INTELLIGENZA ARTIFICIALE, *HUMAN OVERSIGHT* E RESPONSABILITÀ PENALE: PROVE D'IMPATTO A LIVELLO EUROPEO *

Criminalia
Annuario di scienze penalistiche
in disCrimen dal 21.11.2021

Alice Giannini **

ARTIFICIAL INTELLIGENCE, HUMAN OVERSIGHT,
AND CRIMINAL LIABILITY: A EUROPEAN "STRENGTH TEST"?

Artificial intelligence (AI) and criminal law are now an inseparable pair in the scholarly discourse. It is indeed undisputed that AI serves as a strength test for traditional notions of substantive criminal law, such as mens rea *and* actus reus. *Hence, the question that criminal legal scholars are doomed to ask themselves – in the aftermath of yet another advance in technology – seems to be a recurring one: if something goes wrong with these complex systems, should criminal law care? If yes, how? This paper, after a brief digression on the meaning of accountability, responsibility, and liability, will focus on the concept of human oversight and its relationship with the criminal notion of negligence. Particular attention will be paid to the European actor, which is emerging as a key normative player in this field, specifically to the contents of the European Parliament resolution of October 2021 "Artificial Intelligence in Criminal Law" and the proposed European regulation "AI Act".*

KEYWORDS Artificial intelligence – Human oversight – Negligence – New technologies – AI crimes

1. Introduzione

Intelligenza artificiale (IA)[1] e diritto penale rappresentano oggi un binomio in-

* Il presente contributo sviluppa e approfondisce l'intervento presentato al Convegno "La via europea per l'intelligenza artificiale", tenutosi presso l'Università Ca' Foscari di Venezia nei giorni 25-26 novembre 2021

** Dottoranda di ricerca nell'Università degli Studi di Firenze e nell'Università di Maastricht

[1] Non esiste una definizione universalmente condivisa di IA, né una definizione giuridica adottata a livello europeo o internazionale. La discussione del problema è al di fuori del raggio di azione di questo scritto. Per gli scopi di questa riflessione verrà adottata la definizione sviluppata dal gruppo di esperti di alto livello istituito dalla Commissione europea (HILEG), alla quale si rimanda. Cfr., High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and scientific dis-*

separabile, oggetto d'indagine dalla dottrina penalistica sia italiana che internazionale[2]. Vi è infatti unanimità di vedute nel ritenere che la diffusione di sistemi di IA, capaci di agire in modo autonomo ed imprevisto, potrebbe dar luogo a vuoti di responsabilità e che dunque tali macchine sollevino questioni *proprie* del diritto penale[3]. In quest'ottica, lo stesso Parlamento europeo ha affermato, nella Risoluzione

ciplines, 2018. Disponibile presso: https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

[2] Senza pretese di esaustività, si veda all'interno dell'ampia dottrina italiana, *Il sistema penale alla prova dell'intelligenza artificiale*, cap. in Il *sistema penale ai confini delle* hard sciences, a cura di F. Basile et. al, Pacini Giuridica, 2020; R. Borsari, *Intelligenza Artificiale e responsabilità penale: prime considerazioni*, in *MediaLaws,* 3, 2019; A. Cappellini, *Profili penalistici delle* self-driving cars, in *Diritto Penale Contemporaneo,* (2) 2019, 325 – 353 e Machina delinquere potest*? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018; C. Cavaceppi, *L'intelligenza artificiale applicata al diritto penale*, in *Intelligenza artificiale-Algoritmi giuridici:* Ius condendum *o fantadiritto?*, a cura di G. Taddei Elmi e A. Contaldo Pacini, 2020; F. Lagioia, G. Sartor, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, in *Philosophy and Technology*, 33, 2020; M. B. Magro, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica* in *La legislazione penale*, 2020; M. Papa, Future crimes*: intelligenza artificiale e rinnovamento del diritto penale*, in *Criminalia*, 2019; V. Manes*, L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 2020; U. Pagallo, S. Quattrocolo*, The impact of AI on criminal law, and its twofold procedures*, in *Research handbook on the law of artificial intelligence,* a cura di W. Barfield e U. Pagallo, Elgar, 2018; C. Piergallini, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, in *Rivista italiana di diritto e procedura penale* 4, 2020; S. Riondato, *Robot: talune implicazioni di diritto penale,* in *Tecnodiritto. Temi e informatica e robotica giuridica*, a cura di P. Moro e C. Sarra, FrancoAngeli, 2017; I. Salvadori, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in Rivista Italiana di Diritto e Procedura Penale, (1). Nella dottrina internazionale, si veda R. Abbott, A. Sarch, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, in *UC Davis Law Review*, 53, 2019; S. Beck, *Intelligent agents and criminal law—Negligence, diffusion of liability and electronic personhood*, in *Robotics and Autonomous Systems* 86, 2016; K. Burchard*, Künstliche Intelligenz als Ende des Strafrechts? Zur algorithmischen Transformation der Gesellschaft*, in *Normative Orders Working Paper*, 2¸ 2019; S. Gless et al., *If Robots Cause Harm, Who is to Blame? Self-Driving Cars and Criminal Liability*, in *New Crim L. Review* 19, 2016; G. Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, 2015; Y. Hu, *Robot criminals*, in *U. Mich. J.L. Reform*, 52, 2019; M. Simmler, N. Markwalder, *Roboter in der Verantwortung? - Zur Neuaflauge der Debatte um den funktionalen Schuldbegriff*, in *ZTSW*, 2017; G. Seher, *Intelligente Agenten als "Personen im Strafrecht",* in *Intelligente Agenten und das Recht*, Robotik und Recht vol. 9, a cura di S. Gleß e K. Seelman, 2016; P. M. Freitas et al., *Criminal Liability of Autonomous Agents: from the unthinkable to the plausible*, in *AI Approaches to the Complexity of Legal Systems: AICOL 2013. Lecture Notes in Computer Science*, a cura di P. Casanovas et al., vol. 8929, Springer, 2014.

[3] U. Pagallo, *The Laws of Robots: Crimes, Contracts and Torts,* Springer, 2013, 45. Per una disamina a 360 gradi dei quesiti che è fondamentale porsi per analizzare l'impatto dei sistemi di IA sulla parte generale del diritto penale, si veda il questionario sviluppato da L. Picotti per il XXI Congresso dell' International Association of Penal Law, disponibile presso: https://www.penal.org/sites/default/files/Questionnaires%20EN.pdf.

dell'ottobre 2021 "L"intelligenza artificiale nel diritto penale", la necessità di "istituire un regime chiaro ed equo per attribuire la *responsabilità giuridica* e *imputabilità* delle potenziali conseguenze negative prodotte da tali tecnologie digitali avanzate"[4].

Come dovrebbe funzionare questo meccanismo di imputazione e di attribuzione nello specifico campo della responsabilità penale? La Risoluzione, come verrà evidenziato, non fornisce un manuale di istruzioni. Per questo motivo, cercheremo di sviluppare alcune riflessioni in tale direzione, concentrandoci in particolare sulla dimensione europea.

Una cosa è certa: gli interrogativi che i penalisti sono condannati a porsi, all'indomani dell'ennesimo progresso della tecnica, sono sempre i medesimi. Se qualcosa andasse storto con questi complessi sistemi, il diritto penale dovrebbe interessarsene? Se sì, come?

Come noto, non si tratta certo della prima volta che il diritto penale si trova ad avere a che fare con l'innovazione, o, come definito da alcuni, con uno "shock da modernità"[5]. Si pensi, ad esempio, alla responsabilità per danno causato da un prodotto difettoso[6] o ai reati ambientali[7]. Queste sono aree dove spesso la distinzione fra diritto penale, strumento per punire, e altre branche del diritto, strumenti per risarcire, si fa più sottile[8].

Abbott e Sarch, tra i più importanti autori impegnati sul tema, sostengono che i sistemi di IA sollevino un problema di *irriducibilità* per il diritto penale (c.d. "*irreducibility challenge*")[9]. Con questa espressione fanno riferimento a situazioni in cui po-

[4] Risoluzione del Parlamento europeo del 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale (2020/2016(INI)), 13, enfasi aggiunta.

[5] L'espressione è di F. Stella e viene ripresa da F. Basile in *Diritto penale e Intelligenza Artificiale,* Giurisprudenza Italiana – Supplemento 2019, 68. Basile afferma la necessità di discutere le implicazioni per il diritto penale derivanti dall'impiego di sistemi di IA per "scongiurare il rischio che […] il diritto penale soccomba di fronte a quello che si preannuncia essere un nuovo, sconvolgente "shock da modernità"". Cfr. F. Stella, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Giuffrè, 2003, 292.

[6] Per cui si rimanda *ex multis* a A. Madeo, *La tutela penale della salute dei consumatori*, Giappichelli, 2006; C. Piergallini, *Danni da prodotto e responsabilità penale. Profili dogmatici e politico-criminali,* Giuffrè, 2004.

[7] Per cui si rimanda *ex multis* a C. Ruga Riva, *Diritto penale dell'ambiente*, Giappichelli, 2016; N. Pisani, L. Cornacchia, *Il nuovo diritto penale dell'ambiente*, Zanichelli, 2018.

[8] Cfr., C. Wells, *Corporate Criminal Liability in England and Wales*, in *Corporate Criminal Liability: Emergence, Convergence and Risk*, Springer, 2011, 93.

