine qua non e il prezzo del*
suo abbandono. Contributo all'analisi dei rapporti fra causalità e imputazione, in *Riv. it. dir. e proc. pen*.,
2011, n. 2, p. 494 ss.; R. BLAIOTTA, *Causalità giuridica*, cit.

[553] In materia vd., per tutti, C.E. PALIERO, *La causalità dell'omissione*, cit., p. 821 ss.; M. DONINI, *La*
causalità omissiva e l'imputazione "per l'aumento del rischio", in *Riv. it. dir. proc. pen*., 1999, fasc. 1, p.
32 ss., con riferimento ad un contesto "pre-Franzese"; successivamente vd. F. STELLA, *Causalità omissiva,*
probabilità, giudizi controfattuali. L'attività medico-chirurgica, in *Cass. pen*., 2005, 1062 ss.; G. MARINUCCI, *Causalità reale e causalità ipotetica nell'omissione impropria*, in *Riv. it. dir. proc. pen*., 2009, 523 ss.; F. VIGANÒ*, Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza*
Franzese, cit., p. 380 ss.

[554] C. PIERGALLINI*, Intelligenza artificiale*, cit., p. 1760; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 107; F.C. LA VATTIATA, *La responsabilità*

Visto che, per come abbiamo visto, i danni provocati dalla capacità di autonomo apprendimento del sistema di IA non sono idonei ad interrompere il nesso causale *ex* art. 41, comma 2, c.p., bisognerà, allora, individuare di volta in volta il fattore causativo del danno.

I sistemi di IA possono essere dotati di tecnologie avanzate di *event data recorder* (EDR). Trattasi di sistemi che sono in grado di ricostruire la dinamica degli eventi lesivi, ma senza essere in grado di sopperire alla presenza di opacità nelle modalità decisionali. Per esempio, con riferimento ai veicoli, essi possono registrare la velocità di crociera, l'accelerazione, l'attivazione di spie o segnali del veicolo, nonché eventuali input del conducente[555].

Si guardi al caso di un incidente stradale cagionato da un veicolo autonomo. L'EDR può, ad esempio, indicare che il sistema non ha rallentato di fronte al segnale di *stop*. Ma il sistema nullo può dire sulle cause dell'incidente,

---

penale per danni da intelligenza artificiale alla prova del processo, cit., p. 703; D. PIVA*, Machina discere,*
(deinde) delinquere et puniri potest, cit., p. 686; S. PREZIOSI, *La responsabilità penale per eventi generati*
da sistemi di IA o da processi automatizzati, cit., p. 720, che parla di "anomia causale"; D.F. LLORCA E AL.,
Liability regimes in the age of AI: a use-case driven analysis of the burden of proof, submitted to Journal
of Artificial Intelligence Research, 3 novembre 2022, disponibile in *Arxiv* a questo link: https://arxiv.org/abs/2211.01817, p. 6-7; E. KARNER – B.A. KOCH – M.A. GEISTFELD, *Comparative Law*
Study on Civil Liability for Artificial Intelligence – Study requested by the Directorate-General for Justice
and Consumers (EU Commission), novembre 2020, p. 23 ss.; vd. anche C.E.A. KARNOW, *Liability for*
Distributed Artificial Intelligence, cit., p. 191, U. PAGALLO*, The Law of robots*, cit., p. 73 e Y BATHAEE,
The Artificial Intelligence Black Box and the Failure of Intent and Causation, cit., p. 922, i quali, tuttavia,
nel parlare di "*failure of causation*", fanno riferimento alla teoria della *proximate cause*, assimilabile alle
teorie della causalità adeguata, in cui vi è dunque una valutazione della *reasonable foreseeability*
dell'evento.

[555] T. LEHOUILLIER E AL., *The Use of Event Data Recorders in the Analysis of Unintended Acceleration Incidents*, testo della relazione tenuta alla 23e *Conférence canadienne multidisciplinaire sur la sécurité routière Montréal*, Québec, 26-29 mai 2013, disponibile a questo link: https://www.roadsafetyresearch.ca/papers/edr_unintended_acceleration.pdf;

limitandosi a fornire dati grezzi da interpretare, senza essere in alcun modo risolutivo sulla ricostruzione del nesso causale. Non è in grado di dire perché il veicolo non ha rallentato di fronte al segnale di stop, in sostanza.

Per l'accertamento processuale del nesso eziologico, non potranno bastare mere intuizioni del giudice, anche se suffragate dai dati raccolti dall'EDR[556]. Ma quest'ultimo dovrà ricostruire il nesso eziologico tra fatto ed evento lesivo dimostrando tutti i sotto-eventi. Ciò con la consapevolezza che <<*è assai difficile, se non impossibile, che il giudice riesca a spiegare, sulla base di teorie e leggi della scienza universali, l'intero meccanismo di produzione dell'evento lesivo, ricostruendo le catene di subeventi che, partendo dall'evento iniziale, si collegano l'uno con l'altro fino a raggiungere l'effetto finale*>>[557].

I dati forniti dagli EDR altro non sono che uno dei tanti elementi valorizzabili dal giudice nell'accertamento processuale del nesso di causalità, ma non sono risolutivi. E, comunque, l'individuazione dei decorsi eziologici lesivi all'interno dei prodotti complessi è un problema ben noto in dottrina[558]. Alla già nota opacità degli artefatti moderni, l'IA aggiunge la capacità di interazione con ambienti

---

[556] Sull'utilizzo nel processo penale dei dati raccolti dagli *event data recorder* vd. F.C. LA VATTIATA, *La responsabilità penale per danni da intelligenza artificiale alla prova del processo*, cit., p. 707, che, in maniera condivisibile, sostiene che i dati raccolti dall'EDR non potranno essere "auditi" direttamente dal giudice, ma saranno introdotti nel processo in contraddittorio, attraverso consulenza tecnica o perizia. Sul tema della prova scientifica, al centro di ampie riflessioni dottrinali negli ultimi anni, si vedano i contributi raccolti in G. Canzio – L. Lupária Donati (a cura di), *Prova scientifica e processo penale*, Cedam, II ed., 2022; vd. inoltre O. DOMINIONI, *La prova penale scientifica: gli strumenti scientifico-tecnici nuovi o controversi e di elevata specializzazione*, Giuffrè, 2005; F. GIUNTA, *Questioni scientifiche e prova scientifica tra categorie sostanziali e regole di giudizio*, in *Criminalia*, 2014, p. 561 ss.; nonché i contributi raccolti in M. Bertolino – G. Ubertis (a cura di), *Prova scientifica, ragionamento probatorio e decisione giudiziale, Atti del Convegno,* Milano, 10-11 ottobre 2014, *Jovene*, 2015; sulla prova scientifica nel processo civile si vd., per tutti, M. TARUFFO, *La prova scientifica nel processo civile*, in M. Bertolino – G. Forti (a cura di), *Scritti per Federico Stella*, cit., vol. II, p. 1375 ss.

[557] F. STELLA, *Giustizia e modernità*, cit., p. 348.

[558] Vd., per tutti, C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 140 ss.; A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *Leg. pen*., 2011, n. 2, p. 401 ss.

spesso imprevedibili, l'abilità di modificare le proprie performance e di adattarsi in base al contesto e la capacità di prendere decisioni. Quest'ultima capacità, è bene sottolineare, non si basa su un paradigma deterministico come quello dell'*if-this-then-that*. Ma si basa sulla creazione di modelli probabilistici. Ma, bene attenti, questo non sta a significare che gli algoritmi di *machine learning* decidano con coscienza e volontà, come gli esseri umani. Significa che l'assunzione delle decisioni da parte degli algoritmi di *machine learning* avviene sulla base di una quantità talmente grande di dati che l'esito del processo può essere ricostruito solo con enorme difficoltà in sede processuale.

Ora occorre concentrarsi sul decorso eziologico relativo al micro-sistema di *black box* costituito dall'algoritmo di ML. Così facendo, potremo isolare gli specifici problemi relativi all'accertamento del nesso causale in relazione ai sistemi di IA. L'esempio da considerare è il seguente: un sistema di *image recognition* installato su un veicolo autonomo non riconosce un semaforo rosso; tale veicolo, impegnando l'incrocio, si scontra con una moto proveniente dal senso di marcia opposto, provocando il decesso del motociclista. In un caso come questo, l'errore nel riconoscimento del semaforo rosso potrà essere stato causato: i) dall'esecuzione di tecniche di *adversal machine learning* da parte di un gruppo terroristico; ii) dal mancato aggiornamento del software per causa del proprietario; iii) dalle avverse condizioni metereologiche; iv) da un *bug* del codice. Oltre a queste ipotesi di interferenza con il *decision making*, può semplicemente capitare un errore del sistema di IA che scambia, ad esempio, il semaforo rosso per una decorazione.

Parte della giurisprudenza passata, con riferimento a prodotti non intelligenti, ha provato a rendere più flessibile il modello nomologico-deduttivo, eludendo l'accertamento del decorso eziologico concreto[559]. I giudici,

[559] I casi principali nella giurisprudenza internazionale sono, in Germania, quelli relativi all'*Holzshutzmittel (*un preparato utilizzato per la protezione dei mobili di legno, contenente

constatando l'opacità delle vicende causali inerenti alla difettosità dei prodotti, hanno adattato il giudizio di imputazione alla fisionomia del fenomeno della *black box*, ritenendo che, ai fini dell'accertamento causale, potesse essere sufficiente che sussistano un rischio iniziale ed un evento lesivo finale come concretizzazione del rischio; che ci sia il ragionevole convincimento che sussista il nesso eziologico tra rischio ed evento; che non abbiano operato in concreto fattori esterni ed indipendenti rispetto al prodotto. Così facendo, si delinea una nuova causalità: non si conosce il fattore dannoso all'interno del prodotto ma si intuisce soltanto che un prodotto abbia un nesso con determinati danni e si sa che non ci sono fattori esterni che possono avere provocato il danno[560]. I giudici non fanno altro che ricavare induttivamente il nesso di causalità sulla base di un collegamento temporale tra l'utilizzo del prodotto e gli eventi lesivi, nonché sulla base dell'interpretazione della *condicio sine qua non* come condizione necessaria del semplice aumento del rischio e non già dell'evento[561]. A volte, i giudici hanno individuato l'esistenza del nesso di causalità senza neppure scoprire la sostanza lesiva all'interno di un elenco di elementi naturali o chimici "sospetti"[562]. Così comportandosi, i giudici cessano di essere "consumatori" di leggi scientifiche di copertura. Ma cessano anche di esserne produttori[563]

---

un'alta concentrazione di sostanze insetticide) e al *Lederspray* (uno *spray* per pellami presunta causa di edemi polmonari) e, in Spagna, quello relativo all'"olio di colza". In Italia si ricorda il processo delle macchie blu, vd. Trib. Rovereto, 17 gennaio 1969, in *Riv. it. dir. proc. pen*., 1971, p. 1021 ss., su cui vd. F. STELLA, *Leggi scientifiche*, cit., p. 47 ss. e 82 ss. Per una precisa ricostruzione della casistica si rinvia a F. STELLA, *Giustizia e modernità*, cit., p. 224-235; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 50 ss.; sulla giurisprudenza italiana vd. A. DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, cit., p. 455 ss.

[560]C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale*?, cit., p. 359, che riporta le considerazioni di W. HASSEMER, *Produktverantwortung im modernen Strafrecht*, 2. Auf., Heidelberg, C.F. Müller, 1996, p. 53-54.

[561] F. STELLA, *Giustizia e modernità*, cit., p. 237 e 245; vd. anche A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, cit., p. 23 ss.

[562] Vd. in particolare i casi *Holzshutzmittel* e *Lederspray*; C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale*?, cit., p. 359; M. DONINI*, La causalità omissiva e l'imputazione "per l'aumento del rischio",* cit., p. 66.

[563] Sul giudice come "consumatore" di leggi causali vd. F. STELLA, *Leggi scientifiche*, cit., p. 80.

poiché rinunciano alla possibilità di conoscere il fattore eziologico determinante del caso concreto[564].

È più che fondato il rischio che simili scorciatoie processuali si realizzino anche con riferimento ai prodotti intelligenti. Ciò è dovuto al fatto che non esiste, al momento, un consolidato elenco di leggi scientifiche in grado di spiegare il comportamento dei sistemi di IA, tanto è vero che gli stessi scienziati non sono in grado di dominarli.

A ben vedere, identificare la causa dell'errore algoritmico potrebbe non essere utile. Si potrebbe solo dimostrare che la condicio sine qua non dell'evento lesivo sia un mero e generico malfunzionamento del sistema di IA per poter accollare al produttore la responsabilità. Ciò senza neanche identificare il concreto e specifico fattore causativo dell'evento lesivo. Spesso, infatti, i sistemi di IA portano con sé una potenzialità lesiva evidente legata al loro dinamismo. Basti pensare ad un *healthcare robot* che, mentre opera, recide all'improvviso un'arteria del paziente.

Un simile ragionamento riecheggia l'approccio all'accertamento causale basato esclusivamente sull'intuito del giudice. Ragionamento che ha dominato nel nostro Paese almeno fino alla fine degli anni Ottanta del Novecento[565]. La sentenza sul disastro del *Vajont* è, in questo senso, esemplificativa. In essa, i giudici affermarono che <<se nessuno è in grado di spiegare perché la frana si sia verificata, ciò nondimeno non si può

[564] M. DONINI, *La causalità omissiva e l'imputazione "per l'aumento del rischio*", cit., p. 66.
[565] Decisiva per un cambio di passo è stata senza dubbio la sentenza relativa al disastro di Stava, vd. Cass., sez. IV, 6 dicembre 1990, n. 4793, Bonetti; in questo senso vd. G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 264; M. RONCO, *Aporie scientiste e certezza logico razionale: note in tema di nesso causale*, in *Dir. pen. proc.*, 2007, n. 2, p. 247; per alcuni esempi di ragionamento causale, "intuitivo" nella giurisprudenza internazionale in materia di responsabilità per danno da prodotto si rinvia a C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 192 ss.

minimamente dubitare che la frana sia dovuta all'opera dell'uomo>>[566].

Al contrario, l'approccio nomologico-deduttivo, rivisto alla luce dei criteri della sentenza Franzese del 2002, richiede che siano rintracciate le leggi scientifiche in forza delle quali possa spiegarsi l'evento. Se il giudice, pur intuendo un certo legame tra azione ed evento, non sia in grado di "coprire" quel sospetto con leggi scientifiche, allora dovrebbe assolvere l'imputato[567]. Occorrerà, dunque, dimostrare il singolo fatto causativo dell'errata decisione algoritmica, non essendo sufficiente un riferimento ad un generico malfunzionamento.

In poche parole, non può operarsi come nel processo civile in cui l'inscrutabilità della *black box* si aggira tramite la logica del *res ipsa loquitur*. Si dovrà, nel processo penale, individuare il preciso fattore che ha provocato l'evento lesivo.

Da tale individuazione, dipenderà l'identificazione dell'area di responsabilità della società produttrice o degli altri attori coinvolti tra cui il programmatore, il progettista, etc.

4.3.1. Incompatibilità tra causalità e probabilità del *decision making* algoritmico: è possibile individuare delle leggi scientifiche?

Uno degli aspetti più difficili da risolvere, in relazione ai danni provocati dai sistemi di IA, è l'individuazione di leggi scientifiche di copertura. Vi sono ipotesi in cui, l'errore algoritmico può essere spiegato grazie all'insistenza di fattori esterni come, ad esempio, un attacco *hacker* che manda in *tilt* un sistema di riconoscimento delle immagini. In casi come questo, l'ordinario funzionamento degli algoritmi di ML è stato

[566] Trib. L'Aquila, 17 dicembre 1969, in *Rass. Giur. Enel*, 1970, p. 197. Sul carattere "intuitivo" delle argomentazioni adottate dal Tribunale in quell'occasione vd. F. STELLA, *Leggi scientifiche*, cit., p. 75 ss.

[567] C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 254.

deviato da cause esterne. Qui è semplice per periti e consulenti tecnici individuare la legge scientifica che collega il fatto A (attacco *hacker*) al sub-evento B (modifica dei parametri decisionali dell'algoritmo) fino all'evento lesivo finale C (ad esempio uccisione di un pedone se si tratta di veicoli autonomi).

La peculiarità propria dei sistemi di ML di ultima generazione è proprio il modello di ragionamento probabilistico, idoneo di per sé ad indurre l'algoritmo in errore, anche in condizioni di corretto funzionamento[568]. Per come si è visto già, il funzionamento delle tecniche più sofisticate di ML è in parte ignoto. Ci affidiamo ai modelli previsionali degli algoritmi come fossero antichi oracoli: non conosciamo perfettamente il motivo, ma, nella maggioranza dei casi, le previsioni di essi sono corrette e più precise di quelle umane. La differenza è che, se a sbagliare è un oracolo, determinando un evento lesivo, quest'ultimo potrà essere collegato alla decisione dell'oracolo stesso; mentre, se a sbagliare è l'algoritmo, sorge il seguente quesito: come si possono individuare regolarità causali in decisioni che sono di per sé *unpredictable by design*[569].

Una regolarità che può individuarsi è inerente ad un'ipotetica statistica degli errori dell'IA. Riprendendo l'esempio del sistema di riconoscimento delle immagini, si potrà notare che in un caso su mille l'algoritmo commette errori. Questo accade, ad esempio, quando si valuta che un *Clinical Decion Support System* ha una accuratezza del 90% nelle diagnosi di cancro: da ciò diceva che nel 10% dei casi, l'algoritmo commette errori. Ma questa rilevazione statistica nulla dice sui fattori che possono avere provocato l'errore. Dunque, non si tratta di legge scientifica, visto che la statistica non può isolare il motivo alla base dell'errore e

[568] Vd. Z.C. LIPTON, *The Mythos of Model Interpretability*, in *Queue*, May-June 2018, p. 5.
[569] Concorda sull'impossibilità di trovare una legge scientifica di copertura S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 717, che ritiene altresì che la decisione autonoma dell'algoritmo di machine learning possa interrompere il nesso causale, ai sensi dell'art. 41, cpv., c.p.

nulla dice sugli eventi intermedi della catena causale la prova dei quali è necessaria per poter verificare compiutamente l'esistenza di un nesso di causalità.

Pertanto, in assenza di interferenze esterne che conducono l'algoritmo all'errore, sarà difficilissimo risalire ai singoli fattori causali poiché il *decision making* algoritmico è frutto di una stratificazione di stimoli e di esperienze di difficile analisi retrospettiva[570]. Anche nell'ipotesi in cui sia possibile ricostruire *ex post* la concatenazione causale del singolo fatto, magari grazie all'ausilio di sistemi di *white box*, ugualmente non sarà facile trarre una regolarità causale, proprio perché i fattori in base a cui decide l'algoritmo mutano sempre in base al contesto ed all'ambiente in cui opera[571].

Ogni algoritmo "vive" di una interazione con il contesto ambientale e con gli altri soggetti intelligenti che lo influenzano nelle decisioni. Il suo "comportamento" non è, dunque, riconducibile esclusivamente al codice prodotto dallo sviluppatore. Pare potersi fare un paragone con la causalità psichica[572] dell'essere umano: se in essa l'evento non è rappresentato da un accadimento naturalistico, ma da una decisione altrui, lo stesso si può dire con riferimento alla decisione algoritmica, che è frutto

---

[570] P.M. ASARO, *Determinism, machine agency, and responsibility*, in *Pol e soc.*, 2014, n. 2, p. 287; già C.E.A. KARNOW, *Liability for Distributed Artificial Intelligence*, cit., p. 188 ss., prima ancora che fossero sviluppate le tecniche di machine learning, individuava nella complessità dei fattori che potevano incidere su un sistema di intelligenza artificiale un ostacolo all'accertamento processuale del nesso di causalità.

[571] J.A. KROLL E AL., *Accountable Algorithms*, in U. Pa. L. Rev., 2017, p. 661: «*Approaches that attempt to review system failures simply by looking at how the output responds to changes in input are limited by either an inability to attribute a cause to those changes or an inability to interpret whether or why a change is significant*»; vd. anche D. LEHR – P. OHM, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, in *University of California Davis Law Review*, 2017, p. 707, che notano come la logica del *ceteris paribus* sia totalmente estranea ai meccanismi decisionali dei sistemi di *machine learning*. Vd. anche D. SELBST – S. BAROCAS, *The Intuitive Appeal of Explainable Machines*, cit., p. 1094, che parlano di "*inscrutability*" con riferimento alla imperscrutabilità causale tipica degli algoritmi di *machine learning*.

[572] F. D'ALESSANDRO, *Spiegazione causale mediante leggi scientifiche*, a dieci anni dalla sentenza Franzese, in *Criminalia*, 2012, p. 353.

dell'esperienza e della grande mole di dati e stimoli tra i quali sarà difficilissimo individuare la *condicio sine qua non*.

D'altronde, la teoria che nega la possibilità di far rientrare la causalità psichica nel paradigma nomologico-deduttivo non contesta che le azioni umane abbiano una causa, ma si limita ad escludere che la causa in questione possa essere descritta tramite una legge di carattere generale[573].

Con riferimento ai prodotti tradizionali, è stato aspramente criticato l'orientamento giurisprudenziale che faceva uso dell'accertamento della causalità *black box* poiché non conforme all'approccio nomologico-deduttivo[574]. Il meccanismo della *black box* impediva agli esperti di individuare al di là di ogni ragionevole dubbio la causalità tra fatto ed evento. Va detto che tutti i prodotti, quando vengono immessi sul mercato, sono *black box systems*. Ad esempio, all'inizio può esserci incertezza sugli effetti lesivi di una sostanza e, in assenza di una legge scientifica di copertura – che, alla luce della sentenza Franzese, potrebbe anche avere una probabilità statistica bassa – il giudice dovrebbe pronunciare l'assoluzione[575]. Solo con lo studio sistematico degli effetti della sostanza sui consumatori, gli scienziati potrebbero formulare una legge scientifica statistica che ricolleghi l'esalazione della sostanza all'accadimento di eventi lesivi.

Gli eventi lesivi cagionati dai sistemi di IA rischiano di mettere in crisi il diritto penale tradizionale in quanto fondato sul caso concreto e non su fenomeni sistemici caratterizzati da intersecazioni causali imperscrutabili. La principale differenza con la fenomenologia classica della società del rischio consiste nel fatto che il problema

[573] In questi termini, M. RONCO, *Le interazioni psichiche nel diritto penale: in particolare sul concorso psichico*, cit., p. 835.
[574] F. STELLA, *Giustizia e modernità*, cit., p. 233 ss.; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 199 ss.; L. STORTONI, *Angoscia tecnologica ed esorcismo penale*, in *Riv. it. dir. proc. pen*., 2004, fasc. 1, p. 76-77; C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1240; C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 22.
[575] C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 254.

dell'accertamento del nesso di causalità tra fatto e danno algoritmico sconta la discrepanza tra ragionamento causale tipico del modello di imputazione penalistico e *decision making* del *deep learning* che, invece, è fondato sulla probabilità.

Pertanto, o si constata l'incompatibilità tra accertamento causale ed approccio probabilistico dei sistemi di ML e si rinuncia di imputare al produttore reati di evento, o si tenta di rendere più mobili le categorie della causalità al fine di facilitarne l'accertamento processuale. Ma, chiaramente, ciò contrasterebbe con il principio di legalità.

La terza via, ossequiante del principio di precauzione, è quella di vietare la commercializzazione dei sistemi di IA dei quali non sia possibile ricostruire, in via generale, i nessi causa-effetto. Non sembra, tuttavia, che il legislatore europeo si stia muovendo in questa direzione, dal momento che l'*AI Act* – pur richiedendo il rispetto di alcuni, vaghi, requisiti di trasparenza – consente la messa in commercio di sistemi di *deep learning*, per loro natura caratterizzati da "inscrutabilità".

A nostro giudizio, posto che vogliamo delineare un diritto penale dell'IA costituzionale, ci pare che l'unica via percorribile sia la prima, ovverosia quella di constatare l'incompatibilità tra accertamento causale e ragionamento stocastico degli algoritmi, così evitando di imputare l'evento al produttore.

È chiaro che quanto si è precedentemente asserito abbia valore soltanto *rebus sic stantibus*. Ciò significa che, se in un futuro non troppo prossimo si perfezionassero le tecnologie di *explainability* ed *interpretability*, allora si potrebbe addivenire ad una generalizzazione scientifica del rapporto tra *input* ed *output*.

In una simile evenienza, potrebbe sorgere la necessità di abbandonare il modello della sussunzione sotto leggi scientifiche, ma sempre rispettando i principi

costituzionali del diritto penale. D'altro canto, lo scopo dell'accertamento del nesso causale nel processo penale è ricostruire *ex post* un evento singolare e particolare, non già elaborare modelli e leggi generali come nelle scienze naturali[576].

Nel momento in cui si possa risalire a detto evento al di là di ogni ragionevole dubbio, tramite gli appena citati strumenti computazionali di ricostruzione *ex post* della causalità individuale, si potrebbe ritenere provata l'esistenza di una condicio sine qua non *ex* art. 40, comma 2, c.p.

È stata proprio la sentenza Franzese del 2002 a segnare un primo passo per la valorizzazione del metodo induttivo, disponendo che anche le leggi di copertura di incidenza statistica medio-bassa possano essere usate per l'accertamento del nesso di causalità, a patto che non emerga nel processo una spiegazione causale diversa e vi sia <<la sicura non incidenza nel caso di specie di altri fattori interagenti in via alternativa>>.

D'altra parte, in dottrina vi è chi manifesta apertura verso l'uso di modelli induttivi nella spiegazione causale, anche in contesti di assenza di generalizzazioni scientifiche[577]. Ci pare che questa prospettiva possa essere funzionale nel settore dell'IA, laddove è per natura

[576] M. RONCO, *Aporie scientiste e certezza logico razionale: note in tema di nesso causale*, cit., p. 248, al quale si rinvia per gli opportuni riferimenti bibliografici sulla distinzione tra accertamento causale generale, tipico delle scienze naturali, e accertamento individuale, tipico del diritto.

