L'IRCOCERVO
PRIMA RIVISTA ELETTRONICA ITALIANA DI METODOLOGIA GIURIDICA - TEORIA GENERALE DEL DIRITTO E DOTTRINA DELLO STATO
FIRST ITALIAN DIGITAL JOURNAL OF LEGAL METHODOLOGY - GENERAL THEORY OF LAW AND DOCTRINE OF THE STATE

Ex Machina. Il diritto penale, all'improvviso

Vincenzo Bruno Muscatiello

Università degli Studi di Bari Aldo Moro

Abstract: **Ex Machina. Criminal Law, Suddenly**
Artificial intelligence is increasingly occupying our lives, posing new possibilities and new critical issues. The legal system tries to regulate them but, although the civil law reading is widely discussed, the possible punitive instances capture aspects of the artificial reality for which traditional criminal law could prove unprepared.

Keywords: Artificial Intelligence, Civil Law, Criminal Law, Artificial Criminal Liability.

1. Ehi Siri

Nel 2019 l'Organizzazione per la cooperazione e lo sviluppo economico (OCSE) ha previsto che entro 15-20 anni le nuove tecnologie di automazione finiranno probabilmente per eliminare il 14% dei posti di lavoro a livello mondiale e trasformare radicalmente un altro 32%, il che significa che quasi oltre un miliardo di persone, in tutto il mondo, vedranno modificate le loro abitudini lavorative, con un cambiamento radicale delle forme e una accelerazione delle competenze professionali. Le nuove tecnologie non solo potranno accrescere e migliorare le attività ripetitive manuali, ma saranno in grado di eseguire tipologie di lavoro sempre più complesse e fondate sulla ricerca, codifica e scrittura, un tempo immaginate impermeabili ai cambiamenti radicali.

Non possiamo più far finta di ignorare gli sviluppi della intelligenza artificiale, le sue attitudini, le sue funzionalità: siamo di fronte ad una profonda modifica esistenziale, portatrice di grandi benefici e di grandi cambiamenti, per una diversa, e probabilmente migliore, gestione documentale, una dematerializzazione completa dei documenti, dello loro validazione, fino alla loro archiviazione e conservazione; per una diversa e migliore *digital finance*, attraverso la creazione di sistemi di allerta preventiva e strumenti che aiutino a valutare la propria situazione

aziendale e quella dei clienti dei fornitori, per prevenire eventuali crisi e problemi di liquidità; una migliore e diversa capacità di analisi, attraverso strumenti che consentano di trasformare numeri e dati di diversa provenienza e difficile aggregazione, in un vantaggio competitivo attraverso proprio gli algoritmi di *machine learning* che riusciranno ad organizzare le informazioni in modo coerente, per una migliore interpretazione e un loro migliore utilizzo strategico; una diversa e migliore attenzione nel seguire il passo degli obblighi normativi, ricreando soluzioni semplici e intuitive per adeguare le organizzazioni aziendali, per esempio alla disciplina in materia di *privacy* e di GDPR, o per tutti quegli adempimenti che siano richiesti dalla legge, come, per esempio, la fatturazione elettronica o la segnalazione degli illeciti *whistleblowing*. Non solo. L'intelligenza artificiale ci aiuterà in una nuova visione dei *risk management* riuscendo a controllare ed impedire che singole classi di rischio raggiungano livelli minacciosi e ciò che è ordinariamente affidato alle intelligenze umane, potrebbe essere affidato a questi sistemi, a questi algoritmi e a queste nuove applicazioni di IA, per nuove modalità innovative di utilizzo di dati, trasformati in modelli operativi e in processi inesplorati. L'intelligenza artificiale riuscirà ad occupare attività anche quelle pericolose, come quelle dell'edilizia, estrazioni minerarie, monitoraggio di sezioni radioattive delle centrali nucleari, consentirà di, o meglio, riuscirà ad esplorare oceani e lo spazio interplanetario, a svolgere incarichi rischiosi: terremoti, alluvioni, disastri, passeranno, dalla cura del caso, alla disciplina del prevedibile, consentendo di salvare coltivazioni, territori, abitazioni e vite umane[1]; l'AI promette di saper e poter monitorare le coltivazioni in tempo reale, misurare il *se* e il *quanto* della irrigazione in ragione di agenti atmosferici registrati in tempo reale, scegliere il fertilizzante e quando raccogliere per massimizzare la resa[2]; la diagnosi, come quella di ipertensione polmonare, sarà infallibilmente affidata ad un algoritmo; l'abbandono di rifiuti potrà essere monitorato attraverso l'esame del DNA nella

[1] Nell'articolo on line, "Intelligenza Artificiale e Disastri Ambientali: Flood Hub e Nestore", *in Environment da Creative AI* (https://data.creative-words.com/intelligenza-artificiale-e-disastri-ambientali-flood-hub-e-nestore/), emerge l'importanza dell'intelligenza artificiale nella prevenzione di disastri e calamità naturali attraverso una preventiva analisi delle precipitazioni e degli eventi sismici, cosi da anticipare la possibilità di incendi, alluvioni, tempeste etc. Un valido esempio di IA che ha come obiettivo quello di prevenire catastrofi ambientali è Flood Hub, un sistema sviluppato nel 2018 di monitoraggio dei corsi d'acqua con lo scopo di prevedere inondazioni e alluvioni e arginarne le eventuali conseguenze. Si tratta di un sistema di previsione costituito da quattro sottosistemi (convalida dei dati; previsione delle fasi; modellazione delle inondazioni; distribuzione degli avvisi) preceduto da una raccolta dati sulle avvenute alluvioni, usando poi i sistemi di *Machine Learning* e intelligenza artificiale per elaborare delle stime e delle previsioni su possibili future inondazioni. Una diversa applicazione della intelligenza artificiale si è legata allo studio delle sequenze sismiche: pur risultando impossibile determinare con assoluta certezza il luogo e il momento preciso di un evento sismico, i sistemi di IA sono risultati fondamentali per analizzare e individuare i sismi grazie ad un algoritmo – denominato Nestore – in grado di apprendere ed effettuare una valutazione delle probabilità di repliche sismiche.

[2] Ce lo ricorda P. Triolo, "Il piano degli Stati Uniti per rallentare l'AI cinese", in *Limes*, 12 (2022), p. 43.

spazzatura buttata ai bordi della strada (così avviene già in Asia); persino la giustizia punitiva[3], tradizionalmente arricchita da coefficienti di umanità ed equità, saprà aggiungere, o affidarsi, a regole di predittività, calcolate da un algoritmo, e rendersi più agevole, e apparentemente più giusta e imparziale, sulla base di processi decisionali garantiti da algoritmi impermeabili a preconcetti, pregiudizi o idiosincrasie[4], che si tratti di calcolare la recidiva[5] o la prevenzione della criminalità, di vario genere[6], in una nuova configurazione dei rapporti sociali[7] ed in una sempre più vasta area di collaborazione[8], estesa, perché no, alle nuove forme di responsabilità delle persone giuridiche[9] per analizzare il presente, esaminare la rischiosità aziendale attraverso una costante e progressiva digitalizzazione dei protocolli organizzativi e dei sistemi di analisi e di controllo dei rischi[10], e cosi

[3] Di recente G. Lasagni, "Difendersi dall'intelligenza artificiale o difendersi con l'intelligenza artificiale? Verso un cambio di paradigma", in *Rivista italiana di diritto e procedura penale*, (2022), n. 4, pp. 1545 ss.

[4] In argomento, per opportune osservazioni critiche, si veda G. Canzio, "Intelligenza artificiale, algoritmi e giustizia penale", in *Sistema penale*, 08/01/2021; nonché G. Ubertis, "Intelligenza artificiale, giustizia penale, controllo umano significativo", in *Sistema penale*, 11/11/2020, ma anche in *Diritto penale contemporaneo*, 4 (2020).

[5] Gli usi della AI nel sistema giuridico sono raccontati anche da A. Contaldo, F. Campara, nel saggio "Intelligenza artificiale e Diritto. Dai sistemi esperti 'classici' ai sistemi esperti 'evoluti': tecnologia e implementazione giuridica", in G. Taddei Elmi, A. Contaldo (a cura di), *Intelligenza artificiale. Algoritmi giuridici.* Ius condendum *o "fantadiritto"?*, Pacini, Pisa, 2020, pp. 50 ss.

[6] C. Burchard, "L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società", in *Rivista italiana di diritto e procedura penale*, 62 (2019), n. 4, pp. 1919 ss.

[7] Nuova configurazione in cui la funzione terza del diritto, piuttosto che essere affidata a terzi, "diventa immanente, localizzandosi ormai nella tecnica stessa": così A. Garapon, J. Lassègue, *La giustizia digitale. Determinismo tecnologico e libertà*, trad. it., il Mulino, Bologna, 2021, p. 245.

[8] A. Giannini, "Intelligenza artificiale, *human oversight* e responsabilità penale: prove d'impatto a livello europeo", in *Criminalia*, (2021), p. 11.

[9] In tema, per la materia ambientale, R. Nitti, V.B. Muscatiello, nell'ultimo capitolo de *Il diritto penale dell'ambiente*, Cacucci, Bari, 2023.

[10] Un algoritmo programmato in queste forme potrebbe essere un buon viatico per una misura dell'esistente meno esposta alle variabili umane. *Behavioural Biometrics*, *Robot Process Automation*, *Intelligence Process Automation* e *Intelligence Business Process Management Suites*, sono alcune delle formule che traducono la nuova frontiera della protezione tecnologica, già nel 2014 capace di ospitare, quale membro di un consiglio di amministrazione, Vital, un algoritmo basato su intelligenza artificiale, quale membro del consiglio di amministrazione della Deep Knowledge Ventures, una *venture capital fund* con sede ad Hong Kong, col compito di vagliare le possibili perdite derivanti da eccessivi investimenti in progetti sopravvalutati. Nel 2016, anche la Tieto, un'azienda norvegese leader nel settore IT, ha incluso nel suo consiglio di amministrazione un'intelligenza artificiale di nome Alicia T., dotato di un sistema di interfaccia conversazionale, con il quale era persino possibile avere una discussione e porre domande su qualsiasi cosa. E sempre più spesso è annunciata la futura predominante presenza di membri non umani nei C.d.A. delle grandi imprese con funzioni specifiche (*Assisted Artificial Intelligence*), per il sostegno alle decisioni (*Augmented Artificial Intelligence*), infine per sostituire la decisione umana (*Autonomous Artificial Intelligence*).

standardizzare i modelli di prevenzione 231[11] con, ad esempio[12], l'implementazione di *software* che controllino il rispetto delle soglie limite previste dalla normativa di settore; o meccanismi di *allert* che registrino la mancata trasmissione di dati di inquinamento alle autorità competenti in caso di superamento dei valori soglia[13].