[9] R. Abbott, Alex Sarch, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, cit., 330 ss.

trebbe essere estremamente difficile, se non impossibile, *ridurre*[10] un reato commesso da un sistema di IA all'agire di un singolo essere umano[11]. Ciò a causa di quattro caratteristiche possedute dai sistemi di IA: autonomia, ossia la capacità di causare un evento dannoso senza che il sistema sia diretto in tal senso da un agente umano; opacità, ossia l'impossibilità – specie quando si ha a che fare sistemi più avanzati quali quelli basati sul *deep learning*[12] – di poter ottenere una spiegazione su come il sistema, partendo dall'*input* (A), ha ottenuto l'*output* dannoso (B); complessità, ossia il fatto che la creazione di un sistema di IA sia spesso il risultato del contributo di numerosi individui, sviluppatosi in un lungo periodo di tempo, nonché il fatto che il sistema possa essere stato addestrato su banche dati eterogenee e *open source*; ed infine imprevedibilità, ossia la capacità per il sistema di intraprendere attività non previste dalla sua programmazione originale[13]. La questione che si pone con evidenza, dunque, è quella della possibile attribuzione di penale responsabilità all'umano, piuttosto che al sistema intelligente. A questo riguardo occorre infatti interrogarsi: a) in caso di condotta omissiva, sull'esistenza di un dovere giuridico e di un potere fattuale di impedimento dell'evento, e dunque sulla possibilità che l'umano detenga una posizione di garanzia; b) sull'esistenza di una regola cautelare, positivizzata o basata sull'esperienza, e dunque sulla configurabilità di un agente modello; c) sulla conoscibilità di tale regola e la prevedibilità dell'evento; d) sull'esigibilità del comportamento conforme da parte dell'umano. In particolare, questo contributo porrà l'accento, per le peculiari caratteristiche con cui interagisce con l'IA, sul tema della responsabilità a titolo di colpa.

Rispondere a tali quesiti quando vi è il coinvolgimento di un sistema di IA appare particolarmente complesso per una serie di motivi. Innanzitutto, vi è la scarsa, seppur in via di sviluppo, conoscenza scientifica di questa tecnologia, alla quale si somma

[10] Nel loro articolo "*Punishing Artificial Intelligence: Legal Fiction or Science Fiction*" Abbott e Sarch sembrano quasi utilizzare il verbo "*reduce*" nel suo significato più arcaico, ossia quello di "riportare indietro" (dal latino *reducere*). Cfr. Oxford Learner's Dictionaries, disponibile presso: https://www.oxfordlearnersdictionaries.com.

[11] Ibid. Basandosi sulla letteratura nella filosofia morale e giuridica e nell'etica della tecnologia, F. Santoni de Sio e G. Mecacci, *Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them,* in *Philosophy & Technology* 34, 2021, identificano quattro *responsibility gap*, tra cui rientra anche il "*culpability gap*".

[12] Il *deep learning,* o apprendimento profondo, è una sotto disciplina del *machine learning* (un *machine learning* "dopato"). A sua volta il *machine learning* è una sotto disciplina dell'IA. Cfr., "*Deep learning is machine learning on steroids*", K. Hao, *What is machine learning?,* MIT Technology Review, 17 novembre 2018. Disponibile presso: https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/.

[13] R. Abbott, A. Sarch, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, cit., 330 ss.

la quasi assoluta mancanza di usi e convenzioni. Tali regole, come si vedrà, risultano poco conoscibili all'agente umano, anche perché la macchina sulla quale egli dovrebbe esercitare il controllo è stata – in realtà – costruita per prevalere su di lui, e non viceversa. Infine, spesso l'evento nocivo che dovrebbe essere prevenuto è imprevedibile.

Ciò premesso, il contributo verrà sviluppato come segue. Inizialmente si svolgerà una breve digressione sul significato di *accountability*, *leitmotiv* noto a chiunque si approcci al tema dell'*AI-regulation*. Dopodiché si introdurrà la nozione di "area di impatto" dell'IA sulla materia penale[14]. Scopo di questa breve riflessione, tuttavia, non è quello di delineare i *contorni* di questa area – operazione già effettuata da autorevole dottrina[15] – bensì quello di svolgere alcune valutazioni sulla *portata* dell'impatto dell'IA in un settore specifico del diritto penale. In particolare, partendo dai contenuti della già menzionata risoluzione del Parlamento Europeo dell'ottobre 2021 "L'intelligenza artificiale nel diritto penale"[16] e della proposta di regolamento europeo c.d. AI Act[17], ci si concentrerà sul concetto di sorveglianza umana ("*human oversight*") e si discuterà del perché sia possibile individuare una relazione tra tale nozione e quella penalistica di colpa. Pertanto, lo scritto non investirà la tematica relativa alla possibilità di attribuire una responsabilità diretta in capo al sistema intelligente. La questione, seppur connessa a quella oggetto di questa analisi, nonché meritevole di approfondita trattazione in sede ulteriore, è indipendente da quella attinente alla responsabilità del sorvegliante umano[18].

Perché concentrarsi sulla dimensione europea? In via generale, questa ha acquistato rilevanza per la velocità con la quale le istituzioni dell'UE si stanno approcciando alla regolazione dell'intelligenza artificiale e alla realizzazione della c.d. "Good AI society"[19], guadagnando così il primato su altri attori istituzionali. Per

[14] Con tale termine si fa riferimento alla nozione di materia penale accolta nell'ordinamento italiano. Per una disamina della nozione di materia penale sviluppata dalla Corte EDU in rapporto alla qualificazione adottata dalla nostra Corte costituzionale si veda, *ex multis*: AA.VV., *La materia penale tra diritto nazionale ed europeo,* a cura di M. Donini e L. Foffani, Giappichelli, 2018.

[15] Per cui si rimanda, *ex multis*, a F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto Penale e Uomo - DPU*, fasc. 10/2019, 1 segg. (*online*).

[16] Parlamento Europeo, *L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*, cit.

[17] Commissione Europea, Proposta di regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'unione, COM/2021/206, 21 aprile 2021.

[18] Si veda il par. 4. per un breve riassunto del dibattito dottrinale in materia ed in particolare le note 71 e ss.

[19] Per una definizione del termine "*Good AI society*" e un'analisi comparata della strategia europea

quanto attiene, invece, alle specificità della materia penale, preme sottolineare che i sistemi di IA potrebbero avere effetti importanti su quelli che sono gli obiettivi della politica criminale europea. Ne sono la dimostrazione alcune azioni intraprese da varie agenzie europee. Si pensi ad esempio alla recentissima pubblicazione da parte dell'Europol Innovation Lab di un report sui reati commessi utilizzando tecnologie *deepfake*[20], nel quale viene prospettato l'insorgere di nuove forme di reato per il prossimo decennio, che porteranno con sé non poche difficoltà dal punto di vista dell'attribuzione della responsabilità penale[21]; nonché la realizzazione del progetto "*Accountability Principles for Artificial Intelligence*" (AP4AI), che riunisce Eurojust, Europol, CEPOL[22] e il centro di ricerca CENTRIC[23] nella creazione di un quadro di principi per garantire un utilizzo etico, trasparente e responsabile di sistemi di IA da parte delle forze dell'ordine.

2. *Accountability*, *responsibility*, *liability*: istruzioni per la lettura

Prima di iniziare questa riflessione si rende opportuno un chiarimento terminologico-concettuale sul significato di *accountability*[24]. Ad oggi, infatti, la maggior parte delle proposte di regolazione dell'IA hanno come obiettivo la garanzia dell'"*accountability*" e l'affermazione di principi di *soft law* diretti allo sviluppo di un'IA etica o affidabile, piuttosto che l'elaborazione di strumenti di *hard law*[25].

vis-à-vis quella americana si veda H. Roberts et. al, *Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US,* in *Science and Engineering Ethics*, 2021, 27-68.

[20] Con il termine *deepfake* (*deep learning* + *fake contents*) si fa riferimento ad un'immagine o un video sintetico, generato tramite l'utilizzo di tecniche di IA, nel quale si ha la sovrapposizione dell'immagine del viso di un soggetto "*target*" sopra quella di un'altra persona ("*source*"). Il risultato può essere talmente accurato che diventa difficile, se non impossibile, determinare quale è il video originale. Cfr., Thanh Thi Nguyen et al., *Deep Learning for Deepfakes Creation and Detection: A Survey,* 2022, arXiv:1909.11573v4.

[21] Europol (2022*), Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab*, Publications Office of the European Union, Luxembourg, 16.

[22] Agenzia dell'Unione europea per la formazione delle autorità di contrasto.

[23] Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research.

[24] Si veda *infra* per la definizione del termine.

[25] Fa eccezione la modifica al codice della strada francese adottata in Francia con l'ordinanza del Presidente della Repubblica del 14 aprile 2021 (*Ordonnance n° 2021-443 du 14 avril 2021 relative au régime de responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite et à ses conditions d'utilisation*, TRAT2034523R). Per un'analisi approfondita di tale iniziativa in un'ottica comparata, si veda M. Giuca, *Disciplinare l'intelligenza artificiale. La riforma francese sulla responsabilità penale da uso di auto a guida autonoma*, in *Archivio Penale*, Fascicolo n.2, 2022.

L'*accountability* è uno dei sette principi menzionati dall'Artificial Intelligence High-Level Expert Group (AI HLEG) istituito dalla Commissione europea nel documento "Orientamenti etici"[26] e rappresenta un elemento ricorrente nelle iniziative di regolazione europea[27], nonché nel più ampio panorama globale della c.d. *AI ethics*[28].

Secondo alcuni lo "strumento etico" non sarebbe sufficiente per regolamentare l'IA. Il suo utilizzo rappresenterebbe niente di più che l'ennesimo esempio di una "concezione giuridica" di etica[29], una sorta di brutta copia della legge[30]. Altri, in risposta, sostengono che l'etica non sia affatto "sdentata"[31] ma che venga semplicemente utilizzata nel modo sbagliato.

Qual è quindi il significato *prescrittivo* della parola *accountability*[32]? E quali sono i suoi rapporti con il concetto di responsabilità penale?