[577] M. RONCO, *Aporie scientiste e certezza logico razionale: note in tema di nesso causale*, cit., passim; R. BLAIOTTA, *Incidenti da aviolancio, problemi causali nuovi, attività rischiose*, cit., p. 766-767; ID*., Causalità giuridica*, cit., p. 397: «sebbene non accada frequentemente, un'indagine di tipo induttivo, alimentata dalle peculiarità particolarmente caratteristiche e chiare dei dati empirici di una determinata vicenda, può in alcuni casi condurre alla razionale spiegazione dell'evento; anche se non si dispone di informazioni scientifiche ad hoc»; contra R. BARTOLI, *Il problema della causalità penale. Dai modelli unitari al modello differenziato*, Giappichelli, 2010, p. 44-45, che sostiene che la spiegazione causale di un fatto non possa essere fornita in assenza di un sapere nomologico-deduttivo; critico anche A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, cit., p. 413, che coglie una tendenza "anti-scientista" in queste tesi, così come in alcuni sviluppi giurisprudenziali del metodo Franzese.

impossibile individuare regolarità causali nel funzionamento algoritmico.

Pertanto, pure in assenza di generalizzazioni preesistenti, l'accertamento causale può portare ad avere risultati di alta credibilità logico-razionale, quando i fatti siano idonei a fornire una spiegazione convincente e ad escludere la possibilità di decorsi causali alternativi. Anche senza leggi di copertura, la scienza deve essere protagonista nelle aule dei Tribunali poiché soltanto la capacità tecnica di interpretare i dati informatici e la conoscenza del funzionamento dei sistemi di ML possono aiutare a delineare le concause nel caso concreto.

Ovviamente, in assenza di regole consolidate, l'indagine sulla causalità individuale dovrà necessariamente essere basata su elementi probatori solidi.

4.4. Chi è responsabile dei danni cagionati dall'IA? Un *too many hands problem*.

Quando si è fatto riferimento al produttore, non intendevamo esclusivamente il produttore-persona fisica ma anche e soprattutto il produttore-ente. Identificare la singola persona fisica a cui attribuire la responsabilità penale rispetto ad un evento lesivo, all'interno delle società complesse, è stato sempre un problema della responsabilità per danno da prodotto[578] e della criminalità d'impresa[579].

---

[578] C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1241-1242; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 303 ss.; ID., *La responsabilità del produttore: avamposto o Sackgasse del diritto penale*?, cit., p. 370; A. BERNARDI, *La responsabilità da prodotto nel sistema italiano*, cit., p. 25; A. DI MARTINO, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, cit., pp. 464 ss.; F. CENTONZE, *La normalità dei disastri tecnologici*, cit., p. 267 ss.; nello specifico settore della responsabilità per danno da prodotto alimentare vd. L. CORNACCHIA, *Posizioni di garanzia nell'ambito della sicurezza agroalimentare*, in L. Foffani – A. Doval Pais – D. Castronuovo (a cura di), *La sicurezza agroalimentare nella prospettiva europea*, Giuffrè, 2014, p. 457 ss. e, più di recente, V. MONGILLO, *(voce) Industria alimentare e responsabilità colposa*, in *M. Donini (diretto da) Reato colposo*, Enc. dir. – I Tematici, cit., p. 704 ss.

[579] A. ALESSANDRI, *Parte generale*, in C. Pedrazzi e al. (a cura di), Manuale *di diritto penale dell'impresa*, Monduzzi, II ed., 2000, p. 59; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., p. 515 ss.; V. ATTILI, *L'agente-modello "nell'era della*

Plurimi centri decisionali sono coinvolti: nella realtà d'impresa, nessun individuo governa da solo il processo decisionale. Le funzioni decisionali sono frazionate ed investono centri dotati di facoltà di delega e sub-delega. Per questo motivo, si parla di *too many hands problem*[580].

Nonostante la presenza di più persone coinvolte, la responsabilità del produttore nulla ha a che vedere con l'autorìa e la partecipazione che sono il fulcro del concorso di persone nel reato[581]. Piuttosto, trattasi di forme di esecuzione frazionata in cui l'azione tipica è realizzata <<più per sovrapposizione autonoma di decisioni che per comune concerto>>[582].

Dunque, rileva l'agente modello non del singolo individuo, bensì dell'organizzazione: destinatario delle norme cautelari per la gestione del rischio è la persona giuridica, essendo quest'ultima l'unico soggetto in grado di avere la capacità organizzativa ed economica per riconoscere i rischi e minimizzarli adeguandosi alle conoscenze tecnologiche più recenti[583].

Lo scollamento tra struttura costituzionale dell'illecito penale e fenomeno criminale rischia di rendere farraginosa l'attribuzione della responsabilità penale per gli eventi lesivi cagionati dalle scelte aziendali. Ne consegue che la ricerca del soggetto rimproverabile dentro le società si

---

complessità": tramonto, eclissi o trasfigurazione?, in *Riv. it. dir. e proc. pen*., 2006, fasc. 4, p. 1247 ss.

[580] La prima tematizzazione del problema si deve, nella letteratura anglosassone, a D. THOMPSON, *The moral responsibility of public officials: the problem of many hands*, in *American Political Science Review*, 1980, p. 905 ss.; il dibattito è stato poi portato avanti soprattutto attraverso i lavori di K.M. MCGRAW, *Avoiding blame: an experimental investigation of political excuses and justifications*, in *British Journal of Political Science*, 1990, p. 119 ss.; M. BOVENS, *The quest for responsibility: accountability and citizenship in complex organisations*, Cambridge University Press, 1998; R. MASTOP, *Characterising responsibility in organisational structures: The problem of many hands*, in G. Governatore – G. Sartor (eds.), *Deontic logic in computer science*, Springer, 2010, p. 274 ss.

[581] C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1241-1242.

[582] C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1242.

[583]G. MARINUCCI*, Innovazioni tecnologiche e scoperte scientifiche: costi e tempi di adeguamento delle regole di diligenza*, in *Riv. it. dir. proc. pen*., 2005, fasc. 1, p. 56; vd. anche G. DE VERO, *Il nesso causale e il diritto penale del rischio*, cit., p. 695 ss.

trasformi nella ricerca del capro espiatorio[584]. A nostro giudizio, per un modello di responsabilità in senso lato collettiva, come quella della responsabilità per danno da prodotto, si sarebbe potuto introdurre una responsabilità amministrativa dell'ente, in modo tale da poter colpire il centro immediato degli interessi.

Nel novero dei reati-presupposto di cui al d.lgs. n. 231/2001, mancano le fattispecie di pericolo o danno relative alla sicurezza dei prodotti. Se, in futuro, si dovessero introdurre, ai sensi dell'art. 8 dell'appena menzionato corpo normativo, non impedirebbe la configurazione della responsabilità in capo all'ente, in quanto essa sussiste <<*anche quando l'autore del reato non è stato identificato o non è imputabile*>>.

La fattispecie individuata dall'art. 8 parrebbe creare un *tertium genus*, ulteriore rispetto a quello di cui agli artt. 6 e 7, incentrato sulla omessa auto-organizzazione da parte dell'ente ovvero su un inefficace assetto organizzativo nell'ambito delle singole procedure di contenimento del rischio di reato[585]. La responsabilità dovrà comunque essere conforme al principio di colpevolezza, non potendosi configurare una responsabilità dell'ente per il sol fatto di non essere stati in grado di individuare l'autore persona fisica[586]. Dovrà sempre accertarsi una colpa d'organizzazione, ossia la mancata predisposizione di <<un insieme di accorgimenti preventivi idonei ad evitare la commissione di reati del tipo di quello realizzato>>[587].

È proprio la individuazione delle persone responsabili all'interno dell'organigramma societario a costituire uno degli aspetti maggiormente problematici della disciplina in parola. In diverse occasioni, si è affermato un criterio di

[584] Su tale rischio vd. G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 57

[585] Così C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1756.

[586] C. MINELLI, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 6

[587] Cass., sez. VI, 18 febbraio 2010, n. 27735, Scarafia; vd. anche Cass., sez. un., 24 aprile 2014, n. 38343, Espenhahn; Cass., sez. IV, 8 gennaio 2021, 32899, Castaldo

ascrizione della responsabilità derivato dalla mera ripartizione dei ruoli prestabilita dal diritto societario[588]. Una sorta di colpa da posizione che "cozza" con l'attenzione alla misura soggettiva della colpa da affiancarsi alla concezione normativa[589]. Al principio che collega una posizione di garanzia in capo all'amministratore del ramo d'azienda coinvolto nell'evento lesivo, partendo dalla violazione per arrivare alla funzione nell'ambito della quale essa è maturata per poi individuare il soggetto tenuto a governare la stessa funzione, la giurisprudenza in tema di danno da prodotto ha spesso sostenuto la presunta competenza totale del management della società per gli eventuali danni derivanti dal prodotto[590].

Negli ultimi anni, infatti, si è assistito alla ricostruzione dell'obbligo di controllo sulla base della predeterminazione delle aree di competenza nella gestione del rischio[591]. Noi crediamo, però, che tale modo di operare non consideri che ormai le società non operano per compartimenti stagni e che, quindi, non possono valorizzarsi le delimitazioni formalistiche delle aree di competenza, abbandonando la sostanziale attribuzione dei poteri gestionali del rischio[592].

Equiparare la posizione di garanzia alla sfera di responsabilità serve, ancora una volta, per semplificare

---

[588] C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale*?, cit., p. 361; ID. *Danno da prodotto e responsabilità penale*, cit., p. 234, con specifico riferimento al caso *Lederspray*; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse,* cit., passim.

[589] Vd. per tutti M. DONINI, *Teoria del reato. Una introduzione*, cit., p. 348 ss.; D. CASTRONOVO, *La colpa penale*, cit.; ID., *La colpa "penale". Misura soggettiva e colpa grave*, cit.; A. CANEPA, *L'imputazione soggettiva della colpa*, cit.; M. GROTTO, *Principio di colpevolezza, rimproverabilità soggettiva e colpa specifica*, cit.; A.R. DI LANDRO, *Dalle linee guida e dai protocolli all'individualizzazione della colpa penale nel settore sanitario*, cit.

[590] In questo senso C. PIERGALLINI, *La responsabilità del produttore: avamposto o Sackgasse del diritto penale*?, cit., p. 361; C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1241-1242.

[591] L. CORNACCHIA, *Posizioni di garanzia nell'ambito della sicurezza agroalimentare*, cit., 460 ss.; A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., p. 533 ss., che fa riferimento soprattutto alla sentenza della Cassazione sul caso *Thyssenkrupp* (Cass., sez. un., 24 aprile 2014, n. 38343).

[592] A. GARGANI, *Posizioni di garanzia nelle organizzazioni complesse*, cit., p. 536-7.

l'ascrizione della responsabilità nelle organizzazioni complesse, a costo di creare un'insanabile rottura con il principio di legalità.

4.4.1. L'*AI Act* e le aree di competenza nella gestione del rischio.

Nell'ambito della produzione dei sistemi di IA, tutti i problemi appena citati sembrano acuirsi[593]. Il prodotto artificiale, infatti, può essere un artefatto molto complesso dotato di componenti diverse, spesso prodotte da tante aziende differenti. Ci sono *too many hands* coinvolte nel processo di produzione. Può anche darsi che il prodotto abbia un *software open source* che possa essere modificato da chiunque liberamente[594].

La tecnologia *black box* tipica degli algoritmi di ML rende molto complessa la individuazione del processo causale all'origine dell'evento lesivo. Ma, anche fosse agevole in futuro individuare il fatto causativo dell'evento, rimarrebbe il problema di comprendere, all'interno degli organismi societari, quale sia la persona a cui ascrivere l'evento. Il rischio, che per la verità è già presente con riferimento a tutti i fenomeni della criminalità d'impresa, è che la responsabilità venga individuata sulla base della mera posizione che la persona ricopre dentro la società[595].

L'*AI Act*, al riguardo, tenta di fornire un impulso alla tipizzazione di "aree di competenza" nella gestione del

[593] In argomento vd. B. PANATTONI, *Intelligenza artificiale*, cit., p. 338; S. BECK, *Intelligent agents and criminal law*, cit., p. 140; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1754; M.B. MAGRO, *Decisione umana e decisione robotica*, cit., p. 3; F.P. HUBBARD, *Sophisticated Robots: Balancing Liability, Regulation, and Innovation*, cit., p. 1853; ; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 104 ss.; S. AIRES DE SOUSA, *"Não fui eu, foi a máquina": teoria do crime, responsabilidade e inteligência artificial*, cit., p. 70-71; L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma,* cit., p. 828-829, che delinea alcuni dei centri di responsabilità che potrebbero venire in rilievo in relazione alla produzione di *autonomous vehicles.*

[594] R. ABBOTT – A. SARCH, *Punishing Artificial Intelligence*, cit., p. 334; F. CONSULICH, *Flash offenders*, cit., p. 1041.

[595] F.C. LA VATTIATA, *La responsabilità penale per danni da intelligenza artificiale alla prova del processo*, cit., p. 705.

rischio. Esso individua "sfere di responsabilità" in capo al fornitore, al fabbricante, ai rappresentanti autorizzati dei fornitori non aventi sede nell'UE se è impossibile identificare un importatore, agli importatori di sistemi extra-UE, ai distributori ed agli utenti[596].

Il fornitore costituisce il fulcro del sistema della gestione del rischio, essendo definito come <<una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o che fa sviluppare un sistema di IA al fine di immetterlo sul mercato o metterlo in servizio con il proprio nome o marchio, a titolo oneroso o gratuito>>. Sotto la sua responsabilità, vengono espletate le più importanti attività di valutazione della sicurezza del prodotto, tra le quali anche il monitoraggio *post market*. Il fornitore ha l'obbligo di verificare la conformità dei sistemi di IA ai requisiti di sicurezza previsti nel Regolamento, l'obbligo di redigere la documentazione tecnica per i sistemi ad alto rischio, l'obbligo di conservare i *logs* generati automaticamente in caso di incidenti.

Anche il fabbricante costituisce un perno poiché quest'ultimo <<si assume la responsabilità della conformità del sistema di IA al presente regolamento ed ha, per quanto riguarda il sistema di IA, gli stessi obblighi imposti dal presente regolamento al fornitore>>. L'utilizzo della locuzione "persona giuridica" nella definizione di fornitore, rappresentante, distributore, utente, etc., la si deve al fatto che la normativa non è costruita per soddisfare esigenze di personalizzazione proprie del diritto penale, ma ha come obiettivo le generale *risk governance*. Anche nell'ipotesi in cui la gestione del rischio nella singola società sia attribuita ad un organo collegiale,

[596] Li definisce veri e propri "obblighi di controllo" F.C. LA VATTIATA, *Brevi note "a caldo" sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale*, in *Dir. pen. uomo*, 30 giugno 2021, p. 13. A parere di quest'ultimo, l'entrata in vigore del regolamento – immediatamente applicabile negli ordinamenti interni – potrebbe dunque determinare l'immediato sorgere di obblighi di impedimento ai sensi dell'art. 40, co. 2, c.p.

ai fini penalistici occorrerà individuare il soggetto personalmente responsabile.

Anche nel settore dell'IA, dunque, si tende ad individuare prima le aree di competenza per la gestione dei rischi, senza individuare precisi obblighi di controllo *ex* art. 40, comma 2, c.p. A nostro giudizio, nella individuazione concreta delle posizioni giuridiche, i principi costituzionali di legalità e colpevolezza dovrebbero imporre l'identificazione del soggetto personalmente incaricato dell'obbligo di impedire l'evento lesivo, senza che la valutazione giurisprudenziale possa limitarsi alla qualifica formale[597].

4.5. La difficile dimostrazione della colpa del produttore.

Anche qualora si riuscisse a dimostrare che sussiste un fatto tipico eziologicamente ricollegabile alle persone dietro o vicino la macchina, resterebbe il problema della prova dell'elemento soggettivo. Soprattutto, sul piano dell'accertamento della colpa, si coglie la maggiore problematicità della responsabilità per danno provocato da prodotti intelligenti[598].

L'imputazione colposa porta con sé un *deficit* di determinatezza atavico[599]. Nel campo della responsabilità per danno da prodotto, questa assenza di determinatezza è divenuta, per mano dei giudici, estrema flessibilizzazione delle categorie colpose, con il conseguente progressivo affievolimento della rigorosità che il principio di legalità richiederebbe per dimostrare le regole cautelari violate, per dimostrare l'esistenza di una

[597] Per problematiche simili nell'interpretazione del reg. CE 178/2002 in materia di sicurezza alimentare vd. V. MONGILLO*, (voce) Industria alimentare e responsabilità colposa*, cit., p. 710.
[598] In questa sede ci occuperemo del rimprovero colposo, trattandosi della forma di colpevolezza tipica del fenomeno del danno da prodotto. In giurisprudenza, tuttavia, l'imputazione oscilla tra colpa cosciente e dolo eventuale; vd. C. PIERGALLINI, *La responsabilità del produttore: una nuova frontiera del diritto penale*?, cit., p. 1127-1128; ID., *Danno da prodotto e responsabilità penale*, cit., p. 360 ss.
[599] P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 50 ss.; R. BARTOLI, *(voce) Fonti della colpa*, cit., p. 536 ss.

violazione oggettiva delle stesse e per dimostrare l'effettiva esigibilità dell'osservanza di dette regole da parte del produttore[600].

Il *background* di tale flessibilizzazione giurisprudenziale è rappresentato dal conflitto costante tra l'aspirazione solidaristica ad istanze di tutela dei beni fondamentali della giurisprudenza ed il ben più marcato garantismo figlio della concezione liberale del diritto penale anelato dalla dottrina[601].

Lo sviluppo di sistemi intrinsecamente dotati di una imprevedibilità prevedibile[602] pare mettere in discussione la possibilità di muovere un rimprovero colposo al produttore se si pensa che, secondo il criterio della copertura del rischio, affinchè l'evento lesivo sia rimproverabile all'agente, esso deve costituire la tipica realizzazione del rischio che la norma cautelare voleva evitare.

È proprio la crisi della logica del modello nomologico a condizionare l'intera struttura del reato e, nel dettaglio, l'accertamento della colpa del produttore. Le regole cautelari sono regole di esperienza già esistenti al momento della condotta, venute fuori dal consolidamento di giudizi causali di riconoscibilità di precisi e determinati

[600] A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, in *Riv. trim. dir. pen. ec.,* p. 30.

[601] A. PERIN, *La crisi del "modello nomologico" fra spiegazione e prevedibilità dell'evento nel diritto penale*, cit., p. 1387; F. STELLA, *Giustizia e modernità,* cit., passim; cfr. L. STORTONI, *Angoscia tecnologica ed esorcismo penale*, cit., p. 80, che ritiene che l'obiettivo solidaristico sia in ogni caso destinato a fallire, data l'intrinseca inettitudine del diritto penale a rispondere ad istanze di tutela nei confronti di fenomeni diffusivi e sistemici, tipici della società del rischio.

[602] C. PIERGALLINI, *Intelligenza artificiale*, 1762; vd. anche M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1208-1209; S. BECK, *Intelligent agents and criminal law*, cit., p. 139; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 7: «*the behavior of an autonomous agent is predictably unpredictable*»; sostanzialmente anche C. IAGNEMMA, *I 'robot medici': profili problematici in tema di alleanza terapeutica e di responsabilità penale*, cit., p. 449.

rischi e di prevedibilità ed evitabilità di concreti eventi determinati[603].

Qui noi ci troviamo in un contesto ad altissima incertezza scientifica. E, se sussiste ugualmente uno spazio di operatività e conformazione di regole cautelari anche in contesti di incertezza scientifica, la difficoltà propria dei sistemi di IA di identificare generalizzazioni causali in relazione al loro comportamento rende strutturale questa condizione di instabilità delle cautele.

Secondo noi, è realmente impossibile, in questo settore, individuare cautele, rispettate le quali, si eviti del tutto l'evento lesivo. Al massimo, in questo contesto, si potrà ragionare su logiche di individuazione del rischio consentito e di riduzione del rischio tramite misure che influiscano sull'esito del *decision making* algoritmico.

4.5.1. Problemi di colpa specifica: l'individuazione delle regole cautelari.

Ai fini dell'accertamento colposo, per prima cosa, occorre individuare la regola cautelare preesistente che si assume violata[604]. Si assiste da tempo ed in diversi settori a fenomeni di iperregolazione multilivello[605], tramite la formalizzazione di regole preventive, combinata a forme di proceduralizzazione dei processi produttivi e di commercializzazione[606].

Nel settore della sicurezza dei prodotti, vi sono due categorie di normazione: un modello di etero-regolazione

[603] Così, per tutti, D. CASTRONUOVO*, (voce) Colpa penale*, in *M. Donini (diretto da) Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 219.
[604] Sulla necessaria preesistenza della regola cautelare si vd. per tutti, G. FORTI, *Colpa ed evento*, cit., p. 683.
[605] L'espressione è di V. MONGILLO, *(voce) Industria alimentare e responsabilità colposa*, cit., p. 713, che si riferisce tuttavia alla sola normativa in materia di sicurezza dei prodotti alimentari.
[606] D. CASTRONUOVO*, La colpa penale,* cit. p. 307; ID., *Responsabilità da prodotto e struttura del fatto colposo*, cit., p. 322; sul crescente processo di positivizzazione della colpa vd. G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., p. 2.

ed uno di autonormazione[607]. Entrambe le forme si caratterizzano per un forte tecnicismo: solo centri specializzati sul tema e sul settore di riferimento sono in grado di elaborare regole cautelari talmente settoriali e tecniche. Tale fenomeno, di decentramento dei luoghi di produzione normativa, trova il terreno più fertile proprio nel reato colposo. Ciò in quanto, il reato colposo è una fattispecie "aperta" ontologicamente[608], nonché una fattispecie che deve fare i conti con il nuovo contesto globalizzato del rischio tecnologico che ha creato innumerevoli fonti di pericolo.

Le norme cautelari scritte servono ad indirizzare la condotta dei consociati e, soprattutto, ad evitare il verificarsi dell'evento dannoso. Inoltre, servono per la delimitazione dell'area di rischio consentito. L'obiettivo è proprio quello di integrare la fattispecie incriminatrice e non di fungere esclusivamente da parametro di giudizio *ex post* del giudice[609]. Nel momento in cui si scrive, non risultano nello specifico settore dell'IA regole cautelari codificate a cui il produttore possa fare riferimento[610].

L'approvazione ed entrata in vigore dell'*AI Act* costituisce, ovviamente, il principale riferimento per le regole cautelari che andranno sviluppate. Secondo alcuni Autori, lo stesso testo dell'*AI Act* può fungere da regola

[607] D. CASTRONUOVO, *La colpa penale*, cit. p. 307; ID., *Responsabilità da prodotto e struttura del fatto colposo*, cit., p. 322; in generale sul rilievo dei fenomeni di autonormazione nel diritto penale si rinvia a C. PIERGALLINI, *Ius sibi imponere: controllo penale mediante autonormazione*?, in C.E. Paliero e al. (a cura di), *La crisi della legalità. Il «sistema vivente» delle fonti legali, Edizioni Scientifiche Italiane*, 2016, p. 117 ss.; S.B. TAVERRITI, *Autonormazione e prospettive autopoietiche nella gestione della penalità,* in *Riv. it. dir. proc. pen*., 2020, p. 1931 ss.; D. BIANCHI, *Appunti per una teoria dell'autonormazione*, in *Riv. it. dir. proc. pen*., 2019, p. 1477 ss. Critico nei confronti della predisposizione di norme cautelari da parte degli stessi destinatari delle condotte doverose è F. STELLA*, La costruzione giuridica della scienza: sicurezza e salute nei luoghi di lavoro,* in *Riv. it. dir. proc. pen*., 2003, p. 55 ss.

[608] Sul carattere "aperto" della fattispecie colposa vd. C. PIERGALLINI, *(voce) Colpa (diritto penale*), cit., p. 236; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie*", cit., p. 54 ss.; R. BARTOLI, (voce) Fonti della colpa, cit., p. 527.

[609] F. GIUNTA, *La normatività della colpa*, cit., p. 90; V. TORRE, *Produttori di alimenti o produttori di norme*?, cit., p. 509.

[610] I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 108, nota 79; A. MORAITI*, AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, cit., p. 113.

cautelare a cui il giudice potrà far ricorso nell'accertamento dell'elemento normativo della colpa specifica del produttore, in caso di eventi lesivi per il consumatore[611].

L'emanazione di standard tecnici specifici per ogni settore in cui i dispositivi intelligenti saranno messi in commercio è auspicata dalla maggioranza della dottrina, come misura di attenuazione dell'indeterminatezza ed imprecisione della fattispecie colposa[612]. Pare che il nostro Legislatore stia andando e voglia andare in questa direzione. Infatti, con la Legge n. 132/2025 ("Legge italiana sull'IA"), oltre ad aver recepito i fondamentali principi dell'*AI Act*, ha espresso la necessità di prevedere norme cautelari per ogni specifico settore di interesse dei sistemi di IA.

Un guadagno in termini di precisione e determinatezza è possibile solo in caso di positivizzazione di regole di comportamento rigide ed esaustive poiché le regole elastiche, ancorchè scritte, si fondano sullo stesso meccanismo ricostruttivo della colpa generica, ovverosia sull'agente modello[613]. Per tale ragione, è difficile asserire che l'AI Act sia una soluzione soddisfacente, visto che sono molto vaghi ed indeterminati i requisiti minimi per l'accesso dei sistemi di IA sul mercato europeo. Trattasi di norme elastiche. Quando una norma, ad esempio, fa riferimento a categorie come "per quanto possibile" lo fa per rendere elastico il giudizio colposo e fondarlo sul parametro dell'*homo eiusdem professionis et condicionis*. Sono descrizioni di situazioni di fatto di pericolo, senza contenere norme prescrittive comportamentali specifiche,

[611] Sul ruolo della normativa europea nella definizione delle norme cautelari (e sui relativi profili problematici) si vd. l'accurata analisi di C. BUZIO, *Regole cautelari eurounitarie e colpa penale. Spunti per una riflessione*, in *Discrimen*, 17 aprile 2020.