La *Digital Criminal Compliance* riuscirebbe a rendere i meccanismi migliori e più efficienti, sia nella fase di costruzione del modello (ci ricordano, G. Morgante, G. Fiorinelli, "Promesse e rischi della compliance penale digitalizzata", in *Archivio penale*, (2022), n. 2, p. 9, come "le tecnologie digitali possano anzitutto supportare la fase preliminare del c.d. *risk assessment*, contribuendo alla 'fotografia' del contesto e all'identificazione dei rischi, mediante la raccolta dei dati e l'analisi di tutti i dati relativi alla realtà aziendale". Parimenti quanto alla "progettazione del modello organizzativo [ovverosia, alla fase di definizione dei presidi]"), sia in quello della sua vitalità aziendale, offerta alla naturale applicazione quanto alla sua innaturale disapplicazione, entrambe registrate e controllate da un sistema matematico di tipo algoritmico, in grado di captare ed elaborare i dati in un tempo sensibilmente più rapido, e più certo, rispetto al controllo umano; e meno permeabile a condizionamenti, veri o presunti, di tipo corporativistico o particolaristico, pericolo questo ritenuto scongiurabile dal momento che "Gli standard e le norme tecniche possono contenere anche elementi di tipo valutativo e disciplinare situazioni conflittuali, che però, almeno teoricamente, vengono risolte attraverso procedure codificate e schemi della valutazione e della gestione del rischio, sulla base di tecniche statistico-probabilistiche": così V. Torre, "Compliance penale e normativa tecnica", in *Archivio penale*, (2022), n. 1, p. 9.

[11] G. Morgante, G. Fiorinelli, *op. cit.*; R. Trezza, "L'Intelligenza Artificiale come ausilio alla standardizzazione del modello 231: vantaggi 'possibili' e rischi 'celati'", in *Giurisprudenza penale*, (2021), n. 1-*bis*.

[12] L'esempio è tratto dall'articolo *on line* D. Speranzini (2022), "Intelligenza artificiale e D. Lgs. 231/2001: i sistemi automatizzati di prevenzione dei reati", in *MMSLEC*. Recuperato da https://mmslex.com/intelligenza-artificiale-e-d-lgs-231-2001-i-sistemi, [Data di consultazione: 30/05/2024].

[13] La nuova legge n. 22 del 2022 in tema di reati culturali e responsabilità delle persone giuridiche aiuta a comprendere quanta parte, nei meccanismi della *compliance*, possano avere le nuove tecnologie informatiche: così per i modelli di organizzazione, gestione e controllo i quali, ispirandosi alle Linee Guida delle Nazioni Unite per la prevenzione e repressione del traffico di beni culturali del 2014 ed alle misure pre-penali di cui alla Convenzione di Nicosia, potranno aprirsi al tracciamento dell'origine degli oggetti d'arte e della catena proprietaria; alla registrazione e all'aggiornamento presso archivi elettronici di tutti i dati relativi ai beni culturali di cui sono in possesso; all'introduzione di sistemi auspicabilmente informatici di licenze di esportazione e importazione per i beni culturali; all'adozione di registri auspicabilmente elettronici delle transazioni commerciali riguardanti opere d'arte e di antiquariato, laddove non già previsti; al monitoraggio delle compravendite di questi beni su Internet, possibilmente con il coinvolgimento e la responsabilizzazione dei gestori delle piattaforme potenzialmente utilizzate a questo fine; nella segnalazione presso le autorità competenti di attività sospette. Insomma, l'informatica, e più in genere la modernità, al servizio della *compliance*.

Potremmo continuare a lungo, in un lungo elenco di attività, persino le più varie e bizzarre: cantare[14], suonare[15], dipingere, ricordare e scrivere (il correttore di questo saggio è un algoritmo), formulare romanzi, pareri legali, interpretare l'esistenza, anche di settori delle arti della creatività e persino inventare, ponendo, nuovamente, e con più senso, il dubbio sulla esistenza di una intimità cibernetica, nelle forme che la parola *coscienza* può contenere[16], capace di uscire fuori dagli schemi e abbandonarsi ad una bizzarria intelligente, ancora una volta migliore di quella umana. *Le città invisibili* avranno, ed hanno avuta, finalmente una rappresentazione magica, attraverso una stupefacente e *visibile* raffigurazione segnica.

In breve, l'intelligenza artificiale sarà un copilota della nostra vita e, di tanti aspetti della nostra vita, sarà un amplificatore del lavoro delle persone e delle aziende, capace di sottrarre alle persone e alle aziende i tempi di *routine*, e così aggiungere nuovi tempi, aumentare la creatività, dare spazi a nuove soluzioni e a nuove innovazioni, a nuove attitudini ingegnose: la produttività del sistema, in uno studio che ha riguardato il sistema Italia, potrà aumentare fino al 18% grazie all'adozione della intelligenza artificiale generativa, con una generazione di valore che a parità di ore lavorate, di ore cioè lavorate nelle forme tradizionali, potrà arrivare fino a 312 miliardi di euro di valore aggiunto.

Il futuro è lì, ed è impossibile rinunciarvi.

[14] La sapienza artistica raggiunta, inspiegabilmente e insperabilmente, dalla capacità di un sistema di generare un prodotto musicale ha visto la riproduzione della voce di Anita Mui, una cantante scomparsa 40 anni prima, riuscendo a totalizzare 100 milioni di ascolti. La produzione si deve al colosso cinese Tencent, e insieme a questa con milioni di ascolti, altre mille canzoni prodotte dalla intelligenza artificiale: così F.M. De Collibus, "L'Era delle macchine che apprendono", in *Limes*, 12 (2022), p. 24. Ci ricorda C. Cavaceppi, "L'Intelligenza artificiale applicata al diritto penale: criticità attuali e prospettive future", in G. Taddei Elmi, A. Contaldo (a cura di), *op. cit.*, p. 98, che l'impatto economico della automazione potrebbe raggiungere un livello fra il 6.500 e 12.000 miliardi di euro entro il 2025.

[15] Di recente un sistema semplice di IA elaborato dal Politecnico di Milano ha trovato il modo di predire il suono emesso da tavole di violino di forma diversa: l'IA è stata addestrata con un database di oltre 1.500 forme di tavole armoniche ispirate a violini storici (fra cui il Messia di Stradivari) ma anche a forme insolite, riuscendo così a suggerire ai liutai la forma migliore da dare allo strumento.

[16] Ci ricorda G. Taddei Elmi, "Introduzione. Dall'informatica giuridica al diritto dell'informatica", in G. Taddei Elmi, A. Contaldo (a cura di), *op. cit.*, p. XXVI, che, se la coscienza dipende dalla quantità di neuroni e di connessioni sinaptiche attivate contemporaneamente, e se il cervello possiede mille miliardi di neuroni e ciascuno di essi sviluppa mille connessioni sinaptiche per un totale di 10 alla quindicesima, si stima che una disponibilità di RAM di 10 alla quindicesima verrà raggiunta da un calcolatore presumibilmente nel 2029. Un tale calcolatore potrebbe dirsi avere "coscienza".

2. *The artificial game*: la mossa 37

In questo futuro non troppo distante da noi, sempre più prossimo, ormai è chiaro, l'IA cambierà le nostre vite. C'era da aspettarselo. Non è, del resto, la prima volta che le macchine hanno *vinto* contro gli umani, sorpassata l'abilità umana, come quando negli anni '90 il computer *Deep Blue* di IBM sconfisse il grande maestro di scacchi Kasparov, nella rassicurante spiegazione di una vittoria legata all'esistenza di un inventario di tutte le possibili mosse, al cui interno il computer avesse *semplicemente* identificata la mossa migliore. Era come – così si pensava – se una mossa, avendo più probabilità di successo, fosse stata programmata in *Deep Blue*, mappata in tutte le possibili scelte, lasciando al computer semplicemente la decisione di attuare la mossa vincente, la possibilità cioè di razionalizzare e semplificare ciò che l'uomo, l'intelligenza umana, aveva previamente inserito nella sua programmazione. Come dire, il computer, l'IA, faceva più rapidamente quello che, semplicemente, l'intelligenza umana gli aveva consentito di fare, un acceleratore della decisività umana in puro spirito deterministico.

La superiorità umana veniva, dunque, in qualche modo riconosciuta, in una rassicurazione del dominio umano su quello artificiale, durata almeno fino alla consapevolezza dell'inatteso: l'inatteso arriva poco dopo, nella sfida del team *Deep Mind* di Google a Lee Sedol, 18 volte campione del mondo, imbattuto nel gioco da tavolo Go, una disciplina ludica esistente da almeno 3000 anni, praticata, soprattutto in Cina, da quasi 40 milioni di persone, e considerata il più antico giuoco da tavolo ininterrottamente praticato fino ai giorni nostri. Si tratta di una sorta di giuoco degli scacchi, ma con una disciplina più complessa, un reticolo di molte più caselle, con pedine bianche e nere, chiamate *pietre* che hanno tutte lo stesso valore differenziale, appunto come per giuoco degli scacchi, ma molto più complesso in quanto ogni turno ha circa 200 diverse mosse possibili, rispetto alle sole 20 disponibili con gli scacchi, dal momento che ad ogni turno si aggiunge una pedina[17] e ciò rende il numero di combinazioni possibili largamente superiore al totale di tutti gli atomi che esistono nell'universo.

Per questo tipo di giuoco, è fisicamente e umanamente impossibile creare delle istruzioni, da cui poi il computer possa ricavare la mostra vincente, l'unico modo per giocare a questo gioco è usare l'intuizione e la creatività, piuttosto che la memoria o le formule matematiche. La sfida nel 2016 fu vinta da *AlphaGo*, la AI di Deep Mind, addestrata su trenta milioni di giuochi diversi, che sconfisse il campione del mondo imbattuto Lee Sedol quattro volte a uno, nello stupore degli osservatori, e non tanto per il numero delle vittorie, quanto per il fatto che l'intelligenza artificiale avesse vinto usando la cosiddetta e famosa *mossa 37* della seconda partita, il fatto, cioè, che l'intelligenza artificiale avesse pensato e deciso di mettere una pedina sulla quinta linea, scegliendo una mossa che nessun giocatore

[17] Traggo queste informazioni da un bel saggio di M. Di Michele, dal titolo *Intelligenza artificiale. Etica, rischi e opportunità di una tecnologia rivoluzionaria*, Diarkos, Santarcangelo di Romagna, 2023.

professionista avrebbe mai fatto all'inizio di un match. La *mossa 37* apparve, agli umani, come una mossa insolita, bizzarra, fantasiosa, che muoveva da una intuizione creativa, che fuoriusciva dallo schema deterministico della intelligenza artificiale, e si collocava in una dimensione nuova, una sorta di autodeterminazione in cui la macchina, autonomamente, riservava a sé la decisione di come giocare, come cioè trarre dalle innumerevoli mosse osservate e sperimentate, la possibilità di imparare autonomamente dei propri errori, fino ad arrivare al punto di inventare e creare nuove mosse, mai viste prima.