Innanzitutto, quando si ha a che fare con temi di *accountability* la lingua italiana non ci viene in aiuto. Con la parola "*accountability"* si fa riferimento ad un concetto che non coincide con quello di *responsibility* e di *liability*, sebbene vi sia strettamente connesso. Tali sfumature non vengono colte nella nostra lingua, che conosce invece solamente la parola "responsabilità"[33]: questo è fonte di disorientamento per il lettore italofono.

In italiano la parola "responsabilità" ha un significato duplice: il primo, più am-

[26] Gruppo di esperti ad alto livello sull'intelligenza artificiale, *Orientamenti etici per un'IA affidabile*, 2019, 23.

[27] Risoluzione del Parlamento Europeo, *Quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate*, 2020/2012(INL), 20 ottobre 2020, Cfr. artt. 9, 22, 23, 72, 96, 102.

[28] Si può far riferimento, ad esempio, ai principi OCSE (OCSE, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 2019); alle raccomandazioni UNESCO sull'etica dell'IA (UNESCO, *Recommendation on the ethics of artificial intelligence*, SHS/BIO/REC-AIETHICS/2021, 2021). Per una panoramica globale, si vedano gli studi condotti da J. Fjeld et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, Berkman Klein Center for Internet & Security, 2020; A. Jobim et al., *The global landscape of AI ethics guidelines,* in *Nature Machine Intelligence*, volume 1, 2019.

[29] G.E.M. Anscombe, *Modern moral philosophy*, in *Philosophy* 33, 1968, 1-19.

[30] A. Rességuier e R. Rodrigues, *AI ethics should not remain toothless! A call to bring back the teeth of ethics'*, in *Big Data & Society*, 2020, 2.

[31] Ibid.

[32] L'etimologia descrittiva della parola *accountability* ci riporta al verbo *account*, che deriva dal francese antico *acconter*, a sua volta proveniente dal latino *computāre*. *Computāre* significa contare, conteggiare, calcolare, ma anche considerare, dare più o meno importanza, stimare, mettere in conto. Deriva dall'unione di *cum* (insieme) e di *putàre* (tagliare). Il suffisso -*ability*, invece, significa "capacità di", dal latino -*abilis*.

[33] La parola responsabilità deriva dal latino *respondere*, formato dalla particella *re* (addietro, di nuovo) e *spondère* (promettere solennemente, impegnarsi).

pio, è l'addossarsi gli effetti delle proprie azioni o di quelle altrui; il secondo, più ristretto, attiene alla situazione *giuridica* che sorge quale conseguenza della violazione di un obbligo o da un atto illecito. Il termine *accountability* andrebbe tradotto con "rendicontabilità; obbligo/capacità di rendicontazione"[34]. Difatti tale traduzione, frutto di un "calco linguistico", valorizzerebbe il fatto che *accountability* non implichi necessariamente né responsabilità, e cioè l'essere chiamati a dar conto di certi atti e il doverne subire le conseguenze, né rendicontazione, parola che può indicare l'atto "ma non certo l'attitudine [*a dar conto*]"[35]. La parola *accountability*, quale sostantivo astratto, ha dunque un significato duplice: viene utilizzata per far riferimento sia alla *capacità* che all'*obbligo*[36] di qualcuno di rendere conto in un *sistema sociale*[37].

Come già menzionato, un'ulteriore complicazione in questo dialogo italo-inglese deriva dal fatto che la lingua inglese possiede anche due termini ulteriori oltre ad "*accountability*": "*responsibility*" e "*liability*". La dottrina di riferimento non definisce in modo preciso i confini di queste tre nozioni, in particolar modo per quanto riguarda la distinzione fra le ultime due[38]. Si può affermare con certezza che se la *liability* ha a che fare con delle condotte qualificate, l'*accountability* e la *responsibility* rimandano ad un mondo più ampio, che al suo interno contiene anche regole morali e regole sociali, quali le *folksway* e le norme di cortesia.

L'*accountability* può anche essere definita in un'accezione più "procedurale", poiché simboleggia il rituale dell'essere messi sul banco degli imputati (*being 'placed in the dock'*), un banco che però non è quello tipico del procedimento penale, bensì quello figurativo posto dinanzi ad una comunità[39]. In altre parole, secondo questa accezione, l'*accountability*

> is the process aimed at a subsequent public assessment of a person's conduct in a given case in order to evaluate whether this conduct was required and/or justified by this person's *responsi-*

[34] Ibid.

[35] Ibid. (corsivo aggiunto).

[36] Europarole, un'iniziativa promossa dal Dipartimento per le Politiche Europee per facilitare la comprensione di termini o espressioni in lingua inglese abitualmente usati nel linguaggio politico e mediatico, suggerisce di tradurre "*accountability*" con le parole "rendicontabilità; obbligo/capacità di rendicontazione". Si veda: https://www.politicheeuropee.gov.it/it/comunicazione/europarole/accountability.

[37] D. Castiglione, *"accountability",* Encyclopedia Britannica, 22 ottobre 2012, disponibile presso: https://www.britannica.com/topic/accountability.

[38] Per una ricostruzione del dibattito si veda, *ex multis*, R. Mulgan, *Accountability': An Ever-Expanding Concept?*, in *Public Administration*, 78, 2002.

[39] R. S.B. Kool, *(Crime) Victims' Compensation: The Emergence of Convergence*, in *Utrecht Law Review*, Volume 10, Issue 3 (June) 2014, 16.

> *bility* and, once this evaluation is executed, to establish who can be held *liable* for all consequences of the conduct […][40]

Per quanto attiene specificatamente al diritto penale, secondo R.A. Duff la *responsibility* è un concetto "relazionale", poiché riguarda la relazione che si instaura tra una persona responsabile (A), un oggetto (X) per il quale questa persona è responsabile, e un soggetto terzo (B). Responsabilità per Duff significa dunque *"answerability"*, il meccanismo in forza del quale A è *chiamato* a rispondere per X da, e nei confronti di, B[41].

La parola "*liability*"[42] viene utilizzata invece per far riferimento esclusivamente alla situazione in cui un soggetto, nel compiere una certa azione o nello stipulare un determinato contratto, si è assunto un obbligo *giuridico* e risponde quindi delle conseguenze *giuridiche* che da questo ne derivano[43]. Secondo Duff ben vi può essere *(criminal) responsibility* senza vi sia necessariamente *(criminal) liability*: si pensi ad esempio al caso di un reato commesso in assenza di antigiuridicità[44].

Alla luce di tali difficoltà linguistiche, è dunque lecito chiederci se sia previsto all'interno del nostro sistema giuridico un concetto equivalente a quello di *accountability* come sopra delineato. Nel campo del diritto penale forse l'istituto più conferente sarebbe quello dell'imputabilità, intesa nella sua accezione normativa, ossia quale "capacità di rimproverabilità"[45].

Nonostante le differenze suesposte, i tre termini vengono sovente tradotti in italiano – anche dal legislatore europeo – unicamente con la parola "responsabilità"[46].

---

[40] Ivo Giesen, François G.H. Kristen, *Liability, Responsibility and Accountability: Crossing Borders*, *Editorial* in *Utrecht Law Review*, Volume 10, Issue 3 (June) 2014, 6.

[41] R. A. Duff, *Who is Responsible for What, to Whom*?, in *Ohio State Journal of Criminal Law*, vol. 2, 2005, 442. Si veda anche R.A. Duff, *Moral and Criminal Responsibility: Answering and Refusing to Answer*, in J. Coates e N.A. Tognazzini, *Oxford Studies in Agency and Responsibility Volume 5: Themes from the Philosophy of Gary Watson*, 2019, pp. 165-190. R.A. Duff, *Legal and Moral Responsibility*, in *Philosophy Compass*, vol. 4, issue 6, 2009.

[42] L'etimologia della parola riconduce al latino *ligare* (legare, stringere, unire, cementare).

[43] D. Castiglione, *"accountability",* cit.

[44] R. A. Duff, *Who is Responsible for What, to Whom*?, cit. 443.

[45] F. Palazzo, *Corso di diritto penale. Parte generale,* Giappichelli, 2021, pp. 410 e ss.

[46] Ad. es., Parlamento europeo, *Risoluzione sull'intelligenza artificiale: questioni relative all'interpretazione e applicazione del diritto internazionale nella misura in 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)), Considerando I, artt. 14, 24, 25, 49, 52, 70, 72. Si noti come negli *orientamenti etici per un'IA affidabile*, cit., il termine *accountability* non viene tradotto in italiano. Per un'analisi della terminologia utilizzata nella traduzione della risoluzione del Parlamento Europeo dell'ottobre 2021 v. *infra.*

Consapevoli di questi ostacoli, in questo scritto è stata fatta la scelta di riportare i termini *accountability, liability* e *responsibility* senza tradurli.

A questo punto è possibile sviluppare due riflessioni conclusive. Primo, l'*accountability* può presentarsi in una varietà di forme e dimensioni: la responsabilità penale è una di queste[47]. Secondo, l'*accountability* si fonda tradizionalmente su una nozione di *controllo*. Seguendo una massima semplificazione[48], ciò si rispecchia negli elementi costitutivi del reato. Quasi ogni sistema penale moderno, infatti, reclama almeno una qualche forma di condotta dominabile dall'autore[49], nonché un nesso causale, al fine di attribuire la responsabilità penale[50]. Inoltre, la responsabilità penale può essere imposta solo in capo a coloro che sono sufficientemente consapevoli di quello che stanno facendo, nonché delle possibili implicazioni del loro agire (o non agire), in modo da poter ragionevolmente affermare che questi abbiano *scelto* di porre in essere quel determinato comportamento[51].

Ebbene, come reagiscono tali tradizionali nozioni di diritto penale quando messe a confronto con l'agire algoritmico? Tale domanda guiderà il prosieguo di queste riflessioni.

