DIRITTO PENALE
E UOMO

Riflessione

Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine

Artificial intelligence and criminal law: four possible research leads

di Fabio Basile

Abstract. *Gli ambiti all'interno dei quali la rivoluzione tecnologica messa in moto dall'IA potrebbe più significativamente impattare con le pretese di tutela dei beni giuridici, affidate al diritto penale, sono fondamentalmente quattro: le attività di* law enforcement *e, in particolare, di polizia predittiva, dove i sistemi di IA possono fornire un importante contributo per contrastare, o meglio ancora prevenire, la commissione di reati; il possibile impiego di algoritmi decisionali per risolvere vertenze penali, così da operare una sorta di sostituzione, o per lo meno di affiancamento, del giudice-uomo col giudice-macchina; la valutazione della pericolosità criminale affidata ad algoritmi predittivi, capaci di attingere e rielaborare quantità enormi di dati al fine di far emergere relazioni, coincidenze, correlazioni, che consentano di profilare una persona e prevederne i successivi comportamenti, anche di rilevanza penale; infine, le possibili ipotesi di coinvolgimento – come strumento, come autore, o come vittima – di un sistema di IA nella commissione di un reato. Il presente contributo si propone di illustrare tali ambiti, indicando problemi e prospettive connessi all'impiego dei sistemi di IA.*

Abstract. *The areas within which the technological revolution set in motion by the AI could more significantly impact with the claims of protection of legal assets, entrusted to criminal law, are basically four: law enforcement and, in particular, predictive policing, where AI systems can make an important contribution to combating, or better yet preventing, the commission of crimes; the possible use of decision-making algorithms to resolve criminal disputes, in order to operate a sort of replacement, or at least of juxtaposition, of the judge-man with the judge-machine; the crime risk assessment entrusted to predictive algorithms, able to draw and re-elaborate enormous quantities of data in order to bring out relationships, coincidences, correlations, that make it possible to profile a person and predict his or her subsequent behavior, even criminal; finally, the possible hypotheses of involvement - as a tool, as an author, or as a victim - of an AI system in the commission of a crime. This work aims to illustrate these areas, indicating problems and prospects connected to the use of AI systems.*

1. Premessa: limiti e obiettivi dell'indagine.

L'intelligenza artificiale (nel prosieguo, anche "IA") è già presente in molti aspetti della nostra vita quotidiana. È alla base di tutte le ricerche su Internet e di tutte le *app*; è in ogni richiesta fatta al GPS, in ogni videogame o film d'animazione, in ogni banca e compagnia di assicurazione, in ogni ospedale, in ogni drone e in ogni auto a guida autonoma, e in futuro – questa la previsione di una delle massime esperte della materia – «ce la ritroveremo dappertutto»[1]: e, ovviamente, anche in ambiti che hanno immediata rilevanza per il diritto penale. In effetti, tra gli esperti della materia «l'opinione comune è che oggi stiamo vivendo in un mondo sempre più dominato dall'Intelligenza Artificiale [...] grazie alla proliferazione di tecniche di Intelligenza Artificiale che riescono a imparare molto velocemente ed efficacemente, come gli algoritmi di *machine learning*, le tecniche di *mining* e i sistemi predittivi, che sembrano promettere un livello senza precedenti, e forse anche un po' spaventoso, di IA nella nostra vita e nelle nostre società»[2].

Se poi volessimo lanciarci in previsioni di lungo termine, potremmo citare Stephen Hawking, ad avviso del quale «nell'arco dei prossimi cento anni, l'intelligenza dei computer

[1] M.A. Boden, *Intelligenza artificiale*, in J. l-Khalili (a cura di), *Il futuro che verrà*, Bollati Boringhieri, 2018, p. 133.
[2] G.F. Italiano, *Intelligenza artificiale: passato, presente, futuro*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappicchelli, 2018, p. 216. Per analoghe considerazioni, v. J. Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*, Luiss University Press, II ed., 2018, pp. 81 ss., e pp. 193 ss.; L. Floridi, *What the Near Future of Artificial Intelligence Could Be*, in *Philosophy & Technology*, fasc. 32, 2019, pp. 3 ss. (*online* al link https://doi.org/10.1007/s13347-019-00345-yp); M.C. Carrozza, *I Robot e noi*, Il Mulino, 2017, pp. 3 ss. Per una panoramica sul mondo dei sistemi di IA, dalle sue origini a oggi, raccontata da alcuni esperti della materia, si vedano i saggi raccolti in D. Heaven (a cura di), *Macchine che pensano. La nuova era dell'intelligenza artificiale*, Edizioni Dedalo, 2018.

supererà quella degli esseri umani»[3]. E che non si tratti di pura fantascienza, è confermato dal fatto che un'analoga previsione è contenuta anche nei *Considerando* della Risoluzione del Parlamento europeo sulla robotica del 16 febbraio 2017[4], dove si afferma che «è possibile che a lungo termine l'intelligenza artificiale superi la capacità intellettuale umana»[5].

Peraltro, proprio una scorsa a taluni dei *Considerando* di detta Risoluzione può fornire un potente stimolo al giurista ad indagare approfonditamente i possibili scenari delle correlazioni tra IA e diritto in genere, e diritto penale in particolare. Ivi, infatti, si afferma che:

> «A. [...] dal mostro di Frankenstein ideato da Mary Shelley al mito classico di Pigmalione, passando per la storia del Golem di Praga e il robot di Karel Čapek, che ha coniato la parola, gli esseri umani hanno fantasticato sulla possibilità di costruire macchine intelligenti, spesso androidi con caratteristiche umane;
> B. [...] l'umanità si trova ora sulla soglia di un'era nella quale robot, bot, androidi e altre manifestazioni dell'intelligenza artificiale sembrano sul punto di avviare una nuova rivoluzione industriale, suscettibile di toccare tutti gli strati sociali, rendendo imprescindibile che la legislazione ne consideri le implicazioni e le conseguenze legali ed etiche, senza ostacolarne l'innovazione;
> G. [...] l'andamento attuale, che tende a sviluppare macchine autonome e intelligenti, in grado di apprendere e prendere decisioni in modo indipendente, genera nel lungo periodo non solo vantaggi economici ma anche una serie di preoccupazioni circa gli effetti diretti e indiretti sulla società nel suo complesso;
> H. [...] l'apprendimento automatico offre enormi vantaggi economici e innovativi per la società migliorando notevolmente le capacità di analisi dei dati, sebbene ponga nel contempo alcune sfide legate alla necessità di garantire la non discriminazione, il giusto processo, la trasparenza e la comprensibilità dei processi decisionali»[6].

È, quindi, facile intuire che le possibili implicazioni anche di rilevanza penale, derivanti dall'impiego delle tecnologie di IA, potrebbero essere in un prossimo futuro assai numerose e significative, sicché appare opportuno – ed è questo il principale obiettivo perseguito col presente lavoro – procedere in tempi rapidi quanto meno a tematizzare tali implicazioni, cominciare a riflettere su di esse e prospettare questioni e soluzioni, al fine di non aggravare il ritardo del diritto, in particolare del diritto penale italiano, di fronte all'evoluzione tecnologica.

In effetti, come è stato efficacemente rilevato:

> «il progresso irrompe, non chiede permesso. E nel contesto attuale disegnare questo nuovo rapporto tra esseri umani e macchine non è per niente facile. Anche perché le tecnologie digitali hanno una velocità impressionante. Le tecnologie di ieri, come ad esempio la TV, la radio, l'elettricità, l'automobile hanno impiegato più di 50 anni per raggiungere i 50 milioni di utenti. Ci hanno concesso tutto il tempo per abituarci alle loro innovazioni, per avere nuove regole sul loro utilizzo, e per organizzare le nostre vite e le nostre società di conseguenza. Oggi, le tecnologie digitali irrompono molto più velocemente, e non ci danno affatto il tempo per organizzarci e per abituarci alle loro dirompenti innovazioni. Un esempio evidente di questa velocità viene dalle reti sociali: Twitter ha impiegato meno di 3 anni per raggiungere

[3] Intervento di S. Hawking durante la Conferenza *Zeitgeist*, Londra, maggio 2015 (citazione riportata da Redazione, *Do You Trust This Computer?*, in *questa rivista*, 15 maggio 2019; v. la notizia anche su Newsweek (L. Walker, *Stephen Hawking warns artificial intelligence could end humanity*, 14 maggio 2015).
[4] Risoluzione del Parlamento europeo del 16 febbraio 2017, recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).
[5] Secondo J. Kaplan, *Intelligenza artificiale*, cit., p. 32, il "sorpasso", almeno in alcuni ambiti, sarebbe già avvenuto: «i computer già oggi superano le capacità umane in molti compiti, inclusi alcuni che credevamo avessero bisogno di intelligenza umana per essere svolti».
[6] V. nota 4.

> i 50 milioni di utenti; Facebook e Instagram meno di 2 anni. Anche se il record della velocità è quello di Pokemon Go, che è riuscito a raggiungere i 50 milioni di download in soli 19 giorni!»[7].

Anche il diritto penale si deve, quindi, attrezzare per tenere il passo di questa rapidissima evoluzione tecnologica, per non rischiare di soccombere di fronte a quello che si preannuncia essere un nuovo, sconvolgente «*shock* da modernità»[8], che comporterà problemi «analoghi a quelli che hanno contraddistinto altre "transizioni" tecnologiche: verificare l'idoneità delle norme esistenti ad applicarsi alle nuove tecnologie, così da valutare se sia opportuno, per i legislatori, coniare delle regole *ad hoc*, nuove, ovvero persistere, non senza possibili forzature avallate, magari, sul piano giurisprudenziale, nell'applicazione delle norme preesistenti»[9].

Nelle pagine seguenti cercheremo, pertanto, di individuare, senza alcuna pretesa di esaustività, quattro scenari all'interno dei quali la rivoluzione tecnologica messa in moto dall'IA già solleva, o è destinata a sollevare, problemi, dubbi e questioni, rilevanti per il diritto penale:

1. le attività di *law enforcement*, in particolare le attività di c.d. *polizia predittiva*;
2. i c.d. *automated decision systems*, che potrebbero in futuro conoscere un impiego anche all'interno dei procedimenti penali, sostituendo, in tutto o in parte, la decisione del giudice-uomo;
3. i c.d. *algoritmi predittivi*, impiegati per valutare la pericolosità criminale di un soggetto, vale a dire la probabilità che costui commetta in futuro un (nuovo) reato;
4. infine, le possibili ipotesi di coinvolgimento – come *strumento*, come *autore*, o come *vittima* – di un sistema di IA nella commissione di un *reato*.

Nel procedere a sondare tali scenari terremo ben presente il monito di Stephen Hawking, il quale – dopo aver formulato la previsione, già sopra riportata, secondo cui «nell'arco dei prossimi cento anni, l'intelligenza dei computer supererà quella degli esseri umani» – subito dopo avvertiva: «quando questo accadrà, dovremo assicurarci che i computer condividano i nostri stessi obiettivi»[10].

2. Che cosa intendiamo per intelligenza artificiale?

Prima di inoltrarci negli scenari preannunciati nel precedente paragrafo, conviene tuttavia chiarirsi le idee sul concetto di IA, almeno nei limiti del possibile, dal momento che una definizione univoca e universalmente condivisa di IA, come pure di robotica[11] (un ambito in cui innumerevoli sono le applicazioni di IA), non esiste[12].

Esiste, invece, una precisa data di nascita dell'espressione "intelligenza artificiale", utilizzata per la prima volta da John McCarthy, divenuto poi uno dei padri fondatori dell'IA, il

---

[7] G.F. Italiano, *Intelligenza artificiale, che errore lasciarla agli informatici*, in *Agendadigitale.eu*, 11 giugno 2019.
[8] L'espressione è di F. Stella, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Giuffré, 2003, pp. 292 ss., e con essa il grande Maestro intendeva indicare l'affannosa rincorsa del diritto penale alle evoluzioni tecnologiche che si sono succedute nei decenni passati e che, a quanto pare, interverranno anche – e forse ancor più rapidamente – nei decenni futuri.
[9] 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. 334.
[10] Intervento di S. Hawking, citato *supra*, nota 3.
[11] Per un primo inquadramento della tematica della robotica, v. N. Sharkey, *La robotica*, in J. Al-Khalili (a cura di), *Il futuro che verrà*, cit., pp. 189 ss.
[12] Constatano l'assenza di una definizione, tra i tanti, M. B. Magro, *Biorobotica, robotica e diritto penale*, in D. Provolo, S. Riondato, F. Yenisey (a cura di), *Genetics, Robotics, Law, Punishment*, Padova University Press, 2014, pp. 510 s.; R. Calo, *Artificial Intelligence Policy: a Primer and Roadmap*, in *University of Bologna Law Review*, 3:2, 2018, p. 184; C. Trevisi, *La regolamentazione in materia di Intelligenza artificiale, robot, automazione: a che punto siamo*, in *Medialaws*, 21 maggio 2018, p. 1.

quale nell'estate del 1955, nella sua qualità di assistente universitario di matematica al Dartmouth College di Hanover, New Hamphsire, organizzò, insieme ad altri colleghi, un convegno sull'"intelligenza artificiale"[13], descrivendone l'oggetto nei seguenti termini: «lo studio procederà sulla base della congettura che tutti gli aspetti dell'apprendimento o qualsiasi altra caratteristica dell'intelligenza possa essere di principio descritta in modo così preciso che una macchina la possa simulare. Si tenterà di scoprire come si possa fare in modo che le macchine usino il linguaggio, formulino astrazione e concetti, risolvano tipi di problemi ora riservati agli esseri umani, e migliorino sé stesse»[14].

Circa trent'anni dopo, quando ormai il fenomeno dell'IA non solo aveva preso forma ma si era anche significativamente espanso, Roger Schank, uno dei massimi teorici dell'IA e tra i fondatori della linguistica computazionale, più che fornire una definizione o perlomeno una descrizione omnicomprensiva di IA, in un suo saggio del 1987 invitava a riconoscere l'esistenza di una forma di IA sulla scorta della presenza dei seguenti cinque attributi: la capacità di comunicazione; la conoscenza di sé; la conoscenza della realtà esterna; una condotta teleologicamente orientata, ossia tesa al perseguimento di un fine; infine, l'esistenza di un apprezzabile grado di creatività, intesa come capacità di assumere decisioni alternative laddove il piano di azione iniziale fallisca o non sia realizzabile[15].

Già queste prime indicazioni ci consentono, allora, di sgombrare il campo da un paio di equivoci:

– innanzitutto, quando parliamo di IA non dobbiamo necessariamente pensare ad un umanoide simile in tutto e per tutto all'essere umano: l'umanoide può essere, sì, un'applicazione di IA (forse la più eclatante), ma di certo non l'unica e non, almeno nella fase attuale, la più rilevante dal punto di vista pratico[16];

– in secondo luogo, per quanto possa essere suggestivo parlare di *intelligenza* artificiale, occorre rimarcare che, in realtà, «poco, oltre alla speculazione e a un modo di pensare ingenuo, collega il lavoro odierno nel campo dell'IA ai misteriosi meccanismi della mente umana; in realtà, almeno a questo stadio, si tratta di una disciplina ingegneristica con relazioni più che altro metaforiche e di 'ispirazione' con gli organismi biologici»[17], tanto più che l'*intelligenza* (quella degli esseri umani, prima ancora che quella delle macchine), per quanto sia oggetto di numerosissimi studi di psicologi, biologi e neuroscienziati, costituisce ancora un concetto indeterminato[18].

Per questo e per altri motivi talora i ricercatori di IA preferiscono parlare – più che di *intelligenza* – di *razionalità*, laddove per "razionalità" si intende la capacità di scegliere la migliore azione da intraprendere per conseguire un determinato obiettivo alla luce di alcuni criteri di ottimizzazione delle risorse a disposizione[19].