[612] In generale, sul ruolo delle norme cautelari codificate nell'attenuazione del deficit di determinatezza della fattispecie colposa vd. per tutti G. MARINUCCI – E. DOLCINI, *Corso di diritto penale*, Giuffrè, III ed., 2001, p. 41 ss.

[613] D. CASTRONUOVO, *Responsabilità da prodotto e struttura del fatto colposo*, cit., p. 328; F. PALAZZO, *Morti da amianto e colpa penale*, in *Dir. pen. proc*., 2011, n. 2, p. 189

risolvendosi in superflue positivizzazioni della colpa generica.

E allora, saranno fondamentali regole cautelari specifiche per ogni settore. La regola cautelare la cui violazione potrà fondare l'elemento soggettivo del reato colposo deve essere proprio quella finalizzata a prevenire proprio il tipo di evento che si è in concreto verificato. Ad esempio, non può invocarsi la generica regola che impone di adottare misure adeguate di *cybersecurity* in caso di robot sanitario che provochi una lesione personale ad un paziente.

L'assenza di generalizzazioni scientifiche sui nessi eziologici propri del *decision making* algoritmico potrebbe indurci a ritenere impossibile la redazione di *standard* tecnici funzionali al contenimento di una certa tipologia di danni[614]. A dire il vero, anche se non sappiamo le precise modalità con cui detti sistemi decidono, sappiamo che gli errori dei sistemi di ML possono concernere un errore nel codice algoritmico; avere a che fare con l'alimentazione dell'algoritmo con dati errati o non sufficienti, avere a che fare con un addestramento dell'algoritmo inadeguato[615].

A nostro giudizio, le regole cautelari finalizzate all'impedimento di decisioni errate devono concernere la raccolta e la gestione dei dati[616], l'addestramento del sistema di ML e la fase di sperimentazione. Proprio per questo, lo stesso *AI Act* ha previsto spazi di sperimentazione normativa per l'IA con lo scopo di fornire

[614] Lo notava, in relazione ai prodotti tradizionali, C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit., p. 243.

[615] Va tuttavia nuovamente sottolineato che eventi lesivi potranno derivare anche da sistemi di i.a. pienamente conformi alle cautele esistenti, dal momento che i meccanismi probabilistici che ne costituiscono la base di funzionamento sono, per loro natura, fallibili.

[616] In quest'ottica, è di estremo interesse la Proposta di *Data Governance Act* (*Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act),* 25-11-2020, COM(2020) 767 final), che ha l'obiettivo di consentire un maggiore flusso di dati (personali e non) dal settore pubblico al settore privato e viceversa. In argomento vd. G. RESTA, *Cosa c'è di 'europeo' nella Proposta di Regolamento UE sull'intelligenza artificiale*?, in *Dir. inf*., 2022, fasc. 2, p. 335 ss., al quale si rinvia anche per una disamina del rapporto tra tale normativa e le disposizioni previste dal GDPR.

<<un ambiente controllato che facilita lo sviluppo, le prove e la convalida di sistemi di IA innovativi per un periodo limitato prima della loro immissione sul mercato o della loro messa in servizio conformemente a un piano specifico>> (art. 53 *AI Act*).

Poi è chiaro che, in base al settore in cui il sistema di IA insiste, devono applicarsi le regole cautelari, qualora esistenti, nonché le norme cautelari dettate dal codice del consumo in generale per la sicurezza di tutti i prodotti. Ogni produttore ha il dovere di immettere nel mercato prodotti sicuri, definendo "sicuro" qualsiasi prodotto che<<*in condizioni di uso normali o ragionevolmente prevedibili, compresa la durata e, se del caso, la messa in servizio, l'installazione e la manutenzione, non presenti alcun rischio oppure presenti unicamente rischi minimi, compatibili con l'impiego del prodotto e considerati accettabili nell'osservanza di un livello elevato di tutela della salute e della sicurezza delle persone», in funzione di vari elementi, tra i quali la norma considera, in via meramente esemplificativa, le sue «caratteristiche» (n. 1), il suo eventuale «effetto» su altri prodotti (n. 2), la sua «presentazione» (n. 3) e le «categorie di consumatori» cui esso è destinato (n. 4), con particolare riguardo a quelle maggiormente vulnerabili (bambini e anziani)*>>[617].

Inoltre, il produttore detiene, nei confronti dell'utilizzatore, obblighi informativi rispetto alla valutazione ed alla prevenzione dei rischi del prodotto che non siano immediatamente percettibili. Trattasi di norme cautelari elastiche, a nostro giudizio. Esse possono e debbono essere integrate dagli *standard* tecnici dello specifico settore produttivo, nonché da eventuali atti emanati da eventuali autorità amministrative.

Solo in via residuale, il produttore implementa modelli di autonormazione ed autocontrollo. Il produttore che si discosta da questa *congerie* di cautele risponderà

[617] Art.a 103, comma 1, lett. A), cod. cons.

certamente penalmente ai sensi delle contravvenzioni di cui all'art. 112 del codice del consumo e sarà chiamato a rispondere degli eventi di lesione o morte derivanti dal malfunzionamento del prodotto. Restano nell'area del rischio consentito, invece, i rischi residuali.

4.5.2. A proposito di rischio consentito: qual è il rapporto tra regole cautelari scritte e non scritte?

Nel caso in cui il dispositivo intelligente abbia provocato un danno nonostante l'osservanza pedissequa da parte del produttore delle norme cautelari codificate, che succede? Rispettare le cautele positive non sempre conduce ad impedire nel concreto la verificazione di eventi lesivi: già attribuendo capacità decisionale autonoma ad un algoritmo si accetta la verificazione di decorsi eziologici dannosi.

Il rispetto degli standard tecnici basterà per mandare esente da colpe il produttore o il giudice dovrà (anche) valutare la conformità della condotta del produttore rispetto a quella di un agente modello? È un problema vecchio del diritto penale che si ripropone in questo contesto del tutto nuovo.

In dottrina, tradizionalmente, si pensa che il mero rispetto delle regole cautelari positive non sia sufficiente ad esonerare l'agente dal tenere un comportamento diligente proprio dell'agente modello[618]. Nell'ipotesi di attività produttive, il tutto si acuisce poiché le norme cautelari, oltre alla funzione di prevenzione dell'evento lesivo, svolgono funzione di garanzia ed orientamento nei confronti dell'agente, in contesti particolarmente rischiosi e ad alta complessità tecnico-normativa[619].

[618] G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., p. 236; F. MANTOVANI, *(voce) Colpa, in Dig. pen.,* II, Utet, 1988, p. 309; G. FORTI, *Colpa ed evento*, cit., 315 ss.; R. BARTOLI, *(voce) Fonti della colpa*, cit., p. 529-530; V. MILITELLO, *Rischio e responsabilità penale*, Giuffrè, 1988, p. 148.

[619] Nello specifico settore dei sistemi di i.a., vd. S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 719.

Si deve, arrivati qui, ragionare circa l'esistenza di un *erlaubtes risiko*, ovverosia di un'area di rischio consentito[620]. Trattasi di un tema da sempre rilevante per la dottrina che ha affrontato il tema della colpa. Ancora una volta, problemi vecchi che si acuiscono in questo nuovo contesto. Infatti, la teoria del rischio consentito è stata recentemente oggetto di una rinnovata attenzione, proprio con riferimento agli eventi lesivi derivanti dall'IA[621].

[620]La letteratura sul tema è ormai di ampiezza considerevole; vd. G. FORTI, *Colpa ed evento*, cit., p. 250 ss.; ID., *"Accesso alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione*, cit., p. 155 ss.; ID., *La "chiara luce della verità" e l'"ignoranza del pericolo",* cit., p. 573 ss.; C. PIERGALLINI, *Danno da prodotto e responsabilità penale*, cit.; ID., *Il paradigma della colpa nell'età del rischio: prove di resistenza del tipo*, in *Riv. it. dir. proc. pen*., 2005, p. 1670 ss.; V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 55 ss. (che preferisce, tuttavia, l'espressione "rischio adeguato"); F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 227 ss.; F. CONSULICH, *(voce) Rischio consentito*, cit., p. 1102 ss.; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 335 ss.; C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione. La c.d. "flessibilizzazione delle categorie del reato",* in *Criminalia*, 2012, p. 383 ss.; D. CASTRONUOVO, *Principio di precauzione e beni legati alla sicurezza*, in *Dir. pen. cont*., 21 luglio 2011, p. 1 ss.; ID*., Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit.; M. DONINI, *Il volto attuale dell'illecito penale*, cit., p. 119 ss.; A. MASSARO, *Principio di precauzione e diritto penale: nihil novi sub sole?,* cit.; D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit., p. 647 ss.; C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1743 ss.; R. BLAIOTTA, *Incidenti da aviolancio, problemi causali nuovi, attività rischiose,* in *Cass. pen*., 2004, n. 7-8, p. 2377 ss.; E. PENCO, *Limiti-soglia e responsabilità colposa*, cit.

[621] Vd., con vari accenti, S. GLESS – E. SILVERMAN – T. WEIGEND, *If Robots Cause Harm, Who Is to Blame: Self-Driving Cars and Criminal Liability*, cit., pp. 430-431; S. BECK, *Robotics and Criminal Law. Negligence, Diffusion of Liability and Electronic Personhood*, cit., p. 53 ss.; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1750; I. SALVADORI, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 116 ss.; A. FIORELLA, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, cit., p. 656 ss.; V. MANES, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, cit., p. 5; A. CAPPELLINI, *Machina delinquere non potest*?, cit., p. 19; ID., *Profili penalistici delle self-driving cars*, cit. p. 4-5; ID., *Reati colposi e tecnologie dell'intelligenza artificiale*, cit., p. 29 ss.; D. PIVA, *Machina discere, (deinde) delinquere et puniri potest*, cit., p. 681 ss.; S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 718 ss.; G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 73; R. BORGOGNO*, La responsabilità penale nei processi ad elevata automazione*, cit., p. 735; C. MINELLI*, La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, cit., p. 62 ss.; A. MORAITI, *AI Crimes and Misdemeanors: Debating the Boundaries of Criminal Liability and Imputation*, cit., p. 113 ss.; R. BORSARI, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, cit., p. 268; L. D'AMICO, *Intelligenza artificiale e auto a guida autonoma. tra prevenzione primaria, colpa penale e rischio consentito*, cit., p. 593 ss.; J. GÖHLER, *Digital challenges to personal autonomy for German criminal law*, cit., p. 9-10; P.S. GÜNSBERG, *Automated vehicles – is a dilution of human responsibility the answer*?, cit., p. 448 ss.; sostanzialmente anche D. LIMA, *Could AI Agents Be Held Criminally Liable?*, cit., p. 694.

Secondo alcuni Autori, la categoria in parola dovrebbe collocarsi nell'alveo dell'esclusione della tipicità. Mentre, secondo altri, sarebbe elemento di esclusione dell'antigiuridicità o della colpevolezza. Secondo altri ancora, sarebbe un concetto autonomo o meramente descrittivo e pleonastico[622][623]. In linea di massima, la dottrina maggioritaria colloca il concetto di rischio consentito nella trattazione della colpa proprio perché ci fa subito pensare ad attività socialmente utili ma pericolose, laddove il pericolo sopravvive nonostante il rispetto dei prescritti doveri di prudenza.

Per autorevole dottrina, esisterebbero due versioni del concetto in parola. Nella versione, per così dire ampia, si riconducono tutte le attività che, anche se pericolose, sono consentite, a patto che vengano svolte seguendo le adeguate misure cautelari per la minimizzazione del rischio di eventi lesivi. Nella seconda accezione, invece, il concetto di rischio delinea <<un'area (più ristretta e comunque concentrica rispetto alla prima) di condotte pericolose, ammesse dall'ordinamento nonostante che l'adozione di cautele idonee a contrastare i possibili svolgimenti lesivi sia destinata a residuare un certo grado di pericolosità>>[624].

Noi preferiamo questa seconda accezione, che conferisce una valenza autonoma dal punto di vista dogmatico alla nozione di rischio consentito. Il rischio consentito di questa seconda accezione è quello che viene definito rischio consentito "in senso stretto" o "specifico" o, anche, "residuale", intendendosi quel pericolo marginale che le

[622] Su questi aspetti si vedano vd. C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 10; C. EAD., *La legislazione penale tra "diritto penale dell'evento" e "diritto penale del rischio*", cit., 120 (in particolare nota 7); V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 62.

[623] Per una ricostruzione dei diversi orientamenti dogmatici in argomento si rinvia a F. CONSULICH, (*voce) Rischio consentito*, cit., passim.

[624] G. FORTI, Colpa ed evento, cit., p. 456; vd. anche C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione*, cit., p. 389.

misure preventive non sono in grado di disinnescare e che ricade sulla società.

Sarebbero proprio le scelte politiche delle autorità che producono la normativa cautelare a decidere se disinnescare o meno anche quel rischio residuale. Secondo questa accezione, il rischio consentito si innesta sui giudizi di prevedibilità ed evitabilità dell'evento, delineando quelle ipotesi in cui il rispetto di norme cautelari codificate impedisce si possa muovere all'imputato un rimprovero per non aver osservato norme di diligenza, prudenza e perizia non codificate[625].

Se ci pensiamo, sarebbe anche illogico dotare il produttore di un quadro di norme specifiche per il settore in cui commercia e poi punirlo nel caso in cui il pedissequo rispetto delle norme non sia stato sufficiente per impedire gli eventi lesivi. Nella società della post-modernità in cui viviamo, produzione e rischio sono grandezze tra loro direttamente proporzionali. È letteralmente impossibile eliminare del tutto i rischi.

In base a tale consapevolezza, l'agente modello non è chi prevede il rischio e non agisce, ma è chi agisce assicurandosi che la sua condotta sia tale da ridurre al minimo il rischio e, soprattutto, che essa sia tale da rientrare nel rischio consentito. Rischio consentito che, per l'appunto, è un rischio residuale che la società moderna deve accettare[626].

Il rapporto così disegnato tra agente modello e regole codificate sembra l'unico possibile alla luce della concezione normativa della colpa secondo cui <<l'agente modello non è colui che prevede ed evita tutti gli eventi che è possibile prevedere ed evitare, ma colui che prevede

[625] C. PIERGALLINI, *Attività produttive e imputazione per colpa: prove tecniche di "diritto penale del rischio*", in *Riv. it. dir. proc. pen*., 1997, fasc. 4, p. 1492; L. STORTONI, *Angoscia tecnologica ed esorcismo penale*, cit., p. 79; F. PALAZZO, *Morti da amianto e colpa penale*, cit., p. 188

[626] S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale,* cit., p. 358.

ed evita gli eventi che è doveroso prevedere ed evitare>>[627].

La definizione dell'area del rischio consentito è una questione tutta politica di bilanciamento tra esigenze produttive e tutela dei beni giuridici costituzionali che possono essere lesi dallo svolgimento di tali attività[628].

È una scelta politica del legislatore che, sulla base di un bilanciamento costi-benefici, definisce il livello di rischio socialmente accettabile e, quindi, il grado di tutela che si intende assicurare ai beni giuridici potenzialmente interessati dallo svolgimento dell'attività pericolosa. La scelta investe la decisione sulla quota di eventi lesivi che la società è disposta a tollerare in cambio di benefici complessivi derivanti dalla produzione industriale e dal progresso tecnologico. Decidere, infatti, di creare un'area del rischio consentito è una scelta che implica il non esercizio del potere punitivo[629].

Se non si avesse in mente questo, non si comprenderebbe perché la produzione di tabacco sia consentita a discapito di quella dell'amianto, nonostante siano scientificamente certi i danni alla salute cagionati dal fumo. L'istituto del rischio consentito serve proprio per impedire al giudice di addebitare al produttore di sigarette

---

[627] S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 358. Sulla concezione normativa della colpa – ormai dominante nella dogmatica italiana – vd. M. GALLO, (voce) *Colpa penale*, in *Enc. dir.*, VII, 1960, p. 637; G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., spec. p. 113 ss. e 166 ss.; ID., *Il reato come 'azione'. Critica di un dogma*, Giuffrè, 1971, p. 100 e p. 111-121; G. FORTI, *Colpa ed evento*, cit., p. 125 ss.; F. GIUNTA, *Illiceità e colpevolezza nella responsabilità colposa*, cit., p. 169 ss.; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 10 ss.

[628] G. FORTI, Colpa ed evento, cit., p. 255; F. CONSULICH, *(voce) Rischio consentito*, cit., p. 1109; J.M. SILVA SÁNCHEZ, *La expansión del derecho penal*, cit., p. 44; C. PIERGALLINI, *(voce) Colpa (diritto penale*), cit., p. 239; ID., *Danno da prodotto e responsabilità penale*, cit., p. 242; V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 123-124; C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione*, cit., p. 391; F. PALAZZO, *Morti da amianto e colpa penale*, cit., p. 188; E. CORN, *Il principio di precauzione*, cit., p. 158; D. PULITANÒ, *Diritto penale*, cit., p. 308; in generale, sul ruolo di bilanciamento di interessi svolto dalle norme cautelari positive, vd. G. MARINUCCI – E- DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 416.

[629] G. FORTI, *Colpa ed evento*, cit., p. 458; F. CONSULICH, *(voce) Rischio consentito*, cit., p. 1112; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 360.

gli eventi lesivi, rimproverandogli la prevedibilità concreta dei danni, in virtù di una presunta funzione supplente della colpa generica.

Secondo noi, l'unica ipotesi in cui torna a sopravvivere la colpa generica è quella in cui le regole cautelari mostrino segnali univoci di fallimento, cioè siano obsolete o inidonee al raggiungimento degli obiettivi propri della regola scritta. La stessa Corte costituzionale ha avuto modo di stabilire che la decisione sul bilanciamento tra esigenze produttive ed altri beni giuridici è rimessa al legislatore e, nei limiti della ragionevolezza, non può essere disattesa dal giudice[630]. Nessun diritto è tiranno rispetto alle altre situazioni giuridiche protette dalla costituzione[631]. Ma, ciò non esclude che il bilanciamento debba essere svolto secondo <<criteri di proporzionalità, e di ragionevolezza, in modo tale da non consentire né la prevalenza assoluta di uno dei valori coinvolti, né il sacrificio totale di alcuno di loro, in modo tale che sia sempre garantita una tutela sistemica e non frammentaria di tutti gli interessi costituzionali implicati>>[632]. Alla luce di questi principi, la Corte è addivenuta a dichiarare la non illegittimità del primo decreto Salva Ilva che autorizzava il proseguimento dell'attività produttiva, nonostante il sequestro degli stabilimenti industriali, a condizione che fossero rispettate le prescrizioni dell'Autorizzazione Integrata Ambientale. Secondo la Corte, in pratica, il bilanciamento tra gli interessi in gioco era ragionevole ed era frutto di un procedimento amministrativo svoltosi conformemente ai principi di prevenzione, precauzione, informazione e partecipazione.

[630] D.l. 3 dicembre 2012, n. 207, conv. in legge 24 dicembre 2012, n. 231, ritenuto legittimo da C. Cost., sent 9 maggio 2013, n. 85; d.l. 4 luglio 2015, n. 92, ritenuto illegittimo da C. Cost., sent. 7 febbraio 2018, n. 58.

[631] C. Cost., sent 9 maggio 2013, n. 85, § 9.

[632] C. Cost., sent. 7 febbraio 2018, n. 58, § 3.1., che a sua volta cita le sentenze della Corte Cost. n. 63 del 2016 e n. 264 del 2012. Nello stesso senso vd. anche, in dottrina, V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 106.

Non è stata affermata la medesima legittimità del secondo decreto Salva Ilva che faceva dipendere la prosecuzione dell'attività non dalla valutazione di un terzo soggetto ma da un piano di minimizzazione dei rischi predisposto dallo stesso datore di lavoro che era destinatario dei provvedimenti cautelari. Così facendo, il legislatore aveva privilegiato <<in modo eccessivo l'interesse alla prosecuzione dell'attività produttiva>>[633], travalicando i limiti che lo stesso art. 41 Cost. pone all'esercizio dell'attività d'impresa.

Pertanto, concepire così l'istituto del rischio consentito è in linea con il fondamentale principio della separazione dei poteri: il giudice non può sostituire il proprio intuito rispetto alla rilevanza degli interessi costituzionali coinvolti al bilanciamento svolto legittimamente dal legislatore e dai deputati organi amministrativi[634]. Occorre sottolineare che spesso il legislatore non ha le competenze tecniche per dettare le regole cautelari specifiche per ogni settore. Per tale motivo, gli organismi tecnici o le associazioni di categoria, molto spesso, svolgono la funzione di stabilire precisi *standard* tecnici. È quanto accade in UE con gli enti di normalizzazione.

Ritorna quanto si era già affermato nell'*excursus* circa la responsabilità per il danno provocato da prodotto difettoso. Cioè il dubbio circa la possibilità che il prodotto conforme alle normative di *product safety* sia ugualmente difettoso. Lì abbiamo affermato che, in caso di prodotto difettoso ma rispettoso degli standard, ugualmente il produttore abbia l'obbligo risarcitorio nei confronti del consumatore. Mentre, per quanto appena visto con riferimento all'istituto del rischio consentito, parrebbe invece crearsi un'area di impunità nei confronti di chi produca rispettando determinate regole cautelari scritte.

[633] C. Cost., sent. 7 febbraio 2018, n. 58, § 3.3.
[634] S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 362-363.

La differenza ontologica tra le due situazioni è la stessa che intercorre tra responsabilità civile e responsabilità penale. Se nella prima possono esistere imputazioni oggettive, lo stesso non è ammissibile nella seconda. Dunque, gli standard tecnici, nell'ambito della responsabilità penale, oltre a svolgere da norma cautelare integrando la nozione di colpa, fungono anche da garanzia nei confronti dell'agente.

4.5.2.1. L'individuazione delle aree di rischio consentito ed i limiti dell'istituto.

È il momento di individuare i casi concreti in cui la conformità al dovere di diligenza codificato impedisce che la concreta prevedibilità ed evitabilità dell'evento da parte dell'agente possa determinare un rimprovero per colpa generica, nella ipotesi di verificazione degli eventi lesivi[635]. Le regole cautelari improprie dotate di rigidità sono le regole per eccellenza deputate per eccellenza al compito di delineare l'area di rischio consentito. Per regole improprie si intendono quelle che prevedono l'adozione di misure atte a ridurre al minimo il rischio; mentre, al contrario, per regole cautelari proprie s'intendono quelle idonee a neutralizzare completamente il rischio[636]. È ovvio che le regole che delineano il rischio consentito siano per natura di tipo improprio.

Ciò vale di regola, anche se alla fine può capitare che le regole cautelari proprie diventino fallibili, lasciando scoperte aree marginali di rischio consentito. I confini del rischio consentito sono tracciati al meglio da norme rigide, cioè norme che impongono un comportamento specifico al

[635] Per tale nozione di rischio consentito vd. C. PERINI, *Il concetto di rischio nel diritto penale moderno*, cit., p. 527; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 360; F. CONSULICH, *(voce) Rischio consentito*, cit., p. 1116; A. PERIN, *Prudenza, dovere di conoscenza e colpa penale*, *Editoriale Scientifica*, 2020, p. 69.

[636] La definizione è stata proposta da P. VENEZIANI, *Regole cautelari "proprie" e "improprie"*, cit., passim, ed è ormai comunemente accettata, vd. per tutti C. PIERGALLINI, *(voce) Colpa (diritto penale*), cit., p. 229; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 364; C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione*, cit., p. 391.

destinatario della cautela[637]. Esempio classico di regola rigida è la configurazione di limiti-soglia che definiscono i livelli di emissione o assunzione al di sotto dei quali il rischio è sicuramente consentito[638]. Con riferimento alle regole elastiche, anche se vi è chi considera possibile uno spazio di applicazione del rischio consentito anche in relazione ad esse, occorre sottolineare che <<quanto più è indeterminata la regola, tanto più la colpa specifica scolora in quella generica>>[639]. Quel giudizio sulla prevedibilità ed evitabilità in concreto dell'evento che l'area di rischio consentito mira ad escludere si riproporrebbe nella ricostruzione dello specifico contenuto della regola cautelare elastica.

È lo stesso concetto di rischio consentito a basarsi sulla consapevolezza per cui la regola cautelare sia inidonea a prevenire tutti i rischi che è finalizzata a prevenire. Trattasi di una inidoneità, per l'appunto, ricercata: il legislatore è perfettamente consapevole che la cautela a volte fallisce poiché non è in grado di avere efficacia impeditiva nei confronti di tutti gli eventi lesivi. L'istituto ha senso di esistere proprio se si accetta preventivamente che una quota di offese al bene giuridico si verificheranno[640].

L'area di impermeabilità alla colpa generica è circoscritta da ipotesi di fallimento ordinario della norma cautelare che gli stessi organi codificanti prevedono. Fuori da quest'area, non possono esistere zone di impunità nel diritto penale[641]. La valutazione della diligenza, della

---

[637] Sulla distinzione tra regole rigide e regole flessibili vd., nella manualistica, G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 430; G. FIANDACA – E. MUSCO, *Diritto penale. Parte Generale*, cit., p. 583; D. PULITANÒ, Diritto penale, cit., p. 304.

[638] Sulla problematica qualificazione dei limiti-soglia quali norme cautelari vd., tuttavia, E. PENCO, *Limiti-soglia e responsabilità colposa*, cit., p. 211 ss.; sui valori soglia in generale v. F. D'ALESSANDRO, *Pericolo astratto e limiti soglia*, cit., spec. 267 ss., 333 ss.; nella manualistica vd. G. MARINUCCI – E. DOLCINI – G.L. GATTA, *Manuale di diritto penale*, cit., p. 286.

[639] F. GIUNTA, *La normatività della colpa penale*, cit., p. 92.

[640] F. CONSULICH, *(voce) Rischio consentito*, cit., p. 1116.