3. Aree di conflitto ed aree di collaborazione

Il rapporto fra IA e diritto penale si può configurare in termini di conflitto o di collaborazione[52]. Da un lato, dunque, troviamo l'area di conflitto, riconducibile al diritto penale sostanziale, dove i costrutti classici faticano ad indirizzarci verso un soggetto (umano) responsabile. Dall'altro si ha il diritto penale processuale e le procedu-

[47] C. Heyns, *Autonomous weapons system: living a dignified life and dying a dignified death*, in *Autonomous weapons systems. Law, ethics, policy*, a cura di Nehal Bhuta et al., Cambridge University Press, 2016, 11-12.

[48] Questa attività di semplificazione non corrisponde ad una presa di posizione rispetto alle teorie della bipartizione, tripartizione e quadripartizione del reato, per la trattazione delle quali si rimanda, per tutti, a F. Mantovani, *Diritto penale. Parte Generale*, XI edizione, 2020, 111 e ss. e agli autori ivi indicati.

[49] J. Keiler, Actus reus *and participation in European criminal law*, cit., 42.

[50] Cfr. *Comparative Concepts of Criminal Law*, a cura di J. Keiler e D. Roef, Intersentia, 3ª ed., 2019; M. Papa, F. Palazzo, *Lezioni di diritto penale comparato*, 3ª ed., Giappichelli, 2013; J. Keiler, Actus reus *and participation in European criminal law*, cit.; J. Blomsma, Mens rea *and defences in European criminal law*, Intersentia, 2012; A. Ashworth, *Principles of Criminal Law*, Oxford, 6ª ed., 2009.

[51] A. Ashworth, *Principles of Criminal Law*, cit., 101.

[52] I. Vuletić, T. Petrašević, *Is It Time to Consider EU Criminal Law Rules on Robotics*, in *CYELP*, 16, 2020, 228.

re di *enforcement*, le c.d*. aree di collaborazione*[53], dove l'IA si pone quale strumento migliorativo delle prassi attuali. In questo breve commento ci concentreremo sulla prima area individuata, ossia quella di conflitto.

3.1. - *La Risoluzione del Parlamento Europeo "L'intelligenza artificiale nel diritto penale" e l'Artificial Intelligence Act*

Come accennato, il 6 ottobre 2021 il Parlamento europeo ha approvato una Risoluzione dal titolo "L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale"[54]. Pur non avendo forza vincolante, tale Risoluzione rappresenta un punto di partenza importante per svolgere alcune considerazioni su quale sarà l'orientamento dell'Unione europea in relazione ai profili di responsabilità penale collegati all'IA. Se, infatti, dal lato della responsabilità civile il legislatore europeo può già appoggiarsi su solide basi, lo stesso non si può dire per quanto attiene alla responsabilità penale. Una futura regolamentazione a livello europeo in questo settore dovrebbe senza dubbio inserirsi in un dedalo di normative già esistenti, nonché confrontarsi con la più debole competenza dell'Unione europea nella materia penale. Non esiste ad oggi un diritto penale europeo di *parte generale*[55]: questo aspetto è di particolare rilevanza perché, come già sottolineato, l'agire stesso dei sistemi di IA genera problemi trasversali alle categorie dogmatiche tradizionali e alle modalità di disciplina degli elementi del reato da parte dei Codici penali dei singoli Stati membri. Svolte queste premesse, si può ora analizzare il contenuto della Risoluzione quale affermazione programmatica di una futura disciplina penalistica dell'IA a livello eurounitario.

È da segnalare, innanzitutto, una presa di coscienza esplicita dei problemi collegati alla corretta individuazione delle "*legal responsibility and liability*"[56] per i potenziali effetti pregiudizievoli dei sistemi di IA utilizzati nell'ambito della giustizia penale (art. 13). Secondo quanto affermato dal Parlamento europeo, tale constatazione, estensibile anche agli altri settori in cui l'utilizzo di IA è in crescita, è connessa alla complessità dello sviluppo e del funzionamento di tali tecnologie. Infatti, come si legge an-

[53] Ibid.

[54] Parlamento Europeo, *L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*, cit.

[55] Per un'analisi dei c.d. "frammenti" di parte generale del diritto penale europeo, cfr. A. Klip, *European Law. An Integrative Approach*, 4ª ed., Intersentia, 2021, 244 ss.

[56] Nella versione in italiano l'espressione viene tradotta solo con "responsabilità".

che nel Considerando I della Risoluzione, la realizzazione di un sistema di IA richiede "il contributo di molteplici persone, organizzazioni, componenti meccanici, algoritmi, software e utenti umani in ambienti spesso complessi e problematici"[57]. Da ciò deriverebbe l'importanza e la necessità di stabilire un meccanismo attributivo che sia *chiaro* ed *equo*, come sancito dall'art. 13. Non è chiaro, però, come questo modello dovrebbe essere costruito, né come vada identificato il soggetto responsabile.

A questo proposito preme notare due discrepanze nella Risoluzione, ricollegabili alla riflessione sui concetti di *accountability*, *responsibility* e *liability* di cui sopra. La prima attiene al considerando J, dove viene affermato:

> J. considerando che è necessario un modello chiaro per attribuire la *responsabilità* per i potenziali effetti nocivi dei sistemi di IA nel settore del diritto penale […][58]

Nella versione italiana del Considerando, dunque, l'espressione originale "*legal responsibility"* viene tradotta semplicemente con "responsabilità".

Il considerando poi prosegue affermando

> che le norme regolamentari in questo ambito dovrebbero sempre sostenere la *responsabilità umana* e che il loro primo e principale scopo deve innanzi tutto essere la prevenzione di qualunque effetto negativo […] [59]

Tuttavia, nella versione originale in lingua inglese si legge che le disposizioni normative dovrebbero garantire sempre la "*human accountability*"[60]. Non viene utilizzata pertanto né la parola "*responsibility"*, né tantomeno "*liability*". Questi termini, come abbiamo visto, hanno una pregnanza differente. In particolare, l'esser ritenuti *accountable* per un illecito non per forza comporta una responsabilità sul piano giuridico, come invece pare indicare la traduzione italiana di questo passaggio.

La seconda discrepanza nelle traduzioni è contenuta nell'articolo 13, che recita:

> 13. *[il Parlamento europeo]* prende atto delle sfide relative alla corretta individuazione delle *responsabilità giuridica* e *imputabilità* per i potenziali danni, data la complessità

[57] Parlamento Europeo, *L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*, cit., considerando I.

[58] "J. whereas a clear model for assigning legal responsibility for the potential harmful effects of AI systems in the field of criminal law is imperative; whereas regulatory provisions in this field should always maintain human accountability and must aim, first and foremost, to avoid causing any harmful effects to begin with". Parlamento Europeo, *L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*, cit., considerando J (enfasi aggiunta).

[59] Parlamento Europeo, *L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*, cit., considerando J (enfasi aggiunta).

[60] Ibid.

> dello sviluppo e del funzionamento dei sistemi di IA; ritiene sia necessario istituire un *regime chiaro* ed *equo* per attribuire la *responsabilità giuridica* e *imputabilità* delle potenziali conseguenze negative prodotte da tali tecnologie digitali avanzate; sottolinea tuttavia che il primo e principale scopo deve innanzi tutto essere la prevenzione di tali conseguenze[61];

Nel testo originale, tuttavia, i termini utilizzati al posto di "responsabilità giuridica" e "imputabilità" sono, rispettivamente, "*legal responsibility*" e "*liability*". In primo luogo, l'utilizzo di queste parole nella versione inglese della Risoluzione risulta ridondante, posto che il prefisso "*legal*" dinanzi a "*responsibility*" denota esattamente il tipo di responsabilità dalla quale derivano conseguenze giuridiche: la *liability,* per l'appunto. In secondo luogo, "*legal responsibility*" questa volta viene tradotto in italiano con "responsabilità giuridica", invece che solamente con "responsabilità" come nell'antecedente Considerando J. In terzo luogo, il termine "*liability*" viene tramutato nella versione italiana in "imputabilità". Questo comporta l'ingresso nella Risoluzione di un istituto ulteriore, che rappresenta il cuore pulsante del diritto penale[62] e che ha senz'altro un significato diverso da quello di *liability*.

Vi è poi un'ulteriore disarmonia nella Risoluzione, questa volta a livello di contenuto piuttosto che di traduzione. Tenendo come riferimento la versione in inglese, è possibile notare che nel Considerando J viene affermata l'imperatività di garantire sempre una forma di *human accountability.* Al successivo articolo 13, invece, viene sancito l'obbligo di individuare sempre un agente, sia questo persona fisica o giuridica, che sia *legally responsible* e *liable* per le decisioni prese con il supporto del sistema di IA. Le conseguenze della scelta fra prevedere una figura obbligatoria di "*accountable human*" oppure un "*legally responsible and liable human*" non sono di poco conto. Se l'*accountability* può essere infatti garantita con strumenti extra-penali,

[61] "13. Recognises the challenges to the correct location of legal responsibility and liability for potential harm, given the complexity of development and operation of AI systems; considers it necessary to create a clear and fair regime for assigning legal responsibility and liability for the potential adverse consequences produced by these advanced digital technologies; underlines, however, that the aim must, first and foremost, be to prevent any such consequences materialising to begin with; calls, therefore, for the application of the precautionary principle in all applications of AI in the context of law enforcement; underlines that legal responsibility and liability must always rest with a natural or legal person, who must always be identified for decisions taken with the support of AI; emphasises, therefore, the need to ensure the transparency of the corporate structures that produce and manage AI systems". Parlamento Europeo, *L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*, cit., articolo 13 (enfasi aggiunta).

[62] Si veda in tema, *ex multis*, M. Bertolino, *L'imputabilità e il vizio di mente nel sistema penale*, Milano, 1990.

quali ad esempio l'adozione di procedure di *auditing*, garantire che sia sempre individuato *ex ante* un soggetto (penalmente) responsabile comporterebbe, in forza di tale posizione di garanzia, la sua assoggettabilità alle fattispecie di reato previste dai singoli ordinamenti, con tutte le difficoltà applicative che ne derivano e che verranno esposte in seguito.