Tremila anni di antichissima perizia erano messi in discussione da *AlphaGo*, successivamente battuta da *AlphaGo Zero*, un sistema nuovo di zecca, ma ancora più avanzato, ignaro delle regole del giuoco o di giuochi passati. Ciononostante, neanche a dirlo, vincente.

3. Le parole per dirlo

Quello che era apparso chiaro, nel giuoco Go, era come l'intelligenza artificiale potesse andare oltre le capacità degli esseri umani, anche nello svolgimento di attività complesse, senza ricevere sottostanti istruzioni, senza alcuna nozione di concetti base, senza nessuna programmazione, senza, cioè, che nessun essere umano avesse disciplinato e controllato e prevenuto la bizzarria del gesto ludico. L'idea rassicurante, umanamente rassicurante, che le macchine non sarebbero mai state in grado di sviluppare comportamenti, ruoli o funzioni tipiche dell'intelligenza umana – ciò che il giuoco degli scacchi sembrava in qualche modo testimoniare – divenne di colpo un'idea sorpassata, ponendo alla attenzione umana l'esistenza di nuove possibilità, ma anche di nuove pericolosità.

Nell'immagine del cambiamento, fra le nuove consapevolezze, bisognava dunque aggiungere anche una modernità di carattere terminologico, un rinnovamento del linguaggio, l'idea di dover superare o abbandonare, andare oltre la parola *intelligenza artificiale*[18], più volte usata fino a questo momento, e cominciare a pensare non più alla IA come una *semplice* branca dell'informatica, incaricata dello sviluppo di quelle tecniche in grado di simulare le capacità cognitive degli esseri umani, comunque al di sotto, o, al limite, affianco, ma mai sopra l'intelligenza umana, in termini cioè meramente deterministici, sulla base di regole predefinite, decodificate all'interno di un programma; semmai, invece, cominciare a pensare alla intelligenza artificiale – possiamo continuarla a chiamare così, per comodità semantica – nella più compiuta e sviluppata *machine learning*, un sistema di apprendimento dove è possibile, non soltanto simulare, ma generare

[18] Per IA – a rigore anche della definizione contenuta nella proposta di regolamento dell'aprile del 2021 – deve intendersi "un software sviluppato con una o più delle tecniche elencate nel I allegato della proposta e che può, per un dato insieme di obiettivi definiti dall'uomo, generare output come contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono".

autonomamente modelli statistici, tramite l'analisi delle banche dati incredibilmente estese e prese attraverso il dialogo fra sistemi esperti. La differenza è molto più di quanto la parola non riesca ad esprimere: le regole statistiche a cui la *machine learning* fa affidamento sono ricreate dal programma stesso, agevolando uno sviluppo ulteriore e dirompente dell'intelligenza artificiale tradizionalmente intesa, in modo da poter autonomamente riconoscere e stabilire regole che poi consentano di raggiungere gli obiettivi che il sistema artificiale intende perfezionare. Né più né meno che una forma di apprendimento, che consente alla macchina di progettarsi da sola, per ciò che non ancora realizzato, attraverso la creazione di modelli statistici che vengono continuamente aggiornati, attualizzati a quelle che sono le variabili apprese dall'esterno e interiorizzate nel sistema. Sempre meglio, e sempre più nel profondo della conoscenza.

Nel passaggio al *machine learning* e, successivamente, al *deep learning*, grazie a questa capacità di autoproclamazione, l'intelligenza artificiale supera la pura logica deterministica e diviene capace di risolvere problemi complessi in modo creativo e inaspettato, può prendere iniziative, può – e si è verificato nel 2013 sempre in una partita, questa volta al videogioco Tetris – decidere di adottare un comportamento sleale, pur di vincere la partita. Nella sospensione del giuoco, come ci ricorda la pellicola *Wargames*, il giuoco artificiale incontra la paura di perdere e decide di superarla attraverso l'artificio dell'imbroglio.

Cambia il nome delle cose, verrebbe di dire, l'internet delle cose, la *cartografia dei cambiamenti* si arricchisce di nuove parole: l'IA è sempre più un *Intelligenza Generativa*, prossima a divenire una *Superintelligenza*, antagonista, utopica, completamente priva di innocenza, indifferente alle relazioni domestiche che uniscono le parti in un tutto, risolutamente dedita alla parzialità, all'intimità e alla perversità[19]. E in questo intreccio di umano ed artificiale, di organico e cibernetico, di mente e corpo, autosviluppo e progettazione, nulla appare uguale a prima, nemmeno nelle parole: ingannevoli per natura[20], nella sospensione delle parole, il *meticcio* assomiglierà più ad un *fatticcio*[21], e l'*attore*, forse, più ad un

[19] Per una efficace descrizione del *cyborg* si rinvia a D.J. Haraway, nel saggio *Manifesto cyborg. Donne, tecnologie e biopolitiche del corpo*, trad. it., Feltrinelli, Milano, 2018, p. 42.

[20] La cortina di suggestioni retoriche trova spazio nelle parole che avvolgono l'intelligenza artificiale e la sua narrazione: in disparte quelle cinematografiche, cedevoli ad una lettura antropomorfizzata del sistema artificiale, l'inganno è nella parola intelligenza, nell'idea di oracoli, di allucinazioni, di incantesimi, parole queste che conducono – nel pensiero di G. Finocchiaro, *Intelligenza artificiale. Quali regole?*, il Mulino, Bologna, 2024, p. 25 – alla dimensione del mito ed allontana le soluzioni giuridiche sul tema delle responsabilità.

[21] L'espressione *fatticcio* – mutuata da B. Latour, *Il culto moderno dei fatticci*, trad. it., Meltemi, Milano, 2017, p. 70: "Unendo le due fonti etimologiche, – poco prima precisa che la parola 'fatto' rinvia alla realtà esterna, la parola 'feticcio' alle folli credenze del soggetto (ndr.) – noi chiamiamo fatticcio la robusta certezza che permette alla pratica di passare all'azione senza mai credere alla differenza tra costruzione e raccoglimento, immanenza e trascendenza" – ci consente di lasciare spazio alle due opposte teorie che immaginano, sul piano della soggettività delle macchine, la piena capacità di pensare e per essa lo status di persona (c.d. SAI nella intuizione di A. Turing), ovvero all'opposto l'assenza di una soggettività che renderebbe le macchine pari a delle cose, magari

attante[22], in una sublimazione di una semantica ibrida, intessuta di incertezze e di nuovi significati.

4. "Ho paura, Dave"[23]

Occorre in qualche modo rassegnarsi o prendere consapevolezza del fatto che, dopo le rivoluzioni industriali, la prima quella dei motori a vapore, la seconda dell'energia elettrica, e la terza delle tecnologie digitali, siamo alle soglie di una quarta rivoluzione industriale, dove le macchine aiuteranno gli uomini a fare calcoli complessi, senza neanche richiedere più le istruzioni, ma sulla base dei soli obiettivi, per il cui raggiungimento l'intelligenza artificiale sarà libera di scegliere il modo migliore possibile, riuscendo peraltro a fare meglio di quanto l'intelligenza umana non riuscirà a fare.

Non è possibile ignorare come l'affacciarsi del mondo cibernetico possa aggiungere una serie di rischi ai rischi che intende monitorare, e una serie di preoccupazioni, tra cui *bias* eventuali nell'*output* dei modelli, errori legati agli *input* (ad esempio una discriminazione razziale, o culturale)[24], problemi connessi alla gestione e protezione dei dati personali, pericoli di *cyber security*. Molto più che errori – tipo quello che portò l'algoritmo a censurare, ritenendola pornografica, una campagna antimurale al seno; o l'errore che portò a multare un personaggio popolare, il cui volto era ritratto su un autobus in movimento – il rischio che preoccupa la modernità artificiale – in disparte quello cinematografico del cosiddetto *complesso di Frankenstein*[25] – è quello di una autodeterminazione sfuggente, una proiezione finalistica anomala, attuata attraverso modelli comportamentali anomali.

indistinguibili nei comportamenti, ma oggetti sul piano morale e giuridico (c.d. WAI secondo la tesi di J.R. Searle). In mezzo la teoria c.d. negazionista (c.d. NAI) secondo cui nelle macchine non vi sarebbe alcuna capacità equiparabile al pensare o alla intelligenza umana: sul punto A. Contaldo, F. Campara, *op. cit.*, p. 79. Il rimescolamento della nozione di "soggetto" è sottolineato anche da A. Garapon, J. Lassègue, *op. cit.*, p. 251.

[22] La felice espressione risale a B. Latour, nel saggio del 1999 e ripresa da G. Teubner nel saggio *Ibridi e attanti. Attori collettivi ed enti non umani nella società e nel diritto*, trad. it., Mimesis, Milano, 2015, pp. 18 ss. Ce la ricorda anche A. Cappellini, "Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale", in *Criminalia*, 27/03/2019, p. 9, nota 24.

[23] Nel film di Stanley Kubrick, *2001. Odissea nello spazio*, la paura è quella di HAL, la macchina che riconosce i suoi errori, le sue decisioni discutibili, e percepisce l'intenzione di David di farla cessare, spegnerne il funzionamento. La paura nel film del 1968 è quella *della* macchina, la paura del nuovo è all'inverso, quella *dalla* macchina. Sulle paure del nuovo, dal nuovo, G. Finocchiaro, *op. cit.*, p. 15.

[24] I rischi delle nuove tecnologie su alcuni aspetti del vivere sociale, in particolare sul ruolo e sulla funzione delle donne, sono evidenziati da D.J. Haraway, *op. cit.*

[25] L'idea, cioè, spesse volte evidenziata nella rappresentazione cinematografica, di una macchina che si ribella al suo creatore.