---

[13] In argomento, v. J. Kaplan, *Intelligenza artificiale*, cit., p. 37; G. F. Italiano, *Intelligenza artificiale: passato, presente, futuro*, cit., p. 208; L. Floridi, *What the Near Future*, cit., p. 2.
[14] L'intero testo di presentazione del convegno può essere letto *online* al presente *link*.
[15] R.C. Schank, *What's IA, Anyway?*, in *IA Magazine*, Winter 8(4), 1987, pp. 59 ss.
[16] Così C. Trevisi, *La regolamentazione in materia di Intelligenza artificiale*, cit., p. 1.
[17] Così J. Kaplan, *Intelligenza artificiale*, cit., p. 41.
[18] Si noti, per altro verso, che proprio dagli studi sull'intelligenza artificiale stanno pervenendo importanti contributi per scoprire come funziona l'intelligenza umana e il cervello umano. Si veda, ad esempio, un recente progetto europeo di integrale simulazione del cervello umano, realizzato grazie all'impiego di tecniche di IA: Redazione, *Il progetto europeo sul cervello umano*, in *questa rivista*, 2 aprile 2019.
[19] Così, ad esempio, S. Russell, P. Norvig, in quello che è forse il manuale più consultato sull'intelligenza artificiale: *Artificial Intelligence: A Modern Approach*, Prentice Hall, 3ª edizione, 2009, pp. 36 ss.

Ma come fa un sistema di IA a pervenire ad una scelta razionale[20]? Vi perviene percependo tramite *sensori* l'ambiente in cui è immerso, e dunque raccogliendo e interpretando *dati*, *ragionando* su ciò che viene percepito o *elaborando le informazioni* desunte dai dati, *decidendo* quale sia l'azione migliore e agendo di conseguenza attraverso i suoi *attuatori*, eventualmente producendo una modifica del proprio ambiente.

Per meglio comprendere questa descrizione della razionalità dei sistemi di IA occorre altresì considerare che:

– i *sensori* potrebbero essere fotocamere, microfoni, una tastiera, un sito Internet o altri sistemi di immissione dati, nonché sensori di quantità fisiche (ad esempio, sensori di temperatura, di pressione, di distanza, di forza/coppia o sensori tattili);

– i *dati* acquisiti tramite i sensori sono dati digitali, di cui oggi vi è un'immensa disponibilità; e a proposito dei *dati* va fin da subito sottolineato che la qualità del risultato finale dipende, in larga misura, proprio dalla correttezza logica e dalla completezza dei dati raccolti; per contro, se i dati utilizzati per alimentare o addestrare il sistema di IA sono distorti, nel senso che non sono sufficientemente equilibrati o inclusivi, il sistema non sarà in grado di generalizzare in maniera corretta e potrebbe adottare decisioni inique che possono favorire alcuni gruppi rispetto ad altri;

- il *ragionamento* o l'*elaborazione delle informazioni* è un processo operato attraverso un algoritmo che acquisisce come input i suddetti dati per poi proporre un'azione da intraprendere alla luce dell'obiettivo da raggiungere;

– infine, il sistema di IA esegue l'azione prescelta tramite gli *attuatori* a sua disposizione, che possono essere sia *software*, sia elementi fisici (ad esempio, bracci articolati, ruote automatiche), quest'ultimi capaci di intervenire, modificandolo, sull'ambiente circostante[21].

Oggi si riconosce unanimemente che i grandi e rapidi progressi, compiuti dall'IA in tempi recenti, sono stati consentiti dalla felice combinazione di due fattori[22]: *da un lato*, il recente, impressionante aumento delle capacità computazionali, grazie alle quali oggi disponiamo di computer sempre più veloci, potenti, con capacità di memoria (e, quindi, tra l'altro, di archiviazione dati) straordinariamente grandi; *dall'altro lato*, il recente, impressionante aumento di dati digitali, raccolti anche grazie a sensori ad alta definizione e a basso costo: dati provenienti dalla digitalizzazione di documenti o generati da ognuno di noi scattando foto, facendo video o inviando messaggi tramite le reti sociali o altri strumenti di messaggistica, come Whatsapp, Messenger, etc. (c.d. dati *people-to-people*); oppure dati raccolti da istituzioni pubbliche o soggetti privati, inerenti i cittadini o gli utenti, come dati fiscali, sanitari, ricerche sul *web*, transazioni commerciali, bancarie (c.d. dati *people-to-machine*); infine, dati di tipo *machine-to-machine*, generati, automaticamente e indipendentemente dall'intervento di esseri umani, da dispositivi fisici, come ad esempio vari tipi di sensori, dispositivi di geo-localizzazione, *werables*, *smart devices*, tra di loro connessi nell'Internet delle Cose[23].

---

[20] Nel rispondere a questa domanda ci gioveremo principalmente del documento "*Una definizione di IA: principali capacità e discipline scientifiche*", elaborato dal Gruppo Indipendente di 52 esperti ad alto livello, nominato dalla Commissione europea per svolgere a suo favore funzioni di consulenza sull'intelligenza artificiale. Si noti, infine, incidentalmente, che di tale gruppo fa parte anche un professore di diritto penale, il prof. Hilgendorf dell'Università di Würzburg (https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence).

[21] V. il documento cit. nella nota precedente.

[22] Così, tra i tanti, J. Kaplan, *Intelligenza artificiale*, cit., p. 72; G.F. Italiano, *Intelligenza artificiale: passato, presente, futuro*, cit., p. 220; R. Calo, *Artificial Intelligence Policy*, cit., p. 186.

[23] Per un primo inquadramento della tematica dell'Internet delle Cose, v. N. Climer, *Il* cloud *e l'Internet delle cose*, in J. Al-Khalili (a cura di), *Il futuro che verrà*, cit., pp. 133 ss.

La combinazione di tali due fattori – unitamente ad altri progressi nella ricerca – hanno, tra l'altro, consentito di elaborare e di diffondere su larga scala i sistemi di *machine learning* che possiamo, in estrema sintesi, descrivere così: il *software impara* autonomamente dall'ambiente esterno (tramite i dati che immagazzina ed elabora) e modifica le proprie prestazioni adattandole agli esiti del procedimento di apprendimento[24].

Per tentare, infine, una sintesi "ufficiale" di tutte le nozioni e informazioni sopra riportate, possiamo rivolgerci ad una recente Comunicazione del 2018 elaborata dalla Commissione europea, intitolata "*Artificial Intelligence for Europe*"[25], la quale fornisce la seguente definizione di IA:

> «l'intelligenza artificiale (IA) indica sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi. I sistemi basati sull'IA possono consistere solo in *software* che agiscono nel mondo virtuale (ad esempio, assistenti vocali, software per l'analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale), oppure incorporare l'IA in dispositivi *hardware* (ad esempio, in robot avanzati, auto a guida autonoma, droni o applicazioni dell'Internet delle Cose)»[26].

Partendo proprio da tale definizione, il gruppo indipendente di 52 esperti ad alto livello sull'intelligenza artificiale, già sopra ricordato[27], ha a sua volta elaborato un documento contenente "*Una definizione di IA: principali capacità e discipline scientifiche*"[28], in cui, dopo aver fornito alcune ulteriori delucidazioni ed effettuato talune precisazioni (di cui ci siamo già giovati nelle pagine precedenti), formula la seguente *definizione aggiornata* di "Intelligenza artificiale o sistemi di IA", ritenendo che siano tali:

> «sistemi *software* (ed eventualmente *hardware*) progettati dall'uomo che, dato un obiettivo complesso, agiscono nella dimensione fisica o digitale percependo il proprio ambiente attraverso l'acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, ragionando sulla conoscenza o elaborando le informazioni derivate da questi dati e decidendo le migliori azioni da intraprendere per raggiungere l'obiettivo dato. I sistemi di IA possono usare regole simboliche o apprendere un modello numerico, e possono anche adattare il loro comportamento analizzando gli effetti che le loro azioni precedenti hanno avuto sull'ambiente.
> Come disciplina scientifica, l'IA comprende diversi approcci e diverse tecniche, come l'apprendimento automatico (di cui l'apprendimento profondo e l'apprendimento per rinforzo sono esempi specifici), il ragionamento meccanico (che include la pianificazione, la programmazione, la rappresentazione delle conoscenze e il ragionamento, la ricerca e l'ottimizzazione) e la robotica (che comprende il controllo, la percezione, i sensori e gli attuatori e l'integrazione di tutte le altre tecniche nei sistemi ciberfisici)»[29].

---

[24] Sul *machine learning*, v., in una prospettiva tecnica, S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, cit., pp. 634 ss.; L. Floridi, *What the Near Future*, cit., pp. 4 ss.; P. Domingos, *L'algoritmo definitivo: la macchina che impara da sola e il futuro del nostro mondo*, Bollati Boringhieri, 2016, pp. 7 ss.; in una prospettiva giuridica, R. Calo, *Artificial Intelligence Policy*, cit., p. 185; H. Surden, *Machine Learning and Law*, in *Wash. L. Rev.*, 89, 1, 2014, pp. 87 s..

[25] COM(2018) 237 final, del 25 aprile 2018.

[26] Il corrispondente testo in inglese è il seguente: «*artificial intelligence (IA) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. IA-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or IA can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)*».

[27] V. *supra*, nota 20, e testo corrispondente.

[28] V. *supra*, nota 20, e testo corrispondente.

[29] Pag. 6 del documento. Il corrispondente testo in inglese è il seguente: «*artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans3 that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or*

Ebbene, a questo punto, dopo questo rapido "chiarimento di idee" sul concetto di IA, siamo finalmente in grado di descrivere i quattro scenari in cui i sistemi di IA potrebbero assumere, o hanno già assunto, implicazioni di rilevanza penale.

3. Primo percorso d'indagine - IA e attività di *law enforcement.*

Nel documento di presentazione del Convegno annuale di esperti di Polizia, organizzato dall'OSCE, dedicato quest'anno (2019) proprio al tema "*Artificial Intelligence and Law Enforcement*", può leggersi quanto segue:

> «nei loro sforzi per aumentare l'efficienza e l'efficacia e per stare al passo con le innovazioni tecnologiche, le autorità e le agenzie di *law enforcement* di tutto il mondo stanno esplorando sempre più i potenziali dell'IA per il loro lavoro. La crescente quantità di dati ottenuti e archiviati dalla polizia ha anche richiesto metodi e strumenti più sofisticati per la loro gestione e analisi, per l'identificazione di modelli (*pattern*), la previsione dei rischi e lo sviluppo di strategie per allocare le risorse umane e finanziarie dove sono maggiormente necessarie. Anche se l'uso dell'IA nel lavoro delle forze dell'ordine è un argomento relativamente nuovo, alcuni strumenti basati sull'intelligenza artificiale sono già stati testati e sono persino attivamente utilizzati dai servizi di polizia di diversi Paesi del mondo. Questi includono *software* di analisi di video e immagini, sistemi di riconoscimento facciale, di identificazione biometrica, droni autonomi e altri robot e strumenti di analisi predittiva per prevedere le "zone calde" del crimine o anche per identificare potenziali criminali futuri, in particolare i criminali ad elevata pericolosità»[30].

L'impiego di sistemi di IA nelle attività di *law enforcement* è, quindi, già una realtà, e anzi se ne prevede una crescita ed intensificazione nei prossimi anni a vari livelli[31]. Del resto, l'importanza strategica dell'impiego di sistemi di IA nelle attività di *law enforcement* e i preziosi risultati grazie ad essi raggiungibili, sono ben messi in evidenza da un episodio riferito da Giuseppe Italiano in un suo recente saggio:

> «già nel recente passato si sono verificati casi in cui l'utilizzo di opportune (anche semplici) analisi algoritmiche avrebbe potuto prevenire il verificarsi di pericolosi eventi terroristici. Ad esempio, è famoso il caso di Umar Farouk Abdulmutallab, noto anche come il "terrorista delle mutande" (*Underwear Bomber*), che è riuscito a imbarcarsi sul volo Amsterdam-Detroit nel giorno di Natale 2009, con dell'esplosivo cucito all'interno della biancheria intima che indossava, e che ha cercato di farsi esplodere durante il volo. Per una fortunata coincidenza [l'intervento di alcuni passeggeri insospettitisi], l'attacco terroristico non è andato a buon fine [...]. L'*intelligence* aveva dati e informazioni a sufficienza per valutare il grado di pericolosità del terrorista e aveva anche elementi sufficienti per inserirlo nella *black list*, così da negargli la possibilità di imbarco su voli diretti negli Stati Uniti. Ma in quel caso l'*intelligence* non è semplicemente riuscita a connettere le molteplici informazioni,

---

processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.
As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)».

[30] Il documento completo di presentazione del *2019 OSCE Annual Police Experts Meeting: Artificial Intelligence and Law Enforcement: An Ally or an Adversary?, 23-24 September, Wien*, può essere letto su *questa rivista* (Redazione, *Artificial Intelligence and Law Enforcement: an Ally or an Adversary?*, 23 settembre 2019).

[31] In argomento, v. anche il documentato studio di A. G. Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York University Press, 2017, pp. 3 ss.

> provenienti da varie fonti, che erano a sua disposizione. Come dire, nel patrimonio informativo c'erano tutti i dati necessari, ma è semplicemente mancata l'utilizzazione di un buon algoritmo per mettere in correlazione tutti questi dati»[32].

Riflettendo su tale episodio, Giuseppe Italiano rileva, quindi, che:

> «sicuramente, e non soltanto in questa circostanza, tecniche di IA sono e possono essere impiegate con successo nell'analisi delle informazioni disponibili, delle transazioni, dei file di *log*, del traffico sulla rete, e di tutte le "impronte" che ogni individuo lascia in rete e nei sistemi digitali, allo scopo di identificare possibili anomalie e attività sospette, o semplicemente per comporre in una visione coerente le informazioni provenienti da sorgenti multiple ed eterogenee, ed estrarne conoscenza, in modo tale da prendere in maniera automatica decisioni oppure fornire il supporto a decisori umani, che devono essere in grado di reagire sempre più velocemente agli stimoli esterni»[33].

Ebbene, nelle pagine seguenti cercheremo di fornire una sintetica rassegna dei possibili impieghi di sistemi di IA nelle attività di *law enforcement* rivolte alla prevenzione dei reati, dedicando particolare attenzione allo specifico ambito denominato "*predictive policing*" o "polizia predittiva".

3.1. RoboCop: dalla fantascienza alla realtà?

Probabilmente molti di noi ricordano la figura di RoboCop, il poliziotto con un corpo di titanio e kevlar, un cervello informatico e sensori ultrapotenti: se nel 1987, anno di uscita del celebre film, tale immagine apparteneva decisamente alla fantascienza, oggi la realtà ci propone alcune applicazioni delle tecnologie di IA – in uso, per lo più in via sperimentale, presso le forze di polizia di alcuni Stati – che si avvicinano molto a RoboCop[34]: si tratta, nella maggior parte dei casi, di macchine robotiche, non necessariamente umanoidi, utilizzate per una varietà di compiti, come ad esempio attività di pattugliamento, sorveglianza, disinnesco di bombe, individuazione di atteggiamenti sospetti, riconoscimento facciale, etc[35].

Applicazioni di questo tipo, se da un lato hanno il gran merito di preservare da una serie di pericoli gli agenti (umani), e se in talune circostanze assicurano un ottimo livello di efficienza nelle prestazioni erogate, sollevano, dall'altro lato, una serie di problematiche[36]:

– in primo luogo, occorre interrogarsi sulla opportuna ampiezza che il controllo umano deve assumere su tali applicazioni, soprattutto nel caso in cui esse utilizzino sistemi di intelligenza artificiale che ne assicurino ampi margini di autonomia: il controllo dell'uomo si deve limitare alla scelta degli obiettivi, al monitoraggio, o deve essere un controllo più intenso, esercitato anche a costo di compromettere le prestazioni stesse del RoboCop?

---

[32] G. F. Italiano, *Intelligenza artificiale: passato, presente, futuro*, cit., p. 222.
[33] *Ibidem*.
[34] In argomento, v. N. Sharkey, *2084: Big robot is watching you. Report on the future of robots for policing, surveillance and security*, 2008; una versione più breve di tale saggio, intitolata *The robot arm of the law grows longer*, e originariamente pubblicata sulla rivista *Computer*, 2009, p. 113, può essere letta anche a questo indirizzo *web*; v. pure L. Royakkers, R. van Est, *A Literature Review on New Robotics: Automation from Love to War*, in *International Journal of Social Robotics*, Volume 7, Issue 5, 2015, pp. 549 ss.; E.E. Joh, *Policing Police Robots*, in *UCLA Law Rev. Disc.*, 2016, p. 516; L. Pasculli, *Genetics, Robotics and Crime Prevention*, in D. Provolo, S. Riondato, F. Yenisey (a cura di), *Genetics, Robotics, Law, Punishment*, cit., pp. 197 ss.
[35] Un sofisticato programma di riconoscimento facciale – S.A.R.I., Sistema Automatico di Riconoscimento Immagini – è in dotazione anche alla Polizia scientifica italiana, stando a quanto si apprende dalle notizie riportate dall'agenzia giornalistica ANSA: Redazione ANSA, *Ladri individuati grazie al nuovo sistema di riconoscimento facciale*, 7 settembre 2018.
[36] Cfr. Autori citati alla nota 34.