L'articolo 13 della Risoluzione prosegue poi affermando che il Parlamento Europeo

> invita, pertanto, ad applicare con coerenza il *principio di precauzione* per tutte le applicazioni di IA nel contesto delle attività di contrasto;
> sottolinea che la responsabilità giuridica e l'imputabilità *devono sempre ricadere su una persona fisica o giuridica*, che deve sempre essere identificata per le decisioni assunte con il sostegno dell'IA; sottolinea, pertanto, l'esigenza di assicurare la trasparenza delle strutture aziendali che producono e gestiscono i sistemi di IA […][63].

Nella Risoluzione viene quindi ribadito che il primo e principale scopo di queste future norme sarà la prevenzione di effetti negativi, ispirata al principio di precauzione. L'invocazione di questo principio ci riporta immediatamente alle riflessioni in tema di c.d. diritto penale del rischio, che non è possibile affrontare *funditus* in questa sede[64], ma che sono di massimo rilievo alla luce dell'impostazione prescelta dalla proposta di regolamento europeo presentata dalla Commissione nell'aprile 2021 denominata "AI Act", sul quale ci concentreremo brevemente[65].

L'AI Act, infatti, trova le basi proprio in un *risk-based approach*. La proposta di regolamento stabilisce diversi livelli di rischio: i) inaccettabile; ii) rischio alto; rischio basso o minimo[66]. Ad ogni livello di rischio corrispondono poi una serie di requisiti e di obblighi, tra cui, per i sistemi ad alto rischio, l'obbligo di *sorveglianza umana ("human oversight")*. L'intervento umano viene previsto come misura per prevenire, nonché governare, il rischio della realizzazione di effetti pregiudizievoli causati da

[63] Parlamento Europeo, *L'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*, cit., articolo 13 (enfasi aggiunta).

[64] Sul tema si veda, *ex multis*, A. Massaro, *Principio di precauzione e diritto penale:* nihil novi sub sole, in *Diritto Penale Contemporaneo*, 2011; G. Forti, *"Accesso" alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione*, in *Criminalia*, 2006; F. Giunta, *Il diritto penale e le suggestioni del principio di precauzione*, in *Criminalia*, 2006; C. Piergallini, *Il paradigma della colpa nell'età del rischio: prove di resistenza del tipo*, in *Riv. it. dir. proc. pen.*, 2005.

[65] Commissione Europea, *Proposta di regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'unione*, 2021/0106(COD), 21 aprile 2021.

[66] B. Panattoni, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologia all'autonomia artificiale*, cit., 331-333.

un sistema di IA. Allo stesso tempo, possiamo presumere che il "sorvegliante umano" sarà uno dei soggetti responsabili qualora tale rischio si concretizzi ed egli risulti inadempiente rispetto ai propri obblighi di sorveglianza, ossia uno dei soggetti "*legally responsible and liable*" auspicati dal sopramenzionato articolo 13 della sopramenzionata Risoluzione del Parlamento europeo dell'ottobre 2021.

Spostando di nuovo l'attenzione sull'AI Act, nel considerando (48) viene affermato che tra le misure di sorveglianza umana adeguate vi deve essere la garanzia che il sistema "risponda all'operatore umano, e che le persone fisiche alle quali è stata affidata la sorveglianza umana dispongano delle competenze, della formazione e dell'autorità necessarie per svolgere tale ruolo"[67]. L'articolo 14 dell'AI Act è dedicato interamente alla sorveglianza umana e stabilisce, al comma 2, che questa mira "a *prevenire* o *ridurre al minimo* i *rischi* per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile"[68]. Il comma 4 contiene poi un elenco dettagliato delle azioni che dovrebbero essere poste in essere da coloro ai quali viene affidata la sorveglianza umana. Tra queste rientra, ad esempio, intervenire sul funzionamento del sistema di IA ad alto rischio o di interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga (Art. 14, c. 4, lett. e).

In conclusione, nonostante gli strumenti menzionati non abbiano, per ora, forza vincolante, rappresentano senz'altro un indizio di quelli che saranno i prossimi passi dell'Unione europea in questa area di conflitto. La Risoluzione del Parlamento europeo dell'ottobre 2021 e l'AI Act contengono difatti l'invito, forse troppo fiducioso, alla creazione di un quadro regolatorio *forte* e fungono da ottimo punto di partenza per svolgere riflessioni su come tali regole prenderanno forma. Ciò posto, nel proseguo di questo scritto si discuterà se nello specifico campo delle tecnologie basate su IA sia riconoscibile un nesso diretto fra condotte individuali di *human oversight* e l'attribuzione di responsabilità per colpa.

[67] Commissione Europea, *Proposta di regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'unione*, cit., (48).

[68] Commissione Europea, *Proposta di regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'unione*, cit., Art. 14, c. 2 (enfasi aggiunta).

3.2. - Human oversight *e* human in the loop*:* petitio principii*?*

Una delle caratteristiche dei sistemi di IA che può portare ad un malfunzionamento dei meccanismi classici di attribuzione della responsabilità penale è il suo agire autonomo. Se si volge lo sguardo all'*AI Act* pare che una delle soluzioni a tale problema vada individuata nel concetto di "sorveglianza umana" e nell'utilizzo di tecniche c.d. "*Human-In-The-Loop*" (HITL)[69]. Queste consistono nella realizzazione di sistemi di IA in cui il modello (*output*) viene sviluppato tramite l'interazione con un agente umano che, ad esempio, può svolgere il ruolo di "insegnante" nella fase di addestramento del sistema, fornendo un *feedback* alla macchina sul risultato ottenuto[70].

Tali soluzioni, però, presentano alcune problematicità. Come è stato affermato da alcuni autori[71], collocare un essere umano nel *loop* ("circolo") dell'IA, sebbene potrebbe sembrare *prima facie* una soluzione rassicurante, poggerebbe in realtà a sua volta su una logica circolare che distrae dagli usi intrinsecamente dannosi dei sistemi automatizzati. I *policymaker*, infatti, si stanno rivolgendo agli esseri umani per mitigare i rischi posti da sistemi di IA sulla base del presupposto che gli esseri umani siano effettivamente in grado di vigilare sui loro processi decisionali. Come potrà un uomo, nella prassi, supervisionare un sistema che è stato creato per superarlo e per colmare le sue carenze? In altre parole, il concetto di supervisione umana sembra essere usato come una sorta di unguento miracoloso, messo alla bell'e meglio per rimediare ai potenziali rischi posti dall'agire algoritmico. Il vero rischio, però, è quello di fare dell'uomo che supervisiona un capro espiatorio, in evidente contrasto con i principi generali del nostro diritto penale, in particolare con il rifiuto di responsabilità oggettiva.

In conclusione, i quesiti che sorgono sono molteplici. Evidentemente, nel campo dell'IA *human oversight* e *accountability* procedono di pari passo. Ma quale potrebbe essere la relazione tra *human oversight* e *responsabilità penale*? Del resto, i concetti di supervisione, controllo, sorveglianza appartengono senza dubbio al vocabolario del diritto penale. È possibile identificare il sorvegliante umano quale titolare di una posizione di garanzia? Potremmo considerare la disciplina europea in materia di *human oversight* una forma embrionale di nuove regole cautelari di condotta?

---

[69] Commissione europea, *Proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce norme armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e che modifica alcuni atti legislativi dell'Unione*, cit., art. 14.

[70] Per un'indagine sugli approcci HITL applicabili al *machine learning*, si veda X. Wu et al., *A Survey of Human-in-the-loop for Machine Learning*, arXiv:2108.00941, 2021.

[71] B. Green e A. Kak, *The False Comfort of Human Oversight as an Antidote to A.I. Harm*, in *Slate*, 15 giugno 2021.

Partendo da questi quesiti, nel prosieguo di questo scritto tratteremo brevemente i profili relativi alla responsabilità per posizione di garanzia, per poi approfondire i "conflitti" in materia di elemento soggettivo del reato e, in particolare, quelli relativi alla colpa.

3.3. - *Quale conflitto? (1) Posizioni di garanzia*

A questo punto della trattazione diventa rilevante svolgere una breve riflessione sulla possibilità di individuare una *posizione di garanzia* in capo ad un "sorvegliante umano". La dottrina, invero, denuncia da tempo la distorta operazione di sovrapposizione fra l'accertamento della causalità omissiva e quello della colpa[72]. Pertanto, si approccerà il tema con cautela, tenendo ben a mente la distinzione fra inosservanza di un obbligo di garanzia e inosservanza di un obbligo di diligenza[73].

Secondo alcuni autori sarebbe possibile identificare una posizione di garanzia "potenziale", nel campo specifico delle auto a guida semiautonoma[74], in capo al conducente presente all'interno del veicolo[75]. Tale posizione si "attiverebbe" nel mo-

[72] Quest'ultima, infatti, è connaturata da una costante componente omissiva, "anche quando l'addebito concerne una condotta strutturalmente commissiva (chi ha investito un pedone passando con il rosso ha certamente causato la morte di quel pedone ai sensi dell'art. 40 co. 1 c.p., ma anche omesso di adempiere alla norma cautelare che gli imponeva, appunto, di fermarsi al rosso)". Così F. Viganò, *Il rapporto di causalità nella giurisprudenza penale a dieci anni dalla sentenza Franzese,* in *Diritto Penale Contemporaneo,* 2013, 391 (corsivo dell'autore). Sul tema si veda, nella vastissima dottrina, R. Bartoli, *Il problema della causalità penale. Dai modelli unitari al modello differenziato*, Torino, 2010; M. Donini, *Imputazione oggettiva dell'evento. "Nesso di rischio" e responsabilità per fatto proprio*, Giappichelli, 2006; A. Massaro, *La colpa nei reati omissivi impropri*, Aracne, 2011; C. E. Paliero, *La causalità dell'omissione: formule concettuali e paradigmi prasseologici*, in *Riv. It. Med. Leg.*, 1992; K. Sommelier, *Causalità ed evitabilità. Formula della* condicio sine qua non *e rilevanza dei decorsi causali ipotetici nel diritto penale,* ETS, 2013; M. Trapasso, *Imputazione oggettiva e colpa tra azione ed omissione: dalla struttura all'accertamento*, in *Ind. pen.,* 2003.