[641] F. CONSULICH, *(voce) Rischio consentito*, cit., p. 1116; parlano di un eventuale "fallimento" delle norme cautelari G. FORTI, *Colpa ed evento*, cit., p. 671 ss.; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 61 ss., S. ZIRULIA*, Esposizione a sostanze tossiche e*

prudenza e dell'imperizia dell'agente nel prevedere ed evitare l'evento lesivo andrà al di fuori dei confini stretti della cautela positiva ogni volta vi sia un fallimento non preventivato della cautela stessa.

Può trattarsi di un fallimento in concreto se vi siano circostanze del caso concreto che rendono del tutto inadeguata la norma cautelare. Norma che, comunque, continua a "vivere" nella generalità dei casi. Si pensi, ad esempio, a condizioni metereologiche straordinarie che rendano inadeguata la normativa antinfortunistica che chi organizza i cantieri comunque rispetta[642]. Può, anche, trattarsi di un fallimento in astratto, nella ipotesi in cui la norma cautelare si riveli inidonea allo scopo di bilanciamento che gli organi che l'avevano redatta si erano prefissati. E questo può capitare o perché il rischio reale è più elevato o perché si è al cospetto di rischi ulteriori rispetto a quelli che erano stati presi in considerazione ovvero perché sussiste un difetto di tecnica normativa che impedisce di raggiungere la riduzione del rischio che costituiva l'obiettivo della codificazione.

In ambedue le ipotesi, si apre la strada per la valutazione della colpa generica sotto-forma di analisi circa la prevedibilità dell'inadeguatezza della norma cautelare ovvero della conoscenza stessa da parte dell'agente di detta inadeguatezza[643]. Ciò accade poiché, se

---

responsabilità penale, cit., p. 380; E. PENCO, *Limiti-soglia e responsabilità colposa*, cit., p. 234; vd. anche C. PIERGALLINI, *(voce) Colpa (diritto penale),* cit., p. 231, che parla di "inadeguatezza" delle cautele.

[642] L'esempio è di S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 380; cfr. anche G. FORTI, *Colpa ed evento*, cit., p. 673 ss.; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie",* cit., p. 61 ss., che avvalorano una concezione parzialmente diversa di "fallimento" in concreto della norma cautelare, che sarebbe rinvenibile nei casi in cui «l'evento si sia verificato attraverso uno dei decorsi causali di cui si era tenuto conto nella costruzione della norma cautelare (che l'agente avrebbe pertanto dovuto prevedere e contrastare) e tuttavia si accerti che la cautela, prospettabile ex ante, non sarebbe stata in grado di impedire la verificazione dell'evento stesso, rivelandosi appunto, ex post, inefficace» (G. FORTI, Colpa ed evento, cit., p. 674). Secondo tale concezione, si presuppone che la regola cautelare sia stata violata dall'agente, e che soltanto successivamente emerga che le circostanze del caso concreto avrebbero reso in ogni caso inefficacia la norma cautelare violata.

[643] D. CASTRONUOVO, *Principio di precauzione*, cit., p. 167.

escludessimo in queste ipotesi la valutazione circa la colpa generica, trasformeremmo il rischio consentito in una intollerabile forma di immunità[644].

Il Progetto preliminare di riforma del codice penale redatto dalla Commissione Grosso si collocava nell'ottica dello specifico riconoscimento dell'efficacia esimente delle norme cautelari scritte[645]. L'art. 28, cpv, infatti prevedeva che <<il rispetto delle regole cautelari specifiche (…) esclude la colpa relativamente agli aspetti disciplinati da dette regole, salvo che il progresso scientifico o tecnologico, nel periodo successivo alla loro emanazione, non le abbia rese palesemente inadeguate>>.

4.5.2.2. L'*AI Act* e gli obblighi di monitoraggio *post-market*: superamento del rischio consentito?

Gli obblighi di monitoraggio post market previsti tanto dalla normativa generale in tema di sicurezza dei prodotti quanto dallo stesso *AI Act*, a nostro giudizio, incideranno indubbiamente sull'eventuale superamento del rischio consentito. L'art. 60 *AI Act* impone obblighi rigorosi di sorveglianza durante la fase di test, rappresentando una forma di monitoraggio proattivo prima del rilascio definitivo. Il produttore deve registrare le applicazioni di IA ad alto rischio in una banca dati pubblica di livello europeo. Questa registrazione darà la possibilità alle autorità competenti, agli utenti ed alle persone interessate di verificare se il sistema di IA ad alto rischio sia conforme ai requisiti normativi. Dopo l'immissione nel mercato, i fornitori devono istituire un sistema di monitoraggio che <<raccoglie, documenta e analizza attivamente e sistematicamente i dati pertinenti forniti dagli utenti o

[644] S. ZIRULIA*, Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 381; vd. anche V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 95 ss.

[645]COMMISSIONE GROSSO PER LA RIFORMA DEL CODICE PENALE, Articolato approvato dalla Commissione Ministeriale per la Riforma del codice penale nella seduta del 26 maggio 2001, disponibile sul sito del Ministero della Giustizia a questo link: https://www.giustizia.it/giustizia/it/mg_1_12_1.page?contentId=SPS31489&previsiousPage=mg_1_12_1.

raccolti tramite altre fonti sulle prestazioni dei sistemi di IA ad alto rischio per tutta la durata del loro ciclo di vita e consente al fornitore di valutare la costante conformità dei sistemi di IA ai requisiti>>.

I fornitori avranno anche l'obbligo di segnalare ogni tipo di incidente grave o malfunzionamento dei sistemi ad alto rischio. Tale notifica va fatta immediatamente dopo che il fornitore sia stato in grado di stabilire il nesso causale tra il sistema e l'incidente, non oltre quindici giorni dopo esserne venuto a conoscenza.

Le autorità nazionali competenti avranno il compito di indagare sui malfunzionamenti e sugli incidenti, cercando di reperire ogni informazione necessaria che verrà successivamente trasmessa alla Commissione che si occuperà di integrare le informazioni e di pubblicare una relazione di valutazione normativo, cinque anni dopo l'entrata in vigore dell'*AI Act*.

L'autorità di sorveglianza che viene a conoscenza di una difformità del sistema di IA con i requisiti e gli obblighi prescritti può imporre al soggetto l'adozione di misure idonee a far cessare la violazione. Con riferimento agli obblighi di monitoraggio e ritiro delle società produttrici, l'art. 21 *AI Act* prevede che <<i fornitori di sistemi di IA ad alto rischio che ritengono o hanno motivo di ritenere che un sistema di IA ad alto rischio da essi immesso sul mercato o messo in servizio non sia conforme al presente regolamento adottano immediatamente le misure correttive necessarie per rendere conforme tale dispositivo, ritirarlo o richiamarlo, a seconda dei casi. Essi informano di conseguenza i distributori del sistema di IA ad alto rischio in questione e, ove applicabile, il rappresentante autorizzato e gli importatori>>.

Lo scopo è evitare o comunque provare a ridurre al minimo la deviazione del prodotto intelligente rispetto allo spazio di rischio consentito delineato dall'*AI Act*, ma il problema è che tale area non è configurata con la

necessaria chiarezza e precisione. Il codice del consumo e le norme del settore specifico di interesse delineano una serie di obblighi di controllo, monitoraggio e scambi di informazioni con le autorità amministrative ed intervento, nel caso in cui emergano rischi ulteriori rispetto a quelli che la cristallizzazione delle norme cautelari mirava a prevenire. È il caso del rischio da sviluppo che esclude la responsabilità civile del produttore per il danno provocato dal prodotto se <<lo stato delle conoscenze scientifiche e tecniche, al momento in cui il produttore ha messo in circolazione il prodotto, non permetteva ancora di considerare il prodotto come difettoso>>[646].

Detto rischio da sviluppo, sì è vero, esime il produttore dal risarcimento del danno causato dal prodotto difettoso. Ma, nella prospettiva del diritto penale, è un rischio che si colloca al di fuori dell'area di rischio consentito. Dunque, l'insorgenza di un rischio da sviluppo comporta la nascita di un obbligo di attivazione per evitare o ridurre al minimo il rischio sopravvenuto. Ma, comunque, gli obblighi previsti dal codice del consumo volti ad evitare che si verifichino eventi lesivi al di fuori dell'area di rischio consentito originaria sono indeterminati: non sono in grado di chiarire in modo preciso i presupposti né tantomeno il contenuto dei doveri *post market* del produttore.

L'indeterminatezza investe tanto i criteri in forza dei quali stabilire se e quando un rischio ha travalicato l'area del rischio consentito, quanto la scelta tra le varie opzioni individuate dalla norma cautelare. Con riferimento al primo profilo, potrebbero ad esempio presentarsi evidenze empiriche di eventi lesivi che non erano stati considerati durante la definizione dell'area di rischio consentito. Si pensi, ad esempio, alla pubblicazione di studi scientifici che accertino l'esistenza di un nesso eziologico tra l'assunzione di un farmaco e l'insorgenza di una

[646] art. 118, lett. e, cod. cons., e art. 7, lett. e), 85/374/CEE

patologia: in casi come questo, il produttore dovrà adottare le misure necessarie all'eliminazione del rischio.

Determinare i criteri in base ai quali stabilire se e quando un rischio abbia superato l'area del rischio lecito diviene complesso nei contesti di incertezza scientifica, nel momento in cui, una volta immesso nel mercato un prodotto conforme alla regola cautelare, esso sia sospettato di avere una pericolosità ulteriore rispetto a quella presa in esame dagli *standard* cautelari[647]. Secondo autorevole dottrina, non basta un mero sospetto a fondare l'obbligo di adozione di misure più stringenti: servono univoci segnali di incipiente fallimento della regola cautelare, derivanti anche da indici di fatto e non necessariamente dalla dimostrazione scientifica di una provata correlazione tra impiego del prodotto e danno[648].

I criteri che devono fungere da guida nella valutazione circa la sussistenza di detta univocità possono essere considerati, *in primis*, la frequenza e la diffusione di eventi lesivi connessi all'impiego del prodotto, in relazione alla durata della sua commercializzazione ed alla penetrazione sul mercato; *in secundis*, occorrerà verificare se sussistano spiegazioni causali alternative rispetto ai danni verificatisi[649]. Questa prospettiva ha sicuramente il demerito di non essere conforme al principio di determinatezza della fattispecie colposa, ma dà la possibilità di non sottovalutare le istanze di tutela della salute e dell'integrità fisica dei cittadini all'interno della società del rischio.

---

[647] S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 398.

[648] S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 398-399.

[649] C. PIERGALLINI, *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, cit., p. 538 ss. Tali criteri sono in realtà elaborati dall'Autore come presupposti per la configurazione, *de jure condendo*, del cd. illecito di rischio, ossia di un tipo di illecito che prescinde dall'ascrizione nomologica dell'evento, sostituita dalla "propensione al rischio" di una determinata condotta. Tale soluzione consentirebbe di mantenere integre le categorie del diritto penale classico, evitando che queste vengano flessibilizzate – per presunte ragioni di giustizia sostanziale – dalla giurisprudenza. In argomento, per alcuni rilievi critici, vd. G. DE FRANCESCO, *L'imputazione del reato e i tormenti del penalista*, in M. Bertolino – G. Forti (a cura di) *Scritti per Federico Stella*, vol. I, Jovene, 2007, p. 529.

Con riferimento, invece, ai criteri per la scelta della misura più adeguata imposta dalla diligenza, ci si chiede in base a quali criteri dovrà il produttore decidere se occorra ritirare o richiamare il prodotto ovvero se, al contrario, sia sufficiente fornire le informazioni rilevanti al consumatore ed alle autorità amministrative. Lo scopo del produttore, nonché il criterio che dovrà usare il giudice nella ricostruzione dell'agente modello, è ripristinare il livello di sicurezza al quale mirava la regola fallita[650]. Pertanto, al cospetto di una regola cautelare propria, l'obiettivo della condotta deve essere neutralizzare il rischio; mentre, di fronte ad una regola impropria, la condotta consiste nell'adozione di quelle misure che riportano il rischio a quello consentito dalla regola cautelare fallita.

Se mancano del tutto cautele idonee per raggiungere i predetti obiettivi, il produttore deve ritirare o richiamare il prodotto.

4.5.2.3. IA e rischio consentito: approccio quantitativo o qualitativo?

Nel settore della produzione dei sistemi di IA, un fallimento vero e proprio delle regole scritte potrà insorgere eventualmente sul piano astratto. Si afferma questo perché il rischio reale è più elevato di quanto era stato previsto in sede di positivizzazione della regola cautelare, o perché vi sono ulteriori rischi rispetto a quelli originariamente presi in esame, ovvero perché vi è un difetto di tecnica normativa che non consente di raggiungere la riduzione del rischio.

Non è semplice comprendere quali siano i criteri per stabilire se le regole positivizzate siano fallite. Questo perché si mette in conto, con riferimento alle attività produttive autorizzate pericolose, che si verificherà comunque una quota di eventi lesivi. Non è semplice, dunque, capire a partire da quando il produttore ha l'obbligo di attivarsi per

[650] S. ZIRULLA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 383.

fornire informazioni al consumatore ed alle autorità ovvero per ritirare eventualmente il prodotto.

A questi dubbi non è stata ancora fornita un'univoca risposta. In dottrina, tradizionalmente, si ritiene che l'importanza dei beni giuridici in gioco, il numero dei soggetti implicati, l'utilità sociale dell'attività ed il grado di probabilità dell'evento lesivo debbano essere considerati come incidenti sulla liceità del rischio[651]. Quest'ultimo aspetto è fondamentale: non è ammesso lo svolgimento di attività pericolose qualora il rischio di verificazione dell'evento sia molto alto[652].

Già in relazione ai prodotti tradizionali vi era chi ha rilevato che <<*l'indisponibilità di una cornice nomologica idonea a determinare il grado di probabilità dell'evento frustra in radice la possibilità di reperire norme cautelari preventive, funzionali al contenimento di una tipologia di danni, e pregiudica, perciò, la stessa legittimità del riferimento all'istituto del rischio consentito*>>[653].

Nonostante il fenomeno della *black box* sia aggravato nei sistemi di IA, il fatto che si possano registrare le attività dei sistemi di ML aumenta la calcolabilità degli errori e degli eventi lesivi che possono verificarsi. Si tratta di una differenza enorme rispetto ai prodotti tradizionali. Ciò in quanto, dei sistemi di IA già sappiamo che possono prendere decisioni errate e causare danni. Cosa che non si conosce con i prodotti tradizionali. Ciò che non conosciamo rispetto ai sistemi di IA sono gli input che possono determinare un *decision making* fallace, vista l'assenza di leggi scientifiche esplicative.

Per come abbiamo già avuto modo di constatare in precedenza, il decorso causale lesivo dentro l'algoritmo può essere ricostruito *ex post*, anche se con moltissime difficoltà.

---

[651] C. PIERGALLINI, *Danno da prodotto*, cit. p. 243; vd. anche V. MILITELLO, *Rischio e responsabilità penale*, cit., p. 170.
[652] G. FORTI, *Colpa ed evento*, cit., p. 255.
[653] C. PIERGALLINI, *Danno da prodotto*, cit., p. 243.

Ciò che è assente è la possibilità di addivenire ad una generalizzazione del comportamento algoritmico che possa spiegare il motivo per cui l'algoritmo risponda a determinati input in un determinato modo. I sistemi di ML sono, di fatti, *unpredictable by design*.

Una soluzione a tale mancanza potrebbe essere fornita dall'inserimento della possibilità tecnica di registrazione sistematica *ex post* delle decisioni assunte dai sistemi intelligenti. Tali rilevazioni statistiche, irrilevanti ai fini causali, potrebbero comunque costituire un valido parametro per verificare se le norme cautelari codificate siano o meno fallite. Anche la crescita del settore dell'*Explainable AI* agevolerebbe la ricostruzione del decorso causale ex post, nonché renderebbe più agevole la ricostruzione statistica del tasso di affidabilità ed accuratezza dei sistemi.

Si è detto che il decisore pubblico individua una serie di norme cautelari volte a prevenire la verificazione di un determinato rischio. Si è detto anche che l'agente, se emerge un fallimento delle cautele, deve compiere tutte le attività necessarie per ricondurre il rischio al livello che costituiva l'esito del bilanciamento. Uno dei maggiori problemi che abbiamo individuato *supra* concerne proprio il comprendere quando gli eventi lesivi costituiscano la prevedibile estrinsecazione del rischio consentito e quando, al contrario, essi oltrepassino detto limite. A nostro giudizio, in tal senso può essere illuminante la registrazione algoritmica *ex post* degli incidenti proprio per valutare il superamento di tipo quantitativo della soglia di rischio lecito.

La registrazione dovrebbe servire per isolare gli errori algoritmici, senza che sia necessario individuare il preciso input che ha cagionato la verificazione dell'errore algoritmico e, di conseguenza, dell'evento lesivo. Procediamo con una esemplificazione. Mettiamo caso che il legislatore preveda un apparato di norme cautelari rigide per lo sviluppo e la produzione di software di guida autonoma, specificando che la fissazione di queste norme è funzionale a mantenere il

tasso di incidenti sotto la soglia dello 0,0001% rispetto all'utilizzazione del veicolo. Se, nonostante il rispetto delle norme cautelari rigide, il produttore verifichi che il tasso di sinistri derivanti dall'uso del *software* sia maggiore rispetto alla soglia del rischio consentito, egli dovrà adottare cautele ulteriori per la minimizzazione dei rischi e, se del caso, richiamare o ritirare il prodotto.

Il mero fatto del superamento del tasso individuato dal legislatore non determinerebbe l'automatica imputazione al produttore degli eventi lesivi ulteriori, ma dovrebbe essere considerato un segnale di incipiente fallimento delle norme cautelari scritte. Dunque, dovrebbe indurre il produttore ad adottare cautele più pregnanti rispetto a quelle dettate dalla normativa. Approccio simile è stato suggerito anche da una parte della dottrina civilistica con riferimento al tema dell'accertamento di un difetto di *design*[654]. Detto orientamento suggerisce di valutare la difettosità del prodotto intelligente tramite la comparazione tra la prestazione complessiva del sistema di IA che ha provocato un danno e la performance di un modello algoritmico mediamente sicuro. A questo scopo si prospetta l'opportunità di individuare una soglia di riferimento al di sopra della quale l'algoritmo sarebbe difettoso.

Presupposti indefettibili per una simile regolamentazione sono sia che il tasso di accuratezza dell'algoritmo di ML venga valutato prima dell'immissione del prodotto sul mercato, sia vengano condivisi con le autorità competenti i dati inerenti agli eventuali incidenti cagionati dai sistemi di IA[655]. Il vantaggio di una simile soluzione è lo spostamento del fulcro della valutazione sulla sicurezza del prodotto dalla singola prestazione – inspiegabile rispetto al preciso episodio lesivo – alla *performance* complessiva. Ciò consentirebbe di

[654] Vd. in particolare J.S. BORGHETTI, *How can Artificial Intelligence be Defective*?, in S. Lohsse e al. (eds), *Liability for Artificial Intelligence and the Internet of Things*, Nomos, Baden-Baden, 2019, p. 63 ss.

[655] Cfr. G. FORTI, *"Accesso" alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione*, cit., passim.; vd. anche C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, cit., p. 294.

continuare a godere dei benefici derivanti dallo sviluppo e dalla commercializzazione dei sistemi di IA senza che i singoli eventi lesivi possano mettere in dubbio la sicurezza del prodotto.

Soluzioni come queste, di tipo quantitativo, contribuiscono a definire l'area del rischio consentito in altri settori caratterizzati da alti gradi di incertezza. Vogliamo riferirci al settore farmaceutico e nello specifico dei vaccini. Prendendo in esame la popolazione di vaccinabili, si può prevedere che si verifichino eventi infausti, ma non è prevedibile stabilire quando, come e chi li subirà. Per questo esiste la farmacovigilanza. Essa serve con lo scopo di raccogliere ed analizzare gli eventi avversi che seguono l'assunzione dei farmaci e, in caso di vaccini, l'immunizzazione. Se si verificano eventi infausti in un numero maggiore rispetto a quanto emerso nelle fasi di sperimentazione e di autorizzazione del vaccino, il produttore dovrà fornire ulteriori informazioni nel bugiardino ovvero, nei casi più gravi, ritirare il prodotto. Se, invece, gli eventi avversi si mantengono sotto una soglia fisiologica, essi rientreranno nell'area del rischio consentito e non potranno essere imputabili al produttore[656].

Il settore dei vaccini è molto particolare: vi sono questioni di bilanciamento inerenti al diritto alla salute e solo dopo inerenti all'attività economica[657]. A nostro modestissimo giudizio, gli stessi principi si potrebbero riportare nel campo

[656]La farmacovigilanza sui vaccini anti-covid offre, tra l'altro, un interessante modello di valutazione della probabilità circa il nesso di causalità tra immunizzazione e verificazione dell'evento avverso. Il Comitato Consultivo Globale per la Sicurezza dei Vaccini (GACVS) dell'Organizzazione Mondiale della Sanità ha messo a punto un algoritmo che, tenendo conto di una serie di dati (relazione temporale fra la vaccinazione e la reazione segnalata; presenza di possibili spiegazioni alternative; precedenti evidenze di letteratura; frequenza dell'evento segnalato nella popolazione generale, anche non vaccinata; plausibilità biologica) valuta la probabilità della sussistenza del nesso tra assunzione del vaccino ed evento lesivo. Vd. a tal proposito il report pubblicato dalla AGENZIA ITALIANA DEL FARMACO, Tredicesimo rapporto sulla sorveglianza dei vaccini anti-COVID-19 (27/12/2020 – 26/09/2022), disponibile a questo link:
https://www.aifa.gov.it/documents/20142/1315190/Rapporto_sorveglianza_vaccini_COVID-19_13.pdf, p. 2.

[657] Emerge poi il principio di solidarietà (art. 2), nei casi in cui la vaccinazione sia obbligatoria o comunque fortemente raccomandata (vd. il caso del vaccino per il Covid-19).

dell'IA, nei settori in cui la commercializzazione di sistemi di IA può avere un impatto benefico sulla tutela del diritto alla vita ed all'integrità fisica delle persone. Basti pensare, a titolo esemplificativo, all'uso dei CDSS, che sarebbero in grado di identificare patologie in fasi molto precoci ed individuare trattamenti basati sulle specifiche caratteristiche del paziente. Si pensi, ancora, ai veicoli a guida autonoma che potrebbero ridurre fino al 90% gli incidenti stradali.

Il limite che, per noi, pare essere il principale a quanto appena prospettato è costituito dalla politica. Secondo noi, nessun decisore potrebbe manifestare in modo così diretto il bilanciamento tra l'esercizio di un'attività economica e beni di enorme rilievo. Un ulteriore rischio sarebbe che, tramite la individuazione di limiti-soglia, si tenti di perseguire logiche iper-cautelative che comporterebbero la necessità di adottare sempre cautele ulteriori rispetto a quelle positivizzate.

Inoltre, la ultroneità rispetto alla soglia consentita del rischio può sorgere anche dal punto di vista qualitativo. Ad esempio, un'auto a guida autonoma che si comporta in modo del tutto diverso rispetto a come si comportava in fase sperimentale. Questo è ciò che potrebbe essere successo nel novembre 2022, quando un veicolo della casa automobilistica Tesla, percorrendo le strade di una città in Cina, ha accelerato in modo improvviso, causando due morti e tre feriti, per poi scontrarsi contro un immobile[658]. È possibile che, nella fase sperimentale, il software di guida autonoma abbia mostrato un novero di errori all'interno dei quali l'improvvisa accelerazione non rientrava. In un siffatto caso, si potrebbe configurare un segnale di fallimento delle regole cautelari che non contemplavano la minimizzazione del "rischio da accelerazione".

Per come già asserito, sulla individuazione di un eventuale superamento del rischio consentito incideranno

[658] Rai News, a questo link: https://www.rainews.it/video/2022/11/tesla-lanciata-a-folle-velocit-le-telecamere-a-circuito-chiuso-riprendono-lincidente-mortale-06a60958-096a-4005-9213-50c36fcb7641.html (Red. Rai News, Tesla lanciata a folle velocità: le telecamere a circuito chiuso riprendono l'incidente mortale, 14 novembre 2022)

indubbiamente gli obblighi di monitoraggio *post market* previsti dalle norme generali circa la sicurezza dei prodotti, dalla normativa speciale per la tipologia di prodotti e, infine, dall'*AI Act*.

4.5.3. La colpa generica nell'IA.

Si è già accennato che l'impermeabilità dell'area del rischio consentito alla colpa generica viene messa in discussione dalla giurisprudenza che tende a considerare spesso la colpa generica come un <<rinforzo onnipresente di quella specifica, atta a riempire ogni possibile lacuna preventiva che quest'ultima possa presentare>>[659]. E, comunque, anche se si aderisse alla impostazione che mira a creare un'area di impunità per il soggetto che si conformi a regole cautelari scritte esaustive, residuerebbe uno spazio per valutare la prevedibilità e l'evitabilità dell'evento. Questo al ricorrere di una delle due fondamentali circostanze: i) la presenza di norme cautelari elastiche; ii) il fallimento riconoscibile ed evidente di dette norme.

In queste ipotesi, riecheggiano le incertezze che la categoria del rischio consentito aveva l'obiettivo di neutralizzare, ovverosia circa il come individuare l'ampiezza del dovere di diligenza al di fuori del bilanciamento operato dal legislatore e sul tipo di eventi che avrebbero dovuto e potuto essere evitati al contrario di quelli imponderabili[660].

A ciò si deve aggiungere che, nell'era della complessità tecnologica in cui viviamo, non è possibile la creazione di norme cautelari esaustive. Vi sono troppe particolarità nei casi concreti e troppe districazioni di norme cautelari provenienti da diverse fonti. Per tale motivo, si crede indispensabile che il giudice compia un giudizio sulla

[659] F. CONSULICH, *(voce) Rischio consentito*, cit., p. 1109. Vd. ad es. Cass., sez. IV, 20 marzo 2000, n. 3567, in Giur. it., 2001, 1709 ss., che riconosceva la colpevolezza del datore di lavoro, nonostante questi avesse rispettato i valori-soglia in materia di esposizione all'amianto; vd. anche Cass., Sez. IV, 4 novembre 2010, n. 38991. Per una rassegna giurisprudenziale sul rapporto tra colpa generica e colpa specifica v. M. GROTTO, *Principio di colpevolezza*, cit., p. 79 ss.