– In secondo luogo, nient'affatto trascurabile è la questione della *privacy*, in considerazione della gran mole di dati che queste applicazioni (fornite, ad esempio, di sensori e telecamere avanzate) possono acquisire in relazione alla vita, anche privata, dei cittadini: dati che, peraltro, potrebbero essere manipolati abusivamente, sottratti, deformati, con grave pregiudizio per le persone cui essi si riferiscono.

– Da ultimo, occorre considerare che alcune di queste applicazioni sono equipaggiate con armi non letali (ad esempio, il *taser* o lo *spray* al peperoncino) o addirittura letali (ad esempio, classiche armi da fuoco), il che crea indubbie preoccupazioni in ordine al tasso di fallibilità di queste applicazioni e quindi in ordine all'individuazione del responsabile (uomo o macchina?) di eventuali uccisioni o lesioni commesse per errore[37], nonché in ordine alla presumibile assenza, in capo a questi dispositivi robotizzati armati, di doti tipicamente umane – la pietà, l'intuito, la capacità di improvvisazione, il c.d. senso comune[38] – la cui presenza, in operatori della polizia, è sempre auspicabile[39].

Applicazioni di questo tipo, pertanto, andranno monitorate accuratamente, anche dal punto di vista giuridico, tra l'altro al fine di elaborare un preciso quadro normativo che ne regoli il legittimo utilizzo, nel rispetto dei diritti fondamentali delle persone[40].

3.2. Sistemi di intelligenza artificiale e polizia predittiva.

Per "polizia predittiva" possiamo intendere l'insieme delle attività rivolte allo studio e all'applicazione di metodi statistici con l'obiettivo di "predire" *chi* potrà commettere un reato, o *dove* e *quando* potrà essere commesso un reato, al fine di prevenire la commissione dei reati stessi. La predizione si basa fondamentalmente su una rielaborazione attuariale di diversi tipi di dati, tra cui quelli relativi a notizie di reati precedentemente commessi, agli spostamenti e alle attività di soggetti sospettati, ai luoghi, teatro di ricorrenti azioni criminali, e alle caratteristiche di questi luoghi, al periodo dell'anno o alle condizioni atmosferiche maggiormente connesse alla commissione di determinati reati; tra i dati utilizzati a questi fini talora compaiono anche informazioni relative all'origine etnica, al livello di scolarizzazione, alle condizioni economiche, alle caratteristiche somatiche (... una rivincita di Lombroso?), riconducibili a soggetti appartenenti a determinate categorie criminologiche (ad es., potenziali terroristi), etc[41].

In tempi recenti, l'impiego di *software* basati sull'IA ha consentito di fare un salto di qualità nelle attività di polizia predittiva, dal momento che è ora possibile l'acquisizione e la rielaborazione di una mole enorme di dati, scoprendo connessioni prima difficilmente individuabili dall'operatore umano[42].

---

[37] Su questo profilo, v. anche *infra*, parr. 6.2 e 6.3.

[38] Come giustamente sottolinea M. B. Magro, *Biorobotica, robotica e diritto penale*, cit., p. 512, «ai robot dotati di intelligenza artificiale, dotati di conoscenze altamente specialistiche, manca, al di sotto di queste conoscenze, il livello di conoscenze comuni, il c.d. "senso comune", ciò che tutti gli umani posseggono senza aver fatto studi particolari. Il "senso comune" è quello che consente di collegare conoscenze specialistiche di campi diversi e di affrontare i problemi e di risolverli senza la rigidità tipica dell'approccio simbolico dell'intelligenza. Spesso una reazione intelligente ad una certa situazione è quella che, sì, tiene in considerazione il contesto, ma che non è capace di selezionare quale aspetto del contesto sia rilevante».

[39] Sulle c.d. *autonomous weapons*, v. in particolare N. Sharkey, *La robotica*, cit., pp. 195 s.; R. Calo, *Artificial Intelligence Policy*, cit., p. 196, con ulteriori riferimenti.

[40] N. Sharkey, *La robotica*, cit., p. 196, riferisce, ad esempio, dei lavori, in corso presso le Nazioni Unite, per l'adozione di un Trattato internazionale che proibisca lo sviluppo e l'uso delle armi robotizzate.

[41] Per un completo inquadramento della materia della *predictive policing*, v. W.L. Perry, B. McInnis, C.C. Price, S.C. Smith, J.S. Hollywood, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, Rand Corporation, 2013, consultabile a questo *link*.

[42] C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, L. Floridi, *Artificial Intelligence and the "Good Society": the US, EU, and UK approach*, in *Science and Eng. Ethics*, 2018, pp. 505 ss.; L. Bennet Moses, J. Chan, *Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability*, in *Policing and Society*, 2016, pp. 1 ss.; G. Mastrobuoni, *Crime is Terribly Revealing: Information*

I *software* di polizia predittiva – siano essi assistiti o meno da sistemi di IA[43] – possono dividersi fondamentalmente in due categorie:

– quelli che, ispirandosi alle acquisizioni della criminologia ambientale, individuano le c.d. "zone calde" (*hotspots*), vale a dire i luoghi che costituiscono il possibile scenario dell'eventuale futura commissione di determinati reati (*infra*, par. 3.2.1);

– quelli che, ispirandosi invece all'idea del *crime linking*, seguono le serialità criminali di determinati soggetti (individuati o ancora da individuare), per prevedere dove e quando costoro commetteranno il prossimo reato (*infra*, par. 3.2.2).

Va subito detto che, almeno per ora, sia gli uni che gli altri sistemi possono fornire adeguate previsioni solo in relazione a limitate, determinate categorie di reati (ad esempio, reati attinenti alla criminalità da strada, come rapine e spaccio di stupefacenti), e non in via generalizzata per tutti i reati.

3.2.1. Sistemi di individuazione degli *hotspots*.

Rientra nel primo tipo di sistemi il *Risk Terrain Modeling* (RTM): un algoritmo che, rielaborando quantità enormi di dati inerenti i fattori ambientali e spaziali favorenti la criminalità, sembrerebbe consentire la predizione della commissione di reati di spaccio di sostanze stupefacenti in determinate aree urbane[44]. I ricercatori hanno elaborato questo sistema sottoponendo all'algoritmo RTM dati inerenti i fattori ambientali e spaziali più frequentemente connessi alla commissione dei reati suddetti: presenza di luminarie stradali scarse o non funzionanti, vicinanza di locali notturni, di fermate di mezzi pubblici, di stazioni ferroviarie, di snodi di strade ad alta percorribilità, di bancomat, di compro-oro, di parcheggi scambiatori, infine, di scuole. Ciò ha consentito di elaborare una vera e propria "mappatura" di alcune grandi aree metropolitane al fine di individuare le "zone calde" dove più elevato risulta il rischio di spaccio di sostanze stupefacenti, con conseguenti benefici in termini di programmazione e attuazione di interventi di prevenzione della delinquenza connessa allo spaccio[45].

Parimenti finalizzato all'individuazione degli *hotspots* ma in relazione ad un numero più elevato di reati (non solo quelli di spaccio) è anche un *software*, già in uso da alcuni anni negli Stati Uniti e nel Regno Unito, originariamente elaborato da alcuni ricercatori dell'UCLA (Università della California di Los Angeles) in collaborazione con la locale polizia, e oggi venduto, parrebbe con grande successo commerciale, da un'azienda privata americana col marchio *PredPol*, il cui sito pubblicizza tale dispositivo con le seguenti parole:

> «utilizzando solo tre tipologie di dati – tipo di reato, data/ora del reato e luogo del reato – per fare previsioni, la tecnologia *PredPol* ha aiutato le forze dell'ordine a ridurre drasticamente

---

Technology and Police Productivity, 2017, consultabile *online* al presente *link*; per un sintetico quadro, in lingua italiana, dei sistemi di IA finalizzati ad attività di polizia predittiva, v. R. Pelliccia, *Polizia predittiva: il futuro della prevenzione criminale?*, in *cyberlaws.it*, 9 maggio 2019.

[43] Non sempre risulta chiaro se, e in quale misura, i *software* di cui parleremo nelle pagine seguenti si basino su sistemi di IA. Ciò dipende anche dal fatto che alcuni di questi *software* sono di proprietà privata e sono coperti da segreto industriale, sicché i dettagli sul loro funzionamento non sono resi pubblici.

[44] In argomento, v. J.M.Caplan, L.W. Kennedy, J.D. Barnum, E.L Piza, *Crime in Context: Utilizing Risk Terrain Modeling and Conjunctive Analysis to Explore the Dynamics of Criminogenic Behavior Setting*, in *Journal of Contemporary Criminal Justice*, 33(2), 2017, pp. 133 ss.; J.M.Caplan, L.W. Kennedy, *Risk Terrain Modeling: Crime Prediction and Risk Reduction*, Univ. of California Press, 2016; L.W. Kennedy, J.M.Caplan, E.L Piza, *Risk Clusters, Hotspots and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies*, in *Journal of Quantitative Criminology*, 2010, pp. 339 ss.

[45] V. Autori citati alla nota precedente.

> il tasso di criminalità in giurisdizioni di tutti i tipi e di tutte le dimensioni, negli Stati Uniti e all'estero. *PredPol* può vantare una comprovata sperimentazione: il Dipartimento della polizia di Los Angeles ha registrato un calo del 20% dei reati previsti di anno in anno e una divisione della locale polizia ha potuto sperimentare per la prima volta, un'intera giornata senza ricevere denunce di reati. Il Dipartimento dello Sceriffo della contea di Jefferson ha registrato una riduzione del 24% nelle rapine e una riduzione del 13% nei furti con scasso. A Plainfield, New Jersey, da quando si usa *PredPol* si è avuta una riduzione del 54% delle rapine e una riduzione del 69% dei furti di auto»[46].

Sembrerebbe ispirarsi ad una analoga logica predittiva anche un dispositivo in uso presso la polizia italiana: si tratta del sistema informatico *X-LAW*, originariamente predisposto dalla Questura di Napoli, che parrebbe aver già ottenuto ottimi risultati sul territorio italiano nel campo della prevenzione di talune tipologie di reati[47]. Stando alle notizie riferite[48], il *software X-LAW* si basa su un algoritmo capace di rielaborare una mole enorme di dati estrapolati dalle denunce inoltrate alla Polizia di Stato. Tale rielaborazione consente di far emergere fattori ricorrenti o fattori coincidenti, come ad esempio la ripetuta commissione di rapine negli stessi luoghi, da parte di persone con lo stesso tipo di casco o di moto, e con analoghe modalità. Ciò consente di tracciare una mappa del territorio dove vengono evidenziate le zone a più alto rischio fino a raggiungere il livello massimo in determinati orari, così consentendo – nelle zone e negli orari 'caldi' – la predisposizione delle forze dell'ordine per impedire la commissione di tali reati e per cogliere in flagranza i potenziali autori degli stessi.

3.2.2. Sistemi di crime linking.

Si rifà, invece, all'idea del *crime linking*, seguendo le serialità criminali di determinati soggetti (individuati o ancora da individuare), per prevedere dove e quando essi commetteranno il prossimo reato, il *software Keycrime*, originariamente elaborato presso la Questura di Milano, e poi divenuto di proprietà di un'azienda privata[49]. Altri *software* parimenti ispirati all'idea del *crime linking*, e quindi all'individuazione delle persone, più che delle zone calde, sono stati elaborati, e sono in uso, in Germania (*Precobs*)[50], in Inghilterra (*Hart - Harm Assessment Risk Tool*)[51], e negli Stati Uniti[52].

Questi *software* si basano sull'idea di fondo che alcune forme di criminalità si manifesterebbero in un arco temporale e in una zona geografica molto circoscritti (c.d. *near repeat crimes*, o reati a ripetizione ravvicinata): ad esempio, la commissione di una rapina sembrerebbe essere associata ad un elevato rischio di commissione di una nuova rapina, da parte degli stessi autori e in una zona geografica assai prossima al luogo del primo delitto, entro le successive 48 ore e, sia pur con un tasso di rischio decrescente, fino a tutto il mese successivo. Attraverso la

---

[46] https://www.predpol.com/, visitato il 9 agosto 2019.
[47] Notizie riferite da M. Iaselli, *X-LAW: la polizia predittiva è realtà*, in *Altalex.com*, 28 novembre 2018.
[48] *Ibidem*. Ulteriori informazioni e filmati relativi a X-LAW sono facilmente reperibili *online*.
[49] In argomento, v. A.D. Signorelli, *Il software italiano che ha cambiato il mondo della polizia predittiva*, in *Wired.it*, 18 maggio 2019; C. Parodi, V. Sellaroli, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in *Diritto penale contemporaneo*, 2019, fasc. 6, p. 56. Per una descrizione di Keycrime, fornita dal suo stesso ideatore, Mario Venturi, v. *Id.*, *La chiave del crimine*, in *Profiling*, 5, 4, 2014.
[50] Su Precobs, v. l'accurata voce di Wikipedia https://en.wikipedia.org/wiki/Precobs.
[51] Sul *software* HART, v. M. Oswald, J. Grace, S. Urwin, G. Barnes, *Algorithmic risk assessment policing models: lessons from the Durham HART model and "Experimental" proportionality*, in *Information & Communications Technology Law*, 2018, pp. 223 ss.; nella dottrina italiana, v. M. Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, in *Diritto penale contemporaneo*, 29 maggio 2019, pp. 10 ss. Il *software* HART è stato sottoposto a studi di validazione da parte di alcuni ricercatori della Cambridge University: cfr. il presente indirizzo *web*.
[52] Per un progetto pilota avviato nella città di Chicago, v. J. Saunders, P. Hunt, J. S. Hollywood, *Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot*, in *Journal of Experimental Criminology*. 2016, p. 347.

raccolta e l'incrocio di una gran mole di dati, provenienti da varie fonti (ad esempio, immagini riprese da una telecamera o informazioni relative a precedenti analoghi reati), questi *software* cercano, infatti, di "profilare" il possibile autore della serie criminale e prevederne la prossima mossa.

Peraltro, i risultati forniti da questi *software* in alcuni casi potrebbero essere usati non solo a fini predittivi, ma anche per ricostruire la carriera criminale del soggetto profilato, vale a dire per avere una traccia di indagine da seguire per imputargli non solo l'ultimo reato commesso (in occasione del quale egli è stato individuato), ma anche i precedenti reati costituenti la serie criminale ricostruita grazie all'archiviazione e all'elaborazione dei dati.

3.2.3. Considerazioni conclusive sui sistemi di polizia predittiva.

I sistemi di polizia predittiva sopra sinteticamente descritti possono indubbiamente apportare grandi benefici nella prevenzione almeno di alcuni tipi di reati, ma il loro utilizzo suscita più d'una perplessità[53].

Prima di tutto, infatti, occorre rilevare che il loro uso non pare essere stato finora regolato, in nessun Paese, a livello normativo, sicché le condizioni e le modalità del loro utilizzo, nonché la valutazione e la valorizzazione dei loro risultati finiscono per essere affidate alla sola prassi, e quindi all'iniziativa, alla sensibilità, all'esperienza degli operatori di polizia.

Eppure il loro uso potrebbe implicare gravi attriti quanto meno con la tutela della *privacy* (in considerazione della gran mole di dati personali raccolti), e con il divieto di discriminazione (nella misura in cui, ad esempio, identifichino fattori di pericolosità connessi a determinate caratteristiche etniche, o religiose o sociali)[54].

Si tratta, poi, di sistemi che in una certa misura si auto-alimentano coi dati prodotti dal loro stesso utilizzo, col rischio di innescare circoli viziosi: se, ad esempio, un *software* predittivo individua una determinata "zona calda", i controlli e i pattugliamenti della polizia in quella zona si intensificheranno, con inevitabile conseguente crescita del tasso dei reati rilevati dalla polizia in quella zona, che diventerà, quindi, ancora più "calda", mentre altre zone, originariamente non ricondotte nelle "zone calde", e quindi non presidiate dalla polizia, rischiano di rimanere, o di diventare, per anni zone franche per la commissione di reati.