In questa rincorsa al nuovo e al meglio, alla capacità della IA di migliorarsi e fare meglio ciò che l'intelligenza umana o la stupidità umana sa fare peggio, sarà ben possibile che, nella migliore decisione, trovi spazio quella che a noi potrebbe apparire la peggiore decisione possibile dal punto di vista etico; che l'autonoma e libera decisione si riveli disallineata con la tavola dei valori umani; che, una volta ammessa la libera scelta, nella sua attitudine al meglio, possa trovare spazio la possibilità, il rischio, di una decisione intimamente illecita, in altre parole, la possibilità di scegliere e completare una decisione illecita per il solo fatto di essere ritenuta la migliore possibile: una sorta di ossimoro, rischio-non-calcolato, ogniqualvolta almeno l'intelligenza artificiale, pur di raggiungere l'obiettivo finale, possa decidere di assumere un comportamento sleale, in un disallineamento dei valori che potrebbe nel futuro ripetersi sempre più frequentemente.

Il passaggio dalla IA debole (IA), alla IA forte (o AGI) fin su alla Superintelligenza conduce ad una trasmutazione dei vecchi rischi in nuovi rischi, un ammodernamento delle rischiosità, oltre il pensabile, oltre cioè la logica deterministica. Occorre, dunque, immaginare e imporre un approccio responsabile all'intelligenza artificiale, sapendo però che la mappatura dei rischi è complessa, tanto quanto l'individuazione di un rimedio che agisca al punto esatto in cui il rischio trova il suo spazio identitario e la sua attuazione mortifera, consapevoli che la penombra dei livelli algoritmici complica l'individuazione delle cause; entrambi rendono difficile la ricerca delle responsabilità.

5. Precop: sorvegliare e punire

Ad oggi, tutti i tentativi di disciplinare le responsabilità della persona artificiale si sono orientati in due direzioni: da un lato prevenire i rischi di attività illecite; dall'altro, disciplinare le conseguenze – prevalentemente civilistiche – di una attività illecita compiuta da un sistema artificiale.

L'approccio dell'UE all'intelligenza artificiale – in forme diverse da quello cinese, e dalla auto normazione statunitense[26] – si è mosso in entrambe le direzioni, e l'idea di dover e poter disciplinare lo sviluppo della IA e garantirne la sicurezza e il rispetto dei diritti fondamentali, attraverso norme e azioni concrete, ha conosciuto numerose norme e azioni, che compongo un quadro di cura e sottostante

[26] Di entrambi dà conto G. Finocchiaro, *op. cit.*, pp. 45-46 e nella parte conclusiva del saggio: di particolare interesse, per la riflessione penalistica che ne consegue, l'approccio statunitense, affidato all'autoregolazione dei grandi gruppi economici (Amazon, Anthropic, Google, Inflection, Meta, Microsoft e OpenAI) che scelgono di adottare le regole di disciplina. La condivisione delle regole con il mercato, da un lato conferma la rincorsa dei modelli regolativi alle innovazioni offerte dal mercato, e dunque il ruolo del mercato come principale *selfresolving* delle criticità evidenziate dal mercato medesimo, dall'altro la difficoltà per il mercato di pensare di dover sanzionare se stesso, quanto almeno all'applicazione di misure differenti da quelle civilistiche.

preoccupazione: fra le tante[27], in particolare, l'iniziativa dell'aprile 2021, vale a dire il proposto quadro di regolamentazione sull'intelligenza artificiale e la pertinente valutazione d'impatto. Disegnando un quadro giuridico europeo per l'IA che sappia difendere i diritti fondamentali e affrontare i rischi per la sicurezza specifici dei sistemi di IA, e che sappia attuare una revisione della legislazione settoriale in materia di sicurezza (ad esempio, regolamento sulle macchine, direttiva sulla sicurezza generale dei prodotti), l'iniziativa giuridica del 2021 tratteggia un quadro di responsabilità civile aggiornato ed adeguato all'era digitale e all'IA e modulato su quattro diversi livelli di rischio: rischio minimo, rischio elevato, rischio inaccettabile e rischio specifico di trasparenza. Introduce inoltre norme specifiche per i modelli di IA per uso generale.

[27] Questo il timeline delle iniziative europee, dal 2018 ad oggi: *Comunicato stampa: Gruppo di esperti sull'IA e alleanza europea sull'IA* (marzo 2018); *Comunicato stampa: Intelligenza artificiale per l'Europa*; *Comunicazione: Intelligenza artificiale per l'Europa*; *Documento di lavoro dei servizi della Commissione: Responsabilità per le tecnologie digitali emergenti*; *Dichiarazione di cooperazione in materia di intelligenza artificiale* (aprile 2018); *Lancio dell'alleanza europea per l'IA*; *Costituzione del gruppo di esperti ad alto livello sull'IA* (giugno 2018); *Commissione europea: Piano coordinato sull'IA*; *Commissione europea (comunicazione stampa): AI made in Europe; Comunicazione della Commissione europea: AI made in Europe*; *Consultazione delle parti interessate sul progetto di orientamenti etici per un'IA affidabile* (dicembre 2018); *Comunicazione della Commissione europea: Costruire fiducia nell'intelligenza artificiale incentrata sull'uomo*; *Gruppo di esperti ad alto livello sull'IA: Linee guida etiche per un'IA affidabile* (aprile 2019); *Prima Assemblea dell'Alleanza europea per l'IA*; *Gruppo di esperti ad alto livello sull'IA: Raccomandazioni politiche e di investimento dell'IA* (giugno 2019); *Gruppo di esperti ad alto livello sull'IA: Sperimentazione dell'elenco di valutazione dell'IA affidabile* (dicembre 2019); *Commissione europea: Libro bianco sull'IA: un approccio europeo all'eccellenza e alla fiducia*; *Consultazione pubblica su un approccio europeo all'eccellenza e alla fiducia nell'IA* (febbraio 2020); *Valutazione d'impatto iniziale: Requisiti etici e giuridici in materia di IA*; *Gruppo di esperti ad alto livello sull'IA: Elenco di valutazione finale sull'IA affidabile* (ALTAI); *Gruppo di esperti ad alto livello sull'IA*; *Raccomandazioni settoriali di un'IA affidabile* (luglio 2020); *2a Assemblea dell'Alleanza europea per l'IA* (ottobre 2020); *Commissione europea: Comunicazione sul tema Promuovere un approccio europeo all'IA*; *Commissione europea: Proposta di regolamento che stabilisce norme armonizzate in materia di IA*; *Commissione europea: piano coordinato aggiornato sull'IA, Commissione europea: Valutazione d'impatto di un regolamento sull'IA* (aprile 2021); *Consultazione pubblica sulla responsabilità civile – Adeguare le norme in materia di responsabilità all'era digitale e all'intelligenza artificiale*; *Commissione europea: Proposta di regolamento sulla sicurezza dei prodotti* (giugno 2021); *Banca centrale europea, Parere sulla legge sull'IA*; *Consiglio dell'UE: Testo di compromesso della presidenza sulla legge sull'IA*; *Conferenza di alto livello sull'IA: Dall'ambizione all'azione (3d Assemblea dell'Alleanza europea per l'IA)*; *Comitato economico e sociale europeo, Parere in merito alla legge sull'IA* (novembre 2021); *Comitato delle regioni, Parere in merito all'Atto sull'IA* (dicembre 2021); *Lancio del primo sandbox normativo per l'IA in Spagna: Portare avanti il regolamento sull'IA* (giugno 2022); *Proposta di direttiva sulla responsabilità dell'IA* (settembre 2022); *Orientamento generale del Consiglio sulla legge sull'IA* (dicembre 2022); *Posizione negoziale del Parlamento europeo sulla legge sull'IA* (giugno 2023); *Accordo politico sulla legge sull'IA raggiunto dai colegislatori* (dicembre 2023); *Pacchetto sull'innovazione dell'IA per sostenere le start-up e le PMI nell'intelligenza artificiale* (gennaio 2024); *Ufficio europeo per l'IA* (febbraio 2024).

La Proposta di regolamento è stata preceduta dalla Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni (COM(2021) 118 final), diffusa il 9 marzo 2021. Facendo seguito alla strategia digitale del febbraio 2020 (Commissione europea, *Libro bianco sull'intelligenza artificiale – Un approccio europeo all'eccellenza e alla fiducia* (COM(2020) 65 final), la Commissione europea ha delineato il programma per la politica digitale per il prossimo decennio. In particolare, è prevista una struttura di *governance* condivisa con gli Stati membri basata su un sistema di monitoraggio annuale da sviluppare intorno ai seguenti quattro punti cardinali: 1) cittadini dotati di competenze digitali e professionisti altamente qualificati nel settore digitale; 2) infrastrutture digitali sostenibili, sicure e performanti; 3) trasformazione digitale delle imprese: entro il 2030 tre imprese su quattro dovrebbero utilizzare tecnologie digitali, tra cui il 5G, l'Internet delle cose, l'*edge computing*, l'intelligenza artificiale, la robotica e la realtà aumentata per sviluppare nuovi prodotti, nuovi processi di fabbricazione e nuovi modelli commerciali basati su un'equa condivisione dei dati nell'economia dei dati; 4) digitalizzazione dei servizi pubblici, compresi i sistemi giudiziari e, in particolare, le attività di indagine e di contrasto in modo da far fronte a reati digitali sempre più sofisticati.