[660] Così V. ATTILI, *L'agente-modello "nell'era della complessità",* cit., p. 1244-1245.

valutazione della diligenza generica dell'agente. Il legislatore, codificando le norme cautelari, più che vincolare il giudice, funge da guida, essendo in grado di <<definire l'ambito all'interno del quale si muove la problematica della colpa rispetto ad un determinato rischio>>[661].

Nel panorama dell'IA, è molto lontana la precisa definizione di standard cautelari rigidi ed esaustivi, atti a creare confini invalicabili dell'area del rischio consentito. Da ciò deriva la necessità, forzata almeno per i primi tempi, di ricorrere al parametro dell'agente modello.

4.5.3.1. Un'imprevedibilità prevedibile dell'evento lesivo è rimproverabile?

Per usare una locuzione ai più nota, la produzione di sistemi intelligenti costituisce un vero e proprio *shock* da modernità[662]. Il tutto è ampliato nel settore colposo poiché la messa in commercio di prodotti di per sé imprevedibili mette a dura prova la tenuta delle categorie della prevedibilità e dell'evitabilità dell'evento.

Per come abbiamo già detto in precedenza, il funzionamento dell'IA è ignoto anche agli stessi programmatori. La domanda che occorre porsi è la seguente: gli eventi lesivi derivanti dagli algoritmi di IA possono rientrare nell'area di rischio che il produttore deve neutralizzare, visto che sono prevedibili nell'*an* ma imprevedibili nel *quantum* e nel *quomodo*?

Per rispondere a tale questione, occorre comprendere il tipo ed il grado di dettaglio che deve avere il giudizio di prevedibilità dell'evento e se il principio di precauzione può o meno avere un ruolo concreto nell'individuazione delle cautele della colpa generica. A nostro giudizio, la definizione dei confini del giudizio di prevedibilità dell'evento lesivo

[661] R. BARTOLI, *(voce) Fonti della colpa*, cit., p. 531.
[662] F. STELLA, *Giustizia e modernità*, cit., p. 292.

algoritmico è uno degli aspetti più importanti dell'accertamento della colpa del produttore[663].

Le leggi scientifiche servono anche in relazione all'accertamento della colpa. Ma, mentre in relazione alla verifica del nesso causale avviene con valutazione *a posteriori*; con riferimento alla colpa bisogna effettuare una valutazione *ex ante*[664]. Vi è chi ha asserito che, nel primo caso si ha a che fare con una legge descrittiva mentre nel secondo caso con una legge deontica[665]. Bisogna evitare che il giudice possa integrare il precetto colposo *ex post*, individuando una regola cautelare inesistente al momento della condotta sulla base del mero convincimento secondo cui l'imputato avrebbe dovuto comportarsi in maniera differente. E, ovviamente, del <<*senno di poi son piene le fosse*>>[666].

Con riferimento all'oggetto del giudizio di prevedibilità dell'evento, il giudizio potrebbe essere riferito alle precise e specifiche caratteristiche del caso concreto ovvero ad una classe generica di eventi simili. Minore è la precisione con cui si seleziona l'evento assunto come parametro, maggiore sarà il rischio di consentire l'accesso a valutazioni a posteriori della colpa penale, con ciò estendendo l'area di responsabilità. Maggiore è l'attenzione alle caratteristiche concrete del fatto verificatosi e minore sarà lo spazio per il rimprovero penale. Trattasi di una vera e propria proporzionalità inversa.[667]

La difficoltà nella definizione di un univoco criterio sembra l'ennesimo sintomo del generale problema di

---

[663] In questo senso L. PICOTTI, *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, cit., p. 834; C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1762.
[664] Vd. per tutti G. FORTI, *Colpa ed evento*, cit., p. 210; P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 49 ss.
[665] C. BRUSCO, *Rischio e pericolo, rischio consentito e principio di precauzione*, cit., p. 395.
[666] G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 48; in argomento vd. anche P. VENEZIANI, *Regole cautelari "proprie" ed "improprie"*, cit., p. 55; A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato*, cit., p. 403; V. ATTILI, *L'agente-modello "nell'era della complessità"*, cit., p. 1269; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 319-320.
[667] In questo senso, G. CIVELLO, (*voce) Prevedibilità e reato colposo*, in M. Donini (diretto da) *Reato colposo, Enc. dir. – I Tematici*

indeterminatezza della fattispecie colposa. La giurisprudenza, negli ultimi tempi, tende ad espandere l'area della responsabilità colposa. Ciò si verifica, soprattutto, in relazione all'esercizio di attività pericolose in situazioni di incertezza. Spesso, i giudici sono arrivati a creare la regola cautelare utilizzando come base del giudizio di prevedibilità le <<stesse leggi causali utilizzate per stabilire una relazione di causa ed effetto tra eventi>>[668]. Così non si fa altro che confondere l'accertamento del nesso di causalità con quello colposo.

È quanto successo, ad esempio, con la sentenza inerente al disastro del Vajont. In essa, i giudici scrissero che <<nell'esercizio lecito di attività pericolose, ai fini del giudizio di prevedibilità, deve aversi riguardo alla potenziale idoneità della condotta a dar vita a una situazione di danno o di pericolo, e non anche alla specifica rappresentazione *ex ante* dell'evento dannoso, quale si è concretamente verificato in tutta la sua gravità ed estensione. Perché sorga l'obbligo di una particolare prudenza e cautela – proprio perché si tratta di attività essenzialmente e per definizione pericolose – è, in altri termini, sufficiente la prevedibilità di una particolare situazione di danno o di pericolo, senza che si debba necessariamente estendere agli ulteriori effetti che, in ipotesi estremamente pessimistica, potrebbero scaturire dalla situazione potenziale, in particolare quando, come nella specie, l'attività in questione sia profondamente condizionata dalla incidenza di forze naturali, non sempre conoscibili e talora scatenate ed inarrestabili>>[669].

---

[668] G. FORTI, *Colpa ed evento*, cit., 208; vd. anche V. ATTILI, *L'agente-modello "nell'era della complessità"*, cit., p. 1269-1270; F. SGUBBI, *Il diritto penale incerto e inefficace*, in *Riv. it dir. proc. pen*., 2001, p. 1197; A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato*, cit., p. 403; G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 48 ss. Per un interessante studio dell'incidenza dei *bias* cognitivi sui giudizi di prognosi postuma vd. S. GROSSO, *Il giudizio di prevedibilità dell'evento e l'incidenza dell'hindsight bias e dell'outcome bias sul giudizio di responsabilità colposa*, in *Riv. it. med. leg*., 2016, fasc. 2, p. 583 ss.; D. PERRONE, *La prognosi postuma tra distorsioni cognitive e software predittivi*, cit., p. 37 ss.

[669] Cass., sez. IV, 25 marzo 1971, n. 810, Biadone

Affermazioni di principio simili sono confluite nella giurisprudenza sulla responsabilità per esposizione all'amianto[670]. Lì, il datore di lavoro era stato condannato per avere omesso di adottare le idonee misure ad evitare l'esposizione dei lavoratori all'amianto, causa delle morte per mesotelioma pleurico, a trent'anni dal fatto, di un lavoratore. Mentre, all'epoca dei fatti, l'amianto era scientificamente conosciuto come causa possibile di malattie come l'asbestosi, non certo di tumori come il mesotelioma. A ciò si aggiunga quanto affermato con riferimento al caso del Petrolchimico di Porto Marghera. I giudici qui affermarono che per ritenere sussistente la colpa in capo ai vertici della società per la morte per tumore di alcuni dipendenti fosse sufficiente la mera conoscibilità del fatto che la sostanza CVM, usata negli stabilimenti, avrebbe potuto essere idonea a provocare malattie al fegato, non essendo necessario specificare il tipo di malattia, cioè il tumore al fegato[671]. La Cassazione asseriva che, visto che erano in astratto prevedibili gravi danni alla salute dei lavoratori esposti al cloruro di vinile, doveva ritenersi sorto l'obbligo <<anche se fossero mancate regole cautelari di origine normativa>> per il datore di adottare le cautele necessarie per preservare i lavoratori dal rischio in esame, poiché <<la tutela della salute umana costituisce obbligo primario di salvaguardia di un diritto costituzionalmente protetto>>[672]. Non è stata avulsa da critiche dottrinali l'impostazione fatta propria dalla Corte di

[670] Pretura Torino, 9 febbraio 1995, in Riv. it. dir. e proc. pen., 1997, p. 1447; sulla vicenda (e sulla successiva sentenza della Corte d'Appello di Torino) si vd. la nota di C. PIERGALLINI, *Attività produttive e imputazione per colpa: prove tecniche di "diritto penale del rischio"*, cit.

[671] Sulla vicenda giudiziaria che ha coinvolto lo stabilimento petrolchimico di Porto Marghera – la più importante, nel nostro Paese, in materia di danni alla salute e all'ambiente – si rinvia alle approfondite riflessioni di S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 104 ss.; C. PIERGALLINI, *Il paradigma della colpa*, cit., p. 1670 ss.; F. D'ALESSANDRO, *Pericolo astratto e limiti-soglia*, cit., p. 180 ss.; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 132 ss.; A. NAPPI, *La prevedibilità nel diritto penale*, cit., p. 327 ss.

[672] Sulla vicenda giudiziaria che ha coinvolto lo stabilimento petrolchimico di Porto Marghera – la più importante, nel nostro Paese, in materia di danni alla salute e all'ambiente – si rinvia alle approfondite riflessioni di S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 104 ss.; C. PIERGALLINI, *Il paradigma della colpa*, cit., p. 1670 ss.; F. D'ALESSANDRO, *Pericolo astratto e limiti-soglia*, cit., p. 180 ss.; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza.*

Cassazione. Secondo la dottrina, i giudici, ponendo come base del giudizio di prevedibilità un evento generico di danno verificatosi, rendono impossibile l'individuazione di una cautela che sia diversa dalla semplice astensione dall'esercizio dell'attività pericolosa, dal momento che <<non esiste una regola cautelare dotata di un'efficacia preventiva generale>>[673].

Così facendo, non può parlarsi di giudizio di prevedibilità, bensì di semplice responsabilità oggettiva, dato che l'evento viene posto a carico dell'agente sulla sola base della sua causazione materiale. Ad oggi, pare che la giurisprudenza si sia assestata sul principio in forza di cui <<in tema di colpa, la necessaria prevedibilità dell'evento – anche sotto il profilo causale – non può riguardare la configurazione dello specifico fatto in tutte le sue più minute articolazioni, ma deve mantenere un certo grado di categorialità, nel senso che deve riferirsi alla classe di eventi in cui si colloca quello oggetto del processo>>[674].

Secondo noi, la valutazione sulla prevedibilità da parte dell'agente modello dell'evento lesivo algoritmico deve essere netta. Se prevale la concezione espansiva in base a cui è sufficiente la prevedibilità di una determinata classe di eventi, la colpa del produttore sarebbe sempre esistente[675]. Se come ci auguriamo prevale una concezione più restrittiva basata su decorsi eziologici conosciuti, il produttore non sarà mai punibile visto che l'evento per come si realizza non sarà mai prevedibile, posta la intrinseca imprevedibilità dei sistemi di IA. Occorre, dunque, ricercare una via di mezzo tra le due concezioni.

---

[673] C. PIERGALLINI, *Il paradigma della colpa nell'età del rischio*, cit., p. 1692; ID., *Danno da prodotto*, cit., 445; vd. anche G. FORTI, *Colpa ed evento*, cit., 487: «caratteristica della regola cautelare è di prescrivere condotte atte a prevenire un evento descritto assai più concretamente rispetto al modello penale, e dunque caratterizzato da precise modalità o risultato di ben determinati decorsi causali».

[674] Cass., sez. un., 24 aprile 2014, n. 38343, in *Dir. pen. cont*., 19 settembre 2014; conf. Cass., sez. IV, 25 giugno 2013, n. 35309.

[675] Dello stesso avviso sono C. PIERGALLINI, *Intelligenza artificiale*, cit., 1762; M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1208-1209.

4.5.3.2. L'incidenza del principio di precauzione sul giudizio di prevedibilità dell'evento.

L'affermazione del principio di precauzione comporta l'estensione del raggio d'azione del giudizio di prevedibilità dell'evento lesivo. In giurisprudenza, si è arrivati a sovrapporre il piano dell'evento prevedibile con il piano dell'evento non suscettibile di essere escluso, così che sono stati ritenuti prevedibili anche quegli eventi che l'agente non era in grado di escludere al momento della condotta[676].

Anche la sentenza emessa sul caso del petrolchimico di Porto Marghera, nel ritenere che il dubbio scientifico sulla possibilità di eventi avversi ulteriori possa fondare un dovere cautelare, è impregnata di logiche precauzionali. Se, comunque, in quel caso era accertata la dannosità della sostanza usata negli stabilimenti, le sentenze sugli eventi idrogeologici di Messina e del Sarno spostano l'asticella della prevedibilità ancora più in là, ricomprendo nel giudizio tutti i possibili effetti calamitosi che non si possano escludere[677].

Nella decisione della Cassazione sul disastro di Sarno, si può leggere che il giudizio di prevedibilità andava compiuto <<*ipotizzando quindi la più distruttiva ipotesi che potesse verificarsi*>>. Questo in base alla convinzione secondo cui <<*se di un fenomeno naturale (o anche cagionato dall'uomo) non si conoscono le caratteristiche fondamentali – in particolare le cause, le possibilità di evoluzione, gli effetti possibili – la cautela che occorre usare nell'affrontarlo per eliminare o ridurne le conseguenze deve essere ancora maggiore proprio perché non si possono escludere, con valutazione ex ante fondata su conoscenze scientifiche affidabili, gli effetti*

[676]G. CIVELLO, *(voce) Prevedibilità e reato colposo*, cit., p. 1019.

[677] Per un confronto tra la sentenza di legittimità sul caso del petrolchimico di Porto Marghera e quella sul caso del Sarno vd. D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 143 ss.

maggiormente distruttivi>>[678]. A simili affermazioni è addivenuta la Corte nel caso dell'alluvione di Messina, nella quale si è affermato che <<*la valutazione della prevedibilità (…) non può prescindere dalla valutazione su che cosa può avvenire in futuro, a meno che le caratteristiche del fenomeno non siano da sole idonee a fondare un giudizio di esclusione di più gravi conseguenze*>>[679].

Anche con le sentenze circa il terremoto dell'Aquila, pur dando atto che con la scienza sismologica non si possa prevedere precisamente né l'*an*, né il quando, né il *quomodo* degli eventi tellurici, si è detto che dentro il giudizio di prevedibilità rientri anche il rischio di un evento non escludibile[680]. La dottrina ha criticato questo atteggiamento dei giudici, parlando di una sorta di colpa eventuale: l'agente non può concretamente prevedere e quindi evitare l'evento ma può solo rappresentarsi il rischio di un danno, non potendo escludere la sua verificazione[681].

In diverse pronunce, come nella sentenza Bartalini del 2006, la Corte ha espressamente menzionato il principio di precauzione, asserendo che <<*le regole che disciplinano l'elemento soggettivo hanno (…) funzione precauzionale, e la precauzione richiede che si adottino certe cautele anche se è dubbio che la mancata adozione provochi eventi dannos*i>>[682]. Si è anche identificata nell'astensione dall'attività la regola cautelare di chiusura nel caso in cui il garante si renda conto di non essere in grado di incidere sul rischio[683]. Così ragionando, la regola da seguire in caso di incertezza

[678] Cass., sez. IV, 11 marzo 2010, n. 16761.
[679] Cass., sez. IV, 16 febbraio 2012, n. 17069: nel caso di specie si rimproverava ai pubblici amministratori di non aver impedito la circolazione stradale sul greto asciutto di due fiumi, così che, in occasione dell'alluvione, l'acqua aveva travolto cose o persone che ivi si trovavano.
[680] G. CIVELLO, (voce) *Prevedibilità e reato colposo*, cit., p. 1019.
[681] G. CIVELLO, *La "colpa eventuale" nella società del rischio. Epistemologia dell'incertezza e "verità soggettiva" della colpa"*, Giappichelli, 2013.
[682] Cass., sez. IV, 17 maggio 2006, n. 4675, Bartalini, p. 290.
[683]Cass., sez. IV, 14 marzo 2014, n. 22249.

scientifica sarebbe *in dubio pro omissione* e non più *pro actione*[684].

Detto modello precauzionale, che obbliga ad adottare tutte le misure preventive tecnologicamente possibili per evitare il danno, è stato affermato non solo in presenza di regole cautelari elastiche, ma anche con riferimento alle regole rigide. Ciò è quanto succede, ad esempio, con riferimento alla tecnica normativa dei limiti-soglia che viene del tutto resa invana da quel filone giurisprudenziale che ritiene di poter fondare il rimprovero colposo anche in ipotesi di osservanza dei limiti quantitativi di esposizione fissati per legge o regolamento[685].

La maggior parte degli Autori ha criticato l'utilizzo del principio di precauzione come base di valutazione della colpa generica[686]. Poco avrebbe a che vedere con la colpa un istituto che sulla base del mero nesso di causalità collega condotta ed evento, rischiando così di muovere rimproveri colposi per la verificazione di qualsiasi evento lesivo. La grammatica dell'incertezza alla base del principio di precauzione non si concilia con la necessità di determinatezza che dovrebbe costituire la base delle fattispecie incriminatrici. Il semplice rinvio alla precauzione, se non viene formalizzato in norme cautelari

[684] F. CONSULICH, *(voce) Rischio consentito*, cit., p. 1106

[685] Vd. in proposito gli stralci dell'ordinanza pronunciata dal Tribunale di Taranto, in sede di riesame, il 7 agosto 2012, riportati in S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 199 ss., a cui si rinvia per un completo panorama giurisprudenziale in materia. Vd. anche E. PENCO, *Limiti-soglia e responsabilità colposa*, cit., p. 195 ss.

[686] C. PIERGALLINI., *Il paradigma della colpa nell'età del rischio*, cit., p. 1694- 1696; F. STELLA, *Giustizia e modernità*, cit., p. 221; F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 241 ss.; G.P. ACCINNI, *Criteri di imputazione per colpa tra leggi scientifiche e accertamenti giudiziali*, in *Riv. it. dir. proc. pen*., 2006, n. 3, p. 928 ss.; V. ATTILI, *L'agente-modello "nell'era della complessità*", cit., p. 1278 ss.; G. DE FRANCESCO, *"Interpersonalità" dell'illecito penale: un 'cuore antico' per le moderne prospettive della tutela*, in *Cass. pen*., 2015, n. 2, p. 868-869; A. ALESSANDRI, *Attività d'impresa e responsabilità penali*, in *Riv. it. dir. proc. pen*., 2005, fasc. 2, p. 565; D. PULITANÒ, *Colpa ed evoluzione del sapere scientifico*, cit. p. 651; S. ZIRULIA, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 317 ss.; C. PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto "a distanza" tra disvalore d'azione e disvalore d'evento*, cit., p. 238 ss.; da ultimo, F. D'ALESSANDRO, *Precauzione (principio di)*, cit., p. 1001 ss.

positive, lascia aperta la questione circa la individuazione della cautela doverosa: non si conosce se, in assenza di un solido apparato nomologico, ci si debba astenere dalla condotta pericolosa o se si deve adottare cautele aggiuntive.

Trattasi di decisioni che deve necessariamente prendere il decisore pubblico, il quale ha il dovere di prendere atto delle conoscenze specifiche esistenti all'interno delle imprese. Queste ultime devono essere coinvolte nei processi di formazione delle regole cautelari scritte[687] e, secondo alcuni, devono essere destinatarie di un dovere di diffusione delle informazioni possedute sanzionato penalmente[688].

Anche se la maggior parte della dottrina ritiene che l'impiego del principio di precauzione come base del giudizio sulla colpa generica sia contrario alla struttura costituzionale della colpa, vi è comunque chi sostiene un'apertura in questo senso[689]. In materia ambientale, ad esempio, vi è stato un ingresso parziale della logica precauzionale nell'individuazione delle cautele generiche: il principio è stato espressamente menzionato all'art. 191 TFUE. Inoltre, vi è una presa d'atto in forza di cui si crede che il perenne stato di incertezza che caratterizza le attività umane deve indurre ad una maggiore responsabilizzazione nei confronti di coloro che introducono nuovi ed imponderabili rischi nella società.

Ed è proprio sulla base di questa logica che vi è chi propone la valorizzazione della logica precauzionale nell'ambito del giudizio sulla colpevolezza, nei casi in cui sia presente un dato scientifico di un certo rilievo, sia pure

---

[687] F. GIUNTA, *Il diritto penale e le suggestioni*, p. 245.

[688] È la tesi di G. FORTI, *"Accesso" alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione*, cit., passim; per un "aggiornamento" della tesi al contesto produttivo dell'intelligenza artificiale vd. C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1773.

[689] C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., 1743 ss.; E. CORN, *Il principio di precauzione nel diritto penale*, cit., p. 146 ss.; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 143 ss. e 166 ss.

se non consolidato[690]. Il principio di precauzione amplia il raggio d'azione della colpa generica tramite una rivisitazione ermeneutica della nozione di prevedibilità dell'evento. La sola incertezza scientifica, pertanto, non escluderebbe la colpa che dovrebbe essere individuata nella <<*mancata o intempestiva adozione delle cautele suggestive dai signa facti e dalle congetture ricavabili dagli studi e dai dati disponibili nel mondo scientifico, seppure a livello di ipotesi non ancora assurte a rango di leggi scientifiche o non ancora condivise dall'intera comunità scientifica*>>[691].

Il principio di precauzione, quindi, potrebbe condurre ad ampliare lo spettro dei sotto-eventi prevedibili, ivi includendovi anche quelli che presentino omogeneità tra di loro. Ritornando all'esempio dell'esposizione all'amianto, all'agente potrebbe rimproverarsi la morte derivante dal sotto-evento del mesotelioma pleurico, visto che esso è analogo al sotto-evento prevedibile al momento della condotta, ovverosia l'asbestosi. Ciò poiché ambedue le malattie coinvolgono l'apparato respiratorio[692].

4.5.4. IA e rischio consentito: applicazione di quanto fino a qui scoperto.

A noi sembra che nel settore dell'IA si debba identificare un'area di rischio consentito. Delimitare uno spazio di impunità per i produttori che si conformano agli *standard* tecnici del settore ci sembra l'unico modo per poter consentire lo svolgimento dell'attività di ricerca, sviluppo e produzione di sistemi di IA, garantendo agli operatori che la valutazione sulla gestione dei rischi svolta in sede legislativa non possa essere ribaltata in sede giudiziaria.

Far entrare anche il parametro dell'agente modello nella valutazione giudiziale sulla colpevolezza del produttore

[690] C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1754.
[691] C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1775.
[692] C. RUGA RIVA, *Principio di precauzione e diritto penale*, cit., p. 1770 ss. Vd. anche D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 143 ss. e 166 ss., che si mostra "possibilista" in relazione ad un limitato ampliamento della nozione di "prevedibilità".

comporterebbe il rischio di un ampliamento sconsiderato della sfera di responsabilità dei vertici delle imprese produttive e dei dipartimenti di progettazione e sviluppo.

Per come abbiamo avuto modo di comprendere sino ad ora, la caratteristica principale dei sistemi di ML è la loro imprevedibilità prevedibile. Il produttore ed il programmatore di *healthcarebots*, ad esempio, sono consapevoli che il sistema di supporto alla diagnostica ed al trattamento potrà fornire raccomandazioni sbagliate. Lo stesso è a dirsi per i produttori di sistemi di guida autonoma: un *decision making* fondato su decisioni statistiche è destinato naturalmente a sbagliare.

Se il giudizio sulla prevedibilità dell'evento dovesse abbracciare ogni evento appartenente al genere di risultato indesiderato, l'unica regola cautelare da seguire dovrebbe consistere nella rinuncia all'attività produttiva. Se si dovesse applicare il vero fondamento del principio di precauzione in base a cui ogni scelta da compiersi in condizioni di incertezza va valutata in base alla peggiore delle sue conseguenze possibili[693], la diligenza imporrebbe di interrompere lo sviluppo e la commercializzazione di sistemi di IA che sono *in re ipsa* ingovernabili ed imprevedibili[694].

Secondo noi occorre indispensabilmente individuare un limite all'espansione del giudizio sulla prevedibilità generica dell'evento lesivo, al fine di creare un quadro di certezza giuridica per le imprese e di incentivo

[693]A. PERIN, *Prudenza, dovere di conoscenza e colpa penale*, cit., p. 81; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato,* cit., p. 142, che cita, come espressione di tale orientamento, la pronuncia della Cassazione sul disastro di Sarno, Cass., sez. IV, 11 marzo 2010, n. 16761: «il giudizio di prevedibilità andava compiuto […] ipotizzando quindi la più distruttiva ipotesi che potesse verificarsi […]. Se di un fenomeno naturale (o anche cagionato dall'uomo) non si conoscono le caratteristiche fondamentali – in particolare le cause, le possibilità di evoluzione, gli effetti possibili – la cautela che occorre usare nell'affrontarlo per eliminarne o ridurne le conseguenze deve essere ancora maggiore proprio perché non si possono escludere, con valutazione ex ante fondata su conoscenze scientifiche affidabili, gli effetti

[694] Sul punto vd. S. PREZIOSI, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 719.

all'innovazione. Altrimenti, l'alternativa sarebbe la cessazione dello svolgimento dell'attività pericolosa. Scelta, comunque, legittima se assunta dal decisore pubblico, ma che non può essere obbligata dall'autorità giudiziaria.

Ciò che occorre è la creazione di *standard* tecnici rigidi e precisi che non diano la possibilità di ingresso nelle valutazioni al concetto dell'agente modello, al di là dei casi in cui la diligenza imponga proprio la disapplicazione delle cautele positive ovvero ulteriori cautele. Leggendo le norme contenute nell'*AI Act*, si comprende come ci sia ancora molta strada da fare, in quanto trattasi di norme molto generiche ed imprecise. Certo è che si tratta di un testo legislativo di tipo orizzontale che fornisce un quadro normativo per lo sviluppo di tutti i sistemi di IA. Occorre la creazione di un apparato normativo organico e preciso per ogni singolo settore investito dall'IA.