Inoltre, questi sistemi sollecitano una prevenzione dei reati attraverso l'intervento attivo della polizia, attraverso, quindi, una sorta di "militarizzazione" nella sorveglianza di determinate zone o di determinati soggetti, senza invece minimamente mirare alla riduzione del crimine attraverso un'azione rivolta, a monte, ai fattori criminogeni (fattori sociali, ambientali, individuali, economici, etc.).

---

[53] Le considerazioni contenute nel prosieguo del testo rielaborano spunti e riflessioni formulati da L. Pasculli, *Genetics, Robotics and Crime Prevention*, cit., p. 192, e da R. Pelliccia, *Polizia predittiva*, cit., che rinvia, tra l'altro, alle ricerche compiute in materia, e alle relative perplessità espresse, dall'Human Rights Data Analysis Group (Hrdag), raccolte nel sito https://hrdag.org/usa/, alla voce "*The Problem with Predictive Policing*".

[54] Su questi aspetti, v. A. Bonfanti, Big data *e polizia predittiva: riflessioni in tema di protezione del diritto alla* privacy *e dei dati personali*, in *MediaLaws* 24 ottobre 2018; E. Thomas, *Why Oakland Police Turned Down Predictive Policing*, in *Vice.com*, 28 dicembre 2016; J. Kremer, *The end of freedom in public places? Privacy problems arising from surveillance of the European public space*, 2017, in particolare il capitolo 3.4.2, "*Prediction*", p. 269 ss.

Infine, non si deve trascurare il fatto che la maggior parte di questi *software* sono coperti da brevetti depositati da aziende private, i cui detentori sono, giustamente, gelosi dei relativi segreti industriali e commerciali, sicché non si può disporre di una piena comprensione dei meccanismi del loro funzionamento, con evidente pregiudizio delle esigenze di trasparenza e di verificazione indipendente della qualità e affidabilità dei risultati da essi prodotti.

4. Secondo percorso d'indagine - IA e decisione giudiziaria: la macchina-giudice?

Algoritmi basati sull'IA vengono, già da qualche tempo, utilizzati anche a fini decisionali nei più svariati ambiti[55]: si tratta dei c.d. *automated decision systems*, in via di crescente diffusione[56], sia in ambito privato, sia in ambito pubblico[57].

Tra le decisioni che siffatti algoritmi sono in grado di assumere vi sono, ovviamente, anche decisioni finalizzate a comporre, o prevenire, liti e risolvere controversie. Anzi, in quest'ambito, le nuove tecnologie – grazie alla possibilità di attingere a quantità enormi di dati da fonti quali banche-dati giurisprudenziali, legislative, raccolte di precedenti, e simili – hanno già messo a punto dispositivi molto sofisticati, che utilizzano teoria dei giochi, analisi dei risultati positivi e strategie di negoziazione per risolvere le questioni, impiegando, così, una metodologia che i soggetti coinvolti percepiscono come oggettiva e priva di pregiudizi[58]. Si tratta di metodi alternativi di risoluzione delle controversie, spesso gestiti esclusivamente *online*[59], i quali, rispetto ai sistemi tradizionali, comportano riduzione dei tempi e significativi risparmi di spesa sia per i soggetti coinvolti, sia per i soggetti responsabili della decisione[60].

Per ora gli *automated decision systems* sono stati utilizzati prevalentemente per questioni civili (risarcimento danni, gestione di pratiche assicurative, danni da prodotto, etc.)[61]: ad esempio, da notizie di stampa si apprende di un progetto, avviato in Estonia, di creazione di un algoritmo capace di prendere decisioni in ambito civilistico, destinato a risolvere, in primo grado, le controversie di minore entità (del valore fino a sette mila euro)[62]. Nulla esclude, tuttavia, che a breve gli algoritmi decisionali possano trovare impiego anche in ambito penale.

La possibilità di una diffusione di decisioni giudiziarie algoritmiche anche in materia penale ha, però, già richiamato l'attenzione, e destato la preoccupazione, del Consiglio d'Europa il quale, tramite la propria Commissione per l'efficacia della giustizia (CEPEJ), il 4 dicembre 2018 ha adottato la *Carta etica europea per l'uso dell'intelligenza artificiale nei sistemi di*

---

[55] J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullianathan, *Human Decisions and Machine Predictions*, in *Quarterly Journal of Economics*, 2017, p. 237.
[56] D. Reisman, J. Schultz, K. Crawford, M. Whittaker, *Algorithmic Impact Assessments: a Practical Framework for Public Agency Accountability*, 2018.
[57] Sull'impiego, all'interno della pubblica amministrazione, di sistemi decisionali basati sull'IA in Italia e in Argentina, v. ad esempio D.U. Galetta, J.G. Corvalàn, *Intelligenza Artificiale per una Pubblica Amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *federalismi.it, Rivista di diritto pubblico italiano, comparato, europeo*, 6 febbraio 2019, pp. 1 ss.
[58] J. Kaplan, *Intelligenza Artificiale*, cit., p. 137 s.
[59] A.R. Lodder, J. Zeleznikow, *Artificial Intelligence and Online Dispute Resolution*, in A.R. Lodder, J. Zeleznikow, *Enhaced Dispute Resolution through the Use of Information Technology*, Cambridge University Press, 2010, pp. 7 ss. della versione digitale, consultabile *online* al presente *link*.
[60] E. Latifah, A.H. Bajrektarevic, M.N. Imanullah, *Digital Justice in Online Dispute Resolution: The Shifting from Traditional to the New Generation of Dispute Resolution*, in *Brawijaya Law Journal – Journal of Legal Studies*, vol. 6, No. 1, April 2019.
[61] Di recente, sul tema è stata pubblicata una pregevole monografia di un processualcivilista spagnolo, J. Nieva Fenoll, *Inteligencia artificial y proceso judicial*, 2018, trad. in italiano di P. Comoglio, *Intelligenza artificiale e processo*, Giappichelli, 2019, recensita da D. Dalfino, *Stupidità (non solo) artificiale, predittività e processo*, in *Questione Giustizia online*, 3 luglio 2019.
[62] Segnala tale notizia A. Cappellini, Machina delinquere non potest. *Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia* 2019 (*online* su *disCrimen* dal 27 marzo 2019), rinviando all'articolo di C. Lavalle, *In Estonia il giudice sarà un'intelligenza artificiale*, 4 aprile 2019, in *La Stampa*.

giustizia penale e nei relativi ambienti[63]. Come è stato opportunamente evidenziato, si tratta di un «documento di eccezionale rilevanza»[64], poiché è la prima volta che, a livello europeo, «preso atto della crescente importanza dell'intelligenza artificiale nelle nostre moderne società e dei benefici attesi quando questa sarà pienamente utilizzata al servizio della efficienza e qualità della giustizia», vengono individuate alcune fondamentali linee guida, alle quali «dovranno attenersi i soggetti pubblici e privati responsabili del progetto e sviluppo degli strumenti e dei servizi della IA»[65].

In particolare, la Carta etica enuncia i seguenti principi:
1) principio del rispetto dei diritti fondamentali;
2) principio di non discriminazione;
3) principio di qualità e sicurezza;
4) principio di trasparenza, imparzialità e correttezza;
5) principio di garanzia del controllo umano.

Quest'ultimo principio, in particolare, è finalizzato a «precludere un approccio deterministico» e ad «assicurare che gli utilizzatori agiscano come soggetti informati ed esercitino il controllo delle scelte effettuate»[66], al fine di evitare un eccessivo automatismo o una cieca standardizzazione delle decisioni.

Il documento esplicativo, allegato alla Carta etica, ci informa, altresì, che «nel 2018, l'uso di algoritmi di intelligenza artificiale nei sistemi giudiziari europei rimane principalmente un'iniziativa commerciale del settore privato, rivolta a compagnie assicurative, uffici e studi legali, avvocati e privati» (p. 16), pur evidenziando che l'utilizzo di tali algoritmi è meritevole di essere preso in considerazione «nel campo della giustizia civile, commerciale e amministrativa al fine di una risoluzione precontenziosa *online* delle controversie, purché un ricorso successivo al giudice rimanga possibile» (p. 41). Quanto ai procedimenti penali, il documento avverte che «anche se non sono specificamente progettati per essere discriminatori, l'uso di algoritmi basati sull'IA [...] ha mostrato il rischio di favorire la rinascita di teorie deterministiche a scapito delle teorie dell'individualizzazione della pena» (p. 48).

Oltre alla preoccupazione di evitare discriminazioni e automatismi, l'impiego di algoritmi decisionali nell'ambito di un giudizio penale potrebbe risultare particolarmente problematico per tre ordini di ragioni[67]:

– in primo luogo, perché il mezzo di prova più frequentemente usato nel processo penale per l'accertamento dei fatti è la testimonianza ed un *computer* incontrerebbe serie difficoltà nel giudicare se un teste abbia detto la verità, sia stato reticente o abbia mentito;
– in secondo luogo, perché plurimi e non predeterminati sono i criteri di valutazione della prova per cui, specialmente in un processo indiziario, ancor più difficile sarebbe per un algoritmo stabilire se determinati indizi possano essere considerati "gravi, precisi e concordanti" ai sensi dell'art. art. 192, comma 2, c.p.p.;

[63] Consultabile *online* al presente *link*. Per un primo commento ai contenuti della Carta, v. S. Quattrocolo, *Intelligenza artificiale e giustizia: nella cornice della Carta etica europea, gli spunti per un'urgente discussione tra scienze penali e informatiche*, in *www.lalegislazionepenale.eu*, 18 dicembre 2018; C. Barbaro, *Cepej, adottata la prima Carta etica europea sull'uso dell'intelligenza artificiale (IA) nei sistemi giudiziari*, in *Questione Giustizia online*, 7 dicembre 2018.
[64] A. Traversi, *Intelligenza artificiale applicata alla giustizia: ci sarà un giudice robot?*, in *Questione Giustizia online*, 10 aprile 2019.
[65] I virgolettati sono tratti dal testo della Carta (v. *supra*, nota 63).
[66] I virgolettati sono tratti dal testo della Carta (v. *supra*, nota 63).
[67] A. Traversi, *Intelligenza artificiale applicata alla giustizia*, cit., p. 3.

– infine, sembra pressoché impossibile aspettarsi da un algoritmo la capacità di intendere e applicare la regola di giudizio, di cui all'art. 533, comma 1, c.p.p., basata sull'"oltre ogni ragionevole dubbio", dal momento che possiamo immaginare *software* capaci di dare risposte secondo una logica binaria (sì/no; bianco/nero; vero/falso), o anche secondo una logica probabilistica (sì al 70%; bianco all'80%; vero al 90%), ma difficilmente *software* capaci di esprimere valutazioni, nella cui assunzione giochino un ruolo irrinunciabile – per quanto non ponderabile in termini precisi – fattori irriducibilmente umani[68].

Su tali questioni i processualpenalisti, anche in Italia, hanno già avviato un fecondo dibattito[69]. Nelle pagine seguenti noi ci limiteremo, invece, ad esplorare un particolare ambito di utilizzo degli algoritmi in ambito penale, costituito, segnatamente, dagli algoritmi destinati a fornire una prognosi circa la futura commissione di un (nuovo) reato da parte del soggetto sottoposto a valutazione.

5. Terzo percorso d'indagine - IA e valutazione della pericolosità criminale: gli algoritmi predittivi.

5.1. Considerazioni introduttive.

Quali probabilità sussistono che un individuo, avente determinate caratteristiche, possa in futuro commettere un (nuovo) reato?

Si tratta di un quesito la cui risposta è necessaria, tra l'altro, quando si tratta di applicare una misura di sicurezza, una misura cautelare o una misura di prevenzione, o anche per concedere la sospensione condizionale di una pena o l'affidamento in prova al servizio sociale[70]. Ebbene, a tale fondamentale quesito oggi i nostri giudici forniscono risposte per lo più intuitive, affidate esclusivamente alla loro esperienza personale e al loro buon senso, oppure, quando consentito dalla legge, basate su valutazioni cliniche di periti[71], mentre in futuro (e già nel presente di altri ordinamenti giuridici) siffatte valutazioni prognostiche della pericolosità criminale potrebbero essere affidate a specifici algoritmi (*risk assessment tools*, o algoritmi predittivi), capaci di attingere e rielaborare quantità enormi di dati al fine di far emergere *relazioni*, *coincidenze*, *correlazioni*, che consentano di profilare una persona e prevederne i successivi comportamenti, anche di rilevanza penale.

---

[68] Sul punto, v. pure S. Gaboriau, *Libertà e umanità del giudice: due valori fondamentali della giustizia. La giustizia digitale può garantire nel tempo la fedeltà a questi valori?*, in *Questione Giustizia*, fasc. 4, 2018, p. 11.

[69] Tra gli altri, v. G. Canzio, *Il dubbio e la legge*, in *Diritto penale contemporaneo*, 2018, pp. 1 ss.; M. Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei* risk assessment tools *tra Stati Unite ed Europa*, in *ivi*, 2019, pp. 1 ss.; A. Natale, *Introduzione. Una giustizia (im)prevedibile?*, in *Questione Giustizia*, fasc. 4, 2018, pp. 1 ss.; nello stesso fascicolo, v. pure i contributi di C. Costanzi, *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*, e di C. Castelli, D. Piana, *Giustizia predittiva. La qualità della giustizia in due tempi*; vedasi, infine, il fascicolo monografico di *Giurisprudenza Italiana* per i Centosettanta anni della Rivista, dedicato all'argomento dell'Intelligenza Artificiale (di prossima pubblicazione).

[70] Sui plurimi ambiti, all'interno dei quali risulta necessario formulare una prognosi di futura commissione di un (nuovo) reato, sia consentito rinviare a F. Basile, *Esiste una nozione ontologicamente unitaria di pericolosità sociale? Spunti di riflessione, con particolare riguardo alle misure di sicurezza e alle misure di prevenzione*, in *Riv. it. dir. proc. pen.* 2018, pp. 644 s.

[71] Sul cui grado di affidabilità, tuttavia, la dottrina è fortemente scettica: v., per tutti, J. Monahan, *Predicting violent behavior: An assessment of clinical techniques*, SAGE Library of Social Research, 1981.

5.2. La valutazione "attuariale" della pericolosità criminale.

Prima, tuttavia, di entrare nel vivo dell'analisi di questi "algoritmi predittivi" della pericolosità criminale, conviene anteporre alcune considerazioni sulla valutazione c.d. attuariale di siffatta pericolosità, che costituisce il presupposto teorico per l'utilizzo degli algoritmi predittivi.

Negli ultimi anni, infatti, come ben evidenzia Georgia Zara[72], si sta facendo sempre più strada una concezione *evidence-based* di valutazione del rischio individuale di commissione di un (nuovo) reato: una concezione, quindi, basata su riscontri oggettivi, destinata a soppiantare, o quanto meno integrare, le valutazioni intuitive dei giudici, tuttora ampiamente diffuse.

La valutazione *evidence-based* della pericolosità criminale presuppone la previa individuazione di una serie di *fattori di rischio* (o *predittori*) direttamente coinvolti nel comportamento criminoso, fattori che possono, tra l'altro, riguardare:

– l'età,
– il sesso,
– l'origine etnica,
– il livello di scolarizzazione,
– la situazione familiare e lavorativa,
– la posizione sociale,
– i precedenti penali,
– le precedenti esperienze carcerarie,
– i luoghi e le persone frequentati,
– la presenza di autori di reato nella cerchia familiare o nella rete di conoscenze,
– il luogo di residenza,
– le difficoltà di regolazione della rabbia e aggressività,
– il discontrollo degli impulsi,
– una storia di precedente violenza agìta,
– una storia di ospedalizzazione,
– un pensiero pro-criminale,
– alcune variabili contestuali (quali, ad esempio, la mancanza di sostegno familiare e sociale),
– il consumo di sostanze stupefacenti o alcoliche,
– le psicopatie.

Ebbene, tutti questi fattori, una volta raccolti grazie a studi longitudinali prospettici, possono consentire un *approccio di tipo attuariale (o statistico)* alla valutazione della pericolosità criminale. Attraverso, infatti, una loro combinazione, si possono predisporre delle 'scale' che consentono l'attribuzione di un punteggio (*score*) al soggetto preso in esame[73].

---

[72] G. Zara, *Tra il probabile e il certo. La valutazione del rischi di violenza e di recidiva criminale*, in *Diritto penale contemporaneo*, 20 maggio 2016, con riferimento, in particolare, al lavoro di J. P. Singh et al., *A comparative study of violence risk assessment tools: a systematic review and meta-regression analysis of 8 studies involving 25980 participants*, in *Clin Psychol Rev*, 31, 2011, pp. 499 ss.