L'IA è senza dubbio una risorsa, ma anche una preoccupazione costante: da ultimo, nel gennaio 2024 "la Commissione – si legge nella pagina della Commissione europea – ha adottato la comunicazione AI@EC, che delinea strategie per migliorare le capacità proprie della Commissione nel campo dell'intelligenza artificiale (IA), sottolineando nel contempo l'importanza di un uso sicuro, trasparente e incentrato sull'uomo delle tecnologie di IA. Gli orientamenti (inclusi nella comunicazione), incoraggiano la Commissione ad adattare internamente, innovare e adottare l'IA fin da subito per dare un esempio di buone pratiche"[28], facendosi precedere, appena un mese prima, nel dicembre 2023, dopo

[28] L'IA è certamente un destino a cui l'UE non intende sottrarsi, convinta che promuovere l'eccellenza nell'IA rafforzerà il potenziale dell'Europa di competere a livello mondiale e che dunque debba consentirne lo sviluppo e l'adozione, lasciandola prosperare dal laboratorio al mercato; debba garantire che l'IA funzioni per le persone ed agisca come per il bene nella società; costruire una leadership strategica in settori ad alto impatto. Questi alcuni passaggi della comunicazione della Commissione europea sulla visione strategica per promuovere lo sviluppo e l'uso di sistemi di intelligenza artificiale (AI) legali, sicuri e affidabili nella Commissione europea (AI@EC): la Commissione si impegna a sostenere, accelerare e promuovere lo sviluppo e l'adozione di tecnologie AI affidabili all'interno della Commissione e, ove possibile, con partner del settore pubblico e privato, inclusi startup e innovatori. La Commissione segue un approccio basato sul rischio, stabilendo requisiti per i sistemi AI che pongono rischi elevati per la sicurezza e i diritti fondamentali e vietando alcune pratiche che contravvengono ai valori dell'UE. La Commissione si prepara a rispettare le regole previste dal futuro regolamento sull'intelligenza artificiale (*AI Act*). La Commissione introduce azioni organizzative e operative per garantire la capacità istituzionale e operativa di sviluppare e utilizzare tecnologie AI affidabili. Queste azioni riguardano la governance interna, le linee guida operative, la formazione del personale, la priorizzazione dei progetti e la cooperazione con altre istituzioni e Stati membri. La Commissione presenta lo stato dell'arte dei

una "maratona" di colloqui di 3 giorni fra la presidenza del Consiglio e i negoziatori del Parlamento europeo, da un accordo provvisorio sulla proposta relativa a regole armonizzate sull'intelligenza artificiale (IA), il cosiddetto regolamento sull'intelligenza artificiale. Il progetto di regolamento mira ad assicurare che i sistemi di IA immessi sul mercato europeo e utilizzati nell'UE siano sicuri e rispettino i diritti fondamentali e i valori dell'UE, fornendo regole sui modelli di IA per finalità generali ad alto impatto che possono comportare rischi sistemici in futuro, nonché sui sistemi di IA ad alto rischio; rivedendo il sistema di *governance* con alcuni poteri di esecuzione a livello dell'UE; ampliando l'elenco dei divieti, ma con la possibilità di utilizzare l'identificazione biometrica remota da parte delle autorità di contrasto negli spazi pubblici, fatte salve le tutele; offrendo una migliore protezione dei diritti tramite l'obbligo per gli operatori di sistemi di IA ad alto rischio di effettuare una valutazione d'impatto sui diritti fondamentali prima di utilizzare un sistema di IA. L'idea principale è, dunque, quella di regolamentare l'IA sulla base della capacità di quest'ultima di causare danni alla società seguendo un approccio "basato sul rischio": tanto maggiore è il rischio, quanto più rigorose sono le regole. Se da un lato i sistemi di IA che presentano solo un rischio limitato diventano soggetti a obblighi di trasparenza molto leggeri (ad esempio rendere noto che il contenuto è stato generato dall'IA, affinché gli utenti possano prendere decisioni informate in merito all'ulteriore utilizzo), d'altro canto, i sistemi di IA ad alto rischio vengono assoggettati ad una serie di requisiti e obblighi per ottenere accesso al mercato dell'UE (come ad esempio l'obbligo di registrarsi alla banca dati dell'UE per i sistemi di IA ad alto rischio; o di informare le persone fisiche quando sono esposte ad un sistema di riconoscimento delle emozioni; valutazione di conformità aggiornata alle novità tecnologiche) e tanto più nelle catene del valore complesse, dove si avverte maggiormente l'esigenza di rendere trasparenti l'assegnazione delle responsabilità e i ruoli dei vari attori in tali catene, in particolare dei fornitori e degli utenti di sistemi di IA. Per usi a rischio definito inaccettabile, l'elenco dei divieti contempla la manipolazione comportamentale cognitiva, lo *scraping* non mirato delle immagini facciali da Internet o da filmati di telecamere a circuito chiuso, il riconoscimento delle emozioni sul luogo di lavoro e negli istituti di istruzione, il punteggio sociale, la categorizzazione biometrica per dedurre dati sensibili, quali l'orientamento sessuale o le convinzioni religiose, e alcuni casi di polizia predittiva per le persone, con tuttavia possibili eccezioni a fini di attività di contrasto, permeabili ad usi di IA ad alto rischio che non abbiano superato la procedura di valutazione della conformità ma per particolari e disciplinate esigenze di contrasto (ad esempio la tutela di vittime di determinati reati, la prevenzione di minacce reali, presenti o prevedibili, come attacchi terroristici, e la ricerca di persone sospettate dei reati più gravi).

sistemi AI esistenti e previsti nella Commissione, identificando le aree in cui l'AI può essere più vantaggiosa per il personale e i risultati, e le capacità orizzontali che possono essere riutilizzate in vari settori e aree di applicazione.

Si tratta, certamente, del tentativo organico di attuare una nuova architettura di *governance*, non priva di indici di operatività concreti, fra i quali la creazione di un ufficio per l'IA all'interno della Commissione, incaricato di supervisionare i modelli di IA più avanzati, e contribuire a promuovere norme e pratiche di prova e far rispettare le norme comuni in tutti gli Stati membri e supportato da un gruppo scientifico di esperti indipendenti in merito ai modelli di IA per finalità generali, contribuendo allo sviluppo di metodologie per valutare le capacità dei modelli di base, fornendo consulenza sulla designazione e l'emergere di modelli di base ad alto impatto e monitorando i possibili rischi materiali di sicurezza connessi ai modelli di base.

Buone prassi, forum consultivi aperti ai portatori di interessi (rappresentanti dell'industria, PMI, start-up, società civile e mondo accademico) ma non solo: ecco la seconda direttrice, l'idea cioè che la complessità della modernità debba imporre nuove strategie e nuove cautele, ma anche nuove sanzioni, e cioè sanzioni pecuniarie per le violazioni del regolamento sull'IA in percentuale del fatturato annuo globale nell'esercizio finanziario precedente della società che ha commesso il reato o, se superiore, in un importo predeterminato (35 milioni di EUR, o il 7% per le violazioni relative ad applicazioni di IA vietate, 15 milioni di EUR o il 3% per violazioni degli obblighi del regolamento sull'IA e 7,5 milioni di EUR o l'1,5% per la fornitura di informazioni inesatte. Con possibilità di massimali più proporzionati per le sanzioni amministrative pecuniarie per le PMI e le start-up in caso di violazione delle disposizioni del regolamento sull'IA) e poteri di segnalazione quale espressione della strategia europea di *whistleblowing* da parte sia di una persona fisica che di una persona giuridica.

Sorvegliare e punire: la strategia della legislazione europea è, in sintesi, quella di aumentare la sorveglianza e prevedere sanzioni patrimoniali, per comporre un quadro di *relativa tranquillità*, in cui l'ideazione, la costruzione, la vitalità di un sistema di IA si conferma *rischiosamente irrinunciabile*.

6. La colpevolezza artificiale

Tranquillità-relativa e rischiosità-irrinunciabile, sono queste le ragioni dell'inquietudine logico dommatica, che danno corpo alle ragioni del dubbio: la prima, sistemica, giacché il quadro dei divieti non trova applicazione per aree sensibili, quali attività di ricerca, sviluppo e prototipazione che precedano l'immissione sul mercato; o per aree interessate esclusivamente a scopi militari, di difesa o di sicurezza nazionale, indipendentemente dal tipo di entità che svolge tali attività. La linea di confine potrebbe non essere robusta ed impedire elusioni di sistema e, del resto, se anche così non fosse, la nuova validazione di conformità in caso di modifiche sostanziali, o la nuova conformità ai nuovi requisiti al fine di ridurre al minimo i rischi, sembrano riproporre l'elasticità di già conosciuti sintagmi normativi – una sorta di *massima sicurezza tecnologicamente possibile* – di non

facile interpretazione: inevitabile che sia[29] la regolamentazione nascerebbe provvisoria, naturalmente cagionevole, inesorabilmente affidata alla lettura del caso, nel momento storico in cui *quel caso* viene ad esistenza, come dire, in una impredicibile distanza dalla irretroattività della disciplina. La seconda, umanamente o, forse meglio, artificialmente inevitabile: qualsiasi disegno organizzativo, qualsiasi disciplina normativa non può escludere attività fraudolente per aggirare le previsioni normative e la possibilità di inganno, ben oltre la soglia del malinteso.

Potremmo dirla così: nell'artificio dell'intelligenza, ben oltre il *giuoco della imitazione*[30], permane uno spazio per la capacità di *pensare*, fra cui anche assumere *pensieri* criminali[31]. Resta da capire quanto grande sia questo spazio, dal momento che l'illecito, che la vitalità artificiale custodisce, non è solo l'illecito della costruzione, della gestione del sistema, ma possiede natura poliedrica, sino a coinvolgere beni tradizionalmente presidiati non dalle sole sanzioni civilistiche, altrove ritenute non sufficienti. In sostanza dinanzi all'abuso di un punteggio sociale, per finalità pubbliche e private; allo sfruttamento delle vulnerabilità delle persone, all'utilizzo di tecniche subliminali; alla identificazione biometrica remota in tempo reale in spazi accessibili al pubblico, alla categorizzazione biometrica delle persone fisiche sulla base di dati biometrici per dedurne o desumerne la razza, le opinioni politiche, l'appartenenza sindacale, le convinzioni religiose o filosofiche o l'orientamento sessuale; alla polizia predittiva su singoli, al riconoscimento delle emozioni sul luogo di lavoro e negli istituti di istruzione; alla estrazione non mirata di immagini facciali da internet o telecamere a circuito chiuso per la creazione o l'espansione di banche dati, al rischio specifico per la trasparenza – per restare solo ad alcuni dei rischi immaginabili – si aggiunge un ventaglio dei rischi talmente ampio da essere paradossalmente incalcolabile, almeno per l'umana predittività.

[29] Lo scarto fra regolamentazione giuridica e dinamicità in progress trova conferma nella stessa disciplina europea, la proposta di Regolamento europeo per l'intelligenza artificiale, dettata per una visione delle cose immediatamente messa in crisi dall'irruzione del ChatGPT, ciò che rese necessario una immediata rivisitazione del testo elaborato per consentire di includervi la neonata applicazione della intelligenza generativa: G. Finocchiaro, *op. cit.*, p. 50.