Vista la difficoltà di incidere sulla fase esecutiva del funzionamento dei sistemi di IA, è necessaria una regolamentazione preventiva del *design* degli algoritmi di auto-apprendimento. Ruolo cardine nella definizione degli *standard*, in quest'ottica, lo avranno le stesse imprese. Esse, infatti, sono dotate di un'*expertise* tecnologica e di una conoscenza delle fonti di pericolo che non ha paragoni nel settore della vigilanza pubblica[695]. Dunque, un ruolo fondamentale nell'accertamento della colpa per gli eventi lesivi da IA lo avranno gli strumenti di coregolamentazione e di autonormazione che consentono di adeguare l'intervento preventivo sul design alle specifiche esigenze di gestione del rischio dell'impresa[696].

[695] G. FORTI, *"Accesso" alle informazioni*, cit., 192 ss.; C. PIERGALLINI, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, cit., p. 1128; C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, cit., p. 287.

[696] Sulla funzione di cooperative compliance che il diritto penale potrà svolgere nell'ambito della regolazione dei sistemi di i.a. vd. G.P. DEMURO, *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, cit., p. 73.

Inoltre, posta la natura sostanzialmente plurisoggettiva della responsabilità da sviluppo e produzione di IA, sarà fondamentale la responsabilità da reato degli enti[697]. Infatti, la colpa imputabile al produttore è una colpa da disorganizzazione o comunque da mancata o scorretta gestione del rischio. Anche se nel catalogo dei reati-presupposto sono assenti le ipotesi di omicidio e lesioni realizzate in violazione delle norme sulla sicurezza dei prodotti, l'apparato logico-giuridico attualmente formulato potrebbe ugualmente fornirci validi spunti dal punto di vista metodologico.

Le scelte di sviluppo e commercializzazione di un prodotto sono scelte che riguardano la politica d'impresa e che, per poter essere implementate correttamente, richiedono l'adozione di modelli organizzativi e di protocolli decisionali. In questo contesto di grande incertezza circa la lesività dei sistemi di IA, per poter individuare la diligenza necessaria nelle fasi di *design* occorre creare forme di procedimentalizzazione. Così ragionando, la diligenza del produttore persona fisica sarà sempre più simile alla diligenza richiesta all'ente. Ovverosia una diligenza che si sviluppa nella gestione di fattori di rischio imponderabili.

Merita un cenno anche la prospettiva che suggerisce di ricostruire la responsabilità tramite l'interazione tra tecnica giuridica, strategie organizzative e modelli computazionali. Secondo alcuni, in futuro, potremmo essere in grado di tutelare i beni giuridici attraverso una loro traduzione in termini computazionali, formulando impostazioni di *default* per i sistemi di IA. È la prospettiva

[697] B. PANATTONI, *Intelligenza artificiale*, cit., p. 362 ss.

del *Law as a code*[698]. Ovverosia di una *Legal Protection by Design*[699].

Se questa proposta ha il pregio di evidenziare il bisogno di un coinvolgimento del diritto fin dalle fasi del design, allo stesso tempo, non è risolutiva per il nostro tema. Infatti, anche se si implementassero tecniche per una *Legal Protection by Design*, ciò non potrebbe impedire che l'algoritmo intelligente, nel concreto, commetta errori ed imprecisioni. Si riproporrebbero i medesimi problemi già visti in tema di dimostrazione dell'elemento soggettivo circa il come individuare la regola cautelare e la sua esigibilità.

In ottica di reazione penalistica, la progettazione diligente dell'algoritmo rileva sul piano della colpa, ma come uno degli elementi che debbono essere considerati dal giudice nell'accertamento della responsabilità penale.

5. Soluzione forzata allo stato attuale: crisi del diritto penale d'evento e tutela anticipata.

Le difficoltà individuate nei paragrafi precedenti circa l'attribuzione della responsabilità penale in caso di eventi lesivi derivanti dai sistemi di IA sembrano riecheggiare la ormai ben nota crisi del diritto penale d'evento. Detta crisi si è delineata corrispondentemente all'emergenza del diritto penale del rischio[700], provocata dalla necessità di avvalersi di forme anticipate di tutela penale nel settore della tutela dei consumatori.

L'autonomia, la *black box*, l'imprevedibilità dell'agire algoritmico e l'interattività con l'ambiente circostante sono

[698] *Code is law* è il titolo di un celebre articolo del giurista statunitense Lawrence Lessig, vd. ID., *Code is Law. On liberty in Cyberspace*, *Harvard Magazine*, 2000.
[699] M. HILDEBRANDT, *Saved by Design? The Case of Legal Protection by Design*, in *NanoEthics*, 2017, p. 307 ss.; EAD, *Law for Computer Scientists and Other Folk*, cit., p. 251 ss.; cfr. A. GARAPON – J. LASSÈGUE, *Giustizia digitale*, *Il Mulino*, 2021, p. 249.
[700] Vd. per tutti M. DONINI, *Il volto attuale dell'illecito penale*, cit., p. 97 ss.; A. GARGANI, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato*, p. 403 ss.; C.E. PALIERO*, Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, cit., p. 290 ss.

tutti elementi che rendono tortuoso il cammino verso il riconoscimento di una responsabilità penale per il danno derivante dal prodotto. Per tale ragione, riconoscere una responsabilità del produttore – che rispetti i crismi di una responsabilità costituzionale – è molto complicato. In tale contesto, una tutela penale potrebbe manifestarsi in forma anticipata, tramite il ricorso a reati colposi di mera condotta, basati sul pericolo o, addirittura, solo sul rischio[701]. Tutela penale che, per la fondamentale importanza dei beni giuridici, appare indispensabile[702].

Pare che ci stiamo muovendo dal disvalore d'evento al disvalore d'azione[703]. Ciò potrebbe collocarsi nella più ampia posizione della dottrina che ritiene indispensabile un ripensamento della responsabilità colposa tutta. L'evento, nella colpa, costituendo il discrimine tra punizione ed impunità, finisce per rendere eccessivamente casuale la responsabilità stessa. Ciò in quanto la verificazione dell'evento dipende anche da fattori esterni di natura aleatoria[704]. Così facendo chi risponde

[701] È questa la tesi, in prospettiva *de jure condendo*, di Carlo Piergallini, in relazione ai prodotti tradizionali; vd. ID., *Danno da prodotto e responsabilità penale*, cit., p. 534 ss

[702] G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 56; G. FORTI, *"Accesso" alle informazioni sul rischio e responsabilità*, cit., p. 180; contra, com'è noto, F. STELLA, *Giustizia e modernità*, cit., passim, che giunge a sostenere l'impotenza del diritto penale di fronte alla società del rischio.

[703] Per le ragioni culturali che, nel nostro ordinamento, hanno favorito l'affermarsi del primato del disvalore dell'evento – sostanzialmente legate ad un'interpretazione "forte" del principio di offensività – si rinvia a M. MANTOVANI, *Contributo ad uno studio sul disvalore di azione nel sistema penale vigente*, *Bononia University Press*, 2014, passim, spec. 8 ss.; vd. anche N. MAZZACUVA, *Il disvalore di evento nell'illecito penale*, *Giuffrè*, 1983, p. 2 ss., che, tuttavia, nel suo lavoro, si propone di sottolineare proprio una tendenza evolutiva al ridimensionamento del disvalore dell'evento. In argomento vd. anche M. DONINI, *Il volto attuale dell'illecito penale*, cit., p. 197; nello specifico settore dei rapporti tra attività produttiva e principio di precauzione vd. C. PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto "a distanza" tra disvalore d'azione e disvalore d'evento*, cit., passim, spec. p. 240 ss.

[704] Su questi aspetti vd. L. CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, cit., p. 1 ss.; M.C. DEL RE, *Per un riesame della responsabilità colposa*, in *Indice penale*, 1985, p. 31 ss., nell'ambito di una critica complessiva alla colpa quale elemento soggettivo rimproverabile all'agente; cfr. D. CASTRONUOVO, *La colpa penale*, cit., p. 105 e ss., per una ricca ricostruzione, anche nella dottrina straniera, di queste posizioni. In argomento vd. anche G. MARINUCCI, *La colpa per inosservanza di leggi*, cit., p. 121 ss.; F. GIUNTA, *Illiceità e colpevolezza nella responsabilità colposa*, cit., p. 332 ss.; A.R. DI LANDRO, *La colpa medica negli Stati Uniti e in Italia. Il ruolo del diritto penale e il confronto col sistema civile*, Giappichelli, 2009; p. 271 ss. Il dibattito sull'opportunità di mantenere una responsabilità

dell'evento non è il più negligente, bensì il più sfortunato, anche perché il trattamento sanzionatorio stesso è calibrato sulla gravità dell'evento e non sulla gravità della violazione della regola cautelare[705].

Per come abbiamo sempre detto, l'IA acuisce problemi e rende attuali temi già affrontati nel dibattito penalistico e non. Le considerazioni appena svolte acquisiscono, infatti, una pregnanza maggiore in questo settore dato che l'evento lesivo algoritmico altro non è che l'esito finale di una serie di decisioni imponderabili ed imprevedibili in cui svolge un ruolo fondamentale il *decision making* algoritmico[706].

Introdurre reati colposi di mera condotta pare essere una soluzione percorribile per poter tutelare appieno ed in modo effettivo i consumatori. Possono applicarsi, *de jure condito*, le fattispecie di pericolo concreto ed astratto individuate dal codice penale a tutela dell'incolumità pubblica. Ma possono trovare applicazione anche le fattispecie contravvenzionali previste dall'art. 112 del codice del consumo che si applicano <<*se il fatto non costituisce più grave reato*>>. Prima di poter usare detta norma, occorre esser chiari sul concetto di prodotto rilevante ai sensi dell'art. 112 cod. cons. Nozione, questa, concettualmente diversa da quella fornita dall'art. 115 cod. cons. sulla quale ci siamo soffermati *supra*.

La nozione di prodotto generale è fornita dall'art. 3, comma 1, lett. E), codice del consumo. Per prodotto, si intende <<*qualsiasi prodotto destinato al consumatore, anche nel quadro di una prestazione di servizi, o suscettibile, in condizioni ragionevolmente prevedibili, di essere utilizzato dal consumatore, anche se non a lui destinato, fornito o reso disponibile a titolo oneroso o gratuito nell'ambito di un'attività*

penale per colpa è molto acceso negli Stati Uniti; per una lucida sintesi delle varie posizioni vd. L. ALEXANDER – K.K. FERZAN, *Against Negligence Liability*, in P.H. Robinson e al. (eds.), *Criminal Law Conversations, Oxford University Press*, 2011, p. 273 ss.

[705] L. CORNACCHIA, *Responsabilità colposa: irrazionalità e prospettive di riforma*, cit., p. 2; vd. anche L. EUSEBI, *Sistema sanzionatorio e reati colposi*, cit., p. 1201 ss.

[706] F. CONSULICH, *Flash offenders*, cit., p. 1053.

commerciale, indipendentemente dal fatto che sia nuovo, usato o rimesso a nuovo; tale definizione non si applica ai prodotti usati, forniti come pezzi d'antiquariato, o come prodotti da riparare o da rimettere a nuovo prima dell'utilizzazione, purchè il fornitore ne informi per iscritto la persona cui fornisce il prodotto>>.

La norma in parola è stata accusata di circolarità poiché sostanzialmente pare affermare che un prodotto è un prodotto. Ma, andando oltre questo, considerare l'algoritmo di IA come prodotto ai sensi dell'art. 3 appena citato lo si può fare solo con riferimento agli algoritmi incorporati in un bene destinato al consumatore o suscettibile di essere da lui utilizzato[707]. Non pare vi possa essere spazio per un'estensione dell'ambito applicativo della norma ai prodotti digitali, ovverosia ai software che non necessitano di essere incorporati in un hardware. Ciò è quanto, invece, potrebbe accadere con una adeguata interpretazione estensiva in relazione alla definizione fornita dall'art. 115 codice del consumo. Elemento normativo di tutte e tre le fattispecie è la nozione di pericolosità del prodotto che, secondo la dottrina maggioritaria, può essere derivata dalla definizione di prodotto sicuro delineata dall'art. 103, comma 1, lett. A)[708].

Per come già abbiamo avuto modo di illustrare, per "prodotto pericoloso" s'intende <<qualsiasi prodotto che non risponda alla definizione di prodotto sicuro>>, laddove per prodotto sicuro si intende un prodotto che <<in condizioni di uso normali o ragionevolmente prevedibili, compresa la durata e, se del caso, la messa in servizio, l'installazione e la manutenzione, non presenti alcun rischio oppure presenti unicamente rischi minimi,

[707] F. CONSULICH, *Flash offenders*, cit., p. 1039.

[708] D. CASTRONUOVO, *Principio di precauzione e beni legati alla sicurezza*, in *Dir. pen. cont*., 21 luglio 2011, p. 20; contra F. CONSULICH, (voce) *Tutela del consumatore*, in F. Palazzo – C.E. Paliero (diretto da), *Commentario breve alle leggi penali complementar*i, Cedam, II ed., 2007, p. 2980, che ritiene che la nozione di prodotto pericoloso alla quale fa riferimento l'art. 112 è più ristretta di quella stabilita, in via generale, dall'art. 103, lett. b), cod. cons.; la posizione contraria, infatti, determinerebbe la rilevanza penale di condotte di commercializzazione anche di prodotti che recano rischi minimi, in palese contrasto con i principi di offensività e ragionevolezza.

compatibili con l'impiego del prodotto e considerati accettabili nell'osservanza di un livello elevato di tutela della salute e della sicurezza delle persone>>, in considerazione di una serie di elementi, tra i quali la norma considera, in via esemplificativa, le sue <<caratteristiche>>, il suo eventuale <<effetto>> su altri prodotti, la sua <<presentazione>> e le <<categorie di consumatori>> cui esso è destinato, con particolare riguardo a quelle maggiormente vulnerabili, tra i quali i bambini e gli anziani.

Il codice del consumo prevede due presunzioni di sicurezza del prodotto. La prima sussiste in caso di rispetto della legislazione vigente nello Stato membro in cui il prodotto è commercializzato (art. 105, comma 1). Mentre, la seconda vige nel caso in cui vengano osservati gli *standard* tecnici definiti dalle norme armonizzate[709]. In assenza del, la sicurezza del prodotto non può essere presunta, ma è valutata in base <<alle norme nazionali non cogenti che recepiscono norme europee, alle norme in vigore nello Stato membro in cui il prodotto è commercializzato, alle raccomandazioni della Commissione europea relative ad orientamenti sulla valutazione della sicurezza dei prodotti, ai codici di buona condotta in materia di sicurezza vigenti nel settore interessato, agli ultimi ritrovati della tecnica, al livello di sicurezza vigenti nel settore interessato, agli ultimi ritrovati della tecnica, al livello di sicurezza che i consumatori possono ragionevolmente attendersi>> (art. 105, comma 3). Infine, al comma 4 si stabilisce che le Autorità amministrative competenti possono limitare o impedire l'immissione sul mercato del prodotto, o chiederne il ritiro o il richiamo, qualora questo si riveli,

[709] Come già osservato nel precedente capitolo, l'art. 105, co. 2, cod. cons. si riferisce alle norme armonizzate come alle «norme nazionali non cogenti che recepiscono le norme europee i cui riferimenti sono stati pubblicati dalla Commissione europea nella Gazzetta Ufficiale delle Comunità europee a norma dell'articolo 4 della direttiva 2001/95/CE, del Parlamento europeo e del Consiglio, del 3 dicembre 2001».

nonostante la conformità, pericoloso per la salute e la sicurezza del consumatore.

Dunque, anche il prodotto conforme può essere tanto difettoso quanto pericoloso. Pertanto, la mera immissione nel mercato di un prodotto conforme alla normativa vigente non esclude, di per sé, la responsabilità penale del produttore ai sensi dell'art. 112 codice del consumo[710].

5.1. Le fattispecie contravvenzionali di cui all'art. 112 cod. cons.

Il comma 1 dell'art. 112 usa la tecnica ingiunzionale nel punire l'immissione di prodotti pericolosi in violazione del divieto imposto dall'autorità ai sensi dell'art. 107, comma 2, lettera e)[711]. La pena prevista[712] è lievemente più alta rispetto a quella prevista dal comma 2 per l'immissione sul mercato di prodotti pericolosi senza che sia stato emesso un provvedimento amministrativo di divieto[713]. La maggiore offensività della condotta qui punita giustifica il regime sanzionatorio più severo previsto dal primo comma. Infatti, in questo caso, al pericolo per la salute e la sicurezza dei consumatori, si aggiunge anche l'offesa al buon andamento della PA[714].

Il comma 3 dell'art. 112 prevede il trattamento sanzionatorio più lieve[715] nel caso di mancata ottemperanza da parte del produttore o del distributore dei provvedimenti emessi dall'autorità amministrativa competente che siano diversi dal divieto di commercializzazione e consistenti, ad esempio, nell'ordine

[710] F. CONSULICH, (*voce) Tutela del consumatore*, cit., p. 2980; D. CASTRONUOVO, *La colpa penale*, cit., p. 310; A. BERNARDI, *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, cit., p. 16

[711] L'art. 107, co. 2, lett. e) stabilisce che le autorità amministrative preposte alla vigilanza sulla commercializzazione dei prodotti possano vietare l'immissione sul mercato dei prodotti pericolosi e adottare le misure necessarie a garantire l'osservanza dei divieti.

[712] Arresto da sei mesi ad un anno e ammenda da 10.000 a 50.000 euro.

[713] Stessa cornice edittale per l'ammenda ma minimo edittale di 15 giorni per la pena dell'arresto.

[714] F. CONSULICH, *Tutela del consumatore* (voce), cit., p. 2981.

[715] Ammenda da 10.000 a 25.000 euro.

di fornire avvertenze adeguate sui rischi che il prodotto può presentare o di disporre l'adeguamento di un prodotto già commercializzato agli obblighi di sicurezza previsti dal codice del consumo.

La struttura della appena menzionata fattispecie è stata letta dalla dottrina in termini di illecito di pericolo astratto, ma anche di mero rischio ovvero di mera disobbedienza[716]. Non può, di certo, trattarsi di una fattispecie di pericolo concreto, dato che la pericolosità viene valutata sulla base di caratteristiche astratte del prodotto, senza che vi sia la necessità di accertare in concreto le potenzialità lesive della merce[717]. Così come l'intera disciplina generale in materia di sicurezza del prodotto, dette fattispecie contravvenzionali sembrano essere ispirate da logiche precauzionali[718].

A noi pare che queste norme possano essere tranquillamente applicate anche in relazione ai sistemi di IA, anche se le sanzioni ivi previste non sono del tutto idonee ad avere una reale efficacia di deterrenza nei confronti dei comportamenti pericolosi per la salute e l'incolumità dei consumatori[719]. Ovviamente, se prodotti intelligenti dovessero essere immessi nel mercato con la marcatura CE indebitamente replicata, ai produttori saranno applicabili anche le fattispecie di cui agli artt. 514 ss. c.p., eventualmente in concorso con quelle previste dall'art. 112 cod. cons.[720].

5.1.2. Nuove soluzioni: *de jure condendo*.

Si è già detto che le norme appena citate si prestano ad essere applicate anche in relazione ai sistemi di IA ma esse

[716] F. CONSULICH, *Tutela del consumatore* (voce), cit., p. 2981; D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 76.
[717] F. CONSULICH, *Tutela del consumatore* (voce), cit., p. 2981.
[718] D. CASTRONUOVO, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, cit., p. 76.
[719] F. CONSULICH, *Flash offenders*, cit., p. 1039.
[720] Sull'applicabilità degli artt. 514 c.p. ss. nel contesto della responsabilità per danno da prodotto vd. A. BERNARDI, La responsabilità da prodotto nel sistema italiano: profili sanzionatori, cit., p. 20 ss.

rischiano di avere una deterrenza molto limitata, vista la cornice edittale bassa. Potrebbe essere necessaria la introduzione di fattispecie incriminatrici *ad hoc*, idonee a dare specifico rilievo alle condotte pericolose inerenti alla produzione ed all'utilizzo di sistemi di IA.

Lo stesso *AI Act* pare dare una indicazione in tal senso, prevedendo che gli Stati Membri debbano introdurre sanzioni <<effettive, proporzionate e dissuasive>> per il caso di mancato rispetto della disciplina stabilita dal regolamento stesso. L'approccio del decisore europeo pare mostrare un *favor* per le sanzioni amministrative pecuniarie, anche se i singoli Stati membri possono optare per l'introduzione di idonee fattispecie incriminatrici[721].

A nostro giudizio, sarebbe opportuna una tutela anticipata per il mancato rispetto dei requisiti stabiliti dall'*AI Act*. Nel dettaglio, parte della dottrina ha proposto di introdurre reati colposi di mera condotta, indirizzati a criminalizzare l'omessa predisposizione da parte del programmatore di meccanismi preventivi nell'algoritmo che governa il sistema di IA, nonché la disattivazione, l'omessa attivazione o omesso aggiornamento da parte dell'utilizzatore dei presidi di sicurezza di cui invece fosse stato dotato[722].

La prima proposta potrebbe rientrare nella fattispecie di cui all'art. 112 codice del consumo, visto che l'omessa predisposizione da parte del produttore di meccanismi di prevenzione del danno rientra certamente nella fattispecie di pericolosità generica prevista dall'art. 103, comma 1, lett. a), codice del consumo. Magari, vista la maggiore pericolosità degli strumenti di IA, sarebbe il caso di introdurre una nuova fattispecie con una cornice edittale

[721] In argomento vd. C. CASONATO – B. MARCHETTI*, Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, cit., p. 435 ss.
[722] Così F. CONSULICH, *Flash offenders*, cit., p. 1038-1039; vd. anche F. LAGIOIA – G. SARTOR, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective,* cit., p. 459.

più severa di quella al momento prevista dall'art. 112 codice del consumo.

La seconda proposta potrebbe essere in grado di coprire il *vulnus* presente nel nostro sistema, andando a prevenire le conseguenze dannose all'integrità ed all'incolumità fisica che possono derivare dall'utilizzo di sistemi di IA in assenza delle necessarie precauzioni. A nostro giudizio, un parametro di riferimento per la formulazione di tale fattispecie potrebbe essere fornito dall'art. 427 c.p., che criminalizza la rimozione od omissione dolosa di cautele contro infortuni sul lavoro, limitandola ai soli *embodied systems*. Poi, la particolarità dell'imprevedibilità del *decision making* suggerisce di introdurre una fattispecie di pericolo astratto.

Anche il modello ingiunzionale potrebbe essere una valida scelta da seguire, in prospettiva *de jure condendo*[723]. Così facendo, innanzitutto, si favorirebbe il coordinamento tra sanzione penale e controllo amministrativo[724]. Le autorità nazionali competenti ingiungono di assumere misure di sicurezze. La mancata assunzione delle stesse dà luogo ad una condotta *in re ipsa* illecita dal punto di vista penale. Il paradigma ingiunzionale è molto efficace sul piano della tempestività e sul piano dell'individualizzazione delle condotte doverose perché le autorità competenti ingiungono all'impresa di compiere prestazioni di facere individualizzate ed infungibili[725].

6. Facciamo un punto: IA e responsabilità penale.

[723] Di quest'opinione è C. PIERGALLINI, *Intelligenza artificiale*, cit., p. 1773.
[724] In generale, sulla tecnica ingiunzionale vd. G. MARINUCCI, *Innovazioni tecnologiche e scoperte scientifiche*, cit., p. 56 ss.; A. ALESSANDRI, *Parte generale*, in C. Pedrazzi e al. (a cura di), *Manuale di diritto penale dell'impresa*, cit., p. 50 ss.; F. GIUNTA, *Il diritto penale e le suggestioni del principio di precauzione*, cit., p. 247; C. PONGILUPPI, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto "a distanza" tra disvalore d'azione e disvalore d'evento*, cit., p. 252 ss.; C.E. PALIERO, *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, cit., p. 295.
[725] È questo uno dei più apprezzabili vantaggi del modello ingiunzionale; vd. a tal proposito A. ALESSANDRI, *Parte generale*, in C. Pedrazzi e al. (a cura di), *Manuale di diritto penale dell'impresa*, cit., p. 51-52.

Si è avuto modo di comprendere che le peculiari caratteristiche dell'IA – autonomia, interattività ed opacità – sono dirompenti ai fini dell'imputazione penale degli eventi lesivi algoritmici. In quest'ultima parte del nostro lavoro, abbiamo provato ad indagare circa la fattibilità o meno delle prospettive di incriminazione dell'utilizzatore e del produttore di un sistema di IA arrivando a suggerire al decisore di utilizzare tecniche di anticipazione della tutela penale, ricorrendo ai reati di mera condotta.

Con riferimento alla posizione dell'utilizzatore, si è notato che i problemi più importanti sorgono in relazione ai sistemi parzialmente autonomi perché, lì, la funzione di governo sul dispositivo è condivisa tra operatore ed algoritmo di ML. Sorgono, invece, questioni già ben note rispetto agli eventi lesivi provocati dai sistemi di IA completamente autonomi, laddove l'utente si limita a fruire di una attività totalmente svolta dall'algoritmo intelligente.

Abbiamo constatato che l'autonomia dei sistemi di IA comporta certamente la diminuzione degli eventi lesivi provocati dall'errore umano, ma, allo stesso tempo, essa causa nuovi rischi. Siamo stati in grado di individuare tre fattori relativi all'autonomia, in grado di incidere sull'esigibilità della condotta dell'agente confacente alle regole cautelari. Esse sono la distorsione dell'automazione, l'opacità e la ridotta capacità di reazione durante le situazioni di emergenza. L'opacità non consente di riconoscere l'errore algoritmico e di intervenire; la distorsione dell'automazione è la tendenza dell'essere umano ad affidarsi eccessivamente alla scienza tecnica come fosse un oracolo; la ridotta capacità di reazione durante le situazioni di emergenza è collegata alla precedente poiché è chiaro che il livello di attenzione di chi si limita a monitorare l'altrui comportamento è molto più debole rispetto a quello di chi agisce in prima persona.