[73] L. Castelletti, G. Rivellini, E. Straticò, *Efficacia predittiva degli strumenti di* Violence Risk Assessment *e possibili ambiti applicativi nella psichiatria forense e generale italiana*, in *Journal of Psychopathology*, 2014, pp. 153 ss.; G. Rocca, C. Candelli, I. Rossetto, F. Carabellese, *La valutazione psichiatrico forense della pericolosità sociale del sofferente psichico autore di reato: nuove prospettive tra indagine clinica e sistemi attuariali*, in *Rivista Italiana di Medicina Legale (e del Diritto in campo sanitario)*, n. 4, 2012, pp. 1442 ss.

Peraltro, le "scale", utilizzate per la valutazione attuariale della pericolosità criminale, si differenziano tra di loro, in base, tra l'altro, alla *popolazione* in relazione alla quale sono state elaborate (ad esempio, popolazione di adulti, di minori, di maschi, di femmine, di pazienti psichiatrico-forensi, di detenuti o ex-detenuti); in base alla *tipologia di reati implicati* (esistono scale generiche, cioè relative a tutti i reati, e scale specifiche, relative a singole tipologie di reati, come i reati sessuali o i reati violenti); in base alla *temporalizzazione del rischio* (immediato, o a medio o a lungo termine); in base, infine, al *contesto applicativo* (comunità civile, istituti di pena, centri di salute mentale, ospedali psichiatrico-giudiziari)[74].

Ovviamente, poi, non tutti i fattori di rischio sono uguali e non tutti impattano in modo univoco e nello stesso modo. I fattori di rischio, inoltre, hanno un differente tasso di *dinamicità*, nel senso che esistono: *i*) fattori statici, non modificabili (ad esempio, il sesso e l'origine etnica); *ii*) fattori dinamici stabili, che sono modificabili grazie al trattamento terapeutico (ad esempio, il discontrollo degli impulsi); *iii*) infine, fattori di rischio acuti, che cambiano rapidamente e sono associati ad una condizione facilitante la reazione violenta (ad esempio, l'uso di sostanze stupefacenti)[75].

Un altro concetto fondamentale in proposito, infine, è la c.d. *dose-exposure relationship*: se è, infatti, innegabile che più numerosi sono i fattori di rischio, più alta la probabilità di *outcomes* criminali, occorre altresì considerare che «precocità, durata e intensità dell'esposizione a più fattori di rischio che interagiscono in modo cumulativo, equifinale, dinamico, aumentano la probabilità di violenza e manifestazioni criminali»[76].

Pensiamo ora alla possibilità che queste valutazioni attuariali – e prima ancora la raccolta e rielaborazione dei dati che consentono la predisposizione delle "scale" – siano affidate, come di fatto già avviene negli Stati Uniti, a sistemi di intelligenza artificiale, quindi ad algoritmi predittivi, forniti di procedure di autoapprendimento (*machine learning*) e straordinaria capacità e rapidità nel far emergere *relazioni*, *coincidenze*, *correlazioni*, e non sarà difficile immaginare i grandi vantaggi (ma anche, come vedremo, i grandi rischi) che l'evoluzione tecnologica sembrerebbe promettere nella valutazione della pericolosità criminale.

5.3. L'impiego di algoritmi predittivi negli Stati Uniti.

Negli Stati Uniti, in effetti, già da una decina d'anni sono in fase di diffusione algoritmi predittivi della pericolosità criminale. Essi sono, ad esempio, usati nella fase del *parole* (per decidere se un individuo, nelle more della celebrazione del processo, possa essere rilasciato dietro il pagamento di una eventuale cauzione), o per misurare il rischio di recidiva del condannato, ai fini della sua ammissibilità al *probation* o ad altra misura alternativa alla detenzione.

5.3.1. Psa - *Public Safety Assessment*.

Stando a quanto riferito dalla giornalista Ephrat Livni[77], ad esempio, lo Stato del New Jersey (e, in misura minore, anche altri venti Stati), nell'intento di riformare il sistema del *parole*, ha introdotto un sistema algoritmico di valutazione della pericolosità criminale, denominato

[74] G. Zara, *Tra il probabile e il certo*, cit., pp. 17 ss.
[75] G. Zara, *Tra il probabile e il certo*, cit., p. 12.
[76] G. Zara, *Tra il probabile e il certo*, cit., p. 14.
[77] E. Livni, *Nei tribunali del New Jersey è un algoritmo a decidere chi esce su cauzione*, in *Internazionale*, marzo 2017 (trad. F. Ferrone).

Public Safety Assessment – Psa, ideato da una organizzazione *non profit* (la "Laura and John Arnold Foundation") col proposito di fornire ai giudici, impegnati a formulare una prognosi criminale, indicazioni scientifiche ed imparziali in tempi rapidi.

L'algoritmo Psa mette a confronto i fattori di rischio del soggetto sotto valutazione con un *database* di 1,5 milioni di casi provenienti da trecento giurisdizioni di tutti gli Stati Uniti e, in base alle informazioni a disposizione, attribuisce al medesimo un punteggio su una scala da uno a sei.

I fattori di misurazione di rischio presi in esame sono nove (tra cui l'età, i precedenti penali, le passate apparizioni in tribunale e le denunce ricevute in casi precedenti), e tra di essi non compaiono né la razza, né l'origine etnica e geografica.

Da quando il Psa è stato adottato, nello Stato del New Jersey il numero delle persone che sono state rilasciate su *parole* è decisamente aumentato; soprattutto è aumentato il numero delle persone rilasciate *senza* il pagamento di una cauzione. L'algoritmo sembra, quindi, aver favorito i soggetti non-pericolosi non abbienti: si consideri, infatti, che uno studio del 2013 aveva messo in evidenza che quasi il 40 % degli imputati che in teoria avrebbe potuto uscire su cauzione, alla fine restava in prigione perché non aveva abbastanza soldi per il pagamento della cauzione; oggi, invece, gli imputati sono rilasciati, indipendentemente dalle loro condizioni economiche, in base alla valutazione "neutrale" fornita dal Psa (valutazione che, tuttavia, si affianca a quella del giudice, senza sostituirla).

5.3.2. COMPAS - *Correctional Offender Management Profiling for Alternative Sanctions*.

L'algoritmo predittivo di gran lunga più famoso, utilizzato (e controverso) negli Stati Uniti è, tuttavia, quello di COMPAS - C*orrectional Offender Management Profiling for Alternative Sanctions*, un *software* elaborato e commercializzato da una società privata, la Northpointe (da gennaio 2017, ridenominata Equivant)[78].

Per avere un'idea di come funzioni COMPAS è possibile sfogliare *online* una versione del relativo "Manuale operativo", risalente a marzo 2015, da cui tra l'altro si apprende che:

> «COMPAS è uno strumento di valutazione dei rischi [di commissione di reato] e delle esigenze [trattamentali] di quarta generazione. Le agenzie di giustizia penale in tutto il Paese utilizzano COMPAS per assumere decisioni in merito al collocamento, alla supervisione e alla gestione degli autori di reati. COMPAS è stato sviluppato empiricamente con un focus sui predittori di cui è nota l'influenza sulla recidiva. Esso prende in considerazione anche fattori di rischio dinamici e fornisce informazioni su una varietà di fattori di rischio ampiamente convalidati [dalla ricerca scientifica] al fine di agevolare gli interventi correttivi rivolti a ridurre le probabilità di recidivanza [...]. COMPAS è stato sviluppato per la prima volta nel 1998 e da allora è stato rivisto più volte man mano che la base di conoscenze fornite dalla criminologia e dalla prassi correzionale si è evoluta [...]. Continuiamo a apportare miglioramenti a COMPAS sulla base dei risultati di ricerche empiriche e di studi sulla recidiva, condotti in carcere o presso agenzie preposte a seguire la *probation*. COMPAS viene periodicamente aggiornato per stare al passo con le migliori pratiche emergenti e i progressi tecnologici [...].

[78] J. Dressel, H. Farid, *The accuracy, fairness, and limits of predicting recidivism*, in *Science Advances*, fasc. 4, 2018, pp. 1 ss.: gli Autori riportano stime dalle quali risulterebbe che COMPAS, da quando è stato sviluppato nel 1998, è stato utilizzato in più di un milione di casi.

> COMPAS prende in considerazione – nella sua configurazione base – la risposta a 137 domande, concernenti le seguenti voci:
> – precedenti criminali;
> – precedenti illeciti e infrazioni;
> – passato di violenza;
> – violenza attuale;
> – frequentazioni con criminali;
> – abuso di sostanze;
> – problemi economici;
> – difficoltà nell'istruzione e nella formazione professionale;
> – ambiente familiare delinquenziale;
> – contesto sociale;
> – modo di utilizzo del tempo libero;
> – instabilità residenziale;
> – adeguamento sociale;
> – difetti di socializzazione;
> – opportunità criminali;
> – isolamento sociale;
> – pensiero pro-criminale;
> – personalità criminale[79].
> [...] Le risposte alle domande vengono fornite dal soggetto sotto valutazione, oppure vengono ricercate negli archivi o nei registri a disposizione delle procure e della polizia. La razza non è un fattore preso in considerazione. In sede di intervista con l'imputato gli vengono rivolte domande del seguente tipo: "Uno dei tuoi genitori è mai stato in prigione o è attualmente in prigione?". Il questionario chiede inoltre alle persone di essere d'accordo o in disaccordo con affermazioni del seguente tipo: "Una persona affamata ha il diritto di rubare" e "Se le persone mi fanno arrabbiare o perdere la calma, posso essere pericoloso".
> [...] COMPAS si distingue da altri *software* di calcolo attuariale, in quanto tiene in considerazione anche i fattori di rischio dinamici, oltre a quelli statici, e in quanto fornisce indicazioni non solo sul rischio di recidiva, ma anche sul trattamento più adatto per la singola persona per ridurre tale rischio»[80].

Nei confronti dell'impiego di COMPAS sono state tuttavia sollevate – sulla scorta di ricerche indipendenti – alcune critiche in ordine alla sua effettiva validità predittiva (*accuracy*) e alla sua imparzialità (*fairness*).

In particolare, nel maggio del 2016 un gruppo di ricercatori ha pubblicato, su incarico di una Organizzazione Non Governativa (ProPublica), una ricerca in cui si analizzavano le prestazioni di COMPAS su un campione di oltre 7000 persone arrestate nella contea di Broward, in Florida, tra il 2013 e il 2014. Da questa ricerca emergerebbe che le previsioni formulate da COMPAS erano inaffidabili e risentivano di distorsioni su base razziale, favorendo gli imputati bianchi rispetto agli imputati neri, dal momento che si rilevava una sottostima del rischio di recidiva dei primi e una sovrastima del rischio di recidiva dei secondi[81].

---

[79] Come risulta immediatamente evidente, le voci rilevanti per COMPAS coincidono in gran parte, salvo qualche scostamento terminologico, con quelle messe a punto dalla dottrina impegnata a consolidare e diffondere la valutazione attuariale della pericolosità criminale (v. *supra*, par. 5.2).

[80] *Practitioner's Guide to COMPAS Core*.

[81] J. Angwin, J. Larson, S. Mattu, L. Kirchner, *Machine Bias*, in *www.propublica.org*, 23 maggio 2016. In replica a tale studio, l'azienda produttrice di COMPAS ha commissionato una sorta di contro-studio, il quale avrebbe evidenziato una serie di errori nella metodica, nella misurazione e nella classificazione dei dati, commessi dai ricercatori di *ProPublica*: v. A. Flores, K. Bechtel, C. Lowenkamp, *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*, settembre 2016 (si noti, tuttavia, che tutti e tre gli Autori di questo contro-studio fanno parte del team di ricercatori di COMPAS). Per ulteriori riferimenti si può vedere direttamente la presente pagina del sito di Equivant.

Anche uno studio successivo[82] ha sollevato gravi perplessità, evidenziando due profili problematici ulteriori rispetto al rischio di pregiudizio razziale:

– l'algoritmo COMPAS fornirebbe previsioni sostanzialmente equivalenti a quelle fornite da persone prive di conoscenze specifiche in materia: quindi, la sua utilità sarebbe estremamente discutibile;
– il grado di affidabilità delle valutazioni, fornite da COMPAS dopo aver preso in considerazione ben 137 voci, sarebbe sostanzialmente equivalente al grado di affidabilità di valutazioni fornite sulla scorta di solo 2 voci.

Al momento, tuttavia, mancano ulteriori e più approfonditi studi di verifica delle prestazioni di COMPAS, anche a causa del fatto che molte delle informazioni utili per effettuare tali studi sono, in realtà, coperte da segreto industriale, ben custodito dalla società privata che commercializza il relativo *software*.

5.3.2.1. In particolare il caso Loomis e il controverso uso di COMPAS in sede di *sentencing*.

Critiche ancor maggiori nei confronti di COMPAS hanno riguardato il suo possibile utilizzo in sede di *sentencing*, vale a dire a fini di commisurazione della pena dell'imputato riconosciuto colpevole[83].

Tali critiche sono in particolare emerse in relazione al c.d. caso Loomis[84], dal nome di un imputato che aveva fatto ricorso alla Corte Suprema del Wisconsin per contestare l'entità della pena che gli era stata inflitta dalla Corte locale che, in fase commisurativa, si era per l'appunto avvalsa di COMPAS[85]: l'algoritmo predittivo veniva contestato dal ricorrente per la sua predisposizione a seguire pregiudizi basati sul genere e sulla razza, nonché per il difetto di trasparenza relativo al suo meccanismo di funzionamento.

La Corte Suprema del Wisconsin, sollecitata da tali rilievi, ha formulato un *warning* in relazione al futuro uso di COMPAS, mettendo in evidenza:

– la sua natura di prodotto coperto da segreto industriale, che impedisce la divulgazione di informazioni relative al suo metodo di funzionamento;
– il fatto che le valutazioni sono effettuate da COMPAS su base collettiva, di gruppo, e non individuale;
– infine, il rischio di una sovrastima del rischio di commissione di reati a carico di talune minoranze etniche[86].

---

[82] J. Dressel, H. Farid, *The accuracy, fairness*, p. 3, con ulteriori rinvii.
[83] Stando a quanto riferito da J. Angwin, J. Larson, S. Mattu, L. Kirchner, *Machine Bias*, cit., negli Stati di Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington e Wisconsin gli algoritmi predittivi (compreso ovviamente COMPAS) sono usati anche in fase di *sentencing*.
[84] Wisconsin S.C., State v. Loomis, 881, N.W.2d 749 (2016).
[85] Sul caso Loomis esiste una copiosa letteratura, anche fuori dagli Stati Uniti: E. Istriani, *Algorithmic Due Process: Mistaken Accountability and Attribution in State v. Loomis*, in *Harvard JOLT Digest*, 31 agosto 2017; K. Freeman, *Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State v. Loomis*, in *North Carolina Journal of Law & Technology*, Vol. 18, 2016, pp. 75 ss.; Anonimo, *State v. Loomis. Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing*, in *Harvard Law Review*, Vol. 130, 2017, pp. 1530 ss.; nella dottrina italiana, v. C. Costanzi, *La matematica del processo*, cit., p. 234; S. Carrer, *Se l'*amicus curiae *è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin, in Giur. pen. web*,, 24 aprile 2019.
[86] Su questi aspetti, v. in particolare K. Freeman, *Algorithmic Injustice*, cit., p. 76.

Ciò nondimeno, nel caso di specie, la Corte ha respinto il ricorso del sig. Loomis, sulla scorta della considerazione che le valutazioni di COMPAS non erano state "decisive", in quanto erano state pur sempre sottoposte al controllo e alla validazione di un giudice umano[87].

5.4. Considerazioni conclusive.

I sostenitori dell'impiego degli algoritmi predittivi ritengono che questi *software*, grazie all'elaborazione di *big data* e all'apprendimento automatico, rendono le valutazioni di pericolosità criminale più accurate e maggiormente esenti dal rischio di risentire di pregiudizi e condizionamenti culturali.

Per contro, come il dibattito sviluppatosi negli Stati Uniti intorno all'utilizzo di COMPAS ha mostrato, sono state sollevate serie perplessità in ordine all'effettiva validità predittiva (*accuracy*) e all'imparzialità (*fairness*) di questi algoritmi, i quali potrebbero produrre risultati poco affidabili o comunque discriminatori. Il loro uso solleva, altresì, attriti rispetto all'esigenza di una valutazione individualizzata della pericolosità criminale.