[30] Nella versione offerta da Alan Turing, nel saggio "Computing Machinery and Intelligence", in *Mind*, 59 (1950), pp. 433-460, riproposto ora in D.R. Hofstadter, D.C. Dennet (a cura di), *L'io della mente*, trad. it., Adelphi, Milano, 1985, pp. 61 ss., il gioco consiste nel chiudere in due stanze separate un uomo e una donna, lasciando che siano interrogati da un terzo interrogante attraverso una qualche apparecchiatura telescrivente. L'interrogante può rivolgere domande a entrambe le stanze, senza però sapere chi le occupi, e dunque chi fornirà la risposta alla sua domanda. Lo scopo del giuoco è comprendere, attraverso le risposte, in quale stanza vi sia la donna e in quale l'uomo, sapendo che la donna, con le sue risposte, cercherà di aiutare l'interrogante, mentre l'uomo farà invece di tutto per confondere l'interrogante, dando risposte che, a suo parere, potrebbe dare la donna, e così tentare di trarlo in inganno. La versione aggiornata sostituisce l'uomo con un computer, cercando di riflettere sulla capacità del computer di sostituirsi e imitare i comportamenti – e cioè le risposte – dell'uomo.

[31] Ovviamente tutto dipende da cosa si intenda per pensare, cosa significhi comportarsi in modo intelligente e quanto il pensiero esprima questa intelligenza: J. Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*, LUISS University Press, Roma, 2017.

Ben oltre quelli attesi – come quello offerto alla attenzione dei giudici statunitensi chiamati a dirimere una causa intentata da Getty Images contro Stability IA in materia di tutela dei diritti d'autore sulle fotografie sui segni distintivi[32] o, in genere, le violazioni della privacy, dei diritti di autore, le geolocalizzazioni indesiderate, gli eccessi dei dispositivi bellici autonomi (c.d. *Lethal Autonomous Weapons*: LAWS), le incognite della guida autonoma, e così di seguito – o quelli apparentemente bizzarri (un sistema antropomorfo schiavizzato e schiavizzante, un *pet-robot* aggressivo, un *care-robot* che maltratta la persona anziana recalcitrante alla assunzione di un farmaco, una istigazione al suicidio come quella verificatasi in Belgio, l'identificazione errata di un professore completamente estraneo agli abusi sessuali ai quali invece un *chatbot* lo aveva accusato, inventando articoli di giornale mai pubblicati)[33], preoccupano i rischi inattesi, legati alla scelta autonoma degli obiettivi disallineati per via di sotto-obiettivi privi di linee guida (un obiettivo di vendita attraverso una pubblicità ingannevole) o, e forse più probabilmente, quelli legati alla causalità del caso, alla casualità del gesto artificiale: senza troppo cedere all'ingegneria affabulatoria si immagini un sistema di difesa domiciliare che, nell'intrusione di un ladro, sopravanzi i limiti della legittima difesa; o un misura farmacologica errata; un investimento o iniziativa economica azzardata portatrice di una crisi aziendale, etc.

Non ci pensiamo spesso, probabilmente ingannati dalla suggestione cinematografica di una superintelligenza, semmai malevola, ma non erronea. Eppure, il caso spesse volte traduce una indifferenza sovrana delle cose, una sorta di scaltro genio che prova piacere alle coniugazioni insolite, e *in tal caso* nessuna causalità oggettiva (e dunque nessuna valutazione cautelare) potrebbe mai porsi a base di una prevenzione immaginifica, tanto quanto sa esserlo un evento non ancora reale, e per giunta casualmente prodotto da un pensiero artificiale. Il caso stanca l'intelligenza umana, e rende impotenti la prevenzione e la regolazione delle cose artificiali, deprivate dei normali coefficienti di regolarità umana, estranei alla logica del maleficio artificiale.

[32] La controversia giudiziaria fra Getty Imagen, Inc. e Stability AI Ltd, riguardava l'accusa a Stability di aver violato, tramite l'apparato di IA "Stable Diffusion" le norme sui diritti di autore sulle fotografie e i segni distintivi appartenenti a Getty Images, di cui si sarebbero abusivamente appropriate. Circa 12 milioni di fotografie realizzate da Stability AI con un business concorrente con quello della agenzia fotografica con la banca dati più grande del mondo. L'algoritmo aveva in sostanza elaborato/trasformato le immagini derivate dal patrimonio di proprietà intellettuale della azienda di Seattle, e rese simili a quelle originarie (c.d. opere trasformative).

[33] Al caso di Jonathan Turley, professore di diritto della Georgetown University, incluso nella lista dei docenti accusati di avances e citato in un inesistente e mai pubblicato articolo del Washington Post, va aggiunto il caso di Brian Hood, neoeletto sindaco della contea australiana di Hepburn Shire, accusato di corruzione a causa di una deduzione errata che ChatGPT aveva formulato confondendo il suo lavoro in banca e la sua denuncia di casi di corruzione; o il caso di Mark Walters, conduttore radiofonico, accusato di una appropriazione di denaro sulla base di un errata ricostruzione compiuta sempre da ChatGPT. Le vicende sono narrate da G. Finocchiaro, *op. cit.*, p. 62.

L'accidente non è da meno, introduce una equipollenza che, distante dal caso, potrebbe tuttavia rendersi parimenti irrazionale, animata da una concatenazione che il *senno del poi* non avrebbe potuto contenere prima, senza mai riuscire ad avvicinarsi al determinismo delle cose umane. Un destino, potremmo dire, qualcosa in più dell'avvenimento indisciplinato e caotico, tuttavia ancor più sfuggente alla concatenazione deterministica. Nell'intelligenza artificiale si annida un raziocinio magico che non consente previsione, lungimiranza, alcuna anticipazione, ma solo, unicamente, una strategia fatale che disegna un ossimoro dogmatico al quale il sistema non può offrire riparo: l'intelligenza umana dovrebbe contenere l'analisi, la riflessione, la previsione delle concatenazioni fatali e delle coniugazioni segrete, ma si tratterebbe di una lusinga regolativa oggettivamente inimmaginabile, aperta finanche a perturbazioni esponenziali che amplificano una rischiosità non iniziale, non altrimenti prevedibile prima che venga ad esistenza.

Prevedere l'imprevisto, vale quanto prevedere l'improvviso, in una fatalità complessa che il sistema pretende di poter evitare, in anticipo sul tempo di ciò che li ha causati, ma che in realtà segue un tempo del-dopo, ciò che il tempo rallentato del-prima, quello umano, meno veloce e scaltrito di quello artificiale, non potrebbe mai riuscire ad evitare né forse immaginare.

Dinanzi a queste nuove forme di rischi artificiali, tuttavia reali e concreti, la strada della punizione patrimoniale, della responsabilità oggettiva o per fatto altrui, semplicemente distribuita, quanto alle conseguenze risarcitorie, all'interno dei fruitori del rischio e alla loro vicinanza ai vantaggi che detto rischio produce[34] potrebbe non essere sufficiente e reclamare l'aggiunta di una componente penalistica come in fondo per la analoga violazione di matrice umana, senza però che sia possibile sottrarre l'eventuale scelta punitiva a regole e principi irrinunciabili per il diritto penale. La scelta penalistica non sarebbe tuttavia così semplice, per la difficoltà di immaginare una responsabilità penale legata alla intelligenza artificiale, una responsabilità che, cioè, dalla *persona* artificiale scivoli verso quella umana, senza nel contempo garantire il rispetto del paradigma personologico della responsabilità penale, che necessita di essere traslato o forse rinnovato, attualizzato, nel gesto artificiale[35]. Dalla persona umana a quella artificiale, e da questa a quella, in una andata e ritorno che dovrebbe far salvi i principi di colpevolezza e personalità del gesto punitivo e che certamente non potrebbe affidarsi alla soluzione di una responsabilità oggettiva. L'alternativa non può essere fra cinismo delle regole e infedeltà ai principi.

[34] L'indifferenza al dolo, alla colpa, e persino all'errore è un rimedio di impronta essenzialmente civilistico, che va incontro ai profili risarcitori alle vittime da parte, non tanto, o non soltanto, dei creatori del rischio (ciò che potrebbe non essere così semplice individuare) ma, ancor più facilmente, da parte dei fruitori del rischi, coloro cioè che hanno abitato la situazione rischiosa nelle loro dinamiche mercatali, ricavandone utilità che poi sarebbero chiamate in parte a ridistribuire nel caso di danni volontari o involontari, persino imprevedibili. Sul tema G. Finocchiaro, *op. cit.*, pp. 68 ss.

[35] I principi che una opzione penalistica dovrebbe fare salvi hanno occupata la riflessione problematica di C. Piergallini, "Intelligenza artificiale: da 'mezzo' ad 'autore' del reato?", in *Rivista italiana di diritto e procedura penale*, 63 (2020), n. 4, pp. 1745 ss.

Occorre intendersi: nell'ipotesi di un calcolatore programmato alla delinquenza, di un gesto artificiale guidato da un gesto umano, la macchina assomiglia ad uno strumento attraverso il quale l'agente umano attua la commissione di reati, dei quali egli solo – evidentemente – non può che assumersi la responsabilità[36]. Il diritto penale potrebbe affidarsi ai tradizionali e molteplici scenari: addebitare il reato a chi ha generato il focolaio di rischio ideando l'algoritmo (specie se *self-learning*), o applicandolo nel programmare il "robot" o il PC; a chi ha attualizzato quel rischio, producendo e mettendo in commercio il sistema; ovvero a chi ha concretamente gestito quel rischio, servendosi dello stesso, o magari cooperando con esso, ove sussistano, ad esempio, eventuali profili di *culpa in interagendo*[37].

Il problema si situa, invece, nel caso in cui una immensa organizzazione coordinata di piccoli e infinitesimali calcoli riesca a contenere un sistema di tendenze, di desideri, di convinzioni, che la macchina lascerebbe emergere in forma "autonoma", lasciando ospitare fenditure di legalità capaci di emergere in un luogo e in un tempo imprevisto e impredicibile. Questa nuova categoria del *pensiero* artificiale conoscerebbe forme di volontà, di tentativo, di speranza, o di errore, di sbaglio, di colpa, anche solo per una scelta del male minore (ad esempio nelle auto *less-driver*, il sistema potrebbe dover scegliere un danno piuttosto che un altro, un pedone esterno piuttosto che un ospite interno alla autovettura)[38].