[73] Si veda, per tutti, F. Mantovani, *Op. cit.*, 187, n. 51 e i rimandi dottrinali ivi contenuti.

[74] Trattiamo, in particolare, di auto classificate come livello 3 secondo lo standard J3016 dalla SAE (*Society of Automotive Engineers*), ossia di veicoli che possono porre in essere tutte le manovre necessarie per la guida del veicolo (ad esempio l'accelerazione frenata). La persona all'interno del veicolo, quindi, non lo guida a tutti gli effetti, anche se è seduto nel posto del conducente. Egli dovrà intervenire solamente qualora venga effettuata una richiesta esplicita in tal senso dal sistema di IA. Questo tipo di tecnologia viene definito "*hands and feet free but not 'mind free' driving*" da V.A. Banks et al., *Subsystems on the road to full vehicle automation: hands and feet free but not 'mind' free driving",* in *Safety Science*, vol. 62, 2014, pp. 505-514.

[75] Si tratterebbe, in particolare, di una posizione di controllo. Così A. Cappellini, *Profili penalistici delle* self-driving cars*,* cit., 334 e C. Piergallini, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., 1751.

mento in cui il sistema di guida richieda l'intervento del conducente affinché riprenda il controllo della vettura. Seguendo questa impostazione, a partire da quel momento sarebbe possibile muovere un rimprovero nei confronti del conducente per aver omesso il controllo sulla vettura, qualora a tale omissione consegua un evento dannoso[76]. In altre parole, a differenza del conducente di un'auto a guida "normale" – che "attiva" la fonte di rischio accendendo ed immettendo in circolazione il veicolo – nel caso di auto a guida semiautonoma tale situazione di rischio sarebbe latente fino alla *demande de reprise*[77]. Solo allora avverrebbe il trasferimento del controllo dell'attività rischiosa in capo al conducente, fino ad allora svolto dal veicolo[78], e, di conseguenza, solo allora potrebbe esservi responsabilità in capo al conducente per aver malgovernato tali rischi[79]. A tale proposito, ci si può porre il problema – lasciando, per ora, il quesito aperto – della necessità o meno di chiamare in causa fattispecie di responsabilità omissiva in questi specifici scenari. Si potrebbe sostenere, infatti, la responsabilità del conducente per una sua condotta commissiva colposa, in quanto l'attività di controllo sull'auto semiautonoma potrebbe rientrare nell'alveo applicativo delle ordinarie regole sulla circolazione stradale[80], con tutte le ricadute che questo porterebbe[81].

[76] Si veda in merito V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, cit., nota 13, 4, il quale sostiene "[…] è chiaro che se il guidatore non è richiesto di monitorare il traffico sino a una richiesta di riassunzione della funzione di guida, lo stesso non può più essere ritenuto in concreto "human in command" né dunque (penalmente) responsabile di eventuali causazioni lesive occorse sino a quel momento, ove appunto il controllo sulla attività rischiosa era delegato alla macchina AI driven legalmente autorizzata".

[77] Termine utilizzato dal legislatore francese nella *Ordonnance n° 2021-443 du 14 avril 2021 relative au régime de responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite et à ses conditions d'utilisation*, cit. Si veda anche M. Giuca, *Disciplinare l'intelligenza artificiale. La riforma francese sulla responsabilità penale da uso di auto a guida autonoma*, cit., 23 e ss.

[78] Ibid.

[79] A. Cappellini, *Profili penalistici delle* self-driving cars, cit., 334, n. 49.

[80] Come è stato evidenziato da A. Cappellini, è noto che "[n]ella colpa stradale ordinaria […] l'elemento imprudente, anche qualora consista effettivamente in una mancanza di un qualcosa, non rilevi tanto come omissione in sé, quanto piuttosto come un'omissione inserita in una condotta più ampia – la guida del veicolo – avente carattere complessivo indiscutibilmente commissivo". Difatti, continua l'autore, vi è "un momento omissivo in ogni tipo di colpa: ogni violazione cautelare, anche attiva, può essere sempre vista, allo specchio, come l'omissione di una cautela doverosa, del comportamento alternativo lecito". Secondo tale dottrina nel caso di auto a guida-semiautonoma (dotate di livello di automazione pari a 3) si verificherebbe un passaggio "dal modello di evento lesivo cagionato commissivamente per colpa dai conducenti di veicoli tradizionali" a quello "*imputativo alternativamente commissivo od omissivo*". Si avrebbe dunque responsabilità omissiva nel caso in cui la vettura, per un errore o un malfunzionamento, causi un incidente dopo svariate ore di mancata sorveglianza, mentre si avrebbe responsabilità commissiva nel caso in cui il conducente eserciti "maldestramente" il suo potere di controllo ed

Ciò posto, è plausibile affermare che in questo momento la problematica della posizione di garanzia abbia una rilevanza ridotta per una serie articolata di motivi, fra tutti a causa dell'assenza del carattere fondamentale della *giuridicità* della stessa[82]. Difatti, le fonti europee già menzionate e la normativa interna nulla prevedono espressamente in tal senso e non si ravvisa la possibilità di sussumere, in via interpretativa, i sistemi di IA all'interno della sfera applicativa degli obblighi giuridici di garanzia già esistenti, proprio a causa delle già esposte peculiarità di tali tecnologie, nonché delle criticità che verranno evidenziate in seguito[83].

Senz'altro, in un'ottica *de iure condendo*, qualsiasi sforzo in questa direzione dovrebbe essere idoneo a identificare *esattamente* il soggetto garante investito dall'obbligo e titolare di un *effettivo* potere di impedimento, in conformità con i principi fondanti del diritto penale[84]. Ciò è reso difficile, se non impossibile, prima di tutto dalla presenza di una molteplicità di soggetti coinvolti nell'ideazione, produzione, commercializzazione ed utilizzo di sistemi di IA, nonché a causa della parcellizzazione dei ruoli dei singoli agenti umani nel corso di tale processo (il c.d. "*many hands problem*")[85] e delle loro responsabilità[86]. D'altronde, l'introduzione di una po-

incorra così in errori. Cfr. A. Cappellini, *Profili penalistici delle* self-driving cars*,* cit., 335-336, enfasi aggiunta. Questa argomentazione, tuttavia, non attribuisce rilevanza alla richiesta del sistema di riprendere il controllo, che di conseguenza dovrebbe essere esercitato dal conducente fin dall'inizio della circolazione a bordo del veicolo semi-autonomo. Difatti, altra parte della dottrina, come evidenziato poc'anzi, riconosce in tale richiesta il momento dell'"attivazione" della posizione di garanzia. Ciò è rilevante perché, in termini parzialmente difformi, si può porre lo stesso quesito (relativo alla scelta di un paradigma commissivo od omissivo) rispetto alla possibile configurazione di una responsabilità per la *mancata ripresa del controllo* dell'auto a fronte di una richiesta del sistema, nonché per le eventuali conseguenze dannose che ne potrebbero derivare. In altri termini, ci si può chiedere se anche tale specifica condotta sia riconducibile o meno alle più generali norme sulla circolazione stradale già in essere.

[81] Si veda *infra* 3.4.1.

[82] Si veda, per tutti, F. Giunta, *La posizione di garanzia nel contesto della fattispecie omissiva impropria*, in *Diritto penale e processo*, fasc. 5, 1999.

[83] A. Gargani, in *Lo strano caso dell'"azione colposa seguita da omissione dolosa". Uno sguardo critico alla sentenza "Vannini",* in *disCrimen*, 2020, 10 e ss., analizzando la recente sentenza della Corte di Cassazione sul c.d. "caso Vannini" (Cass. Pen., Sez. V, 19 luglio 2021, n. 27905), denuncia "l'ennesima creazione giurisprudenziale di posizioni di garanzia '*extra ordinem*', fondate su (inespresse) ragioni di giustizia sostanziale" quali "l'inosservanza di elementari doveri *etico-morali* di assistenza e di solidarietà umana".

[84] Cfr. *ex multis* F. Mantovani, *Diritto penale. Parte Generale*, cit., 172 e ss.

[85] Il fenomeno è stato dapprima analizzato nell'ambito della filosofia morale da D. F. Thompson, *Designing Responsibility: The Problem of Many Hands in Complex Organizations*, in *The American Political Science Review,* 74(4), 9, 1980. In seguito, lo sviluppo della letteratura è stato esponenziale. Si veda, ad esempio, M. Bovens, *The quest for responsibility. Accountability and citizenship in complex organisations*, Cambridge University Press, 1998; *Moral Responsibility and the Problem of Many*

sizione di garanzia "onnicomprensiva", che comporti cioè l'obbligo esteso per un non meglio identificato "sorvegliante umano" di prevenire ogni danno causato da un sistema di IA, sarebbe in palese contrasto con il principio di tassatività[87].