Vi sono poi problemi di trasparenza di non poco momento: si pensi solo al fatto che gli imputati, ma anche gli stessi giudici, in molti casi (ad esempio, nel caso di COMPAS) non hanno dettagli in ordine al funzionamento interno di questi *software*, giacché tali informazioni sono coperte da segreto industriale.

Oltre a questi profili controversi, dobbiamo inoltre fare i conti, come ben mette in evidenza Giovanni Canzio, con problemi etici e deontologici di deresponsabilizzazione dei giudicanti di non poco momento:

> «il dubbio del giudicante in ordine alla propensione dell'imputato a ripetere il delitto non trova più la soluzione in un criterio metodologico di accertamento del fatto e neppure in una puntuale prescrizione della legge, ma viene affidato a un algoritmo di valutazione del rischio, elaborato da un *software* giudiziario [...]. Considerati i risultati pratici – soprattutto in termini di risparmio – conseguiti dall'impiego del modello matematico-statistico, neppure le cautele e il *warning* delle corti e lo scetticismo degli studiosi, quanto al rispetto delle garanzie del *due process* nella raccolta delle informazioni utili per la valutazione del rischio nel mondo reale e all'eventuale pregiudizio discriminatorio, sono riusciti a frenare l'impetuosa avanzata delle tecniche informatiche di tipo predittivo nel sistema statunitense di giustizia penale. Si è forse agli inizi di uno sconvolgente (e però non auspicabile) mutamento di paradigma della struttura e della funzione della giurisdizione? A fronte della complessità tecnica, dei tempi e dei costi delle faticose operazioni giudiziali ricostruttive del fatto, la postmodernità metterà in crisi l'equità, l'efficacia e le garanzie del modello proprio del razionalismo critico, oppure resterà ben salda e vitale l'arte del giudicare *reasonig under uncertainty*, seppure *by probabilities*?»[88].

Per ora, quanto meno in Europa, gli algoritmi predittivi della pericolosità criminale (e, più in generale, gli *automated decision systems*, descritti nel paragrafo 4), non hanno avuto accesso nelle nostre aule penali, anche perché, a precludere loro l'accesso, si erge l'art. 15 della direttiva 95/46/CE, confluito nell'art. 22 del nuovo Regolamento europeo in materia di protezione dei dati personali, entrato in vigore il 25 maggio 2018. Tale articolo stabilisce, infatti, che ogni

[87] *Ibidem*.
[88] G. Canzio, *Il dubbio e la legge*, in *Diritto penale contemporaneo*, 20 luglio 2018, pp. 3 s. Sul classico ragionamento giudiziario *"by probabilities"*, v. lo stesso G. Canzio, *La motivazione della sentenza e la prova scientifica*: "reasoning by probabilities", in G. Canzio, L. Luparia (a cura di), *Prova scientifica e processo penale*, Cedam, 2018, pp. 3 ss.

persona ha il diritto di non essere sottoposta ad una decisione che produca effetti giuridici o abbia effetti significativi nei suoi confronti fondata esclusivamente su *un trattamento automatizzato di dati* destinati a valutare taluni aspetti della sua personalità.

Ancora in ambito eurounitario, occorre altresì ricordare che la già citata Risoluzione del Parlamento europeo sulla robotica del 2017 pone l'accento proprio sul *principio della trasparenza*, sottolineando la necessità che risulti sempre possibile indicare la logica alla base di ogni decisione, presa con l'ausilio dell'intelligenza artificiale, qualora tale decisione possa avere un impatto rilevante sulla vita di una o più persone.

Possiamo allora riportare a questo proposito le parole di Andrea Natale, il quale, riflettendo sull'utilizzo di algoritmi predittivi in sede giudiziaria, svolge le seguenti, condivisibili, considerazioni di sintesi:

> «(a) il risultato fornito dagli algoritmi predittivi è necessariamente influenzato dalla *qualità* dei dati che vengono posti come *input*; ne discende che è indispensabile prevedere meccanismi che assicurino: (a.1) la qualità del dato; (a.2) l'indipendenza della fonte da cui provengono i dati; (a.3) l'indipendenza dell'autorità che raccoglie i dati; (a.4) l'accessibilità a tutti dei dati posti come *input* dell'algoritmo;
> (b) è necessario scongiurare il rischio che l'algoritmo possa avere un esito discriminatorio fondato su dati personali sensibili, tra cui la razza e l'estrazione sociale [...];
> (c) la verificabilità o meno della struttura dell'algoritmo; un algoritmo ha una sua struttura che non è *neutra* [...]; nel concepire l'architettura di un algoritmo, il programmatore fa delle scelte che, necessariamente, influenzano il *risultato* dell'operazione computazionale; il programmatore può fare degli errori di progettazione; un algoritmo la cui struttura sia protetta da diritti di proprietà intellettuale e non open source è sottratto alla possibilità di controllo, verifica e confutazione da parte della parte processuale e, più in generale, della comunità degli utenti; ciò comporta non pochi problemi, tanto sotto il profilo della validazione dell'affidabilità scientifica del risultato che l'algoritmo restituisce, quanto sotto il profilo del diritto di difesa; si ritiene, pertanto, indispensabile che – laddove si voglia davvero fare un uso processuale di algoritmi predittivi da parte del sistema giudiziario (che è un sistema per sua natura *pubblico*) – nessun segreto possa essere posto sull'architettura degli algoritmi e dei dati che lo alimentano; si deve poi elaborare un meccanismo che assicuri anche l'indipendenza di chi ha elaborato l'algoritmo (che senso ha costituzionalizzare l'indipendenza del giudice e la sua soggezione solo alla legge se non si coltiva analoga pretesa a chi elabora uno strumento decisorio di simile portata?);
> (d) l'algoritmo – anche ove usato non come strumento decisorio esclusivo, ma come mero supporto alla decisione del giudice – richiede formazione; è dunque indispensabile *formare* il personale giudiziario che potrebbe doversene avvalere;
> (e) l'algoritmo predittivo – muovendo da una elaborazione della giurisprudenza e dei casi precedenti – può indicare non 'il risultato' esatto di una certa controversia, ma il suo possibile esito, evidenziando quali siano le linee giurisprudenziali prevalenti e quali gli esiti concreti che si sono dati in casi simili; ciò, però, comporta più di un rischio: (e.1) l'algoritmo non è in grado di "riconoscere" che *quello a lui sottoposto non è un caso simile*; vi sono delle singolarità che un decisore umano, forse rileverebbe e che lo porterebbero ad operare un *distinguishing*; l'algoritmo non è progettato per prevedere questa evoluzione; (e.2) per la stessa ragione, l'algoritmo può favorire quello che Garapon [...] chiama come *effetto moutonnier* (effetto pecora nel gregge): è concreto, in altri termini, il rischio di indurre il giudice pigro ad adagiarsi sulla proposta dell'algoritmo senza assumere su di sé l'autentica responsabilità del giudizio che egli emette; (e.3) per la stessa ragione, l'uso di

algoritmi può favorire una cristallizzazione della giurisprudenza, rendendola meno sensibile ai cambiamenti sociali (e, di fatto, rendendoli meno probabili)»[89].

Le ragioni di perplessità che accompagnano l'utilizzo degli algoritmi predittivi per la valutazione della pericolosità criminale sono state, infine, condivise anche dalla Commissione per l'efficacia della giustizia (CEPEJ) del Consiglio d'Europa che ha elaborato la già menzionata Carta etica. La Carta, infatti, ricorda che l'uso di algoritmi in materia penale al fine di profilare le persone è stato criticato dalle ONG «a causa dei limiti della metodologia utilizzata», e in particolare del loro «approccio meramente statistico», il quale avrebbe «effetti discriminatori e deterministici», sicché esso andrebbe «sostituito da un altro approccio che risulti più rispettoso delle norme europee in materia di sanzioni penali e che salvaguardi le *chances* di riabilitazione e reintegrazione del singolo individuo. Se i sistemi algoritmici riescono a migliorare la raccolta di informazioni per valutazioni inerenti la *probation*, per esempio, e a rendere possibile che le informazioni pertinenti siano raccolte più rapidamente per la successiva elaborazione umana, allora si tratterebbe sicuramente di un progresso (in particolare nei procedimenti sommari). Qualsiasi altro uso è esposto a pregiudizi destinati ad entrare in conflitto con alcuni principi fondamentali, nazionali e sovranazionali»[90].

Alla luce di tutte queste perplessità il presente paragrafo deve, quindi, necessariamente concludersi con un interrogativo: siamo davvero pronti a delegare valutazioni che possono incidere significativamente sui diritti fondamentali di una persona – come la valutazione della pericolosità criminale – ad un *software*?

6. Quarto percorso d'indagine - IA e reato: possibili ipotesi di coinvolgimento – come *strumento*, come *autore*, o come *vittima* – di un sistema di IA nella commissione di un reato.

6.1. Considerazioni introduttive.

Droni che uccidono per le strade urbane, come avvenuto nella città di Dallas nel luglio 2016[91]; auto senza conducente coinvolte nella causazione di incidenti a danno di cose o persone, come nel caso del tragico investimento di un pedone, avvenuto nel marzo 2018 in Arizona[92]; *software* che eseguono, in collaborazione o addirittura in sostituzione dell'uomo, compiti sempre più sofisticati, come pilotare un grosso aereo, ma che qualche volta possono interferire negativamente con la condotta umana, come i recenti disastri aerei dei Boeing 737 Max hanno purtroppo dimostrato[93]: qualora in tali episodi si possa riscontrare un fatto di reato, chi ne risponderà penalmente? il programmatore del *software*, il suo produttore o il suo utilizzatore? e siamo proprio sicuri che in queste ipotesi, e in altre simili immaginabili, il sistema di IA sia

[89] A. Natale, *Introduzione. Una giustizia (im)prevedibile?*, in *Questione Giustizia*, fasc. 4, 2018, pp. 3 ss. Trattasi dello scritto introduttivo, che fa da premessa ad alcuni saggi, raccolti nel cit. fasc. 4, e dedicati, sia pur in prospettive diverse, al tema "*Una giustizia (im)prevedibile?*".
[90] Carta Etica, cit., pp. 67 ss.
[91] N. Sharkey, *La robotica*, cit., p. 197, riferisce di un sospetto cecchino ucciso a Dallas tramite l'intervento di un drone (luglio 2016), commentando con le seguenti parole tale episodio: «in quel caso esisteva una chiara giustificazione e gli esperti di diritto hanno asserito che l'azione era stata legittima, resta il fatto che in quella circostanza si è probabilmente varcato un confine. È giusto proteggere la polizia, e la polizia dovrebbe, fintanto che è possibile, utilizzare mezzi non violenti. Quando questi si dimostrino inefficaci, è certamente necessario elevare il livello della forza impiegata, ma in modo graduale e proporzionale al reato che viene commesso: e sono valutazioni decisamente impegnative per un robot che agisce senza il controllo umano».
[92] L. Butti, *Le auto guideranno da sole, ma con quali responsabilità?*, in *Il Bo Live*, 9 novembre 2018; F. Suman, *Dilemmi morali per le auto a guida autonoma*, in *ivi*, 7 novembre 2018.
[93] G. F. Italiano, *Intelligenza artificiale, che errore lasciarla agli informatici*, in *Agendadigitale.eu*, 11 giugno 2019.

davvero solo uno *strumento* inanimato per la realizzazione del reato o, in considerazione delle sue caratteristiche, esso potrà essere considerato direttamente l'*autore* del reato?

Non vanno d'altra parte trascurate nemmeno le ipotesi in cui i sistemi di IA potrebbero essere considerati alla stregua di *vittime* di attacchi contro loro specificamente rivolti, sicché potremmo interrogarci circa l'opportunità dell'introduzione di nuove figure di reato destinate a punire, ad esempio, chi sottopone ad un malevole logoramento un robot, capace di riprodurre le sembianze (e i sentimenti?) di un cucciolo di cane, utilizzato in un programma di *doll therapy*, o a punire chi compie, senza prima averne acquisito il consenso, atti sessuali con un androide, progettato originariamente non certo per dare sfogo agli istinti libidinosi, ma per svolgere funzioni di *receptionist* all'interno di un albergo.

Si tratta di interrogativi che nella letteratura, soprattutto d'Oltreoceano, sono già affiorati e di cui anche le istituzioni e i soggetti pubblici sembrano avere consapevolezza. Particolare attenzione a questo problema è stata ad esempio dedicata dal Parlamento UE, che nella sua già citata Risoluzione sulla robotica del 2017 dedica un intero paragrafo dei *Considerando* alla voce "responsabilità". Vale la pena riportare alcuni di tali *Considerando*, in quanto essi – benché si riferiscano esclusivamente all'ambito della responsabilità civile e si focalizzino solo sui *software* di IA incorporati in robot – sono in grado di dare un adeguato abbrivio anche ad una riflessione su IA e responsabilità penale:

> «AA. considerando che *l'autonomia di un robot* può essere definita come la capacità di prendere decisioni e metterle in atto nel mondo esterno, *indipendentemente da un controllo o un'influenza esterna*; che tale autonomia è di natura puramente tecnologica e il suo livello dipende dal grado di complessità con cui è stata progettata l'interazione di un robot con l'ambiente;
> AB. considerando che *più* i robot sono autonomi, *meno* possono essere considerati come meri strumenti nelle mani di altri attori (quali il fabbricante, l'operatore, il proprietario, l'utilizzatore, ecc.); che ciò, a sua volta, pone il quesito se le regole ordinarie in materia di responsabilità siano sufficienti o se ciò renda necessari nuovi principi e regole volte a chiarire la *responsabilità legale dei vari attori per azioni e omissioni imputabili ai robot*, qualora le cause non possano essere ricondotte a un soggetto umano specifico, e se le azioni o le omissioni legate ai robot che hanno causato danni avrebbero potuto essere evitate;
> AC. considerando che, in ultima analisi, l'autonomia dei robot solleva la *questione della loro natura* alla luce delle categorie giuridiche esistenti e dell'eventuale necessità di creare una nuova categoria con caratteristiche specifiche e implicazioni proprie [...];
> IA. considerando che [...] l'attuale quadro giuridico non sarebbe sufficiente a coprire i danni causati dalla nuova generazione di robot, in quanto questi possono essere *dotati di capacità di adattamento e di apprendimento che implicano un certo grado di imprevedibilità nel loro comportamento*, dato che imparerebbero in modo autonomo, in base alle esperienze diversificate di ciascuno, e interagirebbero con l'ambiente *in modo unico e imprevedibile*»[94].

6.2. Il sistema di IA quale *strumento* di commissione del reato.

Cominciamo dall'ipotesi relativamente più semplice: quella in cui il sistema di IA costituisce lo *strumento* in mano ad altri – segnatamente, *in mano ad un uomo* – attraverso il quale il reato viene commesso[95]. Le enormi potenzialità dell'IA, infatti, potrebbero – e già lo

[94] Risoluzione del Parlamento europeo del 16 febbraio 2017, cit., corsivi aggiunti.
[95] In proposito, v. già S. Riondato, *Robotica e diritto penale (robot, ibridi, chimere, "animali tecnologici")*, in D. Provolo, S. Riondato, F. Yenisey (a cura di), *Genetics, Robotics, Law, Punishment*, cit., pp. 600 ss.

sono state! – essere asservite anche a scopi criminali, e quindi essere utilizzate per la commissione di reati.

Tra le condotte criminali che più potrebbero essere agevolate dall'impiego di sistemi di intelligenza artificiale vi sono i "crimini informatici, economici ed ambientali, i traffici internazionali di sostanze stupefacenti e di altri prodotti illeciti, la trattata di esseri umani"[96], ma anche le violazioni in materia di *privacy* e trattamento dei dati personali, le violazioni della proprietà intellettuale ed industriale, i reati di diffamazione e le condotte di abuso della credulità popolare, magari commessi attraverso *bot* che creano *fakenews* destinate alla rete, etc.

Due esempi possono forse illustrare le enormi potenzialità "delinquenziali" dei sistemi di IA allorché essi divengano strumenti per la commissione di illeciti attraverso modalità fino a qualche anno fa assolutamente inimmaginabili.