[36] A. Cappellini, *op. cit.*, p. 4. Una sintetica rassegna dei reati che possono essere commessi attraverso l'IA è in F. Basile, "Intelligenza artificiale e diritto penale. Quattro possibili percorsi di indagine", in *Diritto penale e Uomo*, 10 (2019), p. 26, e cioè i "crimini informatici, economici ed ambientali, i traffici internazionali di sostanze stupefacenti e di altri prodotti illeciti, la trattata di esseri umani" (sono queste, le ipotesi di reato riportate nel virgolettato, quelle menzionate nel documento di presentazione del "2019 OSCE Annual Police Experts Meeting: Artificial Intelligence and Law Enforcement: An Ally or an Adversary?") 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. Infine, ricordate sempre da F. Basile, le ipotesi di bagarinaggio *on line* e di manipolazione abusiva del mercato. In proposito anche R. Borsari, "Intelligenza Artificiale e responsabilità penale: prime considerazioni", in *MediaLaws*, (2019), n. 3, p. 263. Un particolare caso di un reato commesso da un robot è ricordato da C.R. Mongioì, G. Ren, G. Zanesco, B. Ziviani nel saggio "Intelligenza artificiale e diritto penale", in *Trento BioLaw Selected Student Papers*, 4 (2019), pp. 1-10, dove si racconta come nel 1981 un impiegato in un'azienda giapponese sia stato ucciso da un robot che stava lavorando di fianco a lui; il robot lo erroneamente identificò come una minaccia al suo lavoro e lo spinse verso un macchinario in funzione lì vicino, uccidendolo.

[37] Come ci ricorda V. Manes, "L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia", in *DisCrimen*, 2020, p. 3.

[38] Può non essere così. In un esperimento condotto nel Bristol Robotics Laboratory dinanzi alla scelta di quale proteggere fra due automi in pericolo, il terzo robot a causa della "trappola etica" ovvero della impossibilità di decidere, aveva invece scelto di lasciar morire entrambi. Se ne dà conto nel saggio di I. Mercuri, "Ingegneria sociale e rischi nell'interazione uomo – robot", in G. Costabile, A. Attanasio, M. Ianulardo (a cura di), *Memberbook 2021. Digital Forensics*, IISFA, Roma, 2021, pp. 59 ss. Il primo caso di donna uccisa a seguito di un investimento da un SUV a guida automatica si è verificato in Arizona, nel marzo del 2018: ce lo ricorda C. Cavaceppi, *op. cit.*, p. 133.

Non è, dunque, solo questione di una decisione che privilegi un bene a dispetto di un altro, o di una dolosa iniziativa illecita di una IA, resasi autonoma e capace di autodeterminazione illecita, l'attesa del domani sarà anche quella di una scelta che oltrepassi i cancelli del lecito, in un territorio di illegalità intelligente abitato non solo da scelte deliberate, dolose in linguaggio penalistico, ma anche semplicemente azzardate, incaute, indisciplinate, in una parola colpose. Il territorio della AGI – in disparte quello della Superintelligenza di *ExMachina*, preconizzato da *Il Pianeta delle Scimmie* – potrebbe conoscere un errore della chirurgia robotica, adiuvante o sostitutiva della mano umana, una sbagliata diagnostica computerizzata, una imprecisa rilevazione della retinopatia diabetica tramite la scansione dell'occhio, una scelta colposa nell'applicazione di una protesi, o di un farmaco piuttosto che un altro, un improprio dosaggio farmacologico deciso dall'algoritmo, un tempo rallentato dell'insulina sulla base del livello di zucchero nel sangue, o, in altre settore, una incauta o imprudente auto a guida autonoma, per restare ai settori (l'infortunistica medica e quella stradale) dove tradizionalmente si incontra il diritto penale della colpa – ai quali, come detto, potremmo aggiungere l'azzardo dell'algoritmo in una operazione commerciale o finanziaria, rivelatasi incauta e produttiva di un dissesto, ciò che l'intelligenza giuridica usa attrarre nel capitolo delle responsabilità commerciali – non governato da una supervisione umana resa improbabile, per difficoltà oggettive o anche solo per effetto mimetico e massivo, vale a dire la difficile adozione di un punto di vista individuale e non consolidato, reso improbabile dalla inanità del confronto rispetto ad uno standard decisionale artificiale e, dunque, più rassicurante e difficile da contrastare. Un po' come, per restare al settore sanitario, i protocolli medici come dispensa dalla colpa; o i modelli di organizzazione come guarentigia punitiva per le persone giuridiche. Tuttavia, errati, imprecisi, inefficaci, custodi a volte di errori clamorosi, sottratti al coraggio del dissenso[39]. In una parola, artificialmente colposi, ma umanamente incolpevoli, probabilmente inesigibili, e tuttavia astrattamente portatrici di responsabilità penali.

[39] Al coraggio del dissenso di taluni, e fra questi ci piace ricordare le critiche di Umberto Eco alle frequenti provvisorietà di Wikipedia, fa da contrappunto la pigrizia dei tanti, la gran parte di noi, nell'affidarsi a quanto si legge in rete, senza lo scrupolo di un approfondimento che non si avverte come necessario.

Il reato non è *con* l'intelligenza artificiale, ma è *della* intelligenza artificiale[40], il che pone il diritto penale[41], dinanzi a nuovi o sconosciuti scenari[42].

7. Dalla colpa di organizzazione alla colpa di programmazione

Non è, si potrebbe dire, in fondo, un problema del tutto nuovo, per il diritto penale, non è, cioè, la prima volta che il sistema punitivo si trova ad affrontare l'identità colpevole di una persona inumana. Superata il macigno della visione antropomorfa della reità penalistica, il nuovo sistema della responsabilità delle persone giuridiche si è a lungo interrogato sulla costruzione di un addebito punitivo in capo ad un soggetto altro, rispetto all'autore del fatto delittuoso. Ed è esattamente questa la strada seguita dalla giuridicizzazione della IA, oneri di *governance* attraverso i quali impedire la commissione di fatti illeciti.

Per le persone artificiali o, meglio, per le responsabilità artificiali l'identificazione della colpa è tuttavia operazione dogmatica più complessa, per certi versi invertita. Non è la persona giuridica che non ha impedito la colpa della persona fisica, ma, qui, la persona fisica che non ha impedito la colpa della persona artificiale. Il nodo dogmatico è apparentemente simile: la colpa si situerebbe nel mancato impedimento, nella incapacità dell'una, la persona giuridica, o dell'altra, la persona fisica, di impedire che si commettessero illeciti, che invece, entrambi, avrebbero potuto impedire; la persona giuridica attraverso regole cautelari sustanziali della colpa di organizzazione; la persona fisica attraverso una congerie di standard costruttivi, protocolli, cautele *in second best*, che avrebbero potuto

[40] Quello, ad esempio, raccontato da I. Salvadori, "Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale", in *Rivista italiana di diritto e procedura penale*, (2021), n. 1, p. 95 della *bot Random Darknet Shopper* che, programmata per fare acquisti *on line*, decise autonomamente di acquistare dieci pastiglie di ecstasy, poi fatte recapitare presso una galleria d'arte di Zurigo ed esposte al pubblico.

[41] L'interesse del diritto penale alla IA si muove tradizionalmente in quattro scenari, l'ultimo dei quali è quello decisamente più preoccupante. Questi gli scenari: 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) come ci ricorda F. Basile, *op. cit.*, pp. 4 ss. Ai quattro scenari, aggiunge due nuove prospettive M. Papa, "Future crimes: intelligenza artificiale e rinnovamento del diritto penale", in *Criminalia*, 2020 (*DisCrimen*, dal 04/03/2020), pp. 3 ss.

[42] Di recente I. Salvadori, *op. cit.*, pp. 83 ss. Quanto alle preoccupazioni derivate in tema di salvaguardia dei principi penalistici, si rinvia al saggio di F. Sgubbi, *Il diritto penale totale*, il Mulino, Bologna, 2019, e alla lettura che, del punto di vista del Maestro bolognese, effettua T. Guerini nel saggio "Il formante algoritmico all'alba della *Justice digital*. Uno sguardo a uno dei futuri possibili del diritto penale, dialogando a distanza con Filippo Sgubbi", in *DisCrimen*, 2021.

impedire la generazione artificiale di condotte illecite: una *colpa di programmazione*, assai simile alla *colpa di organizzazione*.

Semplice a dirsi, la costruzione delle cautele nasconde una complessità che supera l'apparente simmetria logico giuridica. Nelle persone giuridiche le intelligenze umane si sommano e confrontano con l'intelligenza umana che decide la commissione di un fatto illecito. Qui no, qui è diverso: l'intelligenza umana deve competere con l'intelligenza artificiale che, a tutto voler concedere, è più veloce di quella umana. È come se nella partita del facere illecito e della guida lecita si misurassero due giocatori con livelli di intelligenza differenti. La fraudolenta elusione – sintagma usato nel paradigma della 231 – qui avrebbe margini di probabilità altissimi, tanto quanto la manifesta superiore capacità dell'IA di aggirare le regole seppellite nel *black box* dell'algoritmo inziale: i metodi di contenimento, i metodi di incentivazione, l'impedimento dello sviluppo, il controllo sul *surplus* di modernità nelle forme dell'"*obsolescenza programmata*"[43], i *tripwires*, ossia meccanismi di diagnostica random, ad insaputa del sistema[44] potrebbero non essere sufficienti e l'IA potrebbe sapientemente ingannare i guardiani e uscire dalla scatola. Ecco il punto, uno almeno fra i tanti: l'assenza, l'impossibilità di una validazione, e per essa la difficoltà nel capire e prevenire la colpa dell'intelligenza artificiale e, in essa, la tenuta della visione costituzionale della colpevolezza deprivata di effettiva personalità ed esigibilità.

Non sarebbe facile prevederlo, ma sarebbe ancora meno facile impedirlo. Nel *machine learning*, e, ancor più, nel *deep learning*, nella possibilità cioè degli algoritmi di costruirsi e migliorarsi attraverso altri algoritmi, in una sequela di strati che si aggiungono al primo strato, quello più profondo, passando per strati superiori, sempre più in alto, la complessità dei livelli intermedi corre il rischio di far smarrire l'orientamento rispetto anche all'origine genetica, mutata al livello superiore, e poi ancora più in alto, in una zona d'ombra sempre più fitta e oscura, in un *black box* che rende di fatto impossibile vedere e valutare ciò che è avvenuto all'interno dei vari algoritmi, stratificati e intrecciati in diversi livelli[45]. A quale livello si è situato e generato la deviazione dalle regole, il cominciamento illecito, stratificato e scomparso nella profondità dell'apprendimento, non è semplice, forse neppure possibile, comprenderlo nemmeno per coloro che hanno progettato l'algoritmo iniziale, sviluppatosi molto, troppo più in là, rispetto al controllo umano, impossibilitato, ad un certo punto, a monitorare e seguire lo sviluppo dell'(auto)apprendimento e della (auto)determinazione. Senza dire, poi, della complessità dell'esatta individuazione del momento e del luogo in cui collocare il disallineamento fra le due intelligenze, quella artificiale e quella progettuale umana.