Inoltre, laddove si aprisse anche alla responsabilità omissiva, lasciando l'analisi delle questioni specifiche relative all'elemento soggettivo ai paragrafi successivi, sopravviverebbero quesiti di non facile risoluzione. Non si possono ignorare, infatti, le problematicità che i sistemi di IA sollevano per l'accertamento del nesso causale[88] e, di conseguenza, le possibili ricadute sul profilo dell'impedibilità dell'evento. Tali difficoltà sorgono perché la "spiegazione" del nesso causale è legata alla possibilità per l'uomo di dominare completamente (eziologicamente) un certo evento, circostanza che in certi sistemi di IA potrebbe semplicemente non essere possibile. Possiamo dimostrare, infatti, che un sistema di IA, partendo da un insieme di *input,* ha prodotto alcuni *output*, ma potremmo non essere in grado di spiegare né perché, né come. A volte, potremmo anche non riuscire a capire quali *input* hanno avuto un ruolo nell'ottenere l'*output.* In questi casi, dunque, "la distanza tra un'azione umana e le sue conseguenze *[dannose]* aumenta esponenzialmente"[89].

In altre parole, quando si ha a che fare con l'agire di sistemi di IA ci si scontra con la presenza contemporanea di una miriade di fattori causali alternativi, sia umani

Hands, a cura di I.R. Poel et al., Routledge, 2015; D. F. Thompson, *Designing Responsibility: The Problem of Many Hands in Complex Organizations*, in *The Design Turn in Applied Ethics,* a cura di J. van den Hoven et al., Oxford University Press, 2017.

[86] Come affermato da C. Iagnemma, *Il reato omissivo improprio nel quadro di un approccio sistemico all'evento offensivo,* in *Criminalia*, 2020, 3, "l'esito infausto, scaturendo dalle imprevedibili correlazioni tra il fattore organizzativo, strutturale e tecnologico, è di rado fronteggiabile in forma individuale". L'autrice, riprendendo l'opera di F. Sgubbi*, Responsabilità penale per omesso impedimento dell'evento*, CEDAM, 1975, 206, sostiene che nei contesti connaturati da dinamiche organizzative complesse il criterio giuridico-formale per l'individuazione del vincolo di tutela soggetto garante – bene giuridico vacilli: "In tali contesti, infatti, essendo non agevole rinvenire precise indicazioni legislative sulla base delle quali ricostruire i rapporti di tutela intercorrenti tra i vari membri dell'ente e la molteplicità di beni esposti alle diverse fonti di pericolo che caratterizzano il piano organizzativo, strutturale e tecnologico, si corre il rischio … di continuare a interpretare il concetto di *Garantestellung* pur sempre in termini sostanzialistico-funzionali" (p. 10).

[87] Secondo A. Gargani, *Op. cit.*, 12, la giurisprudenza ricorre sempre più spesso all'espediente "abnorme" di "confezionare" posizioni di garanzia "para-giuridiche", fondate sulla cultura del "senso comune", per colmare delle (presunte) lacune della legge.

[88] U. Pagallo, *The Laws of Robots: Crimes, Contracts and Torts*, cit., 73, parla di "*failures of causation*" per descrivere l'effetto destabilizzante dell'autonomia dei sistemi di IA sull'accertamento del nesso causale.

[89] M. Hildebrandt, *Criminal Law and Technology in a Data-Driven Society*, in *The Oxford Handbook of Criminal Law*, a cura di M. D. Dubber T. Hörnle, 190.

che non[90]. Questo rende impraticabile, da un lato, l'individuazione del singolo fattore che non è stato attivato per impedire o interrompere il processo causale già iniziato[91] e, dall'altro, l'esclusione di fattori causali alternativi con la certezza richiesta dal nostro ordinamento.

3.4. - *Quale conflitto? (2) L'elemento soggettivo del reato*

La discussione in materia di IA e colpevolezza si articola su due assi portanti[92]. Il primo asse ha a che fare con la possibilità di concepire un sistema di IA "colpevole". Indubbiamente, "la possibilità di individuare una colpevolezza vera e propria – e non un simulacro della stessa – in capo ad un sistema di IA solleva […] non poche difficoltà, logiche e ontologiche"[93]. In questo dibattito la parte del leone la fanno i filosofi del diritto, piuttosto che i penalisti. Da un lato, vi è il fronte della liberazione robotica[94], a cui appartengono Chopra e White[95], i quali sostengono che prima o poi i robot saranno capaci di sensibilità al comando della norma penale e saranno pertanto rimproverabili mediante pena[96]. Dall'altro, vi è chi sostiene che non si possa parlare di *colpevolezza robotica*, perché i sistemi di IA mancano di autocoscienza (cioè non sono coscienti di essere coscienti), libero arbitrio e autonomia morale[97].

---

[90] Si fa riferimento qui alla commistione del c.d. "*many hands problem*" e dell'imprevedibilità connaturata alle specificità dell'IA, in particolare alle tecniche di *machine learning*. Per capire tale fenomeno è utile l'esempio sviluppato da F. Santoni de Sio e G. Mecacci, *Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them,* cit., 1062, *"… a vehicle may be operated by a driver D1, with the assistance of the automated driving system AS, produced by the car manufacturer X, powered with digital systems developed by the company Y, possibly including some form of machine learning developed by the company Z, and enriched by data coming from different sources, including the driving experience of drivers D2, D3…Dn; vehicles in this system are in principle subject to a standardisation process done by the agency S, the traffic is regulated by the governmental agency G, drivers are trained and licensed by the agency L etc. Second, some specific features of present-day learning AI systems may make this interaction particularly unpredictable – typically when the vehicles' performance is potentially re-designed by the second on the basis of new data acquisition and processing – and opaque, if the reasoning scheme underlying systems' actions is not easily accessible to their controllers, regulators, or even their designers*".

[91] F. Palazzo, *Corso di diritto penale. Parte generale*, cit., 261.

[92] Non affronteremo in questa sede i "conflitti" relativi agli altri elementi del reato.

[93] F. Basile in *Diritto penale e Intelligenza Artificiale*, cit., 72

[94] U. Pagallo, *The Laws of Robots: Crimes, Contracts and Torts*, cit., 54.

[95] S. Chopra, L. F. White, *A Legal Theory for Autonomous Artificial Agents,* Univ. of Michigan Press, 2011.

[96] Fanno parte dello stesso "fronte" anche G. Hallevy e Y. Hu, *supra*, n. 4.

[97] Questa è la posizione condivisa, ad esempio, da gran parte degli autori indicati alla n. 3. Si veda,

Il secondo asse del dibattito riguarda la responsabilità dell'agente umano di volta in volta coinvolto, il c.d. "*human-behind-the-machine*". Come già sottolineato in sede di introduzione, lo scritto si concentrerà solamente su questa seconda tematica, lasciando pertanto da parte questioni relative alla responsabilità diretta del sistema di IA, nonché tutti quei casi in cui il sistema venga utilizzato quale *strumento* per commettere il reato.

Entriamo ora nel regno dell'imputazione colposa.

3.4.1. - *La "sorveglianza umana". Profili di colpa*

Innanzitutto, è fondamentale riconoscere che i sistemi di IA sono già impiegati in attività appartenenti alle c.d. aree di "rischio consentito"[98]: si pensi ad esempio al settore sanitario o a quello dei trasporti[99]. Di conseguenza, operano in contesti dove sono già in essere delle regole di condotta, regole cautelari c.d. "improprie" in quanto determinate a ridurre il rischio di verificazione di un pericolo ineliminabile[100]. Tuttavia – ed è qui che si coglie la peculiarità legata all'IA – si deve anche indagare se la nostra società sia o meno in possesso al momento della corretta conoscenza tecnico-scientifica per applicare regole di condotta preesistenti specificatamente alle nuove casistiche che questi sistemi causeranno, o addirittura per svilupparne di nuove. La risposta a tale quesito parrebbe, per il momento, negativa. In altri termini, sebbene si possa ritenere adeguata l'invocazione del principio di precauzione[101] da parte delle autorità europee, si è d'accordo con chi ritiene che in via generale questo principio

ad esempio, A. Cappellini, Machina delinquere potest*? Brevi appunti su intelligenza artificiale e responsabilità penale*, cit., 14, il quale sostiene che l'agire dei sistemi di IA potrebbe apparire come qualificato da una "tipicità dolosa" che però "nulla dice nulla dice rispetto alla loro possibile colpevolezza, considerato che manca, ancora, una regola d'esperienza generale per cui le IA più avanzate, oltre che dotate di un margine di imprevedibilità, siano anche libere di autodeterminarsi al pari dell'uomo". Tale opinione è condivisa da C. Piergallini, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., pp. 1761 e ss. L'autore ritiene che un sistema di IA sia incapace di azione, di colpevolezza e di pena. L'agire di una macchina, infatti, difetterebbe di *suitas* e di libero arbitrio.

[98] C. Brusco, *Rischio e pericolo, rischio consentito e principio di precauzione. la c.d. "flessibilizzazione delle categorie del reato"*, in *Criminalia*, 2012, 389.

[99] Si vedano le riflessioni svolte *supra* (3.3) relativamente ai profili di responsabilità del conducente di auto a guida semiautonoma.

[100] Ivi, 391.

[101] Relativamente al principio di precauzione, oltre agli autori menzionati alla n. 59, si veda *ex multis*: D. Castronuovo, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, Aracne, 2012; G. Forti, *La "chiara luce della verità" e "l'ignoranza del pericolo". Riflessioni penalistiche sul principio di precauzione*, in *Scritti per Federico Stella*, Napoli, 2007, vol. I, 573.

debba fungere da faro per le scelte di politica legislativa, piuttosto che per quelle di penalizzazione[102]. Ciò non toglie che in futuro si arrivi a superare la soglia richiesta per imporre l'adozione di nuove regole cautelari, anche grazie al progresso nelle leggi scientifiche, in particolare di quelle che si occuperanno di indagare, ricostruire e spiegare il comportamento dei sistemi di IA più complessi[103].

In una direzione leggermente diversa, alcuni autori paventano la creazione di una nuova "colpa da programmazione", che farebbe ricadere dunque la responsabilità per i danni commessi dalla "creatura" sul "creatore"[104]. Sorgono quindi i seguenti interrogativi, ad oggi ancora privi di risposta: come dovrebbe essere configurato questo tipo di colpa? Più specificatamente, come dovrebbero essere caratterizzate le regole cautelari di condotta nella realizzazione e nell'implementazione di un sistema di IA? Qual è la condotta dell'agente umano modello? E soprattutto, chi è l'agente umano modello?