Il primo esempio è costituito dal c.d. *bagarinaggio online*: quando un sito mette in vendita i biglietti per un concerto o un altro evento di grande richiamo per il pubblico, nel giro di pochi minuti una gran quantità di questi biglietti viene accaparrata da pochi soggetti che li acquistano attraverso i *bot*, programmi informatici capaci di eseguire le operazioni di acquisto ad una velocità inaccessibile per qualsiasi essere umano, per poi rimetterli in vendita su un mercato parallelo (*secondary ticketing*) a prezzi, ovviamente, decisamente maggiorati rispetto a quelli originari: un'attività che, tra l'altro, fomenta fenomeni di elusione ed evasione fiscale da parte dei "bagarini"[97].

Il secondo esempio è costituito dalle *condotte di manipolazione abusiva del mercato* che possono essere commesse attraverso sofisticati programmi informatici, a cui è affidata non solo l'esecuzione delle transazioni finanziarie, ma anche la stessa decisione di compierle sulla scorta di un algoritmo che compara, in una frazione di secondo, numerose variabili: si tratta degli HFT (acronimo di *High Frequency Traders*), capaci di eseguire migliaia di operazioni al secondo. Come è stato dimostrato da studi e ricerche, un uso distorto degli HFT può provocare fenomeni di improvvisa e rapidissima oscillazione dei prezzi sui mercati finanziari, anche di rilevanza penale (in termini di aggiotaggio, manipolazione abusiva del mercato, etc.), senza che a tali oscillazioni sia associato alcun mutamento del valore sostanziale del titolo oggetto di contrattazione[98].

Dobbiamo, insomma, prepararci ad un'era in cui la commissione di reati con lo strumento dell'IA potrebbe diventare assai frequente ed incisiva, anche in considerazione dell'accresciuta vulnerabilità di alcuni aspetti della vita umana connessi ad impieghi dell'intelligenza artificiale, a partire dall'impressionante numero di dati sul comportamento e lo stile di vita delle persone che possono essere raccolti tramite vari canali informatici e, da ultimo, tramite l'Internet delle Cose[99], fino all'eventuale instaurazione di rapporti di vera dipendenza, talora anche affettiva, da macchine e sistemi di servizio che si muovono per noi, lavorano per noi (magari in sostituzione di

[96] Le ipotesi di reato riportate nel virgolettato sono menzionate nel documento di presentazione del "*2019 OSCE Annual Police Experts Meeting: Artificial Intelligence and Law Enforcement: An Ally or an Adversary?*", cit. *supra*, nota 30.

[97] Dal 1 luglio 2019 in Italia, con la Legge di Bilancio 2019 (segnatamente, con l'art. 1 comma 1100, legge n. 145 del 2018), è stato introdotto l'obbligo del c.d. biglietto nominativo per talune categorie di eventi con più di 5000 spettatori, proprio al fine di prevenire la pratica illecita del bagarinaggio *online* (v. la presente pagina di Altroconsumo); per un intervento del legislatore californiano, parimenti mirante a contrastare il bagarinaggio *online*, v. P. McGreevy, *California ban on ticket-buying "bots" is signed into law*, in *Los Angeles Time*, 23 settembre 2013.

[98] Il meccanismo degli HFT e dei suoi effetti distorsivi del mercato finanziario è ben illustrato, anche nelle sue implicazioni penali, da F. Consulich, *Il nastro di Möbius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, in *Banca Borsa Titoli di credito* 2018, pp. 195 ss.

[99] S. Tafaro, *Riflessioni sulle intelligenze artificiali. Neutralità della rete*, in *Diritto@Storia*, quaderno 15, 2017, p. 3.

funzionalità umane "disabilitate"), e addirittura accudiscono i nostri anziani o i nostri figli. L'uomo si ritrova in balia della macchina, sguarnito dei presidi tradizionali di protezione, essendo tali presidi tradizionali concepiti e strutturati per proteggerlo da attacchi "umani".

Occorre, allora, mettere in campo nuove fattispecie di reato (o rimodellare quelle già esistenti) al fine di renderle applicabili alla realizzazione di condotte criminose attraverso lo *strumento* dell'IA, offrendo così tutela ai beni giuridici anche da questa nuova fonte di attacchi?

Si tratta di interrogativi che dovranno entrare nell'agenda degli studiosi di diritto penale e, molto presto, anche in quella del legislatore.

6.3. Il sistema di IA quale autore *del reato:* machina delinquere potest?[100]

I vari esempi sopra formulati, che abbiamo presentato come ipotesi in cui il sistema di IA è lo strumento, in mano all'uomo, per la commissione di un reato, potrebbero, tuttavia, prestarsi anche ad una differente lettura. Qualora, infatti, il sistema di IA coinvolto nella commissione del reato fosse un sistema di ultima generazione, fornito di capacità di apprendimento e di autonomia decisionale, potremmo chiederci se non risulti già varcata la frontiera del futuro, tanto da potersi individuare direttamente nel sistema di IA l'*autore* del reato: se così fosse, il sistema di IA dovrà rispondere penalmente di tale reato[101]?

A prescindere dalla risposta che si vorrà fornire alla presente questione, sta di fatto che in tutti i casi in cui la condotta dell'uomo si intreccia e si interseca con l'attività di un sistema di IA, il percorso di attribuzione delle responsabilità indubbiamente si complica, giacché le scelte, le valutazioni, i bilanciamenti, sottesi alla commissione del fatto, non sono più opera esclusiva dell'uomo, ma sono quantomeno condivisi con (se non interamente delegati alla) macchina[102].

Vengono in mente scenari in parte già noti: come si individua il responsabile di un'attività svolta in *equipe*? come si individua il colpevole in quelle ipotesi in cui il procedimento decisionale ed esecutivo è parcellizzato, frazionato e distribuito in capo ad una pluralità di soggetti? La novità sta però ora nel fatto che tra i membri dell'*equipe*, tra i plurimi soggetti coinvolti nel procedimento decisionale ed esecutivo non vi sono più solo esseri umani, ma anche sistemi di IA: e se il sistema di IA presenta autonomia decisionale e capacità di apprendimento e di reazione per effetto della propria esperienza e interazione con l'ambiente, diventa difficile escluderlo del tutto dal meccanismo di attribuzione della responsabilità. In effetti, come è stato acutamente rilevato:

> «il presupposto indefettibile per imputare al programmatore, al fornitore e/o all'utente la responsabilità per danni generati dall'IA coincide con il controllo che questi sono in grado di esercitare sul *software*, un controllo che verrebbe meno specialmente a fronte delle più

---

[100] La suggestiva formula *Machina delinquere non potest* (che noi qui riprendiamo sopprimendo il "non" e aggiungendo il punto di domanda) – formula la quale a sua volta ricalca l'antico brocardo *societas delinquere non potest*, a lungo invocato per precludere una responsabilità da reato a carico degli enti – è stata coniata da A. Cappellini, *Machina*, cit., p. 1.

[101] Oltre agli Autori in seguito citati, si pongono l'interrogativo sulla possibile responsabilità penale dei sistemi di IA, con varietà di risposte, tra gli altri, U. Pagallo, S. Quattrocolo, *The impact of AI on criminal law, and its twofold procedures*, in W. Barfield, U. Pagallo (a cura di), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Pub, 2018, pp. 385 ss.; D. Lima, *Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challanges for Criminal Law*, in *South Carolina Law Review*, 2018, pp. 677 ss.; T. King, N. Aggarwal, M. Taddeo, L. Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in *Science and Engineering Ethics*, 2019; S. Gleß, E. Silverman, T. Weigend, *If robots cause harm, who is to blame? Self-driving cars and criminal liability*, in *New Criminal Law Review*, 2016, pp. 412 ss.; P. Asaro, *A body to Kick, but Still No Soul to Damn: Legal Perspectives on Robotics*, in P. Lin, K. Abney, G.A. Bekey (a cura di), *Robot Ethics*, MIT Press, 2012, pp. 169 ss.

[102] In argomento v. C. Bagnoli, *Teoria della responsabilità*, Il Mulino, 2019, pp. 74 ss.

> evolute e sofisticate tecnologie [...]. Le forme più avanzate di Intelligenza Artificiale, a differenza di altri sistemi innovativi fondati sull'automazione, non permetterebbero di apprezzare *ex ante* e in maniera verosimile il loro funzionamento e le possibili deviazioni rispetto a tali standard. Questo problema troverebbe conforto nello sviluppo di capacità di autoapprendimento che rendono ancora più difficile calcolare le conseguenze dell'utilizzo dell'Intelligenza Artificiale. Proprio la presenza di dinamiche intrinseche al loro funzionamento che permettono alle macchine di apprendere nuove funzioni, acquisendo gradualmente un'autonomia sempre maggiore rispetto a chi ne ha programmato il *software*, segnerebbe un limite importante alla capacità di tenuta degli standard normativi esistenti. Di conseguenza, le nuove intelligenze sarebbero in grado di porre in essere azioni e movimenti indipendentemente dal contributo del programmatore e/o dell'utente, e che non risalgono a comandi precedentemente impartiti, sia nella fase di progettazione che nella fase di utilizzo, e dunque non sono imputabili giuridicamente ad alcuno»[103].

6.3.1. Tra deresponsabilizzazione dell'uomo e responsabilizzazione della macchina.

Prima, tuttavia, di approfondire la questione se "*machina delinquere potest?*", occorre soffermarsi anche su un altro profilo, connesso al precedente. Il coinvolgimento dei sistemi di IA per l'espletamento di talune attività innesca, infatti, inevitabilmente – se non dal punto di vista giuridico, per lo meno a livello di fatto – un processo di «alienazione della responsabilità» dall'agente umano, come efficacemente si esprime Carla Bagnoli[104], che ben descrive il rischio di un allontanamento, di uno scarico della responsabilità (per lo meno, della «responsabilità morale») dall'agente umano, ricorrendo all'esempio dei droni e al loro impiego nell'aeronautica militare:

> «i droni che sostituiscono i piloti dell'aereonautica militare, ovvero i sistemi aeromobili a pilotaggio remoto, [sono] strumenti particolarmente invasivi che semplificano le operazioni militari ed espandono il raggio di azione degli esseri umani, introducendo una sorta di catena di comando molto complessa. Il comando dell'azione rimane all'essere umano, ma l'essere umano non è materialmente l'agente dell'azione; anzi, occupa una posizione distante nel tempo e nello spazio e, talvolta, nemmeno ben identificabile. Il fatto che vi sia una catena di comando cui far risalire la responsabilità dell'azione può far pensare che l'azione sia meno soggetta al caso e all'improvvisazione, ma ogni anello della catena ha il suo punto di fragilità. Data la complessità della struttura gestionale e la distanza dall'azione, è difficile dire chi, alla fine, ha la responsabilità dell'azione. Soprattutto, poiché il comando è così frazionato, l'operatore non si identifica immediatamente con l'agente dell'azione, ciò che induce un tipo di alienazione con pericolosi effetti deresponsabilizzanti. Questa caratteristica ha effetti pesanti anche sul tipo di deliberazione in cui si può impegnare un agente alla catena di comando remota. Proprio perché separato e distante dallo scenario di guerra, non è condizionato allo stesso modo. Per esempio, è plausibile che abbia reazioni emotive differenti, di diversa intensità e che operi con una propensione al rischio alterata rispetto a quella che avrebbe se fosse realmente calato nella situazione. Queste differenze incidono significativamente sul modo in cui gli agenti si assumono la responsabilità delle proprie decisioni [...]. L'argomento principale a sfavore [dell'uso dei droni] è che il drone non consente una interazione con l'avversario e quindi priva l'avversario della relazione

[103] M. Bassini, L. Liguori, O. Pollicino, *Sistemi di Intelligenza Artificiale*, cit., pp. 356 ss. In termini analoghi, v. pure P. Asaro, *The liability problem for autonomous artificial agents*, in *AAIA Symposium on Ethical and Moral Considerations in Non-Human Agents*, 2016, p. 192, il quale ricorre ad un interessante paragone, comparando la posizione del programmatore e dell'utente nei riguardi dei sistemi di IA con quella dei genitori nei confronti dei propri figli una volta che questi siano divenuti "adulti" e abbiano pertanto appreso a sufficienza del funzionamento della realtà per poter sopportare *iure proprio* le conseguenze giuridiche delle proprie condotte.
[104] C. Bagnoli, *Teoria della responsabilità*, cit., p. 77.

personale, limitando drasticamente le modalità di negoziazione e le espressioni di reciprocità»[105].

La deresponsabilizzazione (morale) dell'agente umano rischia di indebolire anche la tutela, garantita dal diritto penale, ai beni giuridici[106]: questo indebolimento potrebbe, allora, essere compensato dall'individuazione di un nuovo soggetto responsabile?

A quanto pare, quanto più si attivano meccanismi, anche solo fattuali, di deresponsabilizzazione dell'agente umano, tanto più abbiamo bisogno di interrogarci sulla possibilità di configurare una responsabilità penale direttamente in capo ai sistemi di IA.

6.3.2. Vacilla il confine tra *machina* e persona?

La questione della possibile attribuzione di responsabilità, anche penale, ad entità diverse dall'uomo non è una novità assoluta nel dibattito giusfilosofico: Platone, ne *Le Leggi*, attribuiva la responsabilità anche ad animali e cose[107]; ancora alle soglie dell'Illuminismo venivano celebrati processi penali a carico di animali "delinquenti"[108]; dal 2001 anche in Italia è stata configurata una responsabilità da reato in capo agli enti (d.lgs. 231 del 2001), a carico, quindi, di "persone" che sono tali solo per effetto di una *fictio* giuridica (per l'appunto, "persone giuridiche").

L'ultima frontiera è segnata dai sistemi di IA: possono essi essere considerati "persone", o quanto meno possono essere equiparati alle "persone", al fine di una attribuzione di responsabilità, anche penale[109]?

Segnali, per ora ancora molto deboli, a favore di una siffatta equiparazione, provengono da alcune iniziative che hanno, in realtà, più che altro il sapore di ardite operazioni di *marketing*, ma che potrebbero anticipare future tendenze[110]: come ad esempio la decisione dell'Arabia Saudita di concedere la cittadinanza al sofisticato androide Sophia[111], oppure la scelta di una municipalità di Tokyo di riconoscere la residenza ad un *chatbot*, Shibuya Mirai, in grado di dialogare, con le competenze e le abilità di un bambino di sette anni, con tutti i suoi "concittadini"[112].

---

[105] C. Bagnoli, *Teoria della responsabilità*, cit., p. 78.
[106] V. in proposito anche F. Consulich, *Il nastro di Möbius*, cit., pag. 204, che efficacemente parla di «disumanizzazione delle offese».
[107] Come ci ricorda da ultimo C. Bagnoli, *Teoria della responsabilità*, cit., p. 72.
[108] Riferimenti in A. Cappellini, *Machina*, cit., p. 20; C. Bagnoli, *Teoria della responsabilità*, cit., p. 73.
[109] Il dibattito in materia è stato inizialmente avviato dai filosofi del diritto e dai filosofi dell'informatica (v., tra gli altri, H. Jonas, *The Imperative of Responsibility. In search of an Ethics for the Technological Age*, University of Chicago Press, 1984; L.B. Solum, *Legal Personhood for Artificial Intelligences*, in *Illinois Public Law and Legal Theory Research Papers*, Series No. 09-13, 2008; L. Floridi, J.W. Sanders, *On the Morality of Artificial Agents*, in *Minds and Machines*, 14/3, 2004, pp. 349 ss.; B.C. Stahl, *Information, Ethics, and Computers: The Problem of Autonomous Moral Agents*, in *Minds and Machines*, 14, 2004, pp. 67 ss.; Id., *Responsible Computers? A Case for Ascribing Quasi-Responsibility to Computers Independent of Personhood or Agency*, in *Ethics and Information Technology*, 8, 2006, pp. 205 ss.; G. Sartor, *Gli agenti* software: *nuovi soggetti del ciberdiritto*, in *Contratto e impresa*, 2, 2002, pp. 57 ss.) e si è di recente acceso anche tra gli studiosi della responsabilità civile (si veda, ad esempio, A. Santosuosso, C. Boscarato, F. Corleo, *Robot e diritto: una prima ricognizione*, in *Nuova giur. civ. comm.*, II, 2012, pp. 497 ss.; A. Santosuosso, *If the agent is not necessarily a human being. Some legal thoughts*, in D. Provolo, S. Riondato, F. Yenisey (a cura di), *Genetics, Robotics, Law, Punishment*, cit., pp. 545 ss., nonché il volume, a cura di Ruffolo U., *Intelligenza artificiale e responsabilità*, Giuffrè, 2018) e tra i costituzionalisti (si veda, ad esempio, il volume a cura di F. Pizzetti, *Intelligenza artificiale, protezione dei dati personali e regolazione*, cit.).
[110] Cfr. C. Trevisi, *La regolamentazione in materia di Intelligenza artificiale*, cit., p. 2.
[111] Si veda quanto riportato a questo *link*.
[112] Si veda quanto riportato a questo *link*.