Abbiamo valutato la possibilità di creare uno scudo penale a favore dell'operatore, ragionando circa la proponibilità, in questo settore, di una causa di non punibilità come quella prevista dall'art. 590-*sexies* c.p. per i medici. Abbiamo anche valutato la possibilità di restringere il campo della colpa penale alle sole ipotesi di colpa grave. E, in verità, occorre sottolineare che tali suggestioni si inseriscono perfettamente nella discussione sempre presente in dottrina circa un eventuale ripensamento del paradigma colposo, a favore della maggior determinatezza della fattispecie e della maggior valorizzazione del momento soggettivo della colpa.

Ancora con riferimento alla posizione dell'utilizzatore, abbiamo analizzato le ipotesi di criminalità dolosa. *Nulla quaestio* con riferimento alle ipotesi in cui il dispositivo intelligente realizzi il piano criminoso voluto dall'agente. Le problematiche sorgono allorquando si verifica una divergenza tra il voluto ed il realizzato. Ci stiamo riferendo alle ipotesi di *aberractio ictus* in ogni sua forma, *aberractio delicti*, preterintenzione, nonché delitti aggravati dall'evento. In queste ipotesi, a nostro giudizio, si deve avere in mente la giurisprudenza costituzionale che subordina l'attribuzione della responsabilità penale alla presenza di una ragionevole prevedibilità ed evitabilità in concreto della deviazione – in questo caso algoritmica - rispetto al piano criminoso.

La concreta integrazione delle fattispecie dolose incontrerà un ostacolo nell'ipotesi in cui l'operatore di un sistema algoritmico si limiti a fornire istruzioni generali per la commissione del delitto, lasciando la scelta sull'*an*, sul *quomodo* e sul *quando* commettere il fatto delittuoso. Qui entra in gioco la scelta politica circa l'irrilevanza del *dolus generalis* che, con riferimento a strumenti come gli *High Frequency Trading* (HFT), rappresenta una lacuna normativa evidente, visto che l'agente potrebbe sfruttare proprio l'interposizione del *decision making* algoritmico per commettere crimini restando impunito.

Ma veniamo alla responsabilità del produttore dei sistemi di IA, fulcro della nostra trattazione. Stante l'autonomia crescente attribuita in capo ai dispositivi intelligenti, lo spazio di intervento dell'utilizzatore è drasticamente ridotto. Ne consegue che il produttore sia il soggetto maggiormente qualificato per intervenire sulla sicurezza dei prodotti da lui immessi nel mercato al fine di prevenire il verificarsi di eventi dannosi.

Collocato all'interno della società del rischio, il reato da intelligenza artificiale altro non è che l'ultima tappa della crisi del diritto penale d'evento già da tempo denunciata dalla dottrina. Il danno da IA non fa altro che riprodurre temi e problemi già noti: la fungibilità tra struttura commissiva ed omissiva del reato, la difficile individuazione dei soggetti personalmente responsabili all'interno delle organizzazioni complesse, l'accertamento del nesso di causalità e della colpa in situazioni di incertezza scientifica. Ciò che aggrava le questioni, rispetto al passato, è proprio l'incerto decidere del *decision making* algoritmico che porta la verifica circa la sussistenza del nesso di causalità a livelli di complessità mai conosciuti fino ad ora. E lo stesso accade con riferimento al requisito della prevedibilità, in concreto, dell'evento lesivo.

È complicato ricondurre i danni da IA al paradigma dei reati d'evento proprio a causa dell'opacità degli algoritmi di ML. La crisi della comprensione dei processi causali che determinano il verificarsi dell'evento lesivo comporta ripercussioni anche sulla struttura del reato e, nel dettaglio, sull'accertamento dell'elemento soggettivo. A differenza dei fenomeni di opacità causale passati in cui il problema della *black box* causale era un problema transitorio poiché risolvibile con l'acquisizione di nuove conoscenze scientifiche, il fenomeno del ML sembra essere refrattario intrinsecamente alle generalizzazioni causali che costituiscono il fondamento del metodo scientifico. È la stessa decisione dei sistemi di auto-apprendimento ad

essere fondata su modelli probabilistici che impediscono di asserire che all'*input* X seguirà l'*output* Y.

Forse, la scienza dell'*explainability* e della *white box* potrà, in futuro, consentirci di rendere più limpido, *ex post*, il decorso eziologico interno al *decision making* algoritmico, anche tramite la predisposizione di appositi strumenti di reverse *engineering*. Ma, anche se si arrivasse a questo, non ci pare possibile che si possa riconoscere una regolarità nel comportamento algoritmico dei sistemi moderni che assumono autonome decisioni[726].

Per questa ragione, abbiamo ritenuto di avanzare l'ipotesi secondo cui l'accertamento della causalità in relazione ai danni provocati dai sistemi di IA possa seguire un paradigma diverso da quello nomologico-deduttivo, così come avviene con riferimento alla causalità psichica, per la quale non sono disponibili leggi scientifiche. I sistemi di ML sono ad oggi capaci di imparare dall'esperienza e di migliorare le proprie performance sulla base dei *feedback* ricevuti, di scegliere le strategie migliori per arrivare all'obiettivo. Per fare ciò, si basano, proprio come fa il cervello umano, sull'induzione e sull'analogia. Non già sull'applicazione deduttiva di principi causalistici.

Stante questa similarità tra processi decisionali umani e quelli algoritmici, si crede che si possa derogare al paradigma nomologico-deduttivo, lasciando spazio alla causalità psichica. È ovvio che, anche nel caso in cui si condividesse una tale asserzione, lo standard rigoroso dell'oltre ogni ragionevole dubbio imporrebbe un vaglio induttivo ancora più preciso sulla causalità individuale, tramite una scrupolosa valutazione delle emergenze probatorie e l'esclusione di decorsi causali alternativi. Se, invece, si pensasse che è inderogabile il criterio della

[726] Non si esclude, tuttavia, un possibile ritorno agli schemi simbolico-causali che hanno caratterizzato lo sviluppo dei primi sistemi di i.a., a metà del secolo scorso. In questo caso, tuttavia, potrebbe essere necessario rinunciare alla stupefacente accuratezza degli attuali sistemi che utilizzano le reti neurali, basati su meccanismi decisionali di tipo induttivo.

sussunzione sotto leggi scientifiche, l'accertamento della causalità in relazione ai danni da IA potrebbe risultare impossibile.

Ciò in quanto, anche nel caso in cui si riuscisse a provare il nesso di causalità, la dimostrazione dell'elemento soggettivo del produttore resterebbe molto complicata. Nel dettaglio, posto che l'attività illecita del produttore, nella maggioranza dei casi, non è volontaria, nel nostro lavoro ci siamo concentrati maggiormente sulla possibilità di muovere allo stesso un rimprovero di tipo colposo.

Abbiamo visto che l'imprevedibilità genericamente prevedibile dei sistemi di IA[727] mette in discussione la stessa possibilità di muovere un rimprovero colposo al produttore. Ciò soprattutto se si considera che, secondo il criterio di copertura del rischio tipico, affinchè l'evento lesivo possa essere rimproverato all'agente, esso deve costituire la realizzazione specifica del rischio che la norma cautelare aveva intenzione di prevenire. Senza conoscenze specifiche ed approfondite riguardo al funzionamento degli algoritmi di ML, gli eventi lesivi derivanti dai sistemi di IA si possono considerare prevedibili con riferimento all'*an*, ma imprevedibili con riferimento al *quantum* ed al *quomodo*. Senza dubbio, questa prevedibilità è scarna: non è tale da determinare un esito positivo nel giudizio di prevedibilità dell'evento lesivo.

Occorre, comunque, rilevare che si è oramai affermato in giurisprudenza il principio in forza di cui il giudizio di prevedibilità non riguarda l'evento realizzatosi in quel determinato modo, quanto, piuttosto, la classe generica di eventi in cui si colloca quello oggetto del processo. Dunque, se prevale questo orientamento, la colpa del produttore potrebbe sempre essere ritenuta sussistente,

[727] C. PIERGALLINI, *Intelligenza artificiale*, 1762; vd. anche M.B. MAGRO, *Robot, cyborg e intelligenze artificiali*, cit., p. 1208-1209.

posto che il produttore di un sistema di IA potrà sempre prevedere una classe determinata di eventi algoritmici lesivi. Mentre, se prevale la concezione restrittiva, difficilmente si potrebbe ritenere esistente una colpa in capo al produttore, posto il carattere di imprevedibilità intrinseca dell'evento lesivo realmente verificatosi.

Come argine alla espansione del giudizio di prevedibilità ed evitabilità dell'evento, si potrà, per come abbiamo notato, ragionare su un'area di rischio consentito. Essa coincide con quel pericolo marginale che le norme cautelari scritte non sono in grado di disinnescare e che, pertanto, ricade sulla società. Anche se è complicato definire regole cautelari in un contesto di incertezza scientifica, riteniamo che una minimizzazione del rischio di decisioni erronee da parte dei sistemi di ML potrà realizzarsi tramite misure di corretta raccolta e gestione dei dati di addestramento, di sperimentazione e nella fase di monitoraggio *post-market*.

Abbiamo notato, a questo riguardo, come i meccanismi di *explainability* potrebbero avere un ruolo fondamentale nel monitoraggio dell'adeguatezza delle cautele positive rispetto agli obiettivi di bilanciamento decisi dal legislatore. Gli sviluppi nella registrazione algoritmica potrebbero permettere di verificare se e quando emergano rischi più elevati o ulteriori rispetto a quelli che le cautela volevano neutralizzare o ridurre. Così consentendo ai produttori di riconoscere i segnali univoci di fallimento delle regole cautelari che comportano l'obbligo di informazione alle autorità ed ai consumatori e, nei casi più gravi, di ritiro o richiamo del prodotto.

Se obiettivo del decisore pubblico è, tramite le norme cautelari rigide, quello di mantenere gli incidenti algoritmici al di sotto di una certa soglia, il superamento di quest'ultima sarà chiaramente un segnale di fallimento delle regole scritte, che dovrà indurre il produttore ad adottare cautele più resistenti.

Concordiamo con quella dottrina secondo cui il diritto penale soffra una vera e propria crisi da incontenibilità con riferimento alle sfide tipiche della società del rischio[728]. Tale crisi è determinata dalla difficoltà di ritenere rientrante nel "tipo" fenomeni complessi ed imperscrutabili. Ecco che, in un contesto siffatto, una tutela penale valida può essere solo in forma anticipata, tramite il ricorso ai reati colposi di mera condotta, eventualmente fondati sulla tecnica ingiunzionale. Potranno applicarsi già le fattispecie previste dall'art. 112 codice del consumo, anche se le cornici edittali risultano troppo basse per dissuadere gli agenti. In futuro, si potranno introdurre norme *ad hoc* per sanzionare il mancato rispetto di alcuni requisiti stabiliti dall'*AI Act*. Si potrà, ad esempio, punire l'omessa predisposizione di determinati presidi di sicurezza, con riferimento alla posizione del produttore. Mentre, con riferimento alla posizione dell'utilizzatore del sistema di IA, si potrà punire la disattivazione dei presidi di sicurezza prima dell'utilizzo.

Sempre in ottica di legislazione futura, potrebbero trovare valorizzazione le tecniche ingiunzionali, in modo tale da poter favorire un coordinamento tra disciplina penale e controllo amministrativo.

Inoltre, vista la natura spesso collettiva del centro decisionale relativo ai sistemi di IA, occorrerebbe, con riferimento alla posizione del produttore estendere i reati-presupposto anche alle fattispecie di pericolo o di danno relative alla sicurezza dei prodotti. Così operando, l'art. 8 d.lgs. n. 231/2001 consentirebbe di muovere un rimprovero alla persona giuridica anche nell'ipotesi in cui non si possa identificare la persona responsabile del reato-presupposto. Lo stesso meccanismo potrebbe risultare inadeguato nel caso in cui la scelta di politica d'impresa non si concretizza tramite l'attività di un dipendente ma con quella di un algoritmo. Qui la mancata identificazione

[728] C.E. PALIERO, *L'autunno del patriarca*, cit., p. 1238.

dell'autore non è causata da una disorganizzazione dell'impresa, quanto, piuttosto, alla stessa incerta configurabilità di un reato, visto che tra la scelta di usare il sistema di IA e l'evento lesivo si verifica la intermediazione del *decision making* algoritmico.

CONCLUSIONI

All'inizio del nostro lavoro, nella descrizione della *black box* tipica dell'IA, abbiamo constatato che l'opacità dei sistemi di IA pare costituire la tappa definitiva di quella che è stata spesso definita come crisi di comprensione e controllo che caratterizza il rapporto post moderno tra uomo e tecnologia. Il dominio sul mondo, ad oggi, si realizza tramite la creazione di agenti intelligenti capaci di eseguire in modo accurato le funzioni loro deputate senza che i processi decisionali che ne regolano il funzionamento possano essere comprensibili appieno da parte dei loro stessi ideatori e produttori.

Il fulcro del nostro lavoro è stato comprendere l'impatto che le nuove forme dell'agire algoritmico possono avere sui crismi dell'imputazione penale degli eventi lesivi cagionati dai sistemi di IA. Per come si è avuto modo di constatare, non si tratta solo di questioni astratte, ma di problemi già presentatisi negli ultimi anni.

In una prima fase, l'impatto dell'IA sulla responsabilità civile è stato oggetto, da parte delle istituzioni europee, di atti di indirizzo e di *soft law*. Dopodiché, la Commissione europea ha presentato una serie di proposte che avevano l'obiettivo di adattare la disciplina sulla sicurezza dei prodotti ed i regimi di responsabilità aquiliana al nuovo contesto tecnologico. Nel dettaglio, l'*AI Act* ha l'obiettivo di stabilire i requisiti minimi per lo sviluppo e la messa in commercio dei sistemi di IA; mentre, la proposta di revisione della direttiva sul danno da prodotto e la proposta di direttiva sulla responsabilità da IA servono per alleggerire l'*onus probandi* del ricorrente che abbia bisogno di ottenere ristoro per i danni derivanti dall'uso di un sistema di IA, nonché a garantire la *disclosure* degli elementi di prova rilevanti da parte del produttore.

L'*AI Act* ha come obiettivo il bilanciamento delle esigenze dell'industria e dell'innovazione con il bisogno di sicurezza dei nuovi prodotti che si immettono sul mercato. Le proposte

in materia di responsabilità cercano di perseguire l'obiettivo di fornire un rimedio effettivo ai soggetti danneggiati, indi per cui si possano superare le difficoltà probatorie insite nell'accertamento del nesso di causalità nell'ambito dei danni causati da sistemi intelligenti.

La responsabilità extracontrattuale si conferma strumento duttile e flessibile poiché in grado di adattarsi ai nuovi fenomeni dannosi, nonché di allinearsi alle decisioni assunte dal decisore politico sulla gestione del rischio. Di fatto, le regole sulla responsabilità per fatto illecito hanno attraversato la rivoluzione industriale e la post-modernità senza subire radicali modifiche e senza tradire il proprio ruolo: esse hanno, anzi, reso ancora più forte il proprio ruolo centrale nella tutela dei danneggiati.

Noi riteniamo, insomma, che il diritto civile sia ben attrezzato per dare risposte alle sfide imposte dall'IA. Ciò lo si deve anche allo *standard* probatorio molto meno rigoroso rispetto a quello proprio del diritto processuale penale. La formula del diritto processuale civile è quella del più probabile che non, mentre nel diritto processuale penale dell'oltre ogni ragionevole dubbio. A nostro giudizio, il diritto civile costituisce proprio il settore dove può essere grande la sperimentazione di soluzioni innovative. Così come può essere terreno di estensione di modelli tradizionali di responsabilità oggettiva alle nuove forme di insidia provocate dall'IA.

Di contro, il diritto penale rischia di subire la fenomenologia del danno provocato dai sistemi di IA. Il diritto penale era già stato messo a dura prova dalla fattispecie del danno da prodotto tradizionale ed era già stato piegato ad esigenze di natura *extra*-penale quali la rassicurazione sociale, il conforto delle vittime etc. Ciò in relazione anche ai danni da esposizione alle sostanze tossiche ed alle calamità naturali. Il diritto penale di fronte all'IA potrebbe trovarsi in un terreno molto insidioso soprattutto

per le possibili frizioni con i principi costituzionali ed il bisogno di pena percepito dalla società.

Lungi dal considerare percorribile l'ipotesi di una responsabilità penale diretta dei sistemi di IA, i dubbi che riguardano la configurabilità di una responsabilità di tipo penale in capo all'utilizzatore ed al produttore risultano fondati. Ciò in quanto vi è il rischio, con riferimento all'utilizzatore, che gli vengano attribuiti obblighi di controllo su sistemi il cui funzionamento è incomprensibile ed imperscrutabile. Con ciò provocando frizioni col principio di colpevolezza. Inoltre, con riferimento alla figura del produttore, l'assenza di conoscenze scientifiche circa il funzionamento dei sistemi di IA mette in crisi il modello nomologico-deduttivo di accertamento del nesso eziologico. Con ciò provocando l'inadeguatezza della valutazione circa la prevedibilità dell'evento, necessaria per l'accertamento della colpa.

A noi pare che il tema dell'IA nel diritto penale ci consenta di osservare, da una posizione privilegiata, talune tendenze del diritto penale che già si erano verificate di fronte a contesti di iper-complessità tecnologica. Inoltre, affrontare le relazioni tra IA e diritto penale ci consente di riflettere sulla necessità di interventi normativi ovvero sulla necessità di rinnovare i tradizionali paradigmi interpretativi.

Ad esempio, con riferimento alla nozione della colpa, nel nostro studio, è emerso, da un lato, il bisogno di valorizzare la dimensione soggettiva della colpa stessa soprattutto con riguardo all'utilizzatore e, dall'altro lato, l'opportunità di creare regole cautelari rigide ed esaustive che siano capaci di delineare un'area di rischio consentito per il produttore dei sistemi di IA. Misura soggettiva e maggiore determinatezza della fattispecie colposa sembrano essere due istanze che, da sempre al centro del dibattito penalistico, trovano terreno fertile nel settore dell'IA.

Così come abbiamo visto che, in relazione ai danni da IA si potrebbe acuire la crisi del diritto penale d'evento già

esistente da quando si è iniziato a parlare di società del rischio. Stante ciò, si apre la strada per una maggiore attenzione al disvalore d'azione, tramite la predisposizione di forme anticipate di tutela penale. A ciò si addiviene ragionando su un dato: se l'evento lesivo algoritmico non è mai preventivabile appieno né tantomeno evitabile da parte di produttore ed utilizzatore, il nucleo del disvalore penale dovrà essere necessariamente incentrato sul fatto della commercializzazione e/o dell'utilizzazione di un prodotto intelligente in assenza delle cautele richieste.

Infine, nel settore della produzione di questi sistemi intelligenti, si nota la disarmonia tra carattere plurisoggettivo degli illeciti e conformazione personalistica della responsabilità penale. Problematica, questa, già riscontrata in generale nei contesti della c.d. criminalità d'impresa. Inserire, nei processi decisionali aziendali, algoritmi che interagiscono con i dipendenti ed agiscono nell'interesse o a vantaggio dell'ente potrebbe amplificare quelle istanze derivanti da una parte della dottrina che ritengono di dover rendere maggiormente autonoma la responsabilità da reato degli enti rispetto a quella della persona fisica, "giocando" sull'art. 8 del decreto legislativo n. 231/2001.

Sullo sfondo della nostra ricerca, si attesta l'interrogativo circa il ruolo del diritto penale in una società sorretta da meccanismi artificiali, il cui funzionamento non è conoscibile del tutto neppure ai loro creatori. La sfida da tenere sempre a mente è quella di salvaguardare il compito del diritto penale di essere una scienza utile per la società senza, però, rinunciare alle garanzie fondamentali dello stesso.

INDICE BIBLIOGRAFICO

ABRAHAM, K.S. – RABA, R.L., *Self-driving cars and tort law*, in *Virginia Law Review*, 2019.

AIRES DE SOUSA, S., *"Não fui eu, foi a macchina": teoria do crime, responsabilidade e inteligência artificial*, in *Inteligência artificial no diritto penal*, Almedina, 2020.

ALBANESE, A., *La responsabilità civile per i danni da circolazione di veicoli ad elevata automazione*, in *Europa e dir. priv.*, 2019.

ALBUS, J.S., *Outline for a theory of intelligence*, in *IEEE Transactions on Systems, Man, and Cybernetics*, 1991.

ALEXE, P. – CUNNINGHAM, F., *AI on track to be formally adopted ahead of the European elections*, in *Lexology*, 2024.

ALPA, G., *Responsabilità dell'impresa e tutela del consumatore*, Giuffrè, 1975.

AMIDEI, A., *Intelligenza Artificiale e product liability: sviluppi del diritto dell'Unione Europea*, in *Giur. it.*, 2019.

AMIDEI, A., *Le responsabilità del produttore di intelligenza artificiale «difettosa»*, in *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 2, il Mulino, 2022.

AMISANO, M., *Prevedere -e non predire- attraverso gli algoritmi e le loro insidie*, in *Arch pen.*, 2022.

AMOROSO, D. – TAMBURRINI, G., *I sistemi robotici ad autonomia crescente tra etica e diritto*, in *BioLaw Journal*, 2020.

ANGWIN, J. e al., *Machine Bias*, in *ProPublica*, 2016.

ASARO, P., *A Body to Kick, but Still No Soul to Damn*, in *Robot Ethics,* MIT Press, 2012.

ASARO, P., *How just could a robot war be?*, in *Robot Ethics*, MIT Press, 2012.

ATHALYE, A. E AL., *Synthesizing Robust Adversarial Examples*, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

AULETTA, K., *Googled: The End of the World as We Know It*, Penguin Press, 2009.

AZZALI, G., *La responsabilità penale del produttore per danni alla salute*, in *Riv. trim. dir. pen. ec.*, 1991.

BAMBERG, D. et alii, *The EU AI Act: an impact analysis*, in *Lexology*, 2024.

BANTERLE, F. – SCHETTINO, A., *Copyright provisions in the AI Act*, in *Lexology*, 2024.

BARTOLI, R., *(voce) Fonti della colpa*, in *Reato colposo, Enc. dir. – I Tematici*, 2022.

BARTOLINI, F., *Auto a guida autonoma e problemi di responsabilità civile*, in *Responsabilità Civile e Previdenza*, 2017.

BARTOLUCCI, M.A., *L'art. 8 d.lgs. 231/2001 nel triangolo di Penrose*, in *Dir. pen. cont.*, 2017.

BASILE, F., *Intelligenza artificiale e diritto penale*, in *Riv. it. dir. proc. pen.*, 2022.

BASILE, F., *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in *Sist. pen.*, 2023.

BASILE, F. – VALLAR, G., *Diritto penale e neuroscienze*, in *Giornale Italiano di Psicologia*, 2016.

BATHAEE, Y., *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, 2018.

BAUM, S.D. E AL., *How long until human-level AI?*, in *Technological Forecasting and Social Change*, 2011.

BECK, U., *La società del rischio. Verso una seconda modernità*, Carocci, 2000 (ed. orig. 1986).

BECKERS, A. – TEUBNER, G., *Three Liability Regimes for Artificial Intelligence*, in *European Review of Private Law*, 2021.

BEKEY, G.A., *Current Trends in Robotics: Technology and Ethics*, in *Ethics and Information Technology*, 2012.

BELVEDERE, A., *(voce) Definizioni*, in *Digesto*, IV ed., 1989.

BELVEDERE, A., *Il problema delle definizioni nel codice civile*, Giuffrè, 1977.

BERNARDI, A., *La responsabilità da prodotto nel sistema italiano: profili sanzionatori*, in *Riv. trim. dir. pen. ec.*, 2003.

BERTOLINI, A., *Artificial Intelligence Does Not Exist! Defying the Technology-Neutrality Narrative*, in *Europa e dir. priv.*, 2022.

BICHI, R., *Intelligenza digitale, giurmetria, giustizia predittiva e algoritmo decisorio*, in *Intelligenza artificiale - Il diritto, i diritti, l'etica*, 2020.

BLAIOTTA, R., *Causalità giuridica*, in *Cass. pen.*, 2007.

BOSTROM, N., *Superintelligenza*, Bollati Boringhieri, 2014.

BRAYNE, S., *Predict and Surveil: Data, Discretion, and the Future of Policing*, Oxford University Press, 2020.

BRICOLA, F., *Responsabilità penale per il tipo e per il modo di produzione*, in *Riv. it. dir. proc. pen.*, 1983.

BRICOLA, F., *Teoria generale del reato*, in *Noviss. dig. it.*, 1973.

BRUSCO, C., *Rischio e pericolo, rischio consentito e principio di precauzione*, in *Cass. pen.*, 2006.

BURCHARD, C., *L'intelligenza artificiale come fine del diritto penale?*, in *Arch. pen.*, 2019.

BURRELL, J., *How the machine 'thinks'*, in *Big Data & Society*, 2016.

BUYERS, J. et alii, *Debate continues over the pros and cons of regulating Artificial Intelligence*, in *Lexology*, 2024.

CADOPPI, A. (a cura di), *Il problema delle definizioni legali nel diritto penale*, Cedam, 1996.

CALABRESI, G. – AL MUREDEN, E., *Driverless Cars*, Il Mulino, 2021.

CALO, R., *Artificial Intelligence Policy: A Primer and Roadmap*, in *University of Bologna Law Review*, 2018.

CALO, R., *Robotics and the Lessons of Cyberlaw*, in *California Law Review*, 2015.

CANEPA, A., *L'imputazione soggettiva della colpa*, Giappichelli, 2011.

CANZIO, G., *Intelligenza artificiale, algoritmi e giustizia penale*, in *Sist. pen.*, 2021.

CAPPELLINI, A., *Machina delinquere non potest?*, in *Studi sulla questione criminale*, 2018.

CAPPELLINI, A., *Profili penalistici delle self-driving cars*, in *Dir. pen. cont. – Riv. Trim.*, 2019.