Possiamo sottolineare, poi, alcune criticità specifiche attinenti al precipuo aspetto dell'esigibilità della condotta doverosa (omessa da parte del "sorvegliante umano"), riconducibili al più ampio tema del c.d. "*human in the loop*" esposto poc'anzi[105]. In particolare, uno degli snodi più problematici consiste nel *livello* di attenzione esigibile dal potenziale supervisore, sia questo il conducente dell'auto a guida semiautonoma o il medico che utilizza un sistema diagnostico basato su IA. Si pensi, ad esempio, all'*automation complacency*, termine coniato in materia di incidenti aerei per far riferimento al fenomeno per cui l'automatizzazione di un qualsiasi compito porta il supervisore umano a confidare che la macchina se ne stia occupando in modo efficace e, di conseguenza, a smettere di prestare attenzione[106]; o all'*automation bias*, ossia la tendenza dell'essere umano a riporre fiducia eccessiva nelle raccomandazioni prodotte da un sistema informatico[107]. Tali fenomeni, amplifi-

[102] Parla di "criterio di politica legislativa" F. Giunta in *Il diritto penale e le suggestioni del principio di precauzione*, cit., 229.

[103] C. Brusco, *Rischio e pericolo, rischio consentito e principio di precauzione. la c.d. "flessibilizzazione delle categorie del reato*, cit., 398.

[104]M. B. Magro, *Biorobotica, robotica e diritto penale*, in *Genetics, Robotics, Law, Punishment*, a cura di D. Provolo, S. Riondato, F. Yenisey, Padova University Press, 2014, 516; V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, cit., 4.

[105] *Supra* 4.1.

[106] L. Smiley, *'I am the Operator': The Aftermath of a Self-Driving Tragedy*, in *WIRED*, 8 marzo 2022. Una delle prime definizioni di *automation complacency* è quella sviluppata da E.L. Wiener, *Complacency: Is the term useful for air safety?*, in *Proceedings of the 26th Corporate Aviation Safety Seminar*, 1981. Tra le più recenti, si veda R. Parasuraman, D. H. Manzey, *Complacency and Bias in Human Use of Automation: An Attentional Integration*, Volume 52, Issue 3, 2010.

[107] Viene definita quale la "la tendenza a usare l'automazione come euristica sostitutiva della ricerca e

cati nel caso di automatizzazione basata su IA riducono sensibilmente la soglia di attenzione dell'agente umano alle prese con la macchina, e, di conseguenza, diminuiscono la soglia dell'esigibilità di una condotta diligente da parte dello stesso.

Ma chi è, esattamente, il supervisore? È possibile sviluppare alcune riflessioni ulteriori partendo da questo quesito. Innanzitutto, esaminando le fonti sopramenzionate, si può notare come venga data poca o addirittura nessuna attenzione ai diversi ruoli che possono essere ricoperti nella catena di sviluppo di un sistema di IA. Viene infatti fatto riferimento esclusivamente alla figura del programmatore, che diventa di conseguenza *epicentro* della responsabilità. Eppure, gli *AI-team* comprendono al loro interno diverse figure, tra cui i *data analyst* e i *data scientist*, che si occupano della raccolta e dell'interpretazione dei dati e si assicurano che siano pertinenti ed esaustivi; i *data* e i *machine learning engineer*, che costruiscono e testano i modelli di *machine learning* ed i programmatori, che implementano la soluzione, trasformando il modello in codice. Tali ruoli si sommano a quello dell'utente, ossia il soggetto che *opera* il sistema di IA. Ci si potrebbe domandare, dunque, se a questi ruoli potrebbero corrispondere equivalenti tipi di agente modello.

In aggiunta, possiamo notare come raramente si discuta della responsabilità collegata ad errori o inadeguatezza dei dati di addestramento del sistema di IA. Sappiamo infatti che i sistemi basati su *machine learning* sono "affamati" di dati: più vengono alimentati con dati etichettati, più accurata sarà la loro previsione. Pensiamo ad un sistema che ha come scopo classificare se, in una foto, è presente un animale, una persona o un'automobile. Il problema è che è impossibile fornire ogni possibile esempio di dati etichettati al sistema. In altre parole, non sarà possibile istruire l'algoritmo sulla base di tutte le immagini di animali, o persone, o automobili nel mondo. Di conseguenza, l'algoritmo dovrà generalizzare tra i suoi esempi per classificare dati che non ha mai visto prima. Inoltre, i database sono colmi di errori[108]. Allo stesso tempo, i sistemi di IA, addestrati su database difettosi, vengono utilizzati in ambienti critici, dove una generalizzazione errata potrebbe rivelarsi fatale. Si potrebbe citare come esempio l'ormai famoso incidente mortale dell'auto a guida autonoma di Uber[109], causato (anche) dall'inabilità del software del veicolo di identificare

dell'elaborazione vigile delle informazioni" da K. Mosier, L.J. Skitka, *Automation use and automation bias*, in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1999, 344.

[108] D. Kang et al., *Finding Errors in Perception Data With Learned Observation Assertions*, Stanford Dawn, 24 gennaio 2022, disponibile presso: https://dawn.cs.stanford.edu/2022/01/24/loa/.

[109] D. Wakabayashi, *Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam*, in *The New York Times online*, 19 marzo 2018, disponibile presso: https://www.nytimes.com/2018/03/

la presenza di un pedone che attraversa sulla strada senza utilizzare le strisce pedonali[110]. Cosa succede, allora, quando l'algoritmo è stato addestrato sulla base di dati difettosi, distorti o sbagliati? Chi è responsabile? Aspetti come la raccolta e la generazione dei dati, l'etichettatura di questi e la valutazione della loro qualità sono essenziali per evitare che il sistema commetta errori, e quindi che si verifichino eventi dannosi. Eppure, non hanno ancora guadagnato sufficiente popolarità nelle analisi di stampo prettamente penalistico. La comunità scientifica, invece, ha iniziato ad analizzare questo problema. Si pensi al movimento "*Data Centric AI*"[111] (DCAI), che mira a spostare l'attenzione dell'ingegneria *machine learning* dalla modellazione ai dati sottostanti utilizzati per addestrare e poi valutare i modelli.

In conclusione, qualora si escludesse l'errore del c.d. supervisore, il danno causato da un sistema di IA potrebbe essere ricondotto ad un difetto di funzionamento dello stesso, sollevando così questioni appartenenti al sistema della responsabilità per danno da prodotto difettoso[112]. L'errore commesso durante l'addestramento dell'algoritmo potrebbe essere parificato ad un errore nell'assemblaggio di una macchina[113]? In questo ambito, come evidenziato da attenta dottrina, le difficoltà sarebbero molteplici: *in primis* ci si chiede come andrebbe condotta l'indagine sull'eziologia del difetto quando il sistema di IA è il risultato della cooperazione di una o più organizzazioni altamente complesse[114].

4. Conclusioni

Il concetto di un'IA criminale non è nuovo. Al contrario, la fantascienza si occupa da tempo di "robot malvagi" che si ribellano agli umani e ne prendono il con-

19/technology/uber-driverless-fatality.html.

[110] "The system never classified her as a pedestrian — or correctly predicted her path — because she was crossing N. Mill Avenue at a location without a crosswalk, and the system design did not include consideration for jaywalking pedestrians". Cfr. National Transportation Safety Board, *Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, Arizona, March 18, 2018,* Highway Accident Report NTSB/HAR-19/03, 16.

[111] Si veda il sito https://datacentricai.org

[112] Per un'analisi del sistema della responsabilità per danno da prodotto difettoso, si veda l'opera di C. Piergallini, *Danno da prodotto e responsabilità penale. Profili dommatici e politico-criminali*, cit.

[113] A. Cappellini, Machina delinquere potest*? Brevi appunti su intelligenza artificiale e responsabilità penale*, cit., 9.

[114] C. Piergallini, *Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?*, cit., 1752. Si veda anche *supra*, 3.3.

trollo, di macchine che "impazziscono" e agiscono in modo imprevedibile[115]. In effetti, si potrebbe sostenere che le tre leggi della robotica di Asimov non sono altro che il più famoso tentativo di regolare forme di condotte illecite dell'IA[116].

Questo breve contributo ha voluto fornire alcune riflessioni – senza alcuna pretesa di esaustività – relativamente ai profili di intersezione fra IA e materia penale. L'IA funge da prova di impatto per gli istituti tradizionali del diritto penale, un test d'urto per il fatto tipico, il nesso causale, il dolo, la colpa. Particolare attenzione è stata prestata all'attore europeo, che si sta affermando quale *key normative player* in questo ambito. Anziché presentare una disamina globale di questa complessa problematica, si è preferito limitare l'analisi alle c.d. *zone di conflitto*, e in particolare ci si è interrogati sui possibili profili di connessione tra il concetto di *human oversight* e quello penalistico di colpa. L'auspicio è che questa attività di diagnosi, nonché i quesiti identificati nel corso della stessa, fungano da strumento utile per l'*AI-criminal scholar* del futuro.

[115] "*Think instead of the false Maria in Metropolis (1927); Hal 9000 in 2001: A Space Odyssey (1968)[…]; C3PO in Star Wars (1977); Rachael in Blade Runner (1982); Data in Star Trek: The Next Generation (1987); Agent Smith in The Matrix (1999) or the disembodied Samantha in Her (2013)*", L. Floridi, *Should we be afraid of AI?*, in *aeon.co*, Disponibile presso: https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible.

[116] S. N. Lehman-Wilzing, *Frankenstein unbound: Towards a legal definition of artificial intelligence*, in *Futures*, vol. 13, n. 6, 1981, 445.