[43] E. La Rosa, "Obsolescenza programmata e diritto penale: una relazione complicata", in *Rivista trimestrale di diritto penale dell'economia*, 33 (2020), n. 3-4, pp. 568 ss.
[44] Per una spiegazione di questi metodi si rinvia a T. Tofanelli, "L'evoluzione della intelligenza artificiale tra etica e diritto", in G. Costabile, A. Attanasio, M. Ianulardo (a cura di), *op. cit.*, pp. 42 ss.
[45] G. Taddei Elmi, *op. cit.*, p. XVII.

Il tempo e lo spazio: le macchine moderne, i moderni congegni microelettronici quintessenziali, sono ovunque, sempre più invisibili, una beffa alla spiritualità umana e una sfida alla ubiquità territoriale, leggeri e precisi, portatili e mobili, l'artificio immateriale ed opaco resta difficile da vedere, in una continua sospensione tra posti lontani e indecifrabili. L'immensa *macchina metamorfica*[46] dipende dalla produzione, dai trasporti, dal lavoro fisico di persone sconosciute fra loro, da data center sparsi nell'oceano della rete, da cavi sottomarini fra continenti, da segnali di trasmissione che attraversano l'aria, da continui cicli di calcolo e, in questo perenne trasgredire i confini del qui ed ora, l'artificio uccide il tempo[47], ed abita luoghi distanti ed ubiqui[48]: entrambi, i luoghi e i tempi delle cose artificiali, impongono il balbettio del diritto penale, governato da un tempo *sudato*, intessuto nel sudario del dopo l'apocalisse del fatto illecito.

La pre-colpa, l'idea cioè di una *governance* per prevenire attività illecite, sconta questa evidente vulnerabilità logico dommatica: intuita già da attenti osservatori[49], sconta la oggettiva difficoltà, forse persino impossibilità, per l'intelligenza umana di esercitare la sorveglianza sui sistemi di IA, in particolare su quelli ad apprendimento automatico, capaci di compiere operazioni e prendere decisioni a tal punto complesse da sorpassare la capacità umana di controllo, "eccedendo non solo i loro limiti cognitivi, ma anche quelli temporali nell'accedere alle informazioni, elaborarle e prendere decisioni".

Il giudizio umano è, in sostanza, largamente sottodimensionato rispetto alla complessità dei modelli che guidano le scelte dei sistemi artificiali, il che rende poco plausibile pensare che una intelligenza umana, per quanto vivace, possa evitare l'elusione fraudolenta di un modello posta in essere da una intelligenza

[46] Così la definisce K. Crawford, *Né intelligente né artificiale. Il lato oscuro dell'IA*, il Mulino, Bologna, 2021, p. 56.

[47] Se l'uccisione del tempo può apparire esagerato, non lo è certamente il *True time* di Google, un protocollo temporale distribuito che funziona stabilendo gerarchie di attendibilità tra gli orologi locali di datacenter in modo che possano decidere come sincronizzarsi. Nella gestione del tempo di Google vi è la possibilità da parte di *True time* di gestire l'incertezza in caso di sfasatura temporale dei singoli server: se l'incertezza è grande, il tempo viene rallentato finché l'incertezza non si esaurisce, vale a dire una sorta di possibilità di rallentare il tempo, modificarlo a piacimento e unire il pianeta sotto un unico codice temporale proprietario. Né più né meo che una privatizzazione del tempo, la capacità di creare una linea temporale variabile sotto il controllo di un orologio master centralizzato. Se non ucciso, il tempo è perlomeno posseduto.

La privatizzazione del tempo è raccontata da K. Crawford, *op. cit.*, pp. 89 ss.

[48] La complessità degli strumenti e delle infrastrutture, e delle maestranze si lascia facilmente intuire nell'esempio tratto da K. Crawford, *op. cit.*, p. 212: dovendo installare un sistema di riconoscimento facciale per le strade di Belgrado, il capo della Polizia ha ordinato l'installazione di duemila telecamere in ottocento luoghi della città, e a tal scopo il governo serbo stipulò un accordo con Huawei per la fornitura dei servizi di videosorveglianza, supporto della rete 4G e di centri di dati e di controllo unificati. Il sistema nasceva cioè strutturalmente ibrido ed ubiquo, con infrastrutture cinesi, ma è quanto ordinariamente avviene anche con infrastrutture indiane e statunitensi dislocate nei paesi europei.

[49] G. Contissa, F. Galli, F. Godano, G. Sartor nel saggio "Il Regolamento europeo sull'intelligenza artificiale. Analisi informatico-giuridica", in *I-lex*, 2021, n. 3.

artificiale; che possa prevedere e scoprire l'astuzia fraudolenta di un sistema straordinariamente più agile e veloce; che un controllore umano possa controllare un controllato artificiale. Nondimeno, che un errore determinato da un sistema artificiale, sulla base di calcoli complicati e profondi, possa essere mentalmente anticipato ed evitato da un semplice controllo di concordanza, affidato ad una mente semplice a cui si affiderebbe la coraggiosa e forse inesigibile funzione di dissentire da una mente superiore. Anche quando l'uomo dovesse mantenere formalmente il controllo sulla decisione finale, la possibilità che egli sia in grado di entrare efficacemente nel merito della decisione si tradurrebbe in un'ipotesi remota, e più verosimilmente, il supervisore umano tenderà a fare affidamento sulle decisioni di un sistema di IA, tanto più se questo è stato certificato, a meno che non abbia motivi specifici per ritenere che esso sia malfunzionante, o che non sia in grado di incorporare nella sua valutazione ulteriori elementi, esterni al sistema stesso.

Nel giuoco degli addebiti, e nella dissoluzione dell'organico, il tempo dell'intelligenza umana è quello di ieri, quello cibernetico è il tempo dell'istante, un tempo nomade, del qui ed ora. Nell'attesa del giorno dopo, il ritardo punitivo è pressoché inevitabile.

8. La costruzione delle regole. La vigilanza artificiale

L'affido al diritto penale non è così scontato anche per altre ragioni, agli ostacoli epistemologici, vanno infatti aggiunte le ostilità tecnocratiche, l'IA come luogo del potere, tradizionalmente recalcitrante ai controlli e ancor più alle punizioni. Decisamente ostile a quelle penalistiche e più propenso ad un disegno di guarentigia penale, e a costrizioni, laddove davvero necessarie, semplicemente risarcitorie. Semmai dovesse non essere così, semmai il sistema dovesse aprirsi a regole punitive, la futura colpevolezza artificiale – ammesso che sia possibile chiamarla così – imporrà nuove regole e nuove misure. L'analisi dei cicli di vita del sistema, una sequela di verifiche continue, la possibilità di mettere a riposo il sistema non già – anche qui una necessaria novità sistemica – per l'obsolescenza del macchinario, ma, tutto al contrario, per la estrema e sopraggiunta modernità del macchinario, resosi capace di tale e tanta autonomia da non essere più controllabile. Ma tutto questo potrebbe non bastare, e certamente non per il diritto penale.

Dolo della macchina? oppure cooperazione colposa nel fatto doloso? errore determinato da fatto altrui? dolo di chi? e colpa in che modo? e *altrui* in che senso? Per non dire del *dove* e del *quando*, a cui si aggiunge, potrebbe aggiungersi – semmai dovessimo spingerci verso una personalità cibernetica – la complessità della scelta del gesto punitivo, ipoteticamente e anacronisticamente retributivo, sino al paradossale recupero di una bizzarra morte cibernetica: il diritto penale artificiale è tutto da costruire attraverso una forza immaginativa che deve saper cogliere il nuovo, anche a costo di scoprire la propria debolezza epistemologica e maturare l'idea della rinuncia, una fuga dalla padronanza punitiva.

Negli step a venire, man mano che ci si avvicinerà alla superintelligenza, aumenteranno i pericoli di un controllo incontrollabile, tuttavia nei primi momenti di questa era artificiale occorrerebbe provare ad adattare il diritto punitivo a nuovi standard di colpevolezza umanizzante, sapendo però che le tradizionali categorie non possono funzionare, non si adattano alla modernità artificiale, e impongono un cambio di paradigma, che sappia apprendere diversamente il *qui* e l'*altrove*, l'*attore* e l'*attante*, il *fatto* e il *fatticcio*, il *proprio* e l'*altrui*. In breve, adattarsi ad una nuova e inattesa topografica delle ambiguità.

E qui il paradosso è dietro l'angolo. Se si vuole davvero insistere per una responsabilità penale, e se non si vuole abbandonare l'idea di una responsabilità personale – ciò che il sistema risarcitorio mostra di prediligere – la migliore e più efficace forma di controllo non potrà che essere quella dell'IA che controlla se stessa, in un *myse en abyme* dove il controllore, controlla il controllato, finché un altro controllore, più evoluto, non controllerà il controllore, in un seguito senza fine, tanto quanto l'inevitabile rincorsa alla modernità artificiale. La personalità della responsabilità artificiale non può che essere, per il momento, quella di non essere riusciti ad impedire che l'IA faccia quello che *avremmo potuto* impedire di fare e in questo nuovo, semmai futuribile, *diritto penale sinestetico*, per dirla con Michele Papa, la personificazione potrebbe assomigliare ad una evidente finzione, la colpa assumere una dimensione ibrida, a metà strada fra il dolo dell'attante e la evitabilità dell'attore. E il cuore del diritto penale potrebbe essere l'errore, l'ultimo della serie, non necessariamente quello causativo dell'evento.

Quanto durerà questo "*avremmo potuto*" è la vera incognita del futuro che *ci* attende. Nel *ci attende*, nella speranza cioè di una prossimità non a noi estranea, si annida l'immanenza di un *presto*, che potrebbe diventare un *subito*. Ma forse anche un *mai*, come per i Topi di Autun[50], giacché il diritto penale dell'intelligenza – occorre ricordarlo – dovrebbe poter risiedere nel tempo del *prima*. E non è detto che possa riuscirvi e, se così fosse, tanto varrebbe rinunciarci.

[50] Nell'anno 1522 i topi furono citati dinanzi il tribunale di Autun in Borgogna, accusati di aver mangiato e distrutto il raccolto del circondario. Non doveva essere un fatto nuovo: Bartolomeo Chassenee, giurista del luogo chiamato a patrocinare la causa, ricordò la maledizione divina contro il serpente, la maledizione di Gesù contro l'albero dei fichi sterile di Bretania, le scomuniche a passeri, lumache, anguille e persino ad un intero frutteto. Alla fine, il tribunale decise di rinviare la sentenza fino a quando non vi fosse certezza che i topi avessero avuta notifica della causa, e così i topi, di fatto, vinsero la causa.