CAPPELLINI, A., *Reati colposi e tecnologie dell'intelligenza artificiale*, in *Arch. pen.*, 2021.

CARLE, M. et alii, *Key considerations regarding the recently passed EU Artificial Intelligence Act*, in *Lexology*, 2024.

CARLETTI, E. – LUCANI, L., *AI Act e copyright*, in *Lexology*, 2024.

CARNEVALI, U., *Prevenzione e risarcimento nelle direttive comunitarie sulla sicurezza dei prodotti*, in *Quadrimestre*, 1989.

CARUANA, R. E AL., *Intelligible Models for Healthcare*, in *KDD '15: Proceedings of the 21th ACM SIGKDD*, 2015.

CASONATO, C. – MARCHETTI, B., *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal*, 2021.

CASTELVECCHI, D., *Can We Open the Black Box of AI?*, in *Nature*, 2016.

CASTRONUOVO, D., *La colpa penale*, Cedam, 2009.

CASTRONUOVO, D., *Principio di precauzione e diritto penale*, Aracne, 2012.

CHAGAL-FEFERKORN, K.A., *The Reasonable Algorithm*, in *University of Illinois Journal of Law, Technology & Policy*, 2018.

CHESINI, F., *"Terminator Scenario"? Intelligenza artificiale nel conflitto armato*, in *BioLaw Journal*, 2020.

CHOPRA, S. – WHITE, L.F., *A Legal Theory for Autonomous Artificial Agents*, University of Michigan Press, 2011.

CITRON, D.K., *Technological Due Process*, in *Washington University Law Review*, 2008.

COCA VILA, I., *Coches autopilotados en situaciones de necesidad*, in *InDret*, 2018.

COMMISSIONE EUROPEA, *Relazione sulle implicazioni dell'intelligenza artificiale... in materia di sicurezza e di responsabilità*, COM(2020) 64 final, 2020.

COMMISSIONE EUROPEA – GRUPPO DI ESPERTI, *Orientamenti etici per un'i.a. affidabile*, 2019.

CONSORTE, F., *Tutela penale e principio di precauzione*, Giappichelli, 2013.

CONSULICH, F., *(voce) Rischio consentito*, in *Enc. dir. – I Tematici*, 2022.

CONTISSA, G. – LAGIOIA, F. – SARTOR, G., *La manopola etica: i veicoli autonomi eticamente personalizzabili e il diritto*, in *BioLaw Journal*, 2020.

COOPER, D.P. et alii, *European Commission Announces New Package of AI Measures*, in *Lexology*, 2024.

CORN, E., *Il principio di precauzione nel diritto penale*, Giappichelli, 2013.

COTUGNO, E.M., *Intelligenza artificiale e tutela del consumatore. I nuovi obblighi informativi*, in *Dir. inf.*, 2021.

CRESPO, E.D., *"Compatibilismo umanista": una proposta per conciliare le neuroscienze e il diritto penale*, in *Riv. it. dir. proc. pen.*, 2022.

CUNNEN, M. E AL., *Artificial Driving Intelligence and Moral Agency*, in *Ethics and Information Technology*, 2019.

D'AGOSTINO, L., *Gli algoritmi predittivi per la commisurazione della pena*, in *Dir. pen. cont.*, 2019.

D'ALESSANDRO, F., *(voce) Precauzione (principio di)*, in *Enc. dir. – I Tematici*, 2022.

D'ALESSANDRO, F., *Pericolo astratto e limiti soglia*, Milano, 2012.

D'AMICO, L., *Intelligenza artificiale e auto a guida autonoma*, in *Cass. pen.*, 2019.

DAVOLA, A. – PARDOLESI, R., *In viaggio col robot*, in *Danno e resp.*, 2019.

DE MARTINO, A., *Danno e rischio da prodotti*, in *Riv. trim. dir. pen. ec.*, 1993.

DELITALA, G., *Il "fatto" nella teoria generale del reato*, Cedam, 1930.

DEMURO, G.P., *Ripartire da principi: sul rapporto del diritto penale con Big Data e intelligenza artificiale*, in *Arch. pen. (web)*, 2019.

DESAI, D.R. – KROLL, J.A., *Trust But Verify: A Guide to Algorithms and the Law*, in *Harvard Journal of Law & Technology*, 2017.

DEVILLE, R. E AL., *Basic Concepts of AI for Legal Scholars*, in *Algorithms and Law*, 2020.

DI GIOVINE, O., *Il judge-Bot e le sequenze giuridiche in materia penale*, in *Cass. pen.*, 2020.

DI GIOVINE, O., *Ripensare il diritto penale attraverso le (neuro-)scienze?*, Giappichelli, 2019.

DI LANDRO, A.R., *Dalle linee guida e dai protocolli all'individualizzazione della colpa penale nel settore sanitario*, Giappichelli, 2012.

DIAKOPOULOS, N., *Algorithmic Accountability*, in *Digital Journalism*, 2015.

DOMINGOS, P., *A Few Useful Things to Know about Machine Learning*, in *Communications of the ACM*, 2012.

DORMEHL, L., *Thinking Machines*, Tarcherperigree, 2017.

DOUMA – PALODICHUK, S.A., *Criminal Liability Issues Created by Autonomous Vehicles*, in *Santa Clara Law Review*, 2012.

EBERS, M., *Regulating AI and Robotics: Ethical and Legal Challenges*, in *Algorithms and Law*, Cambridge University Press, 2020.

EBERS, M., *Standardizing AI*, in *The European Health Data Space*, 2022.

EDMONDS, D., *Uccideresti l'uomo grasso?*, Raffaello Cortina Editore, 2014.

ELISH, M.C., *Moral crumple zones: Cautionary Tales in Human-Robot Interaction*, in *Engaging Science, Technology, and Society*, 2019.

ELISH, M.C. – HWANG, T., *Praise the machine! Punish the human!*, in *Data & Society Working Paper*, 2015.

EYKHOLT, K. E AL., *Robust Physical-World Attacks on Deep Learning Visual Classification*, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

FALLOWS, J., *Here's What Was on the Record About Problems with the 737 Max*, in *The Atlantic*, 2019.

FARBER, D.A., *Coping with Uncertainty: Cost-Benefit Analysis, the Precautionary Principle, and Climate Change*, in *Washington Law Review*, 2015.

FERGUSON, A.G., *The Rise of Big Data Policing*, NYU Press, 2017.

FIANDACA, G. – MUSCO, E., *Diritto penale. Parte Generale*, Zanichelli, VIII ed., 2019.

FINOCCHIARO, G., *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in *Dir. inf.*, 2022.

FINOCCHIARO, V.G., *Intelligenza Artificiale. Quali regole?*, Il Mulino, 2024.

FIORELLA, A., *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, in *Il diritto nell'era digitale*, Giuffrè, 2022.

FLORES, A.W. E AL., *False Positives, False Negatives, and False Analyses*, in *Federal Probation Journal*, 2016.

FLORIDI, L., *Philosophy and computing. An introduction*, Routledge, 1999.

FOGEL, D.B., *Review of "Computational intelligence: Imitating life"*, in *Proceedings of the IEEE*, 1995.

FORTI, G., *"Accesso" alle informazioni sul rischio e responsabilità*, in *Riv. it. dir. proc. pen.*, 1993.

FORTI, G., *Colpa ed evento nel diritto penale*, Giuffrè, 1990.

FORTI, G., *La "chiara luce della verità" e l'"ignoranza del pericolo"*, in *Scritti per Federico Stella*, Jovene, 2007.

FRAGASSO, B., *Intelligenza artificiale e responsabilità penale*, Giappichelli, 2025.

FRANKLIN, S., *History, motivations, and core themes*, in *The Cambridge Handbook of Artificial Intelligence*, 2014.

FUTTER, D. et alii, *The EU ACT is here: What you need to know and what to do next*, in *Lexology*, 2024.

FUZAYLOVA, E., *War torts, autonomous weapons systems, and liability*, in *George Washington International Law Review*, 2019.

GARAPON, A. – LASSÈGUE, J., *Giustizia digitale / Justice digitale*, Il Mulino, 2021 (ed. franc. 2018).

GEISTFELD, M.A., *Reconciling Cost-Benefit Analysis with the Principle That Safety Matters More Than Money*, in *New York University Law Review*, 2001.

GIALUZ, M., *Quando la giustizia penale incontra l'intelligenza artificiale*, in *Dir. pen. cont.*, 2019.

GIANNINI, A., *Intelligenza artificiale, human oversight e responsabilità penale*, in *Cass. pen.*, 2022.

GIANNINI, A., *Lombroso 2.0: On AI and Predictions of Dangerousness in Criminal Justice*, in *International Review of Penal Law*, 2021.

GILLESPIE, T., *The relevance of algorithms*, in *Media Technologies*, MIT Press, 2014.

GIUCA, M., *Disciplinare l'intelligenza artificiale*, in *Arch. pen. (web)*, 2023.

GIUNTA, F., *Il diritto penale e le suggestioni del principio di precauzione*, in *Criminalia*, 2006.

GIUNTA, F., *Illiceità e colpevolezza nella responsabilità colposa*, in *Riv. it. dir. proc. pen.*, 1987.

GLESS, S. – SILVERMAN, E. – WEIGEND, T., *If robots cause harm, who is to blame?*, in *New Genetics and Society*, 2016.

GORSKI, D., *IBM's Watson versus cancer: Hype meets reality*, in *Science-Based Medicine*, 2018.

GRANDI, C. *Neuroscienze e responsabilità penale: nuove soluzioni per problemi antichi?*, Giappichelli, 2017.

GREENS, L., *AI update: Examining the relationship between Artificial Intelligence and Copyright Protection*, in *Lexology*, 2024.

GRIFFIN, A., *Facebook's Artificial Intelligence Robots Shut Down After They Start Talking to Each Other in Their Own Language*, in *The Independent*, 2017.

GROTTO, M., *Principio di colpevolezza, rimproverabilità soggettiva e colpa specifica*, Giappichelli, 2012.

GRUT, C., *The Challenge of Autonomous Lethal Robotics to International Humanitarian Law*, in *Journal of Conflict and Security Law*, 2013.

GUIDOTTI, R. E AL., *A Survey of Methods for Explaining Black Box Models*, in *ACM Computing Surveys*, 2018.

HALM, K.C. et alii, *EU unanimously approved Comprehensive Artificial Intelligence Act*, in *Lexology*, 2024.

HARFORD, K. et alii, *EU Act finally adopted: what's next?*, in *Lexology*, 2024.

HARROLD, L., *Artificial Intelligence and copyright infringement*, in *Lexology*, 2024.

HAWES, C., *The EU'S AI Act: cliff hanger or cliff edge?*, in *Lexology*, 2023.

HEIDEGGER, M., *La questione della tecnica*, in *Saggi e discorsi*, Mursia, 1976 (ed. orig. 1957).

HERZOG, F., *Società del rischio, diritto penale del rischio, regolazione del rischio*, in *Riv. it. dir. proc. pen.*, 1993.

HILGENDORF, E., *Automated Driving and the Law*, in *Robotik und Gesetzgebung*, 2013.

HORGAN, C., *Artificial Intelligence Update*, in *Lexology*, 2024.

HORST, J., *A Native Intelligence Metric for Artificial Systems*, in *Proceedings of the PerMIS Workshop*, 2002.

HUQ, A.Z., *Racial Equality in Algorithmic Criminal Justice*, in *Duke Law Journal*, 2019.

ITALIANO, G.F., *Intelligenza Artificiale: passato, presente, futuro*, in *Ars Interpretandi*, 2022.

ITALIANO, G.F. E AL., *Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni*, in *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1, il Mulino, 2022.

KAPLAN, J., *Artificial Intelligence*, Oxford University Press, 2016.

KAREN, V. et alii, *European Commission Proposed Regulation on Artificial Intelligence*, in *Lexology*, 2021.

KARNOW, C.E.A., *Liability for Distributed Artificial Intelligence*, in *Berkeley Technology Law Journal*, 1996.

KARNOW, C.E.A., *The application of traditional tort law to embodied machine intelligence*, in *Robot Ethics*, MIT Press, 2012.

KEHL, D. E AL., *Algorithms in the Criminal Justice System*, in *Responsive Communities Initiative*, 2017.

KELLEHER, J.D. – TIERNAY, B., *Data Science*, MIT Press, 2018.

KROLL, J.A. E AL., *Accountable Algorithms*, in *University of Pennsylvania Law Review*, 2017.

KURZWEIL, R., *The Singularity Is Near: When Humans Transcend Biology*, Viking Penguin, 2005.

LA VATTIATA, F.C., *Brevi note "a caldo" sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale*, in *Dir. pen. uomo*, 2021.

LA VATTIATA, F.C., *La responsabilità penale per danni da intelligenza artificiale alla prova del processo*, in *Processo penale e giustizia*, 2022.

LAGIOIA, F. – SARTOR, G., *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, in *International Review of Penal Law*, 2016.

LAVORGNA, A. – SUFFIA, G., *La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell'ambito della giustizia penale*, in *Dir. pen. uomo*, 2021.

LEMLEY, M.A. – CASEY, B., *Remedies for Robots*, in *University of Chicago Law Review*, 2019.

LEPORE, A., *I.A. e responsabilità civile. Robot, autoveicoli e obblighi di protezione*, in *Tecnologie e diritto*, 2021.

LIMA, D., *Could AI Agents Be Held Criminally Liable?*, in *South Carolina Law Review*, 2018.

LITVINETS, V., *"A Summary of the European Commission White Paper on Artificial Intelligence"*, in *Lexology*, 2020.

LLORCA, D.F. E AL., *Liability regimes in the age of AI: a use-case driven analysis of the burden of proof*, in *JRC Science for Policy Report*, 2022.

LONDON, A.J., *Artificial intelligence and black-box medical decisions: accuracy versus explainability*, in *The Hastings Center Report*, 2019.

LONGHITANO, L., *L'intelligenza artificiale di Facebook parla una lingua incomprensibile*, in *Wired*, 2017.

MAGRO, M.B., *Robot, cyborg e intelligenze artificiali*, in *Cass. pen.*, 2019.

MAHLER, *Regulating Artificial General Intelligence (AGI)*, in *European Journal of Law and Technology*, 2022.

MANES, V., *L'oracolo algoritmico e la giustizia penale*, in *Discrimen*, 2020.

MARINUCCI, G., *Il reato come 'azione'. Critica di un dogma*, Giuffrè, 1971.

MARINUCCI, G., *La colpa per inosservanza di leggi*, Giuffrè, 1965.

MARINUCCI, G., *Soggettivismo e oggettivismo nel diritto penale*, in *Riv. it. dir. proc. pen.*, 2011.

MARINUCCI, G. – DOLCINI, E. – GATTA, G.L., *Manuale di diritto penale*, Giuffrè, X ed., 2021.

MARKS, D.S. et alii, *European Union Artificial Intelligence Act: An Overview*, in *Lexology*, 2023.

MARRA, W.C. – MCNEIL, S.K., *Understanding "The Loop": Regulating the Next Generation of War Machines*, in *Harvard Journal of Law & Public Policy*, 2013.

MASSARO, A., *Intelligenza artificiale e neuroscienze: l'eterno ritorno del diritto penale?*, in *Derecho penal, inteligencia artificial y neurociencias*, 2023.

MASSARO, A., *Principio di precauzione e diritto penale: nihil novi sub sole?*, in *Indice pen.*, 2012.

MASSI, S., *Affidamento sull'intelligenza artificiale e "disimpegno morale"*, in *Cass. pen.*, 2022.

MAUGERI, A.M., *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, in *Arch. pen. (web)*, 2021.

MCCARTHY, J. – MINKSY, M.L. – ROCHESTER, N. – SHANNON, C.E., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, in *AI Magazine*, 2006 (orig. 1955).

MILITELLO, V., *Rischio e responsabilità penale*, Giuffrè, 1988.

MILLAR, J. – KERR, I., *Delegation, relinquishment, and responsibility*, in *Robot Ethics 2.0*, 2017.

MINELLI, C., *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, in *Arch. pen. (web)*, 2022.

MOBILIO, G., *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica*, in *BioLaw Journal*, 2020.

MONGILLO, V., *Imputazione oggettiva e colpa tra "essere" e normativismo: il disastro di Viareggio*, in *Sist. pen.*, 2021.

MONSEE, C. – RICHTER, S., *The interfaces between artificial intelligence and the Data Act*, in *Lexology*, 2024.

MORRISON – FORESTER, *Draft EU Regolation for Artificial Intelligence*, in *Lexology*, 2021.

MORSE, S.J. – ROSKIES, A.L. (eds.), *A Primer on Criminal Law and Neuroscience*, Oxford University Press, 2013.

MULLER, P., *The AI Act – The Agreement and What It Means*, in *Lexology*, 2024.

MÜLLER, V.C. – BOSTROM, N., *Future Progress in Artificial Intelligence*, in *Fundamental Issues of Artificial Intelligence*, 2016.

MURPHY, K.P., *Machine Learning*, MIT Press, 2012.

NAHRA, K.J. et alii, *The European Parliament Adopts the AI Act*, in *Lexology*, 2024.

PALIERO, C.E., *L'autunno del patriarca*, in *Riv. it. dir. proc. pen.*, 1994.

PALIERO, C.E., *La causalità dell'omissione: formule concettuali e paradigmi prasseologici*, in *Riv. it. dir. proc. pen.*, 1992.

PALIERO, C.E., *Responsabilità per danno da prodotto tra persona fisica ed ente collettivo*, in *Riv. trim. dir. pen. ec.*, 1990.

PANATTONI, B., *AI and Criminal Law: The Myth of 'Control' in a Data-Driven Society*, in *International Review of Penal Law*, 2021.

PANATTONI, B., *Intelligenza artificiale*, in *Enc. dir. – I Tematici*, 2021.

PASQUALE, F., *The Black Box Society*, Harvard University Press, 2015.

PAULESU, P.P., *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Arch. pen. (web)*, 2022.

PELLEGATTA, S., *Automazione nel settore automotive: profili di responsabilità civile*, in *Contratto e impresa*, 2019.

PENCO, E., *Limiti-soglia e responsabilità colposa*, in *Riv. it. dir. proc. pen.*, 2015.

PERIN, A., *Prudenza, dovere di conoscenza e colpa penale*, in *Riv. it. dir. proc. pen.*, 2021.

PERINI, C., *Il concetto di rischio nel diritto penale moderno*, Giuffrè, 2010.

PERIS RIERA, J.M. – MASSARO, A. (a cura di), *Derecho penal, inteligencia artificial y neurociencias*, Roma Tre Press, 2023.

PERRONE, D., *La prognosi postuma tra distorsioni cognitive e software predittivi*, Giappichelli, 2022.

PICOTTI, L., *Cybercrimes and Criminal Relevance of Artificial Intelligence*, in *Ars Interpretandi*, 2022.

PICOTTI, L., *I primi vent'anni della Convenzione di Budapest*, in *Sist. pen.*, 2021.

PICOTTI, L., *Profili di responsabilità penale per la circolazione di veicoli a guida autonoma*, in *Dir. pen. proc.*, 2018.

PIERGALLINI, C., *(voce) Colpa (diritto penale)*, in *Dizionario di diritto pubblico*, 2006.

PIERGALLINI, C., *Danno da prodotto e responsabilità penale*, Giuffrè, 2004.

PIERGALLINI, C., *Il paradigma della colpa nell'età del rischio: prove di resistenza del tipo*, in *Arch. pen.*, 2013.

PIERGALLINI, C., *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Sist. pen.*, 2020.

PIERGALLINI, C., *La responsabilità del produttore*, in *Diritto penale dell'impresa*, 2012.

PIZZETTI, F., *La proposta di Regolamento sull'IA della Commissione Europea presentata il 21.4.2021*, in *Dir. internet*, 2021.

PONGILUPPI, C., *Principio di precauzione e reati alimentari*, in *Riv. trim. dir. pen. econ.*, 2010.

POOLE, D., *Computational Intelligence: A Logical Approach*, Oxford University Press, 1998.

PRICE II, W.N., *Medical Malpractice and Black-Box Medicine*, in *Stanford Law Review*, 2018.

PRIVITERA, W., *Tecnica, individuo e società: cinque lezioni sulla teoria di Ulrich Beck*, FrancoAngeli, 2001.

PULITANÒ, D., *Colpa ed evoluzione del sapere scientifico*, in *Riv. it. dir. proc. pen.*, 2002.

PULITANÒ, D., *Diritto penale*, Giappichelli, IX ed., 2021.

QUATTROCOLO, S., *Artificial Intelligence, Computational Modelling and Criminal Proceedings*, Springer, 2020.

QUATTROCOLO, S., *Intelligenza artificiale e giustizia: nella cornice della Carta Etica europea*, in *Legislazione penale*, 2018.

QUATTROCCOLO, S., *Quesiti nuovi e soluzioni antiche?*, in *Cass. pen.*, 2019.

RAMMOS, V.T. – GLASER, R., *The EU AI Act and general-purpose AI*, in *Lexology*, 2024.

RAVIA, H. et alii, *The AI Act*, in *Lexology*, 2024.

REED, C. E AL., *Responsability, Autonomy and Accountability*, in *Legal Studies*, 2016.

RIONDATO, S., *Robot: talune implicazioni di diritto penale*, in *Dir. inf.*, 2017.

RIONDATO, S., *Robotica e diritto penale*, in *Trattato di diritto penale. Parte generale*, 2019.

ROSENTHAL, V.D., *The EU AI Act and what it means in practice for most companies*, in *Lexology*, 2024.

RUDIN, C., *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*, in *Nature Machine Intelligence*, 2019.

RUGA RIVA, C., *Principio di precauzione e diritto penale*, in *Studi in onore di Giorgio Marinucci*, II, 2006.

RUFFOLO, U., *Artificial Intelligence e responsabilità. «Persona elettronica» e teoria dell'illecito*, in *Giust. civ.*, 2017.

RUFFOLO, U., *La responsabilità da artificial intelligence, algoritmo e smart product*, in *Responsabilità Civile*, 2019.

RUFFOLO, U. – AL MUREDEN, E., *Autonomous vehicles e responsabilità nel nostro sistema ed in quello statunitense*, in *Giust. civ.*, 2019.

RUFFOLO, U. – AMIDEI, A., *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design*, in *Giust. civ.*, 2021.

RUSSEL, S. J. – NORVIG, P., *Artificial Intelligence*, Pearson, IV ed., 2020.

SALANITRO, U., *Intelligenza artificiale e responsabilità*, in *Europa e dir. priv.*, 2019.

SALVADORI, I., *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Arch. pen. (web)*, 2020.

SANTOSUOSSO, A. – SARTOR, G., *La giustizia predittiva: una visione realistica*, in *Giur. it.*, 2022.

SARTOR, G. – OMICINI, A., *The autonomy of technological systems and responsibilities for their use*, in *Autonomous Weapons Systems*, 2016.

SCHARRE, P., *Army of None: Autonomous weapons and the future of war*, W. W. Norton & Company, 2018.

SCHMITT, M.N. – THURNER, J.S., *Out of the Loop: Autonomous Weapons Systems and the Law of Armed Conflict*, in *Harvard National Security Journal*, 2013.

SCHMITT, M.N. – WIDMAR, E., *The Law of Targeting*, in *The Oxford Guide to International Humanitarian Law*, 2020.

SEARLE, J.R., *Minds, brains, and programs*, in *Behavioral and Brain Sciences*, 1980.

SELBST, A.D. – BAROCAS, S., *The Intuitive Appeal of Explainable Machines*, in *Fordham Law Review*, 2018.

SELBST, D., *Negligence and AI's Human Users*, in *Boston University Law Review*, 2020.

SGUBBI, F., *Il diritto penale totale. Punire senza legge, senza verità, senza colpa*, Il Mulino, 2019.

SIGNORATO, S., *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc.*, 2020.

SILVA SANCHEZ, J.M., *L' espansione del diritto penale*, Giuffrè, 2004 (ed. orig. 1999).

SIMMLER, M. – MARKWALDER, N., *Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence*, in *Criminal Law Forum*, 2019.

SMUHA, N.A., *"The EU Approach to Ethical Guidelines for Trustworthy AI"*, in *Computer Law Review International*, 2019.

SOVRANO, F. E AL., *Metrics, Explainability and the European AI Act Proposal*, in *J*, 2022.

STEINKER, L.E. et alii, *Towards safe, reliable and human centered AI*, in *Lexology*, 2024.

STELLA, F., *Giustizia e modernità*, Giuffrè, III ed., 2003.

SUNSTEIN, C., *Laws of Fear*, Cambridge University Press, 2005.

TAMPIERI, M., *L'intelligenza artificiale: una nuova sfida anche per le automobili*, in *Contratto e impresa*, 2020.

TURING, A., *Computing Machinery and Intelligence*, in *Mind*, 1950.

UBERTIS, G., *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Dir. pen. cont. – Riv. Trim.*, 2020.

VAN DER EYCKEN, D. – RIGHINI, E., *Year in review: Artificial Intelligence Law in European Union*, in *Lexology*, 2024.

VEALE, M., *Demystifying the Draft EU Artificial Intelligence Act*, in *Computer Law Review International*, 2021.

VENEZIANI, P., *Regole cautelari "proprie" ed "improprie"*, in *Cass. pen.*, 1996.

VIGANÒ, F., *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese*, in *Dir. pen. cont.*, 2012.

WATERS, P., *Getty Images vs Stability AI*, in *Lexology*, 2023.

WENDEL, W.B., *Economic Rationality and Ethical Values in Design-Defect Analysis*, in *Journal of Law, Economics and Policy*, 2015.

WENDEL, W.B., *Technological Solutions to Human Error and How They Can Kill You*, in *Journal of Law, Economics and Policy*, 2020.

WHITTAM, S., *Mind the compensation gap*, in *New Law Journal*, 2012.

WISCHMEYER, T., *Artificial Intelligence and Transparency: Opening the Black Box*, in *Regulating Artificial Intelligence*, 2020.

ZIRULIA, S., *Esposizione a sostanze tossiche e responsabilità penale*, Giuffrè, 2018.