6.3.3. Una colpevolezza "disumana"?

Come è noto, uno dei requisiti fondamentali della responsabilità penale è la colpevolezza. Mentre gli altri requisiti del reato – la commissione di un fatto storico, corrispondente alla previsione di una norma astratta incriminatrice; l'antigiuridicità; la punibilità[113] – non sembrano comportare gravi ostacoli ad una loro riferibilità, pur con i dovuti adattamenti, anche ad un sistema di IA, la possibilità di individuare una colpevolezza vera e propria – e non un simulacro della stessa – in capo ad un sistema di IA solleva, invece, non poche difficoltà, logiche e ontologiche.

La colpevolezza, infatti, esprime il coinvolgimento soggettivo, personalistico, dell'autore al fatto commesso, e la sua presenza comporta la possibilità di muovergli un rimprovero, sul presupposto che in capo a questo soggetto siano ravvisabili l'imputabilità, il dolo o la colpa, la conoscenza, o per lo meno la conoscibilità della legge penale violata, infine l'assenza di cause di esclusione della colpevolezza (detto in positivo: la normalità del processo motivazionale)[114].

Ebbene: possiamo riferire questi elementi, originariamente concepiti e tradizionalmente riferiti solo all'uomo, anche ad una macchina? In particolare, possiamo parlare di "capacità di intendere e di volere" in relazione ad un *software*? Possiamo configurare una "colpa" (quale mancata osservanza di una regola precauzionale di comportamento) o addirittura un "dolo" (quale volontà consapevole di realizzazione del fatto) dell'algoritmo[115]?

C'è chi dice sì[116]!

Per comprendere le ragioni di una tale risposta affermativa, conviene tornare alle prime pagine di questo nostro lavoro, dove abbiamo descritto brevemente le caratteristiche dei sistemi di IA (v. *supra*, par. 2), e ricordare, in particolare, che «i recenti progressi fatti nella robotica, nella percezione e nel *machine learning*, supportati dai miglioramenti sempre più veloci della tecnologia informatica hanno permesso la messa a punto di una nuova generazione di sistemi capaci di rivaleggiare con le capacità umane in determinati domini o in compiti specifici – o addirittura di superarle. Questi sistemi sono ben più *autonomi* di quanto le persone si accorgano. Sono in grado di imparare dalle loro stesse esperienze e di intraprendere azioni neanche lontanamente contemplate dai loro progettisti. La *frase* di buon senso comunemente accettata secondo la quale "*i computer fanno solo quello che sono programmati a fare*" *non è più vera*»[117].

I sistemi di IA, perlomeno quelli più evoluti e sofisticati, sono quindi capaci di agire in autonomia, di assumere ed eventualmente attuare decisioni proprie, che non erano prevedibili dai loro programmatori. È, pertanto, del tutto ragionevole congetturare che il margine di decisione autonoma di cui dispongono i sistemi di IA costituirà la breccia attraverso la quale potrebbe farsi strada, in un prossimo futuro, una matura teoria della responsabilità, anche penale, dei sistemi di intelligenza artificiale, teoria che dovrà, ovviamente, indicare presupposti e limiti di una siffatta

[113] Sulle differenti concezioni del reato – bipartita, tripartita, o quadripartita – v., *ex pluris*, su posizioni differenti, F. Palazzo, *Corso di diritto penale, parte generale*, VII ed., Giappicchelli, 2018, pp. 197 ss.; G. Marinucci, E. Dolcini, *Corso di diritto penale*, III ed., Giuffrè, 2001, pp. 625 ss.
[114] V. per tutti F. Mantovani, *Diritto penale, parte generale*, X ed., CEDAM, 2017, p. 288.
[115] Su quest'ultimo interrogativo, v. D. Falcinelli, *Il dolo in cerca di una direzione penale. Il contributo della scienza robotica ad una teoria delle decisioni umane*, in *Arch. Pen.*. fasc. 1, 2018, p. 9.
[116] Tra gli scienziati di IA, fornisce una convinta risposta affermativa alle questioni formulate nel testo, J. Kaplan, *Intelligenza artificiale*, cit., p. 153: «un sistema di IA può commettere reati? La risposta è sì»; Id., *Le persone non servono. Lavoro e ricchezza nell'epoca dell'intelligenza artificiale*, Luiss University Press, 2016, p. 80. Tra gli studiosi di diritto penale, la posizione più avanzata è quella sostenuta da Gabriel Hallevy, i cui lavori sono oggetto di una meditata presentazione critica da parte di A. Cappellini, *Machina*, cit., pp. 10 ss., e di M. Bassini, L. Liguori, O. Pollicino, *Sistemi di Intelligenza Artificiale*, cit., pp. 363 ss., ai quali, pertanto, è in questa sede possibile rinviare.
[117] J. Kaplan, *Intelligenza artificiale*, cit., p. 19 (corsivo aggiunto).

responsabilità, i quali non necessariamente dovranno coincidere con quelli validi per gli esseri umani – allo stesso modo in cui, del resto, la vigente disciplina della responsabilità da reato degli enti non coincide in tutto e per tutto con la disciplina da reato degli uomini[118].

Peraltro, la prospettiva di una futura configurabilità di una responsabilità, anche penale, dei sistemi di IA trapela anche in uno dei *Considerando* della già ricordata Risoluzione del Parlamento europeo sulla robotica (per quanto la Risoluzione si occupi esclusivamente di responsabilità civile). Il *Considerando* Z, infatti, così recita:

> «considerando che, grazie agli strabilianti progressi tecnologici dell'ultimo decennio, non solo oggi i robot sono in grado di svolgere attività che tradizionalmente erano tipicamente ed esclusivamente umane, ma lo sviluppo di determinate caratteristiche autonome e cognitive – ad esempio la capacità di apprendere dall'esperienza e di prendere decisioni quasi indipendenti – li ha resi sempre più simili ad agenti che interagiscono con l'ambiente circostante e sono in grado di alterarlo in modo significativo; che, in tale contesto, la questione della responsabilità giuridica derivante dall'azione nociva di un robot diventa essenziale».

Concordiamo pienamente: la questione della responsabilità (anche penale) derivante dall'azione nociva di un robot (*rectius*, un sistema IA) è ormai diventata *essenziale*, anche per l'ambito penale, e dovrà essere oggetto, negli anni a venire, di attenta riflessione in dottrina, al fine di elaborare un quadro concettuale che possa adeguatamente supportare il futuro legislatore.

6.3.4. Quali pene per i sistemi di IA?

Una futura, eventuale teoria della responsabilità penale dei sistemi di IA non dovrebbe, ovviamente, trascurare il profilo delle sanzioni comminabili a tali sistemi. Anzi, si tratta di questione centrale, dal momento che, come è noto, il rilievo *penale* della responsabilità è ampiamente, anche se non esclusivamente, connesso alla tipologia di sanzioni che possono essere applicate all'esito dell'accertamento della responsabilità: insomma, possiamo parlare di responsabilità *penale* (solo) se la sanzione applicata è una *pena*.

Tradizionalmente si riconosce che una pena possa avere, esclusivamente o cumulativamente, una funzione retributiva (viene inflitta la pena per retribuire – far pagare – il male recato con il reato), una funzione general-preventiva (la minaccia della pena funge, per i suoi possibili destinatari, da freno alla spinta criminale) o una funzione special-preventiva (l'inflizione, e l'esecuzione, della pena pone il condannato nella condizione di non commettere più, in futuro, reato)[119].

Di queste tre funzioni, la prima e la terza sembrano realizzabili anche nei confronti dei sistemi di IA: attraverso lo spegnimento definitivo o temporaneo della macchina, o attraverso la sottoposizione della macchina, dotata di congegni di autoapprendimento, ad un nuovo *training*

[118] Per un approccio "possibilista" circa la (futura) configurabilità di una responsabilità penale in capo ai sistemi di IA, v. P.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 (a cura di), *AI Approaches to the Complexity of Legal Systems. AICOL 2013. Lecture Notes in Computer Science*, Springer, 2014, pp. 145 ss.; S. Gleß, T. Weigend, *Intelligente Agenten und das Strafrecht*, in *Zeitschrift für die gesamten Strafrechtswissenschaften - ZStW* 2014, pp. 561 ss.; infine, S. Beck, *Intelligent agents and criminal law - Negligence, diffusion of liability and electronic personhood*, in *Robotics and Autonomous Systems*, 2016, pp. 138 ss.

[119] Sulle "teorie della pena", v. per tutti D. Pulitanò, *Diritto penale*, VII ed. Giappichelli, 2017, pp. 47 ss.; F. Palazzo, *Corso di diritto penale*, cit., pp. 16 ss.

"rieducativo", si potrebbero realizzare, rispettivamente, la funzione retributiva e la funzione special-preventiva della pena[120].

Più difficile, al momento, risulta, invece, immaginare l'esplicazione di una funzione general-preventiva della pena nei confronti dei sistemi di IA, salvo dar spazio a ipotesi per ora fantascientifiche: la formulazione del precetto penale in termini (digitali) tali che possa essere recepito ed elaborato dalla "comunità" dei sistemi di IA, oppure la trasmissione, tramite i canali dell'Internet delle Cose, dell'esperienza della pena subita dal singolo computer, punito in quanto riconosciuto responsabile del reato, anche ai computer consimili, futuri potenziali autori di reati.

Accanto, quindi, al quesito *machina delinquere potest*?, occorrerà presto porsi anche il connesso quesito: (*quomodo*) *machina puniri potest?*

6.4. Il sistema di IA quale vittima *del reato.*

L'ultima tematica, infine, alla quale conviene accennare all'interno di una riflessione dedicato a "intelligenza artificiale e reato", riguarda la possibilità di configurare il sistema di IA quale vittima del reato.

Alcuni degli argomenti che potrebbero, infatti, indurre a considerare il sistema di IA quale "persona" (e, quindi, quale autore di reato), potrebbero fornire elementi anche a favore di un suo riconoscimento quale possibile vittima del reato: non solo, quindi, come "cosa" inanimata che subisce materialmente il reato, ma anche come "persona" o come soggetto equiparato alla persona che soffre del reato subito[121].

Ad un riconoscimento così impegnativo, quasi una sorta di umanizzazione della macchina, sembrerebbe, tuttavia, ostare la considerazione che i sistemi di IA «non hanno, né avranno mai, veri sentimenti»[122].

Ciò non esclude, tuttavia, l'opportunità di riflettere circa la possibile introduzione di nuove figure di reato (o l'eventuale modificazione di figure di reato già esistenti), in modo da rendere punibili anche attacchi rivolti specificamente ai sistemi di IA, che sembrerebbero ad oggi non trovare adeguato e pieno riscontro nelle vigenti disposizioni penali.

Si pensi, ad esempio, a quei robot, riproducenti bambole o animali, di cui si comincia a valorizzare l'utilizzo all'interno di programmi di *doll therapy* o *pet therapy*, rivolti a soggetti autistici, malati di Alzheimer, disabili mentali, rispetto ai quali il paziente potrebbe sviluppare sentimenti e nutrire emozioni (in ciò consistendo lo scopo della terapia!)[123]: ebbene, la distruzione di questi robot, il loro maltrattamento, la loro sottoposizione malevola a logoramento integra un semplice fatto di danneggiamento? o, per lo meno in quei casi in cui i robot siano dotati di capacità cognitiva e siano specificamente destinati a instaurare una relazione affettiva con il paziente, non ci stiamo forse già muovendo in una zona prossima a quella oggi coperta – se non dal reato di

[120] Per alcune suggestioni in tal senso, v. J. Kaplan, *Intelligenza artificiale*, cit., pp. 156 ss.
[121] In proposito, v. S. Riondato, *Robotica e diritto penale (robot, ibridi, chimere, "animali tecnologici")*, in D. Provolo, S. Riondato, F. Yenisey (a cura di), *Genetics, Robotics, Law, Punishment*, cit., pp. 602 ss., secondo il quale, almeno in via di ipotesi ed «entro certi limiti, alcuni raffinati soggetti di intelligenza artificiale potrebbero essere equiparati a "forme di vita" non-umane. Del resto, diversi sistemi di diritto penale tutelano già esseri non-umani da certe offese. L'esempio più evidente è quello della tutela penale degli animali».
[122] J. Kaplan, *Intelligenza artificiale*, cit., p. 126.
[123] Un programma del genere è stato, ad esempio, avviato presso l'istituto Sacra Famiglia di Cesano Boscone, ed è rivolto ai malati di Alzheimer: v. notizia su *Famiglia Cristiana*, n. 35, 2 settembre 2018, , p. 97.

maltrattamenti contro familiari e conviventi (art. 572 c.p.), perlomeno – dal reato di maltrattamento di animali (art. 544 *ter* c.p.)?

Per altro verso, si pensi alle preoccupazioni connesse al c.d. fenomeno dello "stupro robotico", e in particolare agli atti sessuali con robot aventi le dimensioni e le fattezze di minori[124]: se, da un canto, la criminalizzazioni di tali condotte parrebbe un'ulteriore concessione ad una visione moraleggiante e paternalistica del diritto penale[125], dall'altro canto non dovrebbe rimanere priva di ogni rilievo la tematica del "consenso" del robot all'atto sessuale, ovviamente un consenso da ricostruire e interpretare in termini ben diversi da quelli in cui concepiamo il consenso delle persone umane. *Quid iuris*, ad esempio, nel caso di compimento di atti sessuali con un androide, progettato originariamente non per tale scopo, ma per svolgere funzioni di *receptionist* all'interno di un albergo o di un grande magazzino?

Anche tali questioni – sia pur, forse, con minore urgenza rispetto ad altre sopra prospettate – necessitano, quindi, di riflessione in vista di un successivo, eventuale intervento del legislatore.

7. Quale futuro ci aspetta?

Nelle pagine precedenti abbiamo cercato di individuare, senza alcuna pretesa di esaustività (anzi, talora peccando di una certa superficialità, che confidiamo il lettore ci voglia perdonare), quattro scenari all'interno dei quali la rivoluzione tecnologica messa in moto dall'IA già solleva, o è destinata a sollevare, problemi, dubbi e questioni, rilevanti per il diritto penale.

Abbiamo, infatti, indagato le possibile applicazioni di IA nelle attività di *law enforcement*, con particolare attenzione allo specifico ambito della *polizia predittiva*. Siamo poi passati a verificare se i c.d. *automated decision systems* siano già usati, o possano essere in futuro efficientemente e legittimamente usati anche per prendere decisioni all'interno di procedimenti penali. La nostra attenzione si è, quindi, spostata sui c.d. *algoritmi predittivi*, impiegati per valutare la pericolosità criminale di soggetti in vario modo implicati negli ingranaggi della giustizia penale. Infine, ci siamo interrogati sulle possibili ipotesi di coinvolgimento – come *strumento*, come *autore*, o come *vittima* – di un sistema di IA nella commissione di un *reato*.

Tutti questi quattro scenari – per dipingere i quali, a dire il vero, abbiamo dovuto spesso procedere alla formulazione di mere ipotesi e incerte previsioni rivolte al futuro – sembrano accomunati dall'attuale assenza di una regolamentazione normativa, e in particolare di una regolamentazione che prevenga o reprima offese penalmente rilevanti. Le ipotesi e le previsioni, tuttavia, potrebbero presto divenire realtà, e allora quell'assenza normativa comporterebbe conseguenze drammatiche.

[124] In argomento, v. J. Danaher, *Robotic Rape and Robotic Child Sexual Abuse: Should They be Criminalized?*, in *Criminal Law and Philosophy*, 2017, pp. 71 ss.; M. Maras, L. Shapiro, *Child Sex Dolls and Robots: More Than Just an Uncanny Valley*, in *Journal of Internet Law*, 12, 2017, pp. 3 ss.; R. Brown, J. Shelling, *Exploring the implications of child sex dolls*, in *Trends & Issues in Crime and Criminal Justice*, n. 570, 2019, pp. 2 ss.

[125] Esprime giustamente questa preoccupazione A. Cappellini, *Machina*, cit., p. 3.