UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI GIURISPRUDENZA
Pubblicazioni del Dipartimento di Scienze Giuridiche "Cesare Beccaria"

39

INTELLIGENZA ARTIFICIALE

Diritto, giustizia, economia ed etica

a cura di
FABIO BASILE, MARCO BIASI, LUCIO CAMALDO
GAIA CANESCHI, BEATRICE FRAGASSO, DANIELA MILANI

G. Giappichelli Editore

UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI GIURISPRUDENZA
Pubblicazioni del Dipartimento di Scienze Giuridiche “Cesare Beccaria”
39

Comitato di direzione

Direttore:
Prof.ssa Daniela Milani
(Ordinario di Diritto e Religione)

Coordinatore:
Prof. Alessandro Boscati
(Ordinario di Diritto del lavoro)

Componenti:
Prof. Ennio Amodio
(Emerito di Diritto Processuale penale)
Prof. Fabio Basile
(Ordinario di Diritto penale)
Prof. Paolo Di Lucia
(Ordinario di Filosofia del diritto)
Prof. Emilio Dolcini
(Emerito di Diritto penale)
Prof. Vincenzo Ferrari
(Emerito di Filosofia del diritto)
Prof. Gian Luigi Gatta
(Ordinario di Diritto penale)
Prof. Mario Jori
(Emerito di Filosofia del diritto)
Prof. Claudio Luzzati
(Emerito di Filosofia del diritto)
Prof. Luca Micheletto
(Ordinario di Economia politica)
Prof. Carlo Enrico Paliero
(Emerito di Diritto penale)
Prof. Jlia Pasquali Cerioli
(Ordinario di Diritto e Religione)
Prof. Mario Pisani
(Emerito di Diritto Processuale penale)
Prof. Gaetano Ragucci
(Ordinario di Diritto tributario)
Prof. Matteo Rescigno
(Ordinario di Diritto commerciale)
Prof.ssa Daniela Vigoni
(Ordinario di Diritto Processuale penale)

INTELLIGENZA ARTIFICIALE

Diritto, giustizia, economia ed etica

a cura di

FABIO BASILE, MARCO BIASI, LUCIO CAMALDO
GAIA CANESCHI, BEATRICE FRAGASSO, DANIELA MILANI

G. Giappichelli Editore

© Copyright 2025 - G. GIAPPICHELLI EDITORE - TORINO
VIA PO, 21 - TEL. 011-81.53.111
http://www.giappichelli.it

ISBN/EAN 979-12-211-1381-5
ISBN/EAN 979-12-211-6290-5 (ebook)

Volume pubblicato con fondo Unico Dipartimentale - Assegnazione 2025 – Dipartimento di Scienze Giuridiche "Cesare Beccaria", codice F_DOTAZIONE_2025_DIP_004.

Stampa: Stampatre s.r.l. - Torino

Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume/fascicolo di periodico dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941, n. 633.

Le fotocopie effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi, Centro Licenze e Autorizzazioni per le Riproduzioni Editoriali, Corso di Porta Romana 108, 20122 Milano, e-mail autorizzazioni@clearedi.org e sito web www.clearedi.org.

INDICE

pag.

INTELLIGENZA ARTIFICIALE E GIUSTIZIA PENALE

INTELLIGENZA ARTIFICIALE E DIRITTO

PRESENTAZIONE

di Daniela Milani

L'intelligenza artificiale (IA) si colloca indubbiamente tra le invenzioni più rivoluzionarie della storia, anche recente, segnata dall'avvento di Internet, prima, e dello smartphone, poi.

Per l'impatto in grado di generare può essere analogicamente assimilata alla rivoluzione industriale che, grazie all'introduzione delle catene di montaggio, ha notevolmente sviluppato l'efficienza dei sistemi di produzione riducendo tempi e costi. Ma soprattutto, ha modificato l'organizzazione del lavoro, generando altrettante ricadute di ordine economico e sociale.

Al pari dell'avvento della catena di montaggio la rivoluzione digitale promossa dall'IA si candida – grazie all'automazione di processi complessi, all'analisi di grandi quantità di dati e alla creazione di soluzioni innovative – a generare cambiamenti dalla portata eccezionale. Sino al punto da avviare un nuovo sistema di pensiero, il cosiddetto "Sistema 0", che si affianca ai due modelli già esistenti: il Sistema 1, proprio del pensiero intuitivo e il Sistema 2, più analitico e riflessivo. Rispetto a questi due modelli il Sistema 0 si configura come una collaborazione tra intelligenza artificiale ed essere umano, che non dovrebbe annullare completamente le capacità di riflessione e giudizio di quest'ultimo, bensì esternalizzare alcuni processi mentali, come l'analisi di dati complessi, la pianificazione e la predizione, con l'innegabile vantaggio di liberare risorse cognitive da riservare ad altre attività.

Fatto salvo questo indubitabile beneficio non si può sottacere il rischio che un uso non corretto dell'AI possa degenerare in forme di "dipendenza cognitiva". O per meglio dire, il rischio che l'intelligenza artificiale dal fornire un mero supporto esterno passi a integrare ogni fase del processo decisionale, riducendo o compromettendo l'autonomia cognitiva dell'essere umano. Specialmente per quanti, nati tra il 2010 e il 2025 (la cosiddetta generazione Alpha), vivono il Sistema 0 come una realtà normale, fin dalla primissima infanzia.

In tale scenario, a un tempo affascinante e complesso, il diritto è chiamato a svolgere un compito fondamentale: garantire uno sviluppo e un utilizzo dell'IA etico, sicuro e responsabile.

Questa funzione del diritto ben si evidenzia nei contributi pubblicati all'interno del presente volume che raccoglie gli atti del IV Convegno annuale del Dipartimento di Scienze giuridiche "Cesare Beccaria" su "Intelligenza artificiale: diritto, giustizia, economia ed etica". Già a partire dall'analisi che Giovanni Ziccardi propone dell'Artificial Intelligence Act, il regolamento varato dall'UE nel giugno del 2024 al fine di stabilire norme coerenti con i valori dell'Unione a tutela delle persone fisiche, delle imprese, della democrazia, dello Stato di diritto e dell'ambiente.

La tensione tra i vantaggi derivanti dall'impiego dell'intelligenza artificiale e l'esigenza di garantire il rispetto dei diritti fondamentali attraversa tutti i contributi oggetto di pubblicazione, assumendo forme puntuali nelle singole discipline: dal diritto penale al diritto processuale penale; dal diritto del lavoro al diritto commerciale e tributario; dal diritto privato comparato al diritto ecclesiastico. Così come condivisa e ferma è negli autori la convinzione che la supervisione dell'uomo sia indispensabile per consentire un impiego dell'AI giuridicamente orientato al rispetto dei principi, dei valori e dei diritti fondamentali.

Pur sapendo che le riflessioni sono solo agli inizi e che nuovi sviluppi ci attendono in futuro, i contributi raccolti in questo volume si candidano a divenire un punto di riferimento per le riflessioni e i dibattiti interni ai singoli settori disciplinari, con l'ulteriore e innegabile pregio di svilupparsi all'interno di un contesto dialettico e interdisciplinare.

Un particolare ringraziamento va dunque agli organizzatori del IV Convegno annuale del Dipartimento di Scienze giuridiche "Cesare Beccaria": Gaia Caneschi, Fabio Basile, Lucio Camaldo, Marco Biasi e Beatrice Fragasso, che si sono prodigati anche nella raccolta degli atti. Un analogo ringraziamento va agli autori dei singoli contributi che hanno generosamente contribuito alla buona riuscita del convegno e di questo importante volume.

LE AUTRICI E GLI AUTORI

Baccetti Niccolò, Professore Ordinario di Diritto commerciale – *Università degli Studi di Milano.*

Biasi Marco, Professore Associato di Diritto del lavoro – *Università degli Studi di Milano.*

Camaldo Lucio, Professore Associato di Diritto processuale penale – *Università degli Studi di Milano.*

Caneschi Gaia, Ricercatrice di Diritto processuale penale – *Università degli Studi di Milano.*

Castano Silvana, Professoressa Ordinaria di Informatica – *Università degli Studi di Milano.*

Cianitto Cristiana, Professoressa Associata di Diritto e Religione – *Università degli Studi di Milano.*

Charness Gary, già Professore Ordinario di Economia politica – *Università della California Santa Barbara.*

Di Domenico Alessia, Dottoressa di ricerca di Diritto processuale penale – *Università degli Studi di Milano.*

Fasola Marco, Assegnista di ricerca di Diritto tributario – *Università degli Studi di Milano.*

Fragasso Beatrice, Assegnista di ricerca di Diritto penale – *Università degli Studi di Milano.*

Grieco Daniela, Professoressa Associata di Politica economica – *Università degli Studi di Milano.*

Mantovani Letizia, Dottoranda di ricerca di Diritto processuale penale – *Università degli Studi di Milano.*

Milani Daniela, Professoressa Ordinaria di Diritto e Religione – *Università degli Studi di Milano.*

Rescigno Matteo, Professore Ordinario di Diritto commerciale – *Università degli Studi di Milano.*

Ruffini Renato, Professore Ordinario di Organizzazione aziendale – *Università degli Studi di Milano.*

TIMELLINI CATERINA, Professoressa Associata di Diritto del lavoro – *Università degli Studi di Milano*.

TOTARO DAVIDE L., Assistant Professor di Diritto contrattuale comparato – *Università Hitotsubashi di Tokyo*.

VASTA VALENTINA, Assegnista di ricerca di Diritto processuale penale – *Università degli Studi di Milano*.

ZICCARDI GIOVANNI, Professore Ordinario di Informatica giuridica – *Università degli Studi di Milano*.

RELAZIONI INTRODUTTIVE

GIUSTIZIA DIGITALE E INTELLIGENZA ARTIFICIALE: IL PROGETTO *NEXT GENERATION UPP*

di Silvana Castano

1. *Introduzione*

La legge svolge un ruolo cruciale in quasi tutti gli aspetti della nostra vita, sia pubblica che privata. Migliaia di documenti legali vengono costantemente prodotti da enti istituzionali, come Parlamenti e Tribunali, e costituiscono una fonte primaria di informazione e conoscenza, principalmente per giudici, avvocati e altri professionisti del diritto coinvolti nei processi decisionali legali, ma anche per soggetti generici come cittadini o organizzazioni pubbliche e private. Sapere come orientarsi in un contesto così complesso, sia nella sua struttura che nei suoi contenuti, è un'esigenza fondamentale per diverse categorie di utenti: per i professionisti del diritto, a supporto delle loro attività; per gli amministratori, per applicare le procedure legali; e per gli utenti generici/cittadini, per favorire un'esplorazione e un utilizzo efficace delle informazioni legali [1].

La disponibilità di tecnologie e strumenti di intelligenza artificiale per l'estrazione di conoscenza dai documenti legali non è solo auspicabile, ma dunque necessaria [2]. I benefici e i risultati concreti derivanti dalla diffusione di tale tecnologia sono numerosi e variegati, sia per i professionisti del

[1] H. Surden, *Artificial intelligence and law: An overview*, in *Georgia State University Law Review*, 35, 2019, pp. 19-22.

[2] A. Santosuosso, *Intelligenza artificiale e diritto*, Mondadori Università, Milano, 2020.

diritto (ovvero avvocati, giudici e tribunali), sia per le amministrazioni e gli utenti finali. La ricerca giuridica attraverso l'estrazione della conoscenza legale è estremamente importante. Ad esempio, la ricerca giuridica sulla giurisprudenza precedente può essere utile a un avvocato per individuare una decisione emessa in un caso simile a quello in esame, in cui il tribunale si è pronunciato in modo favorevole alla posizione del proprio cliente, oppure una decisione resa in un caso diverso sulla base di un ragionamento che, applicato al caso attuale, porterebbe a un'interpretazione favorevole per il cliente. Quando si conduce una ricerca sulla giurisprudenza, è importante concentrarsi non solo sulla decisione del caso, ma anche sulla motivazione e sul ragionamento (detto "*rationale*") alla base della sentenza. In questo processo, i sistemi di estrazione della conoscenza possono fornire un grande aiuto, specialmente se sono "*context-aware*" (consapevoli del contesto).

Nel contesto delle decisioni amministrative, l'estrazione della conoscenza dai documenti legali potrebbe aiutare le amministrazioni pubbliche a identificare la normativa applicabile a un caso specifico, garantendo una conoscenza approfondita e sempre aggiornata di qualsiasi legislazione rilevante, inclusa quella più specifica. L'estrazione di conoscenza potrebbe anche essere utilizzata per automatizzare, almeno in parte, alcuni processi amministrativi, considerando che in molti paesi europei si è diffuso il principio del "*digital only*", permettendo così anche l'uso di tecnologie di ricerca giuridica[3]. Dal punto di vista degli utenti generici/cittadini, lo sviluppo di strumenti di estrazione della conoscenza potrebbe promuovere trasparenza, accessibilità ed equità nel sistema giuridico, fornendo ai cittadini informazioni e risorse preziose. Facilitando l'accesso ai documenti legali e giuridici, come leggi, decisioni giudiziarie e procedimenti amministrativi, si offrirebbe ai cittadini una migliore comprensione dei propri diritti e delle opportunità disponibili.

Per gestire in modo efficace il crescente volume, la complessità e

[3] D.U. Galetta, G. Pinotti, *Automation and algorithmic decision-making systems in the italian public administration*, CERI-DAPdoi:10.13130/2723-9195/2023-1-7, URL https://ceridap.eu/automation-and-algorithmic-decision-making-systems-in-the-italian-public-administration/; J.P. Schneider, F. Enderlein, *Automated decision-making systems in german administrative law*, CERIDAPdoi:10.13130/2723-9195/ 2023-1-102, URL https://ceridap.eu/automated-decision-making-systems-in-german-administrative-law/; E. Gamero Casado, *Automated decision-making systems in spanish administrative law*, CERIDAPdoi:10.13130/2723-9195/2023-1-119. URL https://ceridap.eu/automated-decision-making-systems-in-spanish-administrative-law/; J. Reichel, *Regulating automation of swedish public administration*, CERIDAPdoi:10.13130/2723-9195/2023-1-112, URL https://ceridap.eu/regulating-automation-of-swedish-public-administration/.

l'articolazione dei documenti legali, le amministrazioni e le organizzazioni a livello nazionale e internazionale stanno dedicando un crescente impegno all'attuazione di processi di trasformazione digitale e di soluzioni evolute per l'estrazione di conoscenza basate su *Natural Language Processing* (NLP), *Machine Learning* (ML) e Intelligenza Artificiale (AI), in grado di affrontare le sfide poste dalla documentazione giuridica, come la complessità del linguaggio, la lunghezza significativa dei testi legali, la scarsa accessibilità ai dataset giuridici – che rende difficile o addirittura impossibile il download su larga scala – e la mancanza di corpora annotati sufficientemente ampi per l'addestramento dei modelli. In questo contesto, si inserisce il progetto *Next Generation UPP* con l'obiettivo di fornire un contributo alla trasformazione digitale del Sistema Giustizia italiano promuovendo nel contempo l'accrescimento nel sistema universitario delle competenze digitali specifiche per il sistema giudiziario.

2. *Il progetto* Next Generation UPP

Per Il progetto *Next Generation UPP: nuovi schemi collaborativi tra Università e uffici giudiziari per il miglioramento dell'efficienza e delle Prestazioni della giustizia nell'Italia nord-ovest*, si inserisce nel PON Governance e Capacità Istituzionale 2014-2020, con particolare riferimento all'Azione 1.4.1 "Azioni di miglioramento dell'efficienza e delle prestazioni degli Uffici Giudiziari attraverso l'innovazione tecnologica, il supporto organizzativo alla informatizzazione e telematizzazione degli Uffici Giudiziari, disseminazione di specifiche innovazioni e supporto all'attivazione di interventi di change management".

Il progetto *Next Generation UPP*, che si è svolto dal 1/4/2022 al 30/9/2023, si propone di migliorare prestazioni della giustizia nell'Italia nord-ovest, di sperimentare nuovi schemi collaborativi tra le università e gli uffici giudiziari in modo da offrire agli addetti all'ufficio del processo (UPP) skill trasversali per garantire l'efficace funzionamento di un moderno sistema giurisdizionale e di fornire supporto al processo di digitalizzazione e innovazione tecnologica.

Ad esempio, l'UPP ha bisogno di avere accesso al patrimonio giurisprudenziale e legislativo per la valorizzazione del patrimonio digitale esistente che raggiunge volumi molto significativi (ad esempio, circa 34 milioni di provvedimenti digitali e 6 milioni di atti telematici depositati da avvocati e altri professionisti nel processo civile telematico dal luglio 2014 al dicembre 2020). Un obiettivo primario è quindi fare in modo che il pa-

trimonio documentale esistente sia accessibile sottoforma di dataset esplorabili con tecniche di estrazione di conoscenza, di ricerca semantica, di *document building* e di massimazione.

L'ambito territoriale di riferimento del progetto *Next Generation UPP* comprende gli uffici giudiziari del Nord-Ovest corrispondenti alla Macro Area 01, quindi alle corti d'appello di Brescia, Genova, Milano, Torino e ai tribunali dei relativi distretti. Al progetto partecipano le 12 università pubbliche della macroarea, ovvero l'Università degli studi di Torino (capofila), l'Università degli studi di Brescia, l'Università degli studi di Bergamo, l'Università degli studi di Genova, l'Università degli studi di Insubria, l'Università degli studi di Milano-Bicocca, l'Università degli studi di Milano Statale, l'Università degli studi di Pavia, il Politecnico di Milano, il Politecnico di Torino, l'Università degli studi del Piemonte Orientale e l'Istituto Universitario di Studi Superiori di Pavia.

Il progetto è organizzato su quattro linee di intervento e relative azioni mirate a: effettuare una ricognizione relativa al funzionamento degli Uffici per il processo già avviati e al funzionamento nel contesto in cui non risultino attivi Uffici per il processo (azioni 1.1. e 1.2); definire un catalogo delle attività e delle procedure per l'attivazione ed il potenziamento degli Uffici per il processo (azione 1.3); individuare nuovi modelli per la gestione dei flussi in ingresso e per la gestione dell'arretrato civile (azione 2.1); istituire una task force per il miglioramento dell'efficienza e della produttività degli uffici giudiziari (azione 3.1); proporre nuovi modelli formativi e nuovi schemi collaborativi tra Università e Uffici Giudiziari nel contesto dei corsi di laurea in discipline giuridiche e dell'offerta post lauream (azione 4.1).

L'Università degli Studi di Milano ha partecipato al progetto con un gruppo di assegnisti e docenti del Dipartimento di Informatica "Giovanni degli Antoni" (coordinatrice prof.ssa Silvana Castano) e un gruppo di assegnisti e docenti dei tre dipartimenti di area giuridica - Dipartimento di Diritto Pubblico Italiano e Sovranazionale, Dipartimento di Diritto Privato e Storia del Diritto, e Dipartimento di Scienze Giuridiche "Cesare Beccaria" (coordinatrice prof.ssa Laura Salvaneschi).

Il presente contributo descrive le attività che, nell'ambito dell'azione 1.3, sono state svolte in collaborazione tra l'unità operativa del Dipartimento Informatica dell'Università degli Studi di Milano (d'ora in poi UNIMI-Informatica) e l'unità operativa IUSS Pavia e i relativi risultati. Le attività previste in generale dalla linea 1.3 riguardavano aspetti diversi che andavano dalla anonimizzazione semiautomatica delle decisioni giudiziarie, agli strumenti di analisi semantica delle sentenze, dalla classificazione automatica dei documenti del processo all'utilizzo di tecniche di intelligenza

artificiale per la gestione dei precedenti, dalla creazione di database giurisprudenziali all'integrazione fra i dati del Ministero e la mole di dati presenti nelle singole corti, fino alla creazione di chatbot per l'interazione con l'utenza.

Le attività svolte da UNIMI-Informatica e IUSS Pavia si sono focalizzate sulla digitalizzazione dell'attività del giudicare e, in particolare, sullo sviluppo di tecniche di intelligenza artificiale per il supporto alla scrittura dei provvedimenti giudiziari (strumento *Document Builder*), con particolare attenzione alla sezione della motivazione, al fine di ridurre i tempi di decisione e di redazione della sentenza.

3. *Il* Document Builder

Il *Document Builder* è uno strumento che assiste e supporta il giudice nella creazione dei provvedimenti giudiziari, con particolare attenzione alla generazione dei contenuti della sezione motivazionale, sfruttando tecniche di intelligenza artificiale.

La creazione di provvedimenti giudiziari e, più in generale, di documenti legali è un processo volto alla produzione di un documento testuale seguendo uno schema predefinito con il supporto di strumenti digitali e automatizzati. Questo compito, anche quando viene svolto interamente da esseri umani, può essere considerato come una sequenza di tre fasi: (1) definizione del formato e della struttura del documento; (2) redazione del contenuto di ogni parte o sezione; (3) modifica e revisione del documento. Ciascuna di queste fasi può beneficiare in misura variabile del supporto di strumenti automatizzati, che spazia da approcci interattivi che prevedono l'intervento umano a servizi completamente automatizzati. Ad esempio, la struttura del documento può essere basata in tutto o in parte su modelli predefiniti, eventualmente adattati alle esigenze dell'utente. Allo stesso tempo, la fase di modifica può essere supportata soluzioni che vanno dal semplice rilevamento degli errori alla riformulazione automatizzata del contenuto testuale fino alla generazione di testo.

L'architettura di *Document Builder* sviluppata nel progetto *Next Generation UPP* è progettata per supportare esperti di ambito giuridico nella redazione di provvedimenti giudiziari (e.g., sentenze) che richiedono (o trarrebbero beneficio da) una struttura ben definita. Un requisito fondamentale nella progettazione di un *Document Builder* per la generazione di provvedimenti giudiziari è quello di preservare il ruolo fondamentale della figura del giudicante, fornendo quindi un supporto automatizzato senza

compromettere l'autonomia decisionale e la capacità di valutazione e giudizio del giudicante nella stesura dei documenti. Di conseguenza, un approccio completamente automatizzato alla generazione di una sentenza, ad esempio, non sarebbe né adeguato né auspicabile: è quindi necessario fornire un ambiente di *document builder* interattivo che mantenga l'esperto giuridico continuamente coinvolto nel processo di generazione dei contenuti. Questo contribuisce anche a ridurre il rischio di un'eccessiva standardizzazione dei provvedimenti giudiziari generati, qualora ci si affidasse ad esempio a strumenti di generazione automatica dei contenuti, che rappresenta un aspetto critico in considerazione della funzione e del valore di tali documenti.

La progettazione del *Document Builder*, condotta nell'ambito della linea 1.3 del progetto *NEXT Generation UPP*, ha visto una stretta collaborazione tra le due istituzioni accademiche UNIMI-Informatica e IUSS Pavia, le quali hanno sviluppato attività distinte ma fortemente interconnesse e complementari. In particolare, l'attività di IUSS Pavia si è concentrata sulla progettazione di un format di provvedimento decisorio (e connessi atti dei difensori) con caratteristiche informatiche tali da essere in grado di ricevere i dati provenienti dalle diverse fonti del Processo Civile Telematico (e relativi registri) e organizzarli in modo appropriato nell'atto decisorio così da renderli ricercabili e utilizzabili per successive applicazioni di machine learning e estrazione di conoscenza[4]. L'attività di UNIMI-Informatica ha riguardato lo sviluppo del *proof-of-concept* di *Document Builder* che, sfruttando il format di provvedimento decisorio elaborato da IUSS Pavia, è in grado di fornire le necessarie funzionalità per supportare la generazione di una nuova sentenza, utilizzando tecniche di intelligenza artificiale per la composizione della parte motivazionale del provvedimento, attraverso la ricerca di contenuti testuali giurisprudenziali provenienti da precedenti simili/analoghi e utili alla decisione vera e propria.

Le attività svolte dalle due istituzioni accademiche hanno richiesto una modalità di lavoro fortemente interdisciplinare con il coinvolgimento e la stretta collaborazione di assegnisti reclutati nell'ambito del progetto in ambito giuridico, in ambito informatico e in ambito di linguistica computazionale sotto la supervisione dei docenti responsabili.

Lo scopo complessivo del *Document Builder* proposto è quello creare

[4] A. Santosuosso, S. D'Ancona, E. Furiosi, *New-generation templates facilitating the shift from documents to data in the Italian judiciary, in Proc. of 2nd Int. Workshop on Digital Justice, Digital Law, and Conceptual Modeling (JUSMOD23) Lecture Notes in Computer Science 14319*, Springer, Cham, 2023.

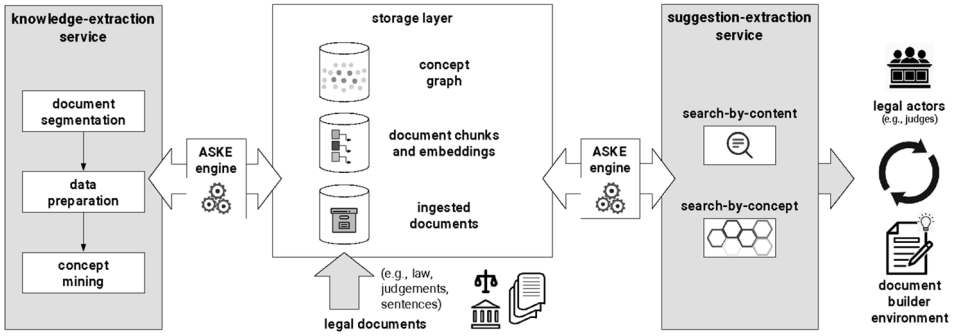
uno strumento operativo di immediato aiuto per gli addetti all'ufficio del processo nella loro attività di studio dei casi e di predisposizione di bozze di provvedimenti, che possa contribuire a una maggiore produttività e a una migliore qualità delle decisioni, al fine di agevolare il giudice nello svolgimento del compito più difficile e di più alto pregio intellettuale e professionale, ovvero quello di selezionare il materiale e la proposta di percorso motivazionale e sfidarla, cambiando o precisando alcuni parametri di interrogazione, o alcuni elementi di fatto e di diritto che contraddicono e cambiano la consequenzialità della proposta del sistema.

3.1. *Architettura del* Document Builder

Il *Document Builder* è pensato per integrarsi e non sovrapporsi agli strumenti già a disposizione del giudice; sfruttando il format di provvedimento decisorio elaborato a cura di IUSS Pavia, le sezioni iniziali della nuova sentenza possono essere popolate ricevendo i dati provenienti dalle diverse fonti del Processo Civile Telematico (e relativi registri), sfruttando ad esempio, funzionalità esistenti nella Consolle del giudice. Le funzionalità basate su intelligenza artificiale del *Document Builder* entrano in gioco nella fase di scrittura della sezione motivazione del provvedimento.

L'architettura del *Document Builder* è mostrata in Figura 1 [5].

Figura 1 – Architettura del *Document Builder*



[5] S. Castano, A. Ferrara, S. Montanelli, S. Picascia, D. Riva, *A Knowledge-Based Service Architecture for Legal Document Building*, in *Proc. of 2nd Int. Workshop on Knowledge Management and Process Mining for Law (KM4LAW), Quebec, Canada*, July 2023.

Il *Document Builder* supporta il giudice nella scrittura della sezione motivazione di una nuova sentenza, fornendo suggerimenti "intelligenti" – ovvero le porzioni di testo più rilevanti/significative per il caso in questione, provenienti da precedenti pertinenti alle ricerche formulate dal giudice – che possono essere importati nell'area di lavoro per essere riusati direttamente e/o per essere personalizzati, integrati e/o raccordati fra loro al fine di formulare la versione finale della motivazione da redigere.

L'interfaccia del *Document Builder* è organizzata in due aree: un'area di lavoro principale in cui l'utente effettua la stesura della motivazione e un'area di suggerimenti testuali forniti dallo strumento in risposta alle ricerche formulate dall'utente per il recupero di testi utili alla stesura della motivazione e alla decisione vera e propria. In particolare, l'utente, ovvero l'autore del documento, inizia il lavoro selezionando un modello tra quelli disponibili e viene assistito nella redazione attraverso suggerimenti forniti dallo strumento sotto forma di blocchi di testo (i c.d. *document chunk*) pertinenti, che possono essere riutilizzati "così come sono" oppure modificati per essere inseriti nella sezione del documento in fase di redazione. I suggerimenti vengono estratti da un corpus di documenti giuridici (ad esempio, sentenze, decisioni giurisprudenziali) precedentemente acquisiti e processati. A tal fine, l'architettura del *Document Builder* prevede una piattaforma di archiviazione per la memorizzazione di: un *database* (*corpus*) di documenti giuridici pre-esistenti acquisiti con i relativi contenuti testuali; i blocchi di testo (*document chunk*) estratti dai documenti e i relativi *embeddings* per la loro classificazione; il grafo dei concetti estratti dai documenti. Il database documentale alla base del *Document Builder* prevede quindi una archiviazione delle sentenze/documenti giuridici con segmentazione delle stesse in sezioni e classificazione delle sezioni risultanti secondo il format di provvedimento decisorio. A ciò si aggiunge la classificazione dei documenti giuridici a livello di singole frasi/paragrafi in base ai concetti estratti [6].

Il *Document Builder* fornisce due servizi principali: il servizio di estrazione della conoscenza e il servizio di estrazione dei suggerimenti.

Il *servizio di estrazione della conoscenza* elabora i documenti del corpus al fine di estrarre un insieme di concetti rappresentativi che forniscono una descrizione tematica dei loro contenuti testuali. I concetti estratti dai documenti vengono organizzati in un grafo, in cui i concetti simili sono colle-

[6] V. Bellandi, S. Castano, S. Montanelli, D. Riva, S. Siccardi, *A Service Infrastructure for the Italian Digital Justice*, in *Proc. of 15th Int. Conf. on Management of Digital Ecosystems (MEDES 2023)*, Springer CCIS, May 2023.

gati mediante archi. Ogni concetto è inoltre associato ai blocchi di testo del/i documento/i da cui è stato estratto, per consentire di recuperare i blocchi di testo pertinenti per un determinato concetto.

Il servizio di estrazione della conoscenza, denominato ASKE, si basa sull'uso di *Large Language Models* (LLM) particolarmente adeguati ad analizzare semanticamente il contenuto testuale dei documenti giuridici, anche elaborati e complessi, indipendentemente dalla forma sintattica utilizzata. Ciò permette non solo di definire quanto due testi sono simili, ovvero trattano degli stessi argomenti, ma consente anche di estrarre da essi i concetti principali rappresentativi del contenuto e utilizzare questi ultimi per indicizzare i documenti stessi a una granularità fine, a livello di blocchi di testo (e.g., singole frasi, paragrafi), abilitando la capacità di fornire come suggerimenti testuali per la composizione della sezione motivazionale frammenti di testo specifici in luogo di intere sentenze[7].

Il *servizio di estrazione dei suggerimenti* fornisce due modalità per individuare i blocchi di testo più pertinenti/rilevanti da proporre all'utente per la redazione della sezione motivazionale del documento:

i) Mediante una *funzionalità di ricerca per contenuto*, l'utente inserisce una query a testo libero, ovvero una frase o locuzione che meglio esprime la questione giuridica di interesse. Come risultato della ricerca, il servizio di estrazione fornisce una lista di testi motivazionali più rilevanti, tratti dalle sentenze del corpus semanticamente simili/pertinenti al testo della query.
ii) Attraverso una *funzionalità di ricerca per materia e concetto di interesse*, l'utente può recuperare frammenti testuali di tipo motivazionale provenienti da precedenti giuridici che risultano maggiormente pertinenti/rilevanti per gli scopi del provvedimento giuridico in via di definizione. Entrambe queste funzionalità di ricerca sfruttano tecniche di elaborazione del linguaggio naturale basate su LLM sul corpus di documenti giurisprudenziali per eseguire un'analisi semantica delle sentenze nel corpus al fine di reperire i frammenti testuali, ovvero i blocchi di testo, effettivamente pertinenti all'oggetto della ricerca.

Per ogni blocco di testo fornito come suggerimento, l'utente ha la possibilità di visualizzare il testo integrale del documento giuridico di provenienza; questa viene considerata una funzionalità essenziale per consentire

[7] S. Castano, A. Ferrara, E. Furiosi, S. Montanelli, S. Picascia, D. Riva, C. Stefanetti, *Enforcing Legal Information Extraction Through Context-Aware Techniques: the ASKE Approach*, in *Computer Law and Security Review*, 52, 2024.

all'utente giudicante di visionare la collocazione del frammento restituito come risultato della ricerca nel contesto dell'intera sezione motivazionale. L'utente ha quindi la possibilità di importare nell'area di lavoro tutti i frammenti testuali che ritiene utili ai fini della stesura della motivazione, utilizzandoli direttamente come citazioni oppure modificandone il contenuto e lavorando ai punti di raccordo fra gli stessi, per arrivare alla migliore formulazione finale della motivazione per la sentenza in questione.

Caratteristica distintiva del *Document Builder* è l'approccio *human-in-the-loop*, che permette all'utente giudicante di esercitare la propria libertà decisionale e il pieno controllo della formulazione motivazionale del provvedimento, selezionando i contenuti testuali proposti dallo strumento, cambiando o precisando alcuni parametri di interrogazione, o alcuni elementi di fatto e di diritto che contraddicono e cambiano la consequenzialità della proposta fornita dallo strumento. Per questo motivo, il *Document Builder* utilizza tecniche di intelligenza artificiale non generativa, ma con le stesse capacità di rappresentazione del significato del testo propria dei *Large Language Models*, al fine di effettuare l'analisi semantica dei documenti ed estrarre concetti e significati da testi giuridici, elaborati e complessi, indipendentemente dalla forma sintattica utilizzata.

3.2. *Applicazione a un caso di studio*

Come esempio di applicazione, nel progetto *Next Generation UPP* è stato definito un caso di studio relativo all'ambito della concorrenza sleale, composto da 50 sentenze che sono state annotate e segmentate secondo il modello di provvedimento decisorio fornito da IUSS Pavia, andando a costituire un dataset di allenamento per un algoritmo di segmentazione automatica delle sentenze.

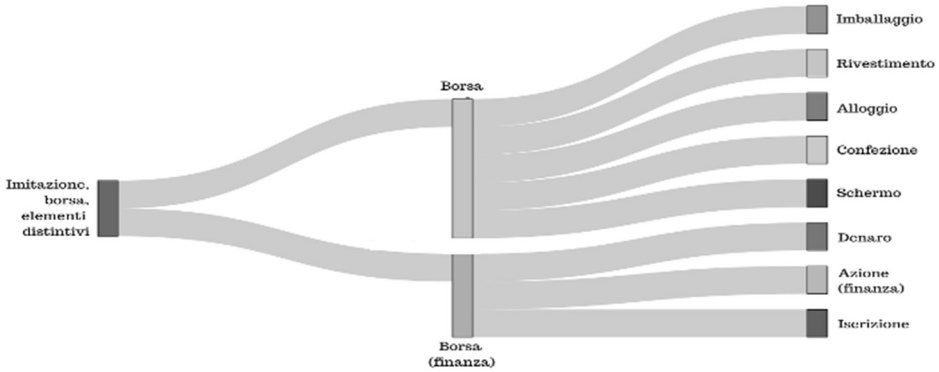
I segmenti delle sentenze sono stati analizzati dall'algoritmo ASKE (*Automated System for Knowledge Extraction*), sviluppato da UNIMI-Informatica e utilizzato per estrarre i concetti rilevanti dai testi giuridici [8]. I concetti estratti sono utilizzati per la classificazione sia delle sentenze sia dei blocchi di testo (*document chunk*) e per la comparazione semantica tra diversi contenuti testuali, che non si basa sull'occorrenza di semplici termini, bensì sul contesto in cui i termini vengono impiegati. In Figura 2 sono riassunti i dati relativi ai concetti estratti e ai blocchi di documento clas-

[8] S. Castano, A. Ferrara, E. Furiosi, S. Montanelli, S. Picascia, D. Riva, C. Stefanetti, *Enforcing Legal Information Extraction Through Context-Aware Techniques*, cit.

sificati utilizzabili per effettuare ricerche di suggerimenti da proporre nella costruzione di una nuova sentenza in ambito concorrenza sleale.

- 50 sentenze civili concorrenza sleale
- 352 concetti, 1552 termini
- circa 4 termini per concetto
- circa 600/800 *chunk* per concetto

Figura 2 – Esempio di grafo dei concetti



Ringraziamenti

Il presente contributo si basa sulle pubblicazioni prodotte in collaborazione con il gruppo di progetto dell'Università degli Studi di Milano e dell'Istituto Universitario di Studi Superiori di Pavia. Un sentito ringraziamento agli assegnisti reclutati sul progetto per il loro prezioso e insostituibile contributo. Un ringraziamento particolare al gruppo informatico del Laboratorio ISLab del Dipartimento di Informatica per il fondamentale contributo allo sviluppo del Document Builder e uno speciale ringraziamento al professor Santosuosso per la proficua e stimolante collaborazione.

Bibliografia

Bellandi V., Castano S., Ceravolo P., Damiani E., Ferrara A., Montanelli S., Picascia S., Polimeno A., Riva D., *Knowledge-Based Legal Document Retrieval: A Case Study on Italian Civil Court Decisions*, in *Proc. of the 1st Int. Workshop on*

Knowledge Management for Law (KM4LAW) – EKAW (Companion) CEUR-WS, Bozen, Italy, 2022.

Bellandi V., Castano, S., Montanelli S., Riva D., Siccardi S., *A Service Infrastructure for the Italian Digital Justice*, in *Proc. of 15th Int. Conf. on Management of Digital Ecosystems (MEDES 2023)*, Springer CCIS, May 2023.

Castano S., Ferrara A., Furiosi E., Montanelli S., Picascia S., Riva D., Stefanetti C., *Enforcing Legal Information Extraction Through Context-Aware Techniques: the ASKE Approach*, in *Computer Law and Security Review*, 52, 2024.

Castano S., Ferrara A., Montanelli S., Picascia S., Riva D., *A Knowledge-Based Service Architecture for Legal Document Building*, in *Proc. of 2nd Int. Workshop on Knowledge Management and Process Mining for Law (KM4LAW)*, Quebec, Canada, July 2023.

Galetta D.U., Pinotti G., *Automation and algorithmic decision-making systems in the italian public administration*, CERI-DAPdoi:10.13130/2723-9195/2023-1-7, URL https://ceridap.eu/automation-and-algorithmic-decision-making-systems-in-the-italian-public-administration/.

Gamero Casado E., *Automated decision-making systems in spanish administrative law*, CERIDAPdoi:10.13130/2723-9195/2023-1-119, URL https://ceridap.eu/automated-decision-making-systems-in-spanish-administrative-law/.

Reichel J., *Regulating automation of swedish public administration*, CERIDAPdoi:10.13130/2723-9195/2023-1-112, URL https://ceridap.eu/regulating-automation-of-swedish-public-administration/.

Santosuosso A., *Intelligenza artificiale e diritto*, Mondadori Università, Milano, 2020.

Santosuosso A., D'Ancona S., Furiosi E., *New-generation templates facilitating the shift from documents to data in the Italian judiciary, in Proc. of 2nd Int. Workshop on Digital Justice, Digital Law, and Conceptual Modeling (JUSMOD23) Lecture Notes in Computer Science 14319*, Springer, Cham, 2023.

Schneider J.P., Enderlein F., *Automated decision-making systems in german administrative law*, CERIDAPdoi:10.13130/2723-9195/ 2023-1-102, URL https://ceridap.eu/automated-decision-making-systems-in-german-administrative-law/.

Surden H., *Artificial intelligence and law: An overview*, in *Georgia State University Law Review*, 35, 2019, pp. 19-22.

UNA LETTURA DELL'ARTIFICIAL INTELLIGENCE ACT: NORME, ETICA, ADEMPIMENTI, ATTUAZIONE

di GIOVANNI ZICCARDI

1. *L'avvento dell'Artificial Intelligence Act*

L'Artificial Intelligence Act (d'ora in avanti: "AI Act") ha rappresentato un passo epocale nella regolamentazione dell'intelligenza artificiale all'interno dell'Unione Europea; è, senza dubbio, il più importante atto legislativo dell'intera Legislatura che è terminata lo scorso anno, con una forte valenza anche *politica*, e non solo normativa.

La sua genesi, e il suo sviluppo, sono il risultato di un lungo e articolato processo di analisi e riflessione, volto a creare un quadro normativo capace di bilanciare le esigenze dell'innovazione tecnologica con la necessità di tutelare i *diritti fondamentali* dei cittadini europei [1].

[1] Per una panoramica introduttiva dal taglio europeo si veda, in lingua inglese, N.T. NIKOLINAKOS, *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies – The AI Act*, Cham, Springer, 2023. Per un primo approccio, sempre in lingua inglese, al tema e alla norma, si veda P. VOIGT, N. HULLEN, *The EU AI Act: answers to Frequently Asked Questions*, Springer, Berlin-Heidelberg, 2024. Per un'introduzione informatico-giuridica all'intelligenza artificiale ormai "classica" ma ancora un punto di riferimento, si veda G. SARTOR, *Intelligenza artificiale e diritto: un'introduzione*, Giuffrè, Milano, 1996. In lingua italiana, invece, si vedano: M. BIASI (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffrè Francis Lefebvre, Milano, 2024; G. VACIAGO (a cura di), *Intelligenza artificiale generativa e professione forense*, Giuffrè Francis Lefebvre, Milano, 2024; M.G. PELUSO, *Intelligenza artificiale e tutela dei dati*,

La crescente diffusione di sistemi di intelligenza artificiale (soprattutto negli ultimi due anni, con l'avvento di ChatGPT e la consapevolezza diffusa di una *dipendenza tecnologica esclusiva* dell'Europa da società nordamericane o cinesi) ha sollevato questioni etiche, legali e sociali, chiamando un intervento normativo che potesse garantire un utilizzo responsabile di tali tecnologie.

I primi segnali di un "interesse istituzionale" nei confronti della regolamentazione dell'intelligenza artificiale sono emersi nel 2018, quando la Commissione Europea ha iniziato ad adottare una *linea politica* (e una conseguente *strategia*) molto chiara dedicata a questo ambito, riconoscendo l'importanza di promuovere lo sviluppo e l'adozione dell'IA in Europa, ma con un'attenzione particolare alla tutela dei valori fondanti dell'Unione.

Tale strategia si proponeva di consolidare la competitività europea nel settore creando, al contempo, un ecosistema "di fiducia" che potesse favorire l'accettazione delle tecnologie da parte dei cittadini e delle imprese. Molta enfasi venne data, sin dall'inizio, all'idea di *trust* e di *trustworthy*, ossia di tecnologia percepita, da molti, come pericolosa, ma che potesse al contrario ispirare sempre più *fiducia* in tutti i cittadini.

In quello stesso anno venne istituito il "Gruppo di Esperti di Alto Livello sull'Intelligenza Artificiale" (di cui si parlerà nel Paragrafo seguente), un organismo composto da accademici e rappresentanti dell'industria e della società civile, con il compito di elaborare linee guida etiche per il corretto sviluppo (e utilizzo) dell'intelligenza artificiale.

Questo gruppo di lavoro ha pubblicato, nel 2019, le "Linee guida etiche per un'IA affidabile", un documento di fondamentale importanza che delinea i principi essenziali per uno sviluppo tecnologico rispettoso della dignità umana e che sarà incorporato nella parte iniziale (considerando 27) dell'AI Act.

Tra i principi individuati, tipici della tradizione dell'Unione Europea, emergono il rispetto della dignità e non discriminazione della persona, dello stato di diritto, della privacy e della protezione dei dati, dell'autonomia individuale, la prevenzione del danno e l'equità e la trasparenza, elementi ritenuti imprescindibili per garantire una crescita sostenibile e socialmente accettabile delle tecnologie basate sull'intelligenza artificiale.

Parallelamente, la Commissione Europea avviò una consultazione pubblica e approfondì l'analisi del contesto tecnologico, sociale ed economico dell'IA: nel 2020 venne, a tal fine, pubblicato un "Libro bianco sull'intelli-

---

Giuffrè Francis Lefebvre, Milano, 2023; O. Pollicino, F. Donati, G. Finocchiaro, F. Paolucci, *La disciplina dell'intelligenza artificiale*, Giuffrè Francis Lefebvre, Milano, 2025.

genza artificiale – Un approccio europeo all'eccellenza e alla fiducia", che segnava un ulteriore avanzamento verso l'elaborazione di un quadro normativo strutturato.

Questo documento proponeva un approccio regolatorio basato sulla *valutazione del rischio*, in base al quale i sistemi di IA dovevano essere sottoposti a differenti livelli di controllo e verifica a seconda del loro potenziale impatto sulla società. L'obiettivo principale era quello di individuare e mitigare i rischi più gravi, senza ostacolare l'innovazione e lo sviluppo tecnologico.

Come è certamente noto al lettore, l'approccio basato sul rischio adottato dall'AI Act è stato oggetto di *numerose critiche*, soprattutto per la sua rigidità, la difficoltà di attuazione e il rischio di frenare l'innovazione. Si tratta, d'altro canto, di un approccio che era già stato adottato in altri provvedimenti precedenti (si pensi al GDPR, il regolamento sulla protezione dei dati) e che aveva dato discreti frutti.

Esistevano, ovviamente, numerosi approcci alternativi che avrebbero potuto essere considerati nella creazione di una normativa per l'intelligenza artificiale. Ognuno di questi modelli presenta, comunque, vantaggi e svantaggi che determinano sempre un diverso equilibrio tra innovazione, tutela dei diritti fondamentali e sicurezza.

Si pensi, *in primis*, a un approccio basato sui *principi* (di solito definito "principle-based regulation"), ossia un modello che definisca una serie di principi generali a cui gli sviluppatori e gli utilizzatori di sistemi di IA si devono attenere, lasciando ampio margine di interpretazione e adattabilità (un esempio di questo approccio è, come noto, il quadro normativo degli Stati Uniti d'America, che si basa su *linee-guida flessibili* piuttosto che su regole rigide).

I vantaggi sarebbero stati, probabilmente, una maggiore adattabilità ai rapidi sviluppi tecnologici e un minore impatto sull'innovazione, oltre a evitare il rischio di una regolamentazione obsoleta a causa della velocità con cui evolve l'IA. Al contempo, però, un approccio basato sui principi può risultare troppo vago e di difficile applicazione concreta, lasciando spazio a interpretazioni divergenti e a una scarsa capacità di *enforcement*.

Ancora, si pensi a un possibile approccio *settoriale* (quella che viene definita comunemente "sector-specific regulation"): invece di classificare i sistemi di IA in base al rischio, si adotta un modello che regolamenta l'intelligenza artificiale in base al *settore di applicazione*, come avviene in alcuni territori come gli Stati Uniti d'America o la Cina. Ciò, forse, avrebbe portato a una maggiore precisione nel regolare l'IA nei settori critici (sanità, finanza e sicurezza pubblica), senza imporre vincoli eccessivi su settori

meno sensibili, ma avrebbe portato, allo stesso tempo, un forte rischio di *frammentazione normativa* (rischio che l'Unione Europea, attraverso l'uso di un *Regolamento*, voleva assolutamente evitare…), con la conseguente necessità di sviluppare molteplici regolamenti settoriali, rendendo assai difficile la gestione di tecnologie trasversali come l'IA.

Al contempo, un approccio basato sulla responsabilità (definito "accountability-based regulation") sarebbe stato un modello capace di porre un'enfasi maggiore sulla *responsabilità legale* di sviluppatori e utenti dell'IA, obbligandoli a dimostrare il rispetto di determinati criteri di sicurezza ed etica, con una maggiore flessibilità rispetto a un approccio rigido e con la possibilità per le aziende di adattare le proprie soluzioni senza sottostare a normative prestabilite. Allo stesso tempo, però, l'efficacia sarebbe dipesa dall'esistenza di *meccanismi di controllo* e di sanzioni adeguati, oltre che dalla capacità di verificare concretamente la conformità.

Un approccio (oggi molto di moda) basato sulla *certificazione volontaria* si sarebbe, a sua volta, presentato come un'alternativa meno vincolante (incentivando la creazione di standard e, appunto, certificazioni volontarie per garantire la sicurezza e l'affidabilità dell'IA, simile a quanto avviene con le certificazioni ISO), con il vantaggio di evitare il rischio di ostacolare l'innovazione e permettendo alle aziende di scegliere *se e come aderire* agli standard, incentivando la compliance senza imporla. Al contempo, però, vi era all'orizzonte, chiaro, il rischio di *scarsa adesione*, soprattutto da parte di aziende "meno etiche", e la difficoltà nel garantire un'applicazione uniforme a livello europeo.

Infine, un ipotetico approccio ispirato alla *regolamentazione dei prodotti* (definita "product safety regulation") avrebbe "trattato" i sistemi di IA come *prodotti fisici* soggetti a regolamentazioni sulla sicurezza e la conformità, proprio come avviene per i dispositivi medici o i macchinari industriali, legandosi così a un modello consolidato e già applicato a molte tecnologie (e magari permettendo di integrare l'IA in framework normativi esistenti). Si noti, però, che l'intelligenza artificiale non è un semplice "prodotto fisico", ma una tecnologia in continua evoluzione, rendendo difficile una regolamentazione *statica* basata su standard fissi.

In questa fase preliminare di "nascita" dell'AI Act, quindi, possiamo affermare che l'*approccio basato sul rischio* è stato scelto con la speranza di garantire un equilibrio tra innovazione e protezione dei cittadini, ma può presentare limiti significativi, soprattutto in termini di applicabilità pratica e di impatto sulla competitività europea.

Altri approcci, come quello basato sulla responsabilità o sulla certificazione volontaria, avrebbero forse potuto garantire maggiore flessibilità,

mentre un modello settoriale avrebbe permesso di affrontare con più precisione le specificità dei diversi usi dell'IA.

Tuttavia, nessun approccio è esente da criticità, e la sfida principale rimane quella di conciliare la necessità di regolamentare l'intelligenza artificiale senza *soffocare* il progresso tecnologico.

Dopo queste discussioni preliminari sull'approccio da adottare, e la scelta del *rischio*, il 21 aprile 2021 segnò una tappa cruciale con la presentazione della proposta legislativa dell'Artificial Intelligence Act da parte della Commissione Europea.

L'approccio basato sul rischio prevedeva, come conseguenza immediata, una suddivisione delle applicazioni dell'IA in diverse categorie.

I sistemi considerati *ad alto rischio*, come quelli impiegati in ambiti critici quali la sanità, l'occupazione, l'istruzione o il sistema-giustizia, erano sottoposti a obblighi stringenti di conformità e trasparenza.

I sistemi di IA con un rischio limitato, come i chatbot o i filtri di contenuti online, devono rispettare alcuni requisiti di trasparenza, mentre i sistemi con rischio minimo possono essere utilizzati senza particolari restrizioni.

Infine, alcune applicazioni dell'IA, considerate di rischio inaccettabile in quanto potenzialmente lesive dei diritti fondamentali, venivano *vietate*. Tra queste rientrano, ad esempio, i sistemi di sorveglianza biometrica indiscriminata da remoto in tempo reale in luoghi pubblici, o quelli in grado di manipolare il comportamento umano in modo dannoso.

L'AI Act prevede, a tal fine, una serie di obblighi specifici per i sistemi di IA ad alto rischio, tra cui l'adozione di misure di gestione del rischio, una documentazione tecnica dettagliata, la trasparenza verso gli utenti e la previsione di un monitoraggio umano che permetta di intervenire in caso di anomalie o malfunzionamenti.

Inoltre, viene introdotto un meccanismo di *governance* che prevede l'istituzione di un Comitato europeo per l'intelligenza artificiale, incaricato di garantire l'applicazione coerente del regolamento in tutti gli Stati membri e di facilitare la cooperazione tra le diverse autorità nazionali.

Dopo la presentazione della proposta, il processo legislativo dell'AI Act ha seguito le consuete fasi di discussione e negoziazione all'interno delle istituzioni europee, in alcuni momenti anche abbastanza accese e con una forte attività di lobbying da parte delle (poche) grandi multinazionali tecnologiche che producono IA.

Il Parlamento Europeo e il Consiglio dell'Unione hanno, in queste fasi, apportato modifiche e integrazioni al testo iniziale, con l'obiettivo di affinare il bilanciamento tra innovazione e tutela dei diritti fondamentali.

L'entrata in vigore del regolamento, il primo di agosto del 2024, segna un momento storico, in quanto si tratta del primo tentativo globale di regolamentare l'intelligenza artificiale in modo organico e sistematico, ponendo l'Unione Europea nel dibattito internazionale sulla governance dell'IA.

L'AI Act non solo stabilisce, a nostro avviso, un precedente normativo, ma rappresenta anche un *modello* potenzialmente replicabile a livello globale: il suo approccio basato sul rischio, unito a una forte attenzione alla protezione dei diritti fondamentali, potrebbe costituire un punto di riferimento per altri Paesi e organizzazioni internazionali che stanno valutando l'introduzione di normative analoghe.

Tuttavia, l'efficacia della regolamentazione dipenderà in larga misura dalla sua attuazione pratica e dalla capacità delle istituzioni europee di monitorare e far rispettare le disposizioni contenute nell'atto.

Il cammino verso un'intelligenza artificiale affidabile, sicura ed eticamente responsabile non si esaurisce, ovviamente, con l'adozione dell'AI Act, ma richiederà un costante aggiornamento delle norme e una collaborazione continua tra governi, imprese, ricercatori e società civile, per affrontare le sfide in continua evoluzione che l'intelligenza artificiale pone alla nostra società.

2. *L'incorporazione dei principi di computer ethics nell'AI Act*

L'evoluzione dell'intelligenza artificiale nell'Unione Europea, accanto alle considerazioni di politica legislativa che abbiamo illustrato nel paragrafo precedente, ha sempre seguito un percorso guidato anche da *principi etici*, e da un costante equilibrio tra innovazione tecnologica e protezione dei diritti fondamentali[2].

La volontà di garantire uno sviluppo responsabile dell'IA ha portato, nel 2018, alla costituzione del "Gruppo di Esperti di Alto Livello sull'Intelligenza Artificiale" ("AI HLEG"), incaricato di delineare un insieme di principi che potessero fungere da base per la regolamentazione futura.

Il risultato principale di questo lavoro è stato la pubblicazione, nel 2019, delle "Linee guida etiche per un'IA affidabile", un documento che identifica i principi fondamentali a cui qualsiasi sistema di intelligenza arti-

[2] Sul tema dell'etica dell'intelligenza artificiale si veda P. Benanti, S. Maffettone, *Noi e la macchina. Un'etica per l'era digitale*, LUISS University Press, Roma, 2024. Si veda anche P. Benanti, *Human in the loop. Decisioni umane e intelligenza artificiale*, Mondadori, Milano, 2022.

ficiale dovrebbe attenersi per essere ritenuto affidabile, sicuro ed eticamente sostenibile.

Secondo l'AI HLEG, affinché un sistema di IA possa essere considerato affidabile, esso deve rispettare tre componenti essenziali: i) deve essere *legale*, rispettando tutte le normative vigenti; ii) deve essere *etico*, garantendo il rispetto dei principi morali e dei diritti fondamentali; iii) deve essere *solido* dal punto di vista tecnico, evitando malfunzionamenti o vulnerabilità che potrebbero generare danni.

All'interno di questa cornice generale, sono stati individuati quattro principi fondamentali che devono guidare lo sviluppo e l'implementazione dei sistemi di IA.

Il primo principio è il *rispetto dell'autonomia umana*, che implica la necessità di garantire che i sistemi di IA supportino le capacità decisionali degli esseri umani, senza sostituirle o manipolarle. Questo principio pone l'accento sulla necessità di preservare la libertà di scelta degli individui, evitando che le tecnologie possano essere utilizzate per influenzare le decisioni umane in modi che riducano il libero arbitrio.

L'AI Act ha incorporato questo principio prevedendo obblighi di trasparenza e garanzie contro la manipolazione ingannevole, in particolare per i sistemi di IA che interagiscono *direttamente* con gli utenti o che impiegano tecniche di persuasione avanzate, come la personalizzazione algoritmica e il riconoscimento delle emozioni.

Il secondo principio è la *prevenzione del danno*, che richiede che i sistemi di IA siano progettati per garantire la sicurezza, evitando qualsiasi tipo di danno fisico o psicologico agli esseri umani.

Questo principio si riflette nell'AI Act attraverso una *categorizzazione* dei sistemi di IA in base al rischio che comportano, vietando quelli considerati di rischio inaccettabile e imponendo rigorosi requisiti di sicurezza per quelli classificati come ad alto rischio. Il regolamento europeo stabilisce, ad esempio, che i sistemi impiegati in ambiti critici come la sanità, i trasporti o la giustizia debbano essere sottoposti a valutazioni di conformità dettagliate, dimostrando la loro affidabilità prima di poter essere utilizzati.

Il terzo principio riguarda l'*equità e la non discriminazione*, elementi essenziali per evitare che i sistemi di IA rafforzino o amplifichino pregiudizi esistenti.

Poiché gli algoritmi di apprendimento automatico si basano su dati storici che possono contenere *bias*, è fondamentale che i modelli siano progettati per ridurre al minimo le distorsioni e per garantire trattamenti equi per tutti gli utenti. L'AI Act affronta questa problematica imponendo obblighi

specifici in termini di trasparenza e auditing per i sistemi di IA utilizzati in settori sensibili come l'occupazione, il credito e l'istruzione, assicurando che le decisioni automatizzate siano *spiegabili* e prive di discriminazioni sistematiche.

Il quarto principio è la trasparenza, che prevede che i sistemi di IA siano comprensibili e accessibili agli utenti e agli stakeholder coinvolti. La *trasparenza* implica che sia chiaro quando una persona interagisce con un sistema di intelligenza artificiale e che siano disponibili informazioni sul funzionamento del modello e sui criteri che guidano le decisioni automatizzate.

Nell'AI Act, questo principio si traduce in obblighi di *spiegabilità e tracciabilità* per i sistemi di IA ad alto rischio, che devono fornire documentazione tecnica dettagliata, consentendo la verifica e il controllo delle loro decisioni. Inoltre, vengono introdotte misure per garantire che gli utenti siano sempre informati quando interagiscono con sistemi di IA, specialmente in ambiti come la sorveglianza biometrica o la creazione di contenuti generati artificialmente.

La *spiegabilità* di un sistema di intelligenza artificiale ("AI explainability") è un aspetto fondamentale anche per i giuristi, che merita un rapido approfondimento. Consiste, in particolare, nella capacità di rendere *comprensibile* agli esseri umani il funzionamento di un modello di IA, permettendo di comprendere come e perché un sistema prenda determinate decisioni.

Questo concetto è essenziale per garantire la trasparenza, la fiducia e la responsabilità nell'uso dell'intelligenza artificiale, soprattutto nei settori ad alto impatto sociale come la sanità, la finanza e la giustizia.

La spiegabilità, semplificando molto, può essere suddivisa in quattro diverse componenti chiave: i) la *interpretabilità* (riguarda la possibilità di comprendere le relazioni tra input e output di un modello di IA, e alcuni modelli, come le reti neurali profonde, sono difficili da interpretare, mentre altri, come gli alberi decisionali, sono più trasparenti); ii) la *trasparenza* (si riferisce alla capacità di esaminare il *funzionamento interno* di un sistema di IA, come i pesi e i parametri di un modello di apprendimento automatico, e anche in questo caso alcuni algoritmi sono considerati "scatole nere" perché il loro processo decisionale è opaco e complesso); iii) la *giustificabilità* (implica la possibilità di fornire motivazioni coerenti e logiche che spieghino una decisione presa dal sistema, in modo comprensibile per gli esseri umani), e iv) la *riproducibilità* (significa che un sistema di IA dovrebbe produrre risultati simili quando vengono forniti dati e condizioni simili, permettendo di verificare e validare il suo funzionamento).

Oltre a questi quattro principi fondamentali, il Gruppo di Esperti di Alto Livello ha individuato *sette requisiti chiave* per l'implementazione di un'IA affidabile, tra cui la governance dei dati, la robustezza e la sicurezza, la supervisione umana e la responsabilità.

Si legga, a tal proposito, il testo del Considerando 27:

> "Sebbene l'approccio basato sul rischio costituisca la base per un insieme proporzionato ed efficace di regole vincolanti, è importante ricordare gli orientamenti etici per un'IA affidabile del 2019 elaborati dall'AI HLEG indipendente nominato dalla Commissione. In tali orientamenti l'AI HLEG ha elaborato sette principi etici non vincolanti per l'IA che sono intesi a contribuire a garantire che l'IA sia affidabile ed eticamente valida. I sette principi comprendono: intervento e sorveglianza umani, robustezza tecnica e sicurezza, vita privata e governance dei dati, trasparenza, diversità, non discriminazione ed equità, benessere sociale e ambientale e responsabilità. Fatti salvi i requisiti giuridicamente vincolanti del presente regolamento e di qualsiasi altra disposizione di diritto dell'Unione applicabile, tali orientamenti contribuiscono all'elaborazione di un'IA coerente, affidabile e antropocentrica, in linea con la Carta e con i valori su cui si fonda l'Unione. Secondo gli orientamenti dell'AI HLEG con «intervento e sorveglianza umani» si intende che i sistemi di IA sono sviluppati e utilizzati come strumenti al servizio delle persone, nel rispetto della dignità umana e dell'autonomia personale, e funzionano in modo da poter essere adeguatamente controllati e sorvegliati dagli esseri umani. Con «robustezza tecnica e sicurezza» si intende che i sistemi di IA sono sviluppati e utilizzati in modo da consentire la robustezza nel caso di problemi e resilienza contro i tentativi di alterare l'uso o le prestazioni del sistema di IA in modo da consentire l'uso illegale da parte di terzi e ridurre al minimo i danni involontari. Con «vita privata e governance dei dati» si intende che i sistemi di IA sono sviluppati e utilizzati nel rispetto delle norme in materia di vita privata e protezione dei dati, elaborando al contempo dati che soddisfino livelli elevati in termini di qualità e integrità. Con «trasparenza» si intende che i sistemi di IA sono sviluppati e utilizzati in modo da consentire un'adeguata tracciabilità e spiegabilità, rendendo gli esseri umani consapevoli del fatto di comunicare o interagire con un sistema di IA e informando debitamente i deployer delle capacità e dei limiti di tale sistema di IA e le persone interessate dei loro diritti. Con «diversità, non discriminazione ed equità» si intende che i sistemi di IA sono sviluppati e utilizzati in modo da includere soggetti diversi e promuovere la parità di accesso, l'uguaglianza di genere e la diversità culturale, evitando nel contempo effetti discriminatori e pregiudizi ingiusti vietati dal diritto dell'Unione o nazionale. Con «benessere sociale e ambientale» si intende che i sistemi di IA sono sviluppati e utilizzati in modo sostenibile e rispettoso dell'ambiente e in modo da apportare benefici a tutti gli esseri umani, monitorando e valutando gli impatti a lungo termine sull'individuo, sulla società e sulla democrazia.

> L'applicazione di tali principi dovrebbe essere tradotta, ove possibile, nella progettazione e nell'utilizzo di modelli di IA. Essi dovrebbero in ogni caso fungere da base per l'elaborazione di codici di condotta a norma del presente regolamento. Tutti i portatori di interessi, compresi l'industria, il mondo accademico, la società civile e le organizzazioni di normazione, sono incoraggiati a tenere conto, se del caso, dei principi etici per lo sviluppo delle migliori pratiche e norme volontarie".

Questi elementi elencati come "non vincolanti" nel Considerando citato sono, poi, stati incorporati "indirettamente" nell'AI Act attraverso una serie di disposizioni che mirano a garantire che lo sviluppo dell'intelligenza artificiale avvenga in modo etico e responsabile.

Ad esempio, il regolamento prevede che i sistemi di IA ad alto rischio siano soggetti a *verifiche periodiche* e a *meccanismi di controllo continuo*, mentre per le aziende che sviluppano o utilizzano tali tecnologie vengono introdotti obblighi di *conformità* e *tracciabilità*.

L'integrazione dei principi etici dell'AI HLEG nell'AI Act rappresenta una delle caratteristiche distintive dell'approccio europeo alla regolamentazione dell'intelligenza artificiale.

A differenza di altri modelli normativi, che si concentrano prevalentemente sugli aspetti *economici* e di *sicurezza*, il quadro europeo pone al centro della regolamentazione la protezione dei *diritti fondamentali*, cercando di prevenire gli impatti negativi dell'IA sulla società. Questo approccio ha influenzato anche il dibattito internazionale sulla governance dell'intelligenza artificiale, spingendo altre giurisdizioni a considerare l'adozione di *misure simili* per garantire uno sviluppo tecnologico etico e sostenibile.

L'AI Act si presenta, quindi, non solo come una norma giuridica, ma vuole rappresentare un vero e proprio tentativo di definire un modello di riferimento per *l'uso responsabile dell'intelligenza artificiale* a livello globale.

L'incorporazione di principi etici nel regolamento europeo non solo contribuisce a creare un ecosistema di fiducia, ma aiuta anche a promuovere un'innovazione che sia al servizio dell'umanità, piuttosto che una mera espressione di progresso tecnologico privo di considerazioni morali.

La sfida principale, tuttavia, rimarrà quella *dell'attuazione pratica* di queste disposizioni e della capacità delle istituzioni europee di monitorare e garantire il rispetto dei principi etici nel lungo termine, in un contesto tecnologico in continua evoluzione.

3. *Un esempio pratico di carta per la governance dell'AI: il Decalogo della Statale*

Un esempio interessante di fusione di principi etici, giuridici, tecnologici e di buon senso in una *carta programmatica* per una realtà molto complessa quale una grande università pubblica è un documento di governance dell'AI presentato ai primi di febbraio del 2025 dal gruppo di lavoro sull'intelligenza artificiale dell'Università di Milano. Si tratta di un vero e proprio "decalogo" teorico-pratico che si propone di orientare tutta la popolazione dell'Ateneo nell'uso di questi strumenti così innovativi.

Come si legge nel preambolo, l'Università degli Studi di Milano muove dall'idea di *sostenere* l'uso di strumenti di intelligenza artificiale a sostegno di *tutte* le sue attività, mantenendo fermo, prima di tutto, il principio di porre *la persona* al centro di ogni iniziativa. Lo scopo, dunque, non è quello di *vincolare* ma, piuttosto, di promuovere un utilizzo etico, legittimo e consapevole di strumenti di intelligenza artificiale, recependo i più recenti documenti normativi in materia e le *buone pratiche* già promosse da molte università e centri di ricerca a livello nazionale e internazionale.

I principi generali del decalogo saranno, poi, corredati da specifiche linee guida in tema di ricerca e terza missione, didattica, e attività amministrative, che comprenderanno anche indicazioni sugli strumenti il cui utilizzo è avvallato dell'ateneo (tali linee guida, in particolare, saranno definite attraverso *modalità partecipative* che coinvolgeranno, per ciascun ambito, rappresentanze di tutte le componenti dell'ateneo).

Il primo punto trattato è quello della "AI Literacy" (l'alfabetizzazione in tema di IA resa *obbligatoria* dall'art. 4 del regolamento, adempimento già in vigore dal febbraio 2025). In particolare, vi è la previsione di *percorsi di alfabetizzazione* in tema di intelligenza artificiale e di eventi culturali, tecnici, giuridici, scientifici e divulgativi al fine di garantire un innalzamento del livello di alfabetizzazione e di competenze digitali in tema di intelligenza artificiale nell'Università degli Studi di Milano. Ciò è stato programmato al fine di accrescere le conoscenze di tutti, di rispettare gli obblighi di legge e le migliori pratiche e di sviluppare lo spirito critico della comunità accademica, con l'obiettivo ambizioso di portare l'intera comunità universitaria, entro il 2030, a un livello di conoscenza che consenta un uso affidabile, responsabile e condiviso degli strumenti di intelligenza artificiale.

Viene affrontato, subito dopo, il tema della *data protection*, prevedendo un obbligo costante di protezione dei dati personali e di data governance. In questo caso si richiede un utilizzo di strumenti di intelligenza artificiale

che avvenga sempre nel rispetto dei principi, delle buone pratiche e delle norme in materia di tutela dei dati personali e di riservatezza delle persone e delle informazioni a loro riferite.

In particolare, viene evidenziato il principio di *minimizzazione dei trattamenti*, che impone grande cautela nel trasmettere agli strumenti di intelligenza artificiale *unicamente* le informazioni necessarie per ottenere i risultati attesi. Allo stesso tempo, ogni strumento di intelligenza artificiale che non sia stato preventivamente *avallato* o "certificato" dall'Ateneo e dagli uffici preposti alla ricognizione tecnologica, agli acquisti e alla amministrazione dei sistemi, richiederà apposite valutazioni (da parte del Comitato Etico, di commissioni *ad hoc* o in base a norme contenute in specifici regolamenti) prima di poter essere utilizzata.

Nel prosieguo del documento viene illustrato, poi, il già visto (e fondamentale) principio di *trasparenza e tracciabilità* delle operazioni. Più precisamente, si stabilisce come l'utilizzo di strumenti di intelligenza artificiale per lo svolgimento delle rispettive attività (amministrative, di ricerca, di studio e redazione elaborati o tesi) debba essere dichiarato e trasparente. In particolare, è necessaria l'indicazione (tra gli altri) dei seguenti elementi: *i*) la base di dati utilizzata, *ii*) le chiavi di ricerca immesse o i percorsi di interrogazione fatti, *iii*) gli strumenti concretamente utilizzati. Deve anche essere garantita, come richiesto dalla regolamentazione europea, la *tracciabilità* delle operazioni effettuate dal sistema di intelligenza artificiale.

Particolare attenzione, in linea con i contenuti del piano strategico di Ateneo, è poi data alla *sostenibilità ambientale*: gli strumenti di intelligenza artificiale sono in grado di svolgere molteplici funzioni utili, ma sono anche molto *energivori*. Ciò comporta che, prima di adottare o adoperare uno strumento di intelligenza artificiale, sia opportuno *valutare l'impatto* ambientale dello stesso, limitandone l'utilizzo ai casi in cui vi sia un'effettiva esigenza di supporto ad attività lavorative, di studio o di ricerca.

Con riferimento alla *qualità dei dati* utilizzati e alla inclusività, i risultati dell'attività di elaborazione svolta dagli strumenti di intelligenza artificiale devono sempre essere *attentamente analizzati* al fine di evitare o mitigare possibili risultati discriminatori correlati ad elementi caratteristici di un soggetto o di un gruppo di persone, soprattutto se rientranti nella categoria dei *soggetti vulnerabili*. A tal fine, fondamentale è la garanzia della qualità dei dati (in ingresso, di training, di validazione) utilizzati dal sistema.

Centrale, poi, è la questione di *cybersecurity*: come qualsiasi ritrovato informatico, anche gli strumenti di intelligenza artificiale possono essere vittime di *attacchi informatici*. Per questo, l'attenzione alla protezione dei dispositivi, alla legittimità delle basi di dati, alla possibilità di avvelenamento

degli stessi e alla affidabilità del produttore o del distributore del sistema di IA deve sempre essere elevata.

Nella parte finale del Decalogo ci si concentra, invece, su *accountability* e *supervisione umana*. L'Ateneo si impegna a rispettare tutti gli obblighi di legge, con particolare riferimento a quelli relativi alla protezione dei dati, alla cybersecurity, al rispetto del diritto d'autore, alla sicurezza delle infrastrutture, alla regolamentazione europea dell'intelligenza artificiale e al plagio di contenuti altrui. Si chiede, a tal fine, la *collaborazione* di tutti affinché non vi siano utilizzi dell'intelligenza artificiale che non siano rispettosi di tali principi e delle regole professionali e deontologiche applicabili.

Per tutti i sistemi di intelligenza artificiale usati in Ateneo dovrà poi essere garantita, come richiesto dalla normativa vigente, una *supervisione/sorveglianza umana* in grado di correggere i risultati, re-interpretarli o ribaltarli, disattivare o bloccare il sistema di intelligenza artificiale.

Con riferimento, invece, alla *governance*, l'Ateneo s'impegna a promuovere i processi *data-driven* nel rispetto dei valori etici e giuridici. In particolare, è *proibita* l'immissione in sistemi di intelligenza artificiale di dati personali "sensibili" e di dati che possono porre dei rischi per l'Ateneo (verbali riservati, informazioni soggette a riservatezza, informazioni con elevato valore economico, informazioni legate alla tutela della proprietà intellettuale e industriale dell'Ateneo), a meno che non sia strettamente necessario per lo svolgimento dei propri compiti. In questo caso, l'Ateneo svolge delle apposite valutazioni preliminari prima di consentire tali procedure.

Infine, la valutazione circa l'*opportunità* di adottare strumenti di intelligenza artificiale, e l'adozione degli stessi, viene lasciata, in un'ottica di responsabilità, a ciascun utente, che deve impegnarsi a un uso consapevole e attento dello strumento e dei suoi risultati.

4. *Gli adempimenti previsti e le concrete difficoltà*

Muovendo, ora, a un approccio meno teorico e più pratico e informatico-giuridico, si anticipava come l'Artificial Intelligence Act abbia previsto un quadro normativo *rigoroso* per la regolamentazione dei sistemi di intelligenza artificiale nell'Unione Europea, con un approccio basato sulla valutazione del rischio.

La conformità alle disposizioni dell'AI Act varia, quindi, in base alla classificazione del rischio dei sistemi di IA, con obblighi più *stringenti* per

le applicazioni considerate ad alto rischio e requisiti più *leggeri* per quelle a rischio limitato o minimo [3].

I sistemi di intelligenza artificiale classificati come ad alto rischio sono soggetti, in particolare, a un regime di conformità rigoroso, in quanto il loro utilizzo potrebbe avere un impatto significativo sui diritti e sulla sicurezza delle persone. Questi sistemi includono, ad esempio, quelli impiegati nei settori della sanità, dell'occupazione, dell'istruzione, delle infrastrutture critiche e del sistema-giustizia.

Per tali applicazioni, l'AI Act prevede una serie di adempimenti specifici, tra cui, *in primis*, l'obbligo di implementare un sistema di gestione del rischio che consenta di individuare, valutare e mitigare potenziali minacce derivanti dall'utilizzo dell'IA.

Questo processo deve essere *continuo* e *aggiornato*, in modo da garantire che i rischi emergenti vengano prontamente affrontati. Il livello di difficoltà di tale adempimento, a nostro avviso, si può presentare *elevato*, in quanto richiede la definizione di strategie di gestione del rischio, la formazione del personale e l'integrazione di strumenti di monitoraggio avanzati.

Un altro requisito essenziale per i sistemi ad alto rischio riguarda la *documentazione tecnica*, che deve essere dettagliata e dimostrare la conformità del sistema alle disposizioni dell'AI Act.

Tale documentazione deve contenere informazioni sul funzionamento dell'algoritmo, sui dati utilizzati per l'addestramento e sui meccanismi di mitigazione del rischio. Questo adempimento, sebbene tecnicamente complesso, può essere, a nostro avviso, gestito attraverso un'integrazione efficiente di procedure di *audit* e *certificazione*, garantendo che ogni fase dello sviluppo del sistema sia tracciabile e verificabile.

La trasparenza è, si diceva poco sopra, un ulteriore pilastro della conformità per i sistemi ad alto rischio.

Gli utenti devono essere informati in modo chiaro e comprensibile sul funzionamento dell'IA e sulle modalità con cui le decisioni vengono prese, e questo obbligo si traduce, in pratica, nella necessità di fornire spiegazioni dettagliate, utilizzando un linguaggio accessibile e strumenti grafici, per consentire agli utenti di comprendere come, e perché, una determinata decisione sia stata adottata dal sistema. Questo adempimento, pur non essendo tecnicamente complesso, può presentare difficoltà nel garantire una *comunicazione* efficace tra i fornitori di IA e gli utenti finali.

---

[3] Per una guida interessante agli adempimenti si veda, in lingua inglese, T. Myklebust, T. Stålhane, D.M.K. Vatn, *The AI Act and The Agile Safety Plan*, Springer, Cham, 2025.

L'AI Act richiede, inoltre, che i sistemi di IA ad alto rischio siano sottoposti a *supervisione umana*, in modo da garantire che le decisioni automatizzate possano essere *controllate* e *corrette* in caso di errori o risultati imprevisti. Questo requisito impone alle aziende di implementare *procedure* che permettano l'intervento umano in tempo reale, il che può risultare complesso in applicazioni ad alta automazione. L'adozione di interfacce di monitoraggio intuitive e di protocolli di revisione periodica rappresenta la soluzione più efficace per adempiere a questo obbligo, che vediamo come il *più complesso* del Regolamento.

Infine, la *robustezza* e la *sicurezza* dei sistemi ad alto rischio devono essere garantite attraverso test approfonditi e verifiche periodiche, che dimostrino l'affidabilità delle soluzioni adottate e la loro resistenza a potenziali attacchi o malfunzionamenti.

Questo adempimento, che potremmo definire genericamente di "cybersecurity", è altamente tecnico e può richiedere *ingenti risorse*, sia in termini di tempo che di investimenti tecnologici, rendendo fondamentale la collaborazione tra sviluppatori, esperti di sicurezza e organismi di certificazione.

Per quanto riguarda i sistemi di IA a rischio *limitato* o *minimo*, l'AI Act prevede requisiti *meno stringenti*, volti principalmente a garantire la trasparenza nei confronti degli utenti.

Questi sistemi comprendono, ad esempio, chatbot, assistenti virtuali e algoritmi di raccomandazione, che non presentano un impatto significativo sui diritti fondamentali o sulla sicurezza delle persone.

Per tali applicazioni, l'adempimento principale riguarda l'*obbligo di informare gli utenti* del fatto che stanno interagendo con un sistema di intelligenza artificiale. Questa misura di trasparenza è, a nostro avviso, relativamente semplice da implementare (richiama molto l'idea di *informativa tanto* cara alla protezione dei dati) e può essere soddisfatta attraverso messaggi chiari e visibili all'interno delle interfacce utente.

Un ulteriore requisito per i sistemi a rischio limitato riguarda la prevenzione della *manipolazione ingannevole* (anche con tecniche subliminali), che impone ai fornitori di IA di garantire che i loro sistemi non influenzino in modo subdolo il comportamento degli utenti.

Questo obbligo può essere rispettato attraverso la progettazione di interfacce-utente etiche, che evitino pratiche di persuasione occulta o oscura, anche note come "dark patterns". Anche se questo adempimento non presenta particolari difficoltà tecniche, richiede un'attenzione specifica nella fase di progettazione dei sistemi di interazione.

Le modalità migliori per garantire la conformità all'AI Act dipendono,

in definitiva, dalla tipologia del sistema di IA e dal livello di rischio ad esso associato.

Per le applicazioni ad alto rischio, è consigliabile adottare un approccio *integrato* che combini audit interni ed esterni, strumenti di monitoraggio automatico e verifiche periodiche condotte da enti indipendenti. L'adozione di *standard internazionali* per la certificazione della sicurezza e dell'affidabilità dell'IA può rappresentare un ulteriore vantaggio competitivo, facilitando la dimostrazione della conformità alle autorità regolatorie.

Per i sistemi a rischio limitato, invece, la conformità può essere garantita attraverso l'implementazione di meccanismi di *trasparenza* e *informazione* all'utente, che non richiedono necessariamente un controllo esterno ma devono essere integrati efficacemente all'interno dell'esperienza d'uso del sistema. Un'adeguata formazione dei gruppi di sviluppo e l'adozione di *linee guida* chiare per la progettazione etica possono contribuire significativamente alla riduzione del rischio di non conformità.

Sebbene il livello di complessità degli adempimenti vari in base al rischio associato al sistema, la chiave per garantire la conformità risiede in un *approccio proattivo*, che integri la gestione del rischio, la trasparenza e la supervisione umana fin dalle prime fasi di sviluppo dell'IA. In tal modo, sarà possibile non solo soddisfare i requisiti normativi, ma anche rafforzare la fiducia degli utenti e promuovere un'adozione responsabile delle tecnologie basate sull'intelligenza artificiale.

5. *Alcune conclusioni: i limiti dei tempi di attuazione e l'attuale situazione geo-politica*

L'Artificial Intelligence Act, si è visto, rappresenta indubbiamente un ambizioso tentativo di regolamentare l'intelligenza artificiale a livello europeo, ma non è privo di *limiti* e *criticità*, che si manifesteranno con sempre maggior evidenza nei prossimi cinque anni (termine previsto per un'attuazione *completa* di tutte le parti del Regolamento).

Nonostante il suo impianto normativo dettagliato, e la sua enfasi sulla protezione dei diritti fondamentali, l'efficacia della sua attuazione e il suo impatto globale sono oggetto, in questi mesi, di un dibattito assai vivace. In un contesto geopolitico in cui l'intelligenza artificiale è divenuta un terreno di *competizione* tra grandi potenze, il modello regolatorio europeo si distingue nettamente dagli approcci adottati da Stati Uniti e Cina, sollevando però, al contempo, interrogativi sulla sua capacità di *incidere* efficacemente nel panorama internazionale.

Uno dei principali limiti dell'AI Act riguarderà, probabilmente, la sua *implementazione pratica*. La complessità degli adempimenti richiesti, soprattutto per i sistemi di IA ad alto rischio, potrebbe rappresentare un *ostacolo* significativo per le aziende, in particolare per le piccole e medie imprese. Il regolamento impone una serie di *oneri burocratici*, tra cui la documentazione dettagliata, i meccanismi di trasparenza e le verifiche periodiche, che potrebbero risultare gravosi per le realtà con minori risorse a disposizione.

A ciò si aggiunge il rischio, sollevato da più parti, che l'AI Act possa *rallentare l'innovazione* in Europa, creando un ambiente normativo troppo rigido rispetto ad altri contesti internazionali più permissivi.

Questo aspetto è stato oggetto di critiche non solo da diverse parti politiche in Europa ma, anche, da parte dell'*industria tecnologica*, che teme un'eccessiva regolamentazione capace di ostacolare la competitività delle imprese europee rispetto ai giganti tecnologici statunitensi e cinesi.

A livello geopolitico, l'AI Act si colloca, oggi, in un panorama frammentato, caratterizzato da approcci differenti alla regolamentazione dell'intelligenza artificiale e da un uso bellico dell'IA legato alle *guerre* in corso (soprattutto i conflitti Russia-Ucraina e Israele-Hamas).

Gli Stati Uniti d'America, ad esempio, hanno adottato un modello basato principalmente sull'autoregolamentazione e sull'intervento *mirato* solo in determinati settori, come la sicurezza nazionale e la protezione dei dati personali. Un simile approccio, meno vincolante rispetto a quello europeo, consente un'accelerazione dell'innovazione e una maggiore flessibilità per le aziende ma, al contempo, presenta il rischio di una minore protezione per gli utenti.

La Cina, dal canto suo, ha sviluppato una strategia che combina una forte regolamentazione statale con un uso intensivo dell'IA per scopi governativi e di sorveglianza. Il governo cinese ha introdotto norme stringenti su alcune applicazioni dell'IA, come il *riconoscimento facciale* e gli *algoritmi di raccomandazione*, ma con un obiettivo principalmente orientato al controllo sociale piuttosto che alla tutela dei diritti individuali.

Questa divergenza negli approcci normativi solleva una questione di fondo sulla capacità dell'AI Act di influenzare il panorama globale, in un periodo storico di *insofferenza diffusa*, nella classe politica, dell'idea di regolamentazione e per le attività delle autorità di controllo indipendenti.

Se, da un lato, il regolamento europeo si propone giustamente come modello di riferimento per una governance etica dell'IA, con un intento *nobile* che è chiarissimo, dall'altro viene ventilato il rischio che l'Europa rimanga *isolata* nel definire standard che non vengano adottati da altre potenze tecnologiche.

La mancanza di un quadro normativo armonizzato a livello internazionale potrebbe, così, portare a una *frammentazione* del mercato dell'IA, con imprese europee costrette a rispettare norme più severe rispetto ai concorrenti stranieri, con conseguenti svantaggi competitivi.

Un'altra critica mossa all'AI Act nei primi mesi di attuazione riguarda la sua capacità di affrontare le *sfide emergenti* dell'intelligenza artificiale in un contesto in *continua evoluzione*.

La rapidità con cui si sviluppano nuove tecnologie basate sull'IA rende difficile una regolamentazione che possa anticipare tutti i possibili rischi e scenari futuri: si pensi all'avvento *improvviso* prima di ChatGPT (nella primavera del 2023) e, poi, della IA cinese DeepSeek agli inizi del 2025.

Il regolamento, pur adottando un approccio basato sul rischio, potrebbe risultare inadeguato di fronte a innovazioni impreviste e così rapide, creando la necessità di continui aggiornamenti normativi. Un'eccessiva rigidità nelle regole potrebbe portare a problemi di applicazione, rendendo il sistema normativo rapidamente obsoleto rispetto ai progressi tecnologici.

Un ulteriore punto critico riguarda, a nostro avviso, la (futura) capacità delle istituzioni europee di *far rispettare* le disposizioni dell'AI Act.

L'attuazione completa del regolamento richiede la creazione di autorità di vigilanza nazionali e meccanismi di controllo che potrebbero rivelarsi difficili da gestire, soprattutto considerando le differenze tra gli Stati membri in termini di risorse e competenze tecniche. Inoltre, il rispetto degli obblighi di conformità dipenderà in gran parte dalla capacità delle aziende di adeguarsi alle nuove regole, il che potrebbe generare ritardi e incertezze nella fase di implementazione.

Nonostante queste possibili criticità, che saranno valutate con attenzione da tutti nei prossimi cinque anni, l'AI Act rappresenta un passo fondamentale verso una regolamentazione responsabile dell'intelligenza artificiale.

La sfida principale per l'Unione Europea sarà quella di trovare un *compromesso* tra la necessità di proteggere i cittadini e il desiderio di promuovere l'innovazione, evitando che il quadro normativo diventi un ostacolo alla crescita del settore.

Il futuro dell'AI Act dipenderà, in definitiva, dalla sua capacità di adattarsi ai rapidi cambiamenti tecnologici e di influenzare il dibattito globale sulla governance dell'intelligenza artificiale, cercando di promuovere un modello che possa essere adottato anche a livello internazionale.

L'Unione Europea si trova, in conclusione, di fronte a un *bivio*. Da un lato, ha concretamente la possibilità di diventare un punto di riferimento globale per una regolamentazione etica e responsabile dell'IA. Dall'altro, appare all'orizzonte il rischio di *restare indietro* rispetto a potenze tecnolo-

giche che privilegiano un approccio più flessibile e orientato alla competitività.

La sfida, a nostro avviso, sarà quella di garantire che l'AI Act non si trasformi in un freno all'innovazione, ma che possa invece fungere da stimolo per la creazione di un ecosistema di intelligenza artificiale sicuro, affidabile e conforme ai valori europei.

INTELLIGENZA ARTIFICIALE
E GIUSTIZIA PENALE

INTELLIGENZA ARTIFICIALE E CRISI DEL DIRITTO PENALE D'EVENTO: PROFILI DI RESPONSABILITÀ PENALE DEL PRODUTTORE DI SISTEMI DI I.A.*

di Beatrice Fragasso

1. *L'intelligenza artificiale, a metà via tra* res cogitans *e* res extensa

Che le tecniche di produzione moderna creino sistematicamente rischi che la società non è in grado di controllare ed assorbire è questione studiata e dibattuta almeno dalla pubblicazione di *La società del rischio* di Ulrich Beck [1], uno dei testi che ha esercitato, e ancora esercita, maggiore influenza tra coloro che sono dediti allo studio dei rapporti tra diritto penale e società (post) moderna [2].

---

* Il presente contributo è già stato pubblicato sul fascicolo 1/2024 della *Rivista italiana di diritto e procedura penale*. Rispetto alla versione contenuta nella *Rivista*, il testo qui riportato è stato opportunamente aggiornato, tenendo in considerazione, in particolare, l'approvazione e la pubblicazione del testo definitivo del Regolamento dell'Unione europea sull'Intelligenza artificiale (Reg. 2024/1689).

[1] U. Beck, *La società del rischio. Verso una seconda modernità*, Carocci, Bari, 2000 (ed. ted., *Risikogesellschaft. Auf dem Weg in eine endere Moderne*, Suhrkamp, 1986); sul rapporto tra rischio e società post moderna v. anche N. Luhmann, *Sociologia del rischio*, Mondadori, Milano, 1996 (ed. ted., *Soziologie des Risikos*, De Gruyter, Berlino, 1991).

[2] Sul c.d. "diritto penale del rischio" si v., senza alcuna pretesa di esaustività, e limitandoci

Il circolo vizioso insito in tale prospettiva è ben noto: da un lato, i progressi tecnico-scientifici permettono di affrontare pericoli che, in epoca pre-moderna, avrebbero potuto comportare il verificarsi di danni incommensurabili; dall'altro, tuttavia, lo sviluppo capitalistico crea nuovi rischi globali, che sfuggono alla comprensione e al dominio dell'uomo – tanto che i "disastri" derivanti dall'esercizio dell'attività produttiva ed industriale hanno finito per essere considerati inevitabili, financo *normali*[3]. Di qui il paradosso: la tecnica – da strumento di *comprensione* e *dominio* sulla forza cieca e schiacciante della natura – diventa essa stessa un elemento misterioso, imperscrutabile, incontrollabile.

L'intelligenza artificiale, in questo contesto, apre un nuovo capitolo nel rapporto tormentato tra tecnologia e *agency* umana. In questo caso, infatti, l'imprevedibilità non costituisce un difetto del sistema, ma è piuttosto il risultato voluto e ricercato dagli stessi programmatori, poiché consente di raggiungere i risultati più performanti[4]: in questo senso, si dice che il sistema di i.a. è *unpredictable by design*[5].

Già da questa prima affermazione, si può cogliere come l'accezione di intelligenza artificiale a cui s'intende aderire in questa sede sia piuttosto ristretta, e funzionale a cogliere il *novum* di alcuni *AI systems* rispetto alle

---

alla letteratura in lingua italiana, F. Stella, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, III ed., Giuffrè, Milano, 2003; C.E. Paliero, *L'autunno del patriarca. Rinnovamento o trasmutazione del diritto penale dei codici?*, in *Riv. it. dir. proc. pen*., 1994, p. 1220; M. Donini, *Il volto attuale dell'illecito penale*, Giuffrè, Milano, 2004, p. 97 ss.; C. Piergallini, *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, Giuffrè, Milano, 2004; A. Gargani, *La "flessibilizzazione" giurisprudenziale delle categorie classiche del reato di fronte alle esigenze di controllo penale delle nuove fenomenologie di rischio*, in *Legis. pen*., 2011, n. 2, p. 403 ss.; v. anche i contributi pubblicati in L. Stortoni, L. Foffani (a cura di), *Critica e giustificazione del diritto penale nel cambio di secolo. L'analisi critica della Scuola di Francoforte*, Giuffrè, Milano, 2004.

[3] C. Perrow, *Normal accidents*, Princeton University Press, Princeton, 1984. L'autore, in particolare, si riferiva ad incidenti verificatisi nell'ambito di impianti chimici e centrali nucleari o nella gestione del traffico aereo e marittimo; per una prospettiva penalistica v. F. Centonze, *La normalità dei disastri tecnologici: il problema del congedo dal diritto penale*, Giuffrè, Milano, 2004. Il tema dell'imprevedibilità delle tecnostrutture più avanzate è trattato anche, più recentemente, da S. Abersman, *Overcomplicated. Technology at The Limits of Comprehension*, Gildan Media Corporation, Seattle, 2016, che, a tal proposito, parla di *entanglement* (groviglio).

[4] J. Millar, I. Kerr, *Delegation, relinquishment, and responsibility: The prospect of expert robots*, in R. Calo e al. (eds.), *Robot Law*, Edward Elgar Publishing, Cheltenham, 2016, p. 107: "*a feature and not a bug*"; v. anche M.A. Lemley, B. Casey, *Remedies for Robots*, n 86 *University of Chicago Law Review*, 2019, p. 1334 ss.

[5] R. Calo, *Robotics and the Lessons of Cyberlaw*, in 103 *Calif. L. Rev*., 2015, p. 542; J. Millar, I. Kerr, *Delegation, relinquishment, and responsibility*, cit., p. 107.

tecnologie precedenti [6]. In particolare, nel riflettere sul possibile impatto dell'intelligenza artificiale sulle categorie penalistiche, ciò a cui ci interessa dare rilievo è la possibilità che taluni sistemi di i.a., nel perseguire gli obiettivi definiti dai programmatori, si comportino in base a processi di *decision making* autonomi, che non sono interamente prestabiliti in fase di progettazione e che, talvolta, non sono nemmeno comprensibili da parte dell'operatore umano (c.d. *opacità tecnologica* o *black-box*). La nostra attenzione, in breve, sarà concentrata sugli algoritmi di *machine learning* (e, nella loro versione più sofisticata, di *deep learning*), i quali non recano istruzioni *nozionistiche* (lo schema deterministico dell'*if-this-then-that*), ma sono piuttosto dotati delle informazioni necessarie su *come apprendere*, *come classificare*, *come generalizzare* [7].

Se non propriamente "agenti" – poiché le "finalità" delle loro azioni parrebbero pur sempre etero-definite [8] – tali sistemi non possono nemmeno essere considerati come meri "strumenti" nelle mani dell'essere umano, poiché non reagiscono in maniera deterministica agli *input* forniti dall'uomo e non sono da questi pienamente dominabili. Ci troveremmo di fronte, allora, ad una sorta di "terza *res*" – a metà tra la *res cogitans* e la *res extensa* di cartesiana memoria –, che parrebbe mettere in discussione quel modello dualistico (oggetto-soggetto) che ha caratterizzato la storia del pensiero moderno occidentale e che costituisce il fondamento delle odierne strutture sociali, comprese quelle normative.

A scanso di equivoci, evidenziamo fin da subito che con tali affermazio-

---

[6] È impossibile, in questa sede, dare conto del vastissimo dibattito che si è sviluppato intorno alla questione della definizione dell'intelligenza artificiale. Per una classificazione di alcune definizioni di i.a. proposte da scienziati, legislatori, e dottrina giuridica, tra gli anni '50 del Novecento e il 2019, v. S. Samoili e al., *AI WATCH. Defining Artificial Intelligence*, EUR 30117 EN – Publications Office of the European Union, 2020; v. anche A. Bertolini, *Study on Artificial Intelligence and Civil Liability*, Study requested by the JURI Committee, July 2020, p. 22 ss. Per un approccio penalistico al problema definitorio v. A. Giannini, *Criminal behavior and Accountability of Artificial Intelligence Systems*, Eleven, 2023, p. 25 ss.

[7] V. per tutti S. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Pearson College Div., 4th ed., 2020, p. 651; P. Domingos, *A Few Useful Things to Know about Machine Learning*, in *Communications of the ACM*, nov. 2012, vol. 55, n. 10, p. 79. In altra sede, abbiamo tentato di individuare una nozione "penalisticamente orientata" di intelligenza artificiale, che comprendesse soltanto i sistemi di i.a. "imprevedibili", e, di conseguenza, potenzialmente dirompenti per le categorie penalistiche, v. B. Fragasso, *La responsabilità penale del produttore di sistemi di intelligenza artificiale*, in *Dir. pen. cont. – Riv. Trim.*, 2023, n. 1, p. 29 ss.

[8] Alludiamo, qui, alla concezione kantiana di "autonomia", intesa come potere della ragione umana di dare a sé stessa una legge morale, v. I. Kant, *Critica della ragion pratica*, Rizzoli, Segrate, 1992 (ed. ted. originale, *Kritik der praktischen Vernunft*, 1788).

ni non intendiamo congetturare una sorta di *capacità morale* o di *autodeterminazione* in capo agli agenti algoritmici, che potrebbe addirittura aprire la strada ad ipotesi di *responsabilità penale diretta* dei sistemi di i.a. [9]: teoria senza dubbio suggestiva, e che meriterebbe ben più spazio di quello che in questa sede possiamo dedicarle, ma che, per l'attuale funzionamento dei sistemi algoritmici – e, dal punto di vista funzionalistico, per la *percezione* che di essi hanno i consociati – non ci pare accoglibile [10]. Piuttosto, più limitatamente, ciò che ci preme sottolineare è che i sistemi di *machine learning* hanno la capacità, già oggi, di emanciparsi dalle informazioni ricevute in sede di programmazione [11], giungendo a soluzioni *innovative* e addirittura, secondo alcuni, *creative* [12], con conseguente fuoriuscita degli *output* artificiali dalla sfera di dominio (e di comprensione) di programmatori e sviluppatori.

Dal punto di vista penalistico, ciò potrà tradursi, evidentemente, in una messa in crisi dei criteri ascrittivi del fatto illecito, che sono tradizionalmente

---

[9] Il principale propugnatore della responsabilità penale diretta in capo ai sistemi di i.a. è il penalista israeliano Gabriel Hallevy; v., tra i vari lavori sul tema, G. Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems*, Springer, Cham, 2015; nello stesso senso, ma con accezioni spesso molto diverse tra loro, v. Y. Hu, *Robot Criminals*, in *U. Mich. J. L. Reform*, vol. 52, 2019, p. 487 ss.; C. Mulligan, *Revenge Against Robots*, in *South Carolina Law Review*, vol. 69, 2018, p. 579; M. Simmler, N. Markwalder, *Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence*, in *Crim. Law Forum*, n. 30, 2019, p. 1 ss.; F. LaGioia, G. Sartor, *AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective*, in *Philosophy & Technology*, vol. 33, 2020, p. 433 ss. Per una completa rassegna di tali posizioni v. A. Giannini, *Criminal behavior and Accountability of Artificial Intelligence Systems*, cit., p. 45 ss.

[10] Per una più approfondita confutazione della tesi della responsabilità penale diretta delle macchine v., per tutti, A. Cappellini, Machina delinquere non potest?, *Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018; C. Piergallini, *Intelligenza artificiale, da 'mezzo' ad 'autore' del reato?*, in *Riv. it. dir. proc. pen.*, n. 4, 2020, p. 1766; L. Greco, *Direito penal para roboˆs? Só se poderá falar em uma pena para roboˆs quando o direito penal deixar de se interessar pelo ser humano*, in *Jota info*, 5 novembre 2021; M.E. Florio, *Il dibattito sulla responsabilità penale diretta delle i.a.: "molto rumore per nulla"?*, in *Sist. pen.*, n. 2, 2024; F.C. La Vattiata, *AI Systems Involved in Harmful Events: Liable Persons or Mere Instruments? An Interdisciplinary and Comparative Analysis*, in *BioLaw Journal*, n. 1, 2023, p. 485 ss.; T. Weigend, *Convicting Autonomous Weapons? Criminal Responsibility of and for AWS under International Law*, in *Journal of International Criminal Justice*, 2023.

[11] Sull'"autonomia artificiale" come capacità di prendere decisioni in situazioni di incertezza v. A. Beckers, G. Teubner, *Three Liability Regimes for Artificial Intelligence*, Bloomsbury, Londra, 2022; G. Teubner, *Soggetti giuridici digitali. Sullo status privatistico degli agenti software autonomi*, Edizioni Scientifiche Italiane, Napoli, 2019, p. 55 ss.

[12] A. Beckers, G. Teubner, *Three Liability Regimes*, cit., p. 40.

fondati sul mancato dominio, da parte dell'agente, di un fatto offensivo effettivamente dominabile [13]. Il fatto che la questione sia delicata sotto il profilo penalistico è, tra l'altro, testimoniato dall'attenzione che l'*Association Internationale de Droit Pénal* sta riservando al tema, tanto da decidere di dedicare proprio ai rapporti tra diritto penale e intelligenza artificiale i lavori del Congresso internazionale tenutosi a Parigi nel giugno 2024 [14].

I problemi di ascrizione della responsabilità penale si porranno soprattutto con riferimento a quei sistemi di i.a., che, nel lasso di tempo in cui sono attivi, sono in grado di agire sul mondo circostante (reale o virtuale) *direttamente*, e che possono, di conseguenza, provocare offese a beni giu-

[13] La letteratura che analizza l'impatto dell'intelligenza artificiale sulla responsabilità penale è ormai di ampiezza considerevole. Senza pretese di esaustività, e limitandoci soltanto ai lavori in lingua italiana, v. C. Piergallini, *Intelligenza artificiale*, cit.; F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, fasc. 10, 2019; F. Consulich, Flash offenders. *Le prospettive di* accountability *penale nel contrasto alle intelligenze artificiali devianti*, in *Riv. it. dir. proc. pen.*, 2022, fasc. 3, p. 1015; I. Salvadori, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, in *Riv. it. dir. proc. pen.*, n. 1, 2021, p. 100; A. Fiorella, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale. Rischio permesso e limiti di autonomia dell'Intelligenza Artificiale*, in R. Giordano e al. (a cura di), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia*, Giuffrè, Milano, 2022, p. 651 ss.; D. Piva, Machina discere, (deinde) delinquere et puniri potest, *ivi*, p. 681 ss.; S. Preziosi, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, *ivi*, p. 713 ss.; A. Cappellini, Machina delinquere non potest*?*, cit.; B. Panattoni, *Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale*, in *Dir. inf.*, fasc. 1, 2021, p. 317; M.B. Magro, *Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica*, in *Leg. pen.*, 10 maggio 2020; A. Giannini, *Intelligenza artificiale,* human oversight *e responsabilità penale: prove d'impatto a livello europeo*, in *Discrimen*, 2022; C. Minelli, *La responsabilità "penale" tra persona fisica e corporation alla luce della Proposta di Regolamento sull'Intelligenza Artificiale*, in *Dir. pen. cont. – Riv. trim*., 2022, n. 2, p. 50 ss.; L. D'amico, *Colpa, precauzione e rischio. le tensioni penalistiche nella moderna era tecnologica*, in *Leg.. pen.*, 21 ottobre 2023; L. Romanò, *La responsabilità penale al tempo di ChatGPT*, in *Dir. pen. cont. – Riv. Trim.*, 2023, n. 1, p. 70; volendo, v. anche B. Fragasso, *La responsabilità penale del produttore di sistemi di intelligenza artificiale*, cit., p. 28; nel settore specifico della circolazione di veicoli a guida autonoma v. M. Lanzi, Self-driving cars *e responsabilità penale*, Giappichelli, Torino, 2023; in ambito medico, N. Amore, *L'effetto della robotica e dell'IA nell'imputazione giuridica degli eventi infausti*, in N. Amore, E. Rossero, *Robotica e intelligenza artificiale nell'attività medica. Organizzazione, autonomia, responsabilità*, Il Mulino, Bologna, p. 244 ss.

[14] Il Congresso ha avuto ad oggetto sia i profili di diritto penale sostanziale, sia quelli di diritto penale processuale. Sui primi – che sono quelli che maggiormente interessano in questa sede – v. la *Resolution on Traditional Criminal Law Categories And AI*, contenuta, insieme al rapporto generale (a cura di L. Picotti) e ad una selezione dei rapporti nazionali, in L. Picotti, B. Panattoni (ed.), *Traditional Criminal Law Categories and AI: Crisi or Paligenesis?*, in *Revue Internationale de Droit Pénal*, 1, 2023; in argomento v. anche L. Picotti, *Intelligenza artificiale e diritto penale: le sfide*, in *Dir. pen. proc.*, n. 3, 2024, p. 293 ss.

ridici tutelati penalmente *senza la mediazione di un essere umano*. È questo il caso, per quanto riguarda le offese all'integrità fisica, degli *embodied AI-systems* (macchine a guida autonoma, droni, *healthcare robots*, ma anche elettrodomestici che utilizzano modelli di *machine learning*), così come di quegli algoritmi di i.a. che – pur non avendo propriamente un *hardware* – incidono su un'infrastruttura fisica (ad es., un sistema di i.a. deputato alla regolazione delle temperature all'interno dell'altoforno di un'acciaieria). Ma tali offese "immediate" possono concernere anche beni giuridici *diversi* dall'integrità fisica: si pensi, ad esempio, alla pubblicazione di post diffamatori o di istigazione all'odio *online* ad opera di *social bots*, o alle sensibili e artificiose alterazioni dei prezzi di strumenti finanziari che possono essere provocate dai sistemi di *algorithmic trading* [15]. In questi esempi, l'imponderabile agire del sistema di i.a., interponendosi tra la condotta del produttore e l'evento lesivo, potrebbe mettere in discussione la stessa integrazione degli elementi del reato.

Di converso, non ci pare che possano avere un effetto altrettanto dirompente sulle categorie del reato quegli algoritmi intelligenti che si limitano a fornire *output* e *raccomandazioni* – pur sempre caratterizzati da un elevato grado di *imprevedibilità* rispetto all'*input* di partenza – che necessitano tuttavia di un filtro umano per incidere concretamente sui beni giuridici. Si pensi, ad esempio, in ambito medico, ai c.d. *Clinical Decision Support Systems* (CDSS) [16], dispositivi di *machine learning* che, analizzando enormi quantità di dati concernenti la *medical knowledge* e la storia clinica del paziente, sono in grado di fornire indicazioni diagnostiche e terapeutiche, utilizzabili dai medici competenti per la decisione clinica. In questo caso, si porrà il problema di stabilire *fino a che punto* il medico dovrà affidarsi alla raccomandazione algoritmica: un interrogativo che ricorda da vicino il dibattito sui limiti dell'osservanza alle linee guida e che, pur nella sua complessità, non ci sembra tale da poter paralizzare strutturalmente l'imputazione colposa [17].

---

[15] Sulle implicazioni penalistiche del fenomeno dell'*algorithmic trading* v. F. Consulich, *Il nastro di Möbius. intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato*, in *Banca borsa*, 2018, vol. 71, n. 2, p. 195 ss.; A.F. Tripodi, *Uomo*, societas, machina, in *Leg.. pen.*, 10 maggio 2023.

[16] In argomento v. W.N. Price II, *Regulating Black-Box Medicine*, 116 *Mich. L. Rev.*, 2017, p. 421; J.S. Allain, *From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artifificial Intelligence Systems*, in 73 *La. L. Rev.*, 2013, p. 1049; F. Lagioia, G. Contissa, *The Strange Case of Dr. Watson: Liability Implications of AI Evidence-Based Decision Support Systems in Health Care*, in *European Journal of Legal Studies*, vol. 12, n. 2, 2020, p. 245 ss.

[17] Per un primo inquadramento del problema v. M. Caputo, (voce) *Colpa medica*, in M. Donini (diretto da), *Reato colposo*, *Enc. dir. – I Tematici*, Giuffrè, Milano, 2022, p. 197 ss.; A. Pe-

Si pensi, poi, sempre in relazione a sistemi intelligenti che non hanno capacità di azione *immediata*, ai sistemi di i.a. *generativa* [18]: emblematica, in questo senso, è la casistica relativa alla creazione e alla diffusione di *porn deepfakes* o all'utilizzo di codici *malware* predisposti, su specifico *input*, da ChatGPT [19]. L'effetto dirompente dell'intelligenza artificiale, in questi casi, si apprezza comunque, ma in termini *quantitativi*, invece che *qualitativi*, nella misura in cui determina un incremento esponenziale delle capacità criminali degli agenti, offrendo loro sofisticatissimi strumenti tecnologici a basso costo: una "democratizzazione" del crimine che ha già destato allarme tra le autorità di *enforcement* [20], e che richiederà sicuramente un impegno preventivo da parte delle società che sviluppano e gestiscono tali servizi, nell'ambito di un'ormai imprescindibile politica di collaborazione tra *big tech* e autorità di controllo nella prevenzione degli illeciti [21].

Le osservazioni contenute nei prossimi paragrafi saranno svolte pensando alla prima delle categorie citate – quella che concerne il *ruolo immediato* dei sistemi di i.a. nell'offesa di beni giuridici tutelati penalmente – e si concentreranno, più in particolare, sull'ascrivibilità al produttore [22] dei

---

RIN, *Standardizzazione, automazione e responsabilità medica. Dalle recenti riforme alla definizione di un modello d'imputazione solidaristico e liberale*, in *BioLaw Journal*, 2019, n. 1, p. 207.

[18] I sistemi di i.a. generativa sono modelli di *deep learning* in grado di generare contenuti (testi, immagini, video, musica etc.), in risposta agli *input* degli utenti (c.d. *prompt*); v. A. ZEWE, *Explained: Generative AI*, in *MIT News*, 9 novembre 2023.

[19] In argomento v. G. FIORINELLI, *Il concorrente virtuale: la prevenzione dell'uso di ChatGPT per finalità criminali tra etero- e auto-regolazione*, in *Riv. it. med. leg.*, n. 2, 2023

[20] EUROPOL, *ChatGPT. The impact of Large Language Models on Law Enforcement*, 19 dicembre 2023.

[21] Ci riferiamo, in particolare, al Regolamento (UE) 2022/2065 del 19 ottobre 2022 (c.d. *Digital Service Act*), che obbliga le piattaforme *online* a collaborare con le istituzioni nel contrasto della pubblicazione e la diffusione di contenuti illegali; in argomento v. i contributi pubblicati nella sezione monografica contenuta nel fascicolo 3/2023 di *Medialaws* (a cura di A. Gullo), *Il* Digital Services Act *e il contrasto alla disinformazione: responsabilità dei* provider*, obblighi di* compliance *e modelli di* enforcement, p. 13 ss. È dubbio, in ogni caso, che tale Regolamento sia applicabile a fornitori di servizi come ChatGPT, v. in proposito C. BURCHARD, *Das Pro und Contra für Chatbots in Rechtpraxis und Rechtsdogmatik*, in *Computer Recht*, n. 2, 2023, n. 2, p. 132 ss., trad it. a cura di V. MANES, *I pro e i contro dei chatbot nella prassi legale e nella dogmatica giuridica. Un contributo critico sulla funzione del diritto e della scienza (giuridica): state ancora argomentando o state già chattando?*, in *Ordines*, n. 1, 2023, p. 339.

[22] Premettiamo fin da ora che, per comodità espositiva, in questo contributo faremo sempre riferimento alla figura del "produttore", intendendo tuttavia includere, in tale espressione, tutte quelle persone che, a vario titolo, contribuiscono ai processi di *sviluppo, progettazione e commercializzazione* dei dispositivi intelligenti. Va da sé, ovviamente, che l'individuazione del soggetto personalmente responsabile – tra coloro che fanno parte del ciclo *lato sensu* produttivo –

reati colposi causalmente orientati (omicidio e lesioni colpose). Ciò non toglie che molte di tali considerazioni possano essere estese anche alla posizione del produttore di *recommendation systems* e di sistemi di i.a. generativa.

2. *L'approccio europeo all'imprevedibilità algoritmica, tra normativa sulla sicurezza e modelli di responsabilità oggettiva per il produttore*

Innanzitutto, può essere utile accennare, in maniera inevitabilmente cursoria, agli indirizzi politici che hanno caratterizzato, fino ad ora, la regolamentazione dell'intelligenza artificiale.

Di fronte a sistemi tecnologici di cui non si conoscono integralmente i rischi, e i cui *output* non sono pienamente comprensibili, il decisore pubblico avrebbe potuto adottare un approccio precauzionale "puro"[23], imponendo un'astensione dalla produzione e dalla commercializzazione dei sistemi di i.a., fino a che non fossero emerse maggiori garanzie sulla sua "affidabilità" e, soprattutto, controllabilità[24]. Come noto, invece, così non è stato – e non è certo questa la sede per interrogarsi sulla concreta praticabilità di una scelta alternativa, in un mondo sempre più interconnesso e globalizzato, in cui è in atto una vera e propria "corsa all'intelligenza artifi-

---

incontrerà difficoltà specifiche, determinate dal problematico accertamento della specifica *causa* dell'evento lesivo (es. difetto di programmazione o di addestramento o di installazione, etc.) e della persona responsabile all'interno delle organizzazioni complesse.

[23] In generale, sul principio di precauzione come criterio di gestione del rischio v. C. Sunstein, *Laws of Fear. Beyond the precautionary principle*, Cambridge University Press, Cambridge, 2005. M.A. Geistfeld, *Reconciling Cost-Benefit Analysis with the Principle That Safety Matters More Than Money*, in 76 *New York University Law Review*, April 2001, p. 173 ss. Sul rapporto tra principio di precauzione e diritto penale si vedano, per tutti, gli studi monografici di D. Castronuovo, *Principio di precauzione e diritto penale. Paradigmi dell'incertezza nella struttura del reato*, Aracne, Roma, 2012; E. Corn, *Il principio di precauzione nel diritto penale. Studio sui limiti all'anticipazione della tutela penale*, Giappichelli, Torino, 2013; F. Consorte, *Tutela penale e principio di precauzione*, Giappichelli, Torino, 2013; v. anche C. Ruga Riva, *Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica*, in E. Dolcini, C.E. Paliero (a cura di), *Studi in onore di Giorgio Marinucci*, II, Giuffrè, Milano, 2006, p. 1743 ss.

[24] Si v., a tal proposito, la proposta di moratoria, firmata da decine di esperti del settore, sullo sviluppo di sistemi di i.a., Future of Life Institute, *Pause Giant AI Experiments: An Open Letter*, 22 marzo 2023; sul rapporto tra principio di precauzione e sviluppo dei sistemi di i.a., v. N. Amore, *L'effetto della robotica e dell'IA*, cit., p. 244 ss.; L. D'amico, *Colpa, precauzione e rischio*, cit., *passim*.

ciale" [25]. Piuttosto, ciò che qui ci preme evidenziare – pur nell'apparente banalità di tale constatazione – è che i decisori pubblici, in Europa così come altrove, hanno permesso lo sviluppo e la commercializzazione di tecnologie [26] che, nelle parole di alcuni dei più importanti esperti del settore, "nemmeno i loro creatori sono in grado di comprendere, prevedere o controllare in modo affidabile" [27]. Non ci interessa, qui, esprimere un giudizio di valore su tale scelta; ci sembra, tuttavia, rilevante mettere in luce come le autorità pubbliche, nella definizione del bilanciamento di interessi tra esercizio di attività pericolose utili e protezione dell'integrità personale dei cittadini, abbiano implicitamente accettato il possibile verificarsi di *eventi lesivi* derivanti dall'*imprevedibilità* dei sistemi di i.a.

Proprio il concetto di "rischio" è, d'altra parte, al centro dal Regolamento sull'intelligenza artificiale dell'Unione Europea (c.d. *AI Act*) [28], che detta i requisiti di sicurezza che ciascun sistema di i.a. dovrà soddisfare per poter essere introdotto sul mercato. In particolare, l'AI Act divide i dispositivi di intelligenza artificiale in tre categorie – a seconda del rischio che questi pongono per la sicurezza degli utenti ed il rispetto dei diritti fondamentali dei cittadini –, individuando per ciascuna categoria un regime giuridico applicabile: così, i sistemi di i.a. che presentano rischi considerati "inaccettabili" non potranno essere introdotti nel mercato europeo (art. 5); i sistemi ad alto rischio (artt. 6 e 7) potranno essere commercializzati, purché rispettino una serie di requisiti (artt. 8-15); per i sistemi a rischio limi-

---

[25] C. Weise, K. Metz, *The race to dominate A.I.*, in *The New York Times*, 8 dicembre 2023.

[26] L'Unione Europea, già a partire dalla Comunicazione del 2018 intitolata "L'intelligenza artificiale per l'Europa" (SWD(2018) 137 final}, 25 aprile 2018), ha tentato di assumere un ruolo di *leadership* nella regolamentazione dei sistemi di i.a. La *policy* europea, in particolare, trova il suo fulcro nel Regolamento 2024/1689 sull'intelligenza artificiale (c.d. *AI Act*), pubblicato sulla Gazzetta Ufficiale dell'Unione Europea il 12 luglio 2024 ed entrato in vigore il primo agosto 2024. Il provvedimento inizierà ad applicarsi il 2 agosto 2026 (salvo che per alcune norme, che si inizieranno ad applicare prima o dopo, v. art. 113, *AI Act*). Anche al di fuori dell'Unione Europea non mancano le iniziative legislative in materia di intelligenza artificiale: si pensi, ad esempio, all'*Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* firmato dal Governo statunitense il 30 ottobre 2023, o all'*Artificial Intelligence (Regulation) Bill [HL]* che è in corso di discussione al Parlamento inglese, o, ancora, ai vari provvedimenti che sono stati adottati dal Governo cinese (v. a tal proposito M. Sheehan, *Tracing the Roots of China's AI Regulations*, pubblicato sul sito del *Carnegie Endowment for International Peace*, 27 febbraio 2024).

[27] Future of Life Institute, *Pause Giant AI Experiments*, cit., in cui si parla di «*powerful digital minds that no one – not even their creators – can understand, predict, or reliably control*».

[28] *Regolamento che stabilisce regole armonizzate sull'Intelligenza artificiale (Legge sull'intelligenza artificiale)*, cit.

tato (art. 50), infine, basterà il rispetto di alcuni obblighi di trasparenza. Disposizioni specifiche sono inoltre previste per i *modelli di i.a. per finalità generali* (art. 51 ss.), ossia per quei modelli che sono «in grado di svolgere con competenza un'ampia gamma di compiti distinti» (art. 3, n. 63).

Rinviamo al testo del Regolamento – e all'ampia letteratura che ne ha già dato un primo commento [29] – per la precisa individuazione dei criteri che consentono di distinguere le varie categorie di sistemi di i.a., nonché dei requisiti di commercializzazione per i sistemi a rischio alto e limitato. Il rilevo che qui ci pare importante mettere nuovamente in evidenza è che la normativa europea – nonostante costituisca la strategia di *governance* dell'intelligenza artificiale che, a livello globale, maggiormente si pone l'obiettivo di tutelare le persone dai rischi derivanti dallo sviluppo tecnologico – *non potrà impedire*, in assoluto, che i sistemi di intelligenza artificiale, *pur conformi* rispetto alla normativa sulla sicurezza, cagionino offese a beni giuridici [30].

Dal punto di vista civilistico, di tali danni potrà essere chiamato a rispondere il produttore – sempre che il prodotto sia difettoso, ovvero presenti un'anomala condizione di pericolosità [31]. In particolare, è appena sta-

---

[29] Senza pretese di esaustività, v. L. Floridi, *The European Legislation on AI: a Brief Analysis of its Philosophical Approach*, in *Philosophy & Technology*, n. 34, 2021, p. 215 ss.; M. Ebers, *Standardizing AI, The Case of the European Commission's Proposal for an 'Artificial Intelligence Act'*, in L.A. DiMatteo e al. (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, , 2022; nella letteratura italiana v. G. Alpa, *Quale modello normativo europeo per l'intelligenza artificiale?*, in *Contr. impr.*, n. 4, 2021, p. 1003 ss.; U. Ruffolo, A. Amidei, *La regolazione ex ante dell'intelligenza artificiale tra gestione del rischio by design, strumenti di certificazione preventive e «autodisciplina» di settore*, in A. Pajno e al. (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. 1, Il Mulino, Bologna, 2022, p. 489 ss.; per i profili di interesse penalistico v. F.C. La Vattiata, *Brevi note "a caldo" sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale*, in *Dir. pen. uomo*, 30 giugno 2021. Va comunque sottolineato che l'AI Act si applicherà ad una gamma di tecnologie molto più ampia di quella oggetto del presente contributo. In particolare, a marcare – in senso estensivo – l'ambito di applicazione del Regolamento è la norma che definisce il di "sistema di i.a." come «un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali» (art. 3, § 1, punto 1).

[30] Per una lettura dell'*AI Act* come perimetro di riferimento del "rischio consentito" v. J. Laux e al., *Trustworthy artificial intelligence and the European UnionAI act: On the conflation of trustworthiness and acceptability of risk*, in *Regulation & Governance*, vol. 18, 2024, p. 3 ss.

[31] Vi è tuttavia una parte della dottrina che, in ambito di responsabilità *civile* per danno da prodotto, ritiene che il rispetto degli *standard* tecnici di sicurezza del prodotto basti a far considerare un prodotto "non difettoso", e, dunque, ad escludere la responsabilità oggettiva del

ta approvata una riforma della direttiva sulla responsabilità per danno da prodotto difettoso [32], che, da un lato, estende la nozione di "prodotto" anche ai sistemi di i.a., ampliando, di conseguenza, il regime di responsabilità oggettiva ivi previsto, e, dall'altro, alleggerisce l'onere della prova per il danneggiato, attraverso meccanismi di *disclosure* obbligatoria in capo al produttore. La tendenza, insomma, quantomeno sul piano civilistico, parrebbe quella di cercare di *accollare economicamente al produttore il rischio dell'imprevedibilità del sistema di i.a.*, garantendo così, al consumatore, un rimedio risarcitorio efficace e facilmente esperibile [33].

3. *Sistema di i.a. conforme alla normativa sulla sicurezza: l'efficacia esimente del rischio consentito in materia penale*

È scontato sottolineare che le semplificazioni viste nel precedente paragrafo – sia che si collochino sul piano soggettivo, sia che riguardino lo *standard* probatorio – non possono essere considerate ammissibili nell'ambito dell'accertamento della responsabilità penale. Le offese derivanti dall'attivazione di sistemi di i.a. *conformi* alla normativa di settore dovranno dunque, a nostro parere, essere considerate come rientranti nel perimetro

produttore, in caso di eventi lesivi. In particolare, è Enrico Al Mureden che si è fatto esplicito promotore, in Italia, della c.d. *preemption doctrine* di matrice statunitense; v. E. Al Mureden, *La sicurezza dei prodotti e la responsabilità del produttore*, Giappichelli, Torino, 2017; *contra* E. Bellisario, *Il danno da prodotto conforme tra regole preventive e regole risarcitorie*, in *Eur. dir. priv.*, 2016, fasc. 3, p. 841 ss.

[32] Commissione europea, *Proposta di Direttiva del Parlamento Europeo e del Consiglio sulla responsabilità per danno da prodotti difettosi*, COM (2022) 495 final, 28 settembre 2022, adottata definitivamente dal Parlamento europeo in data 12 marzo 2024. A tale intervento normativo si aggiunge poi la *Proposta di Direttiva relativa all'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale (direttiva sulla responsabilità da intelligenza artificiale)*, COM (2022) 496 final, 28 settembre 2022, ancora in discussione, che prevede un meccanismo di alleggerimento dell'onere della prova e di *disclosure* assimilabile a quello previsto dalla proposta di riforma della direttiva sul danno da prodotto. Per una panoramica sulle due proposte v. E. Bellisario, *Il pacchetto europeo sulla responsabilità per danni da prodotti e da intelligenza artificiale. Prime riflessioni sulle Proposte della Commissione*, in *Danno e responsabilità*, n. 2, 2023, p. 153 ss.

[33] Parte della dottrina civilistica ha tuttavia sostenuto che le tecniche normative predisposte dal legislatore europeo non siano all'altezza degli obiettivi di semplificazione del contenzioso. In argomento v. A. Bertolini, *La responsabilità civile derivante dall'utilizzo di sistemi di intelligenza artificiale: il quadro europeo*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffrè, Milano, 2024.

del "rischio consentito" – salvo che in alcuni casi eccezionali, su cui ci soffermeremo tra un attimo.

Non è di certo possibile, in questo breve intervento, cercare di dare conto di decenni di sofisticate riflessioni teoriche circa la sussistenza e i confini di un perimetro di rischio consentito in materia penale [34]. Basterà qui sottolineare che tale nozione, quale che sia la ricostruzione ermeneutica alla quale si aderisca, risponde all'esigenza di identificare un'area di esenzione dalla responsabilità penale in relazione a condotte che, pur lesive di beni giuridici tutelati penalmente, si ritengono consentite, in forza di un bilanciamento politico, effettuato a priori, tra rischi e benefici dell'esercizio delle attività pericolose. Il concetto di rischio consentito è effettivamente il perno attorno al quale ruota la gran parte degli studi sin qui dedicati al rapporto tra intelligenza artificiale e responsabilità penale [35], offrendo un armamentario ermeneutico già pronto – benché non immune da divergenze interpretative – per mettere al riparo il produttore da un'eccessiva dilatazione dell'imputazione colposa.

In particolare, in una delle concezioni maggioritarie in dottrina, l'area

---

[34] La letteratura sul tema è davvero sterminata; si v., per tutti, G. Forti, *Colpa ed evento nel diritto penale*, Giuffrè, Milano, 1990, p. 250 ss.; Id., *"Accesso" alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione"*, in *Criminalia*, 2006, p. 155 ss.; C. Piergallini, *Danno da prodotto e responsabilità penale*, cit., *passim*; Id., *Il paradigma della colpa nell'età del rischio: prove di resistenza del tipo*, in *Riv. it. dir. proc. pen.*, p. 1684 ss.; V. Militello, *Rischio e responsabilità penale*, Giuffrè, Milano, 1988, p. 55 ss. (che preferisce, tuttavia, l'espressione "rischio adeguato"); F. Consulich, (voce) *Rischio consentito*, in M. Donini (diretto da), *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 1102 ss.; S. Zirulia, *Esposizione a sostanze tossiche e responsabilità penale*, Giuffrè, Milano, 2018, p. 335 ss.; M. Donini, *Il volto attuale dell'illecito penale*, cit., p. 119 ss.; A. Massaro, Principio di precauzione e diritto penale: *nihil novi sub sole?*, in *Dir. pen. cont.*, 2011; D. Pulitanò, *Colpa ed evoluzione del sapere scientifico*, in *Dir. pen. e proc.*, 2008, p. 647.

[35] V., con vari accenti, L. Picotti (ed.), *Resolution on Traditional Criminal Law Categories And AI*, cit., §§ 12.b e 17; S. Gless, E. Silverman, T. Weigend, *If Robots Cause Harm, Who Is to Blame?*, in *New Criminal Law Review*, 2016, pp. 430-431; C. Piergallini, *Intelligenza artificiale*, cit., p. 1750; I. Salvadori, *Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale*, cit., p. 116 ss.; V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in U. Ruffolo (a cura di), *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Giuffrè, Milano, 2020, p. 5; A. Cappellini, Machina delinquere non potest*?*, cit., p. 19; A. Fiorella, *Responsabilità penale dei Tutor e dominabilità dell'Intelligenza Artificiale*, cit., p. 656 ss.; D. Piva, Machina discere, (deinde) delinquere et puniri potest, cit., p. 681 ss.; M. Lanzi, *Self-driving cars* e responsabilità penale, cit., p. 266 ss.; L. D'Amico, *Colpa, precauzione e rischio*, cit., *passim*; C. Grandi, *Positive Obligations (Garantestellung) Grounding Criminal Responsibility for not Having Avoided an Illegal Result Connected to the AI Functioning*, in L. Picotti, B. Panattoni (ed.), *Traditional Criminal Law Categories and AI*, cit., p. 73.

di "rischio consentito" delinea quelle ipotesi in cui il rispetto delle norme cautelari *codificate* impedisce che possa essere mosso all'imputato un rimprovero per non aver osservato norme di diligenza, prudenza, perizia *non codificate*[36]. La necessità di una tale limitazione nella valutazione dell'elemento colposo parrebbe giustificata dal fatto che, in settori particolarmente rischiosi e ad alta complessità tecnico-normativa, gli *standard* cautelari positivizzati, oltre che una funzione di *prevenzione* rispetto al verificarsi dell'evento lesivo, esercitano altresì una funzione di *garanzia* e di *orientamento* nei confronti dell'agente, che risulterebbe frustrata laddove al produttore potesse contestarsi la violazione di una norma cautelare non scritta e, quindi, per sua natura *indeterminata*.

Ovviamente, tuttavia, l'emanazione di regole cautelari specifiche per la messa in commercio di dispositivi intelligenti potrebbe acquisire la citata funzione di orientamento soltanto qualora la norma positivizzata fornisse *standard* di comportamento rigidi ed esaustivi, dal momento che le cautele elastiche si fondano sul medesimo meccanismo ricostruttivo utilizzato per la colpa generica, ossia il riferimento all'agente modello[37]. L'*AI Act*, da questo punto di vista, non pare fornire soluzioni del tutto soddisfacenti, dal momento che diverse norme fanno riferimento ad obbligazioni che il produttore dovrà adempiere "nella misura del possibile" o "per quanto possibile"[38] – una vaghezza, d'altronde, che appare inevitabile, trattandosi di un testo che regola, con approccio orizzontale, un settore in continua trasformazione.

Piuttosto, la definizione di norme cautelari rigide e settoriali sarà demandata agli organismi di standardizzazione. Tra questi, un ruolo fondamentale sarà sicuramente svolto, da un lato, da CEN (Comitato europeo di normalizzazione) e CENELEC (Comitato europeo di normalizzazione elettrotecnica), gli enti di normalizzazione ai quali la Commissione europea ha richiesto la standardizzazione, sulla base dei criteri generali espressi nell'*AI*

---

[36] G. Forti, *Colpa ed evento*, cit., p. 457; C. Piergallini, *Attività produttive e imputazione per colpa: prove tecniche di "diritto penale del rischio"*, in *Riv. it. dir. proc. pen.*, fasc. 4, 1997, p. 1492; L. Stortoni, *Angoscia tecnologica ed esorcismo penale*, in *Riv. it. dir. proc. pen.*, fasc. 1, 2004, p. 79; F. Palazzo, *Morti da amianto e colpa penale*, in *Dir. pen. proc.*, n. 2, 2011, p. 188.

[37] D. Castronuovo, *Responsabilità da prodotto e struttura del fatto colposo*, in *Riv. it. dir. proc. pen.*, 2005, p. 328; F. Palazzo, *Morti da amianto e colpa penale*, cit., p. 189.

[38] V., ad esempio, l'art. 9, § 5, lett. a), che stabilisce che il sistema di gestione dei rischi debba garantire l'eliminazione o la riduzione dei rischi "per quanto possibile"; o ancora, l'art. 10, § 3, che prevede che «[i] set di dati di addestramento, convalida e prova sono pertinenti, sufficientemente rappresentativi e, nella misura del possibile, esenti da errori e completi nell'ottica della finalità prevista».

Act[39], e, dall'altro, dal *focus group* congiunto costituito da ISO (*International Organization for Standardization*) e IEC (*International Electrotechnical Commission*), finalizzato alla pubblicazione di *standard* per lo sviluppo e la produzione di sistemi di intelligenza artificiale[40].

In ogni caso, un margine di operatività di un rimprovero a titolo di colpa generica dell'agente dovrebbe comunque ritenersi configurabile nel caso in cui le regole cautelari manifestino *segnali univoci* di fallimento "non preventivato"[41] – ossia quando esse si rivelino evidentemente inidonee al raggiungimento degli obiettivi di contenimento del rischio prefissati dal legislatore. In questi casi, la diligenza dovrebbe imporre di adottare cautele ulteriori; il che, nella prospettiva del produttore di un sistema di i.a., potrebbe significare: il fornire informazioni aggiuntive al consumatore, l'"aggiornamento" dell'algoritmo, o, nei casi più gravi, il ritiro o il richiamo del prodotto dal mercato[42].

Non possiamo non sottolineare, a tal proposito, come sia tutt'altro che peregrina la prospettiva che le regole cautelari scritte si rivelino, in concreto, manifestamente inadatte a soddisfare gli scopi di minimizzazione del rischio fissati dal decisore pubblico, data la rapidità con la quale continuano ad emergere nuovi rischi associati all'utilizzo dei sistemi di i.a. Se è vero che la regolazione si è spesso trovata a rincorrere i progressi scientifici, è infatti sotto gli occhi di tutti come l'attuale sfasatura tra velocità dell'evoluzione tecnologica e passo pachidermico del processo normativo (specie a

---

[39] La Commissione Europea, il 5 dicembre 2022, ha pubblicato la bozza di "richiesta di standardizzazione" nei confronti di CEN e CENELEC. Una volta adottata la proposta, gli *standard* dovrebbero essere pubblicati entro il 31 gennaio 2025 (dunque prima della data in cui comincerà ad applicarsi l'*AI Act*). In argomento v. il report pubblicato dall'Oxford Information Labs, a cura di M. McFadden e al., *Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation*, dicembre 2021. Sul funzionamento del processo europeo di standardizzazione si rinvia a Commissione europea, *La guida blu all'attuazione della normativa UE sui prodotti 2022*, 29 giugno 2022, 2022/C 247/01.

[40] Il *focus group* è denominato ISO/IEC JTC 1/SC 42. Tra gli *standard* di più recente pubblicazione v. ISO/IEC 5339:2024 (*Information technology – Artificial intelligence – Guidance for AI applications*) e ISO/IEC TR 5469:2024 (*Artificial intelligence – Functional safety and AI systems*).

[41] Parlano di un eventuale "fallimento" delle norme cautelari, tra gli altri, G. Forti, *Colpa ed evento*, cit., p. 671 ss.; P. Veneziani, *Regole cautelari "proprie" ed "improprie" nella prospettiva delle fattispecie colpose causalmente orientate*, Cedam, Padova, 2003, p. 61 ss., S. Zirulia, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 380.

[42] Sulle peculiari connotazioni che potrebbe assumere la valutazione del rischio consentito – e del suo "fallimento" – nel settore della produzione di sistemi di i.a., sia consentito il rinvio a B. Fragasso, *La responsabilità penale del produttore di sistemi di intelligenza artificiale*, cit., p. 35 ss.

livello europeo) rischi di frustrare ogni tentativo di gestione *ex ante* del rischio algoritmico [43].

4. *Sistema di i.a. difforme rispetto alla normativa sulla sicurezza: quale spazio per l'imputazione dell'evento lesivo imprevedibile?*

Quanto detto fino ad ora dovrebbe suggerire che *al di fuori* del perimetro normativo relativo allo sviluppo e alla commercializzazione dell'intelligenza artificiale possano delinearsi delle ipotesi di responsabilità, in capo al produttore di sistemi di i.a., per i reati di omicidio o lesioni colpose. Questa, effettivamente, dovrebbe costituire la vera ragione d'essere di una riflessione sul rischio consentito: l'esercizio di un'attività pericolosa che si svolge *all'interno dei limiti di rischio socialmente accettati* – così come definiti a livello politico-normativo – dovrebbe essere considerata tendenzialmente esente da profili di responsabilità penale; l'attività industriale-produttiva che, invece, è esercitata *in violazione di tali limiti* potrà invece essere soggetta ad un rimprovero penale, qualora ne ricorrano i presupposti. L'identificazione di un'area di rischio consentito, insomma, dovrebbe esercitare una funzione di *incentivo* al rispetto degli *standard* cautelari; e, parallelamente, la possibilità di essere soggetti ad una sanzione penale dovrebbe avere una funzione di *disincentivo* (o meglio, per utilizzare la terminologia consueta in ambito penalistico, di *deterrenza*).

Il problema è che, anche al di fuori del perimetro del rischio consentito legalmente determinato, le concrete possibilità di accertare una responsabilità penale del produttore per l'evento lesivo algoritmico si preannunciano – a meno di non voler flessibilizzare oltre misura le categorie del diritto penale – piuttosto scarse, con conseguente rischio di vanificare la funzione general-preventiva dell'ordinamento penale [44]. Se è vero che di una "crisi"

---

[43] Basti pensare che, stando alla prima versione dell'*AI Act*, i *Large Language Models* (come, ad esempio, ChatGPT) sarebbero rientrati nella categoria dei sistemi di i.a. a rischio moderato, con conseguente applicazione di un regime di adempimenti minimi. Dopo l'emersione, sul finire del 2022, degli enormi rischi legati all'utilizzo di ChatGPT, il legislatore europeo è corso ai ripari, e l'*AI Act* prevede oggi un intero Capo (il quinto, artt. 51 ss.), dedicato agli obblighi per i fornitori di "modelli di i.a. per finalità generali", tra i quali rientrerà anche ChatGPT. Non è, tuttavia, difficile immaginare che nel prossimo futuro possano emergere sistemi di i.a. altrettanto *disruptive* come lo è stato ChatGPT, con il rischio che l'*AI Act* possa rivelarsi presto obsoleto.

[44] È quello che in dottrina è conosciuto come "*responsibility gap*", v. A. MATTHIAS, *The responsibility gap: Ascribing responsibility for the actions of learning automata*, in *Ethics and Information Technology*, vol. 6, 2004, p. 175 ss.

del diritto penale d'evento si parla ormai da tempo in dottrina – già in relazione alla prima emersione di quei fenomeni sistemici, globalizzati ed incerti che caratterizzano la post modernità [45] –, il rischio è che tale crisi sia accelerata e aggravata dal diffondersi di prodotti "intelligenti", che, per loro natura (!), presentano caratteri di imponderabilità e imperscrutabilità.

Già sul piano oggettivo, la caratteristica opacità dell'intelligenza artificiale potrebbe ostacolare la verifica circa la sussistenza di un nesso eziologico tra condotta umana ed evento lesivo. Ad oggi, infatti – e, stando, alle dichiarazioni degli scienziati su un'incipiente crisi di comprensione dei modelli computazionali più avanzati, sempre di più anche in futuro – non esiste un consolidato corredo nomologico che consenta di *spiegare* il comportamento dei sistemi di i.a., tanto che sono gli stessi scienziati, talvolta, a riferirsi alle attività algoritmiche come ad un fenomeno quasi magico [46]. Questo, ovviamente, potrebbe porre problemi di compatibilità con il paradigma nomologico-deduttivo dell'accertamento causale, che, anche nella sua versione *post Franzese*, richiede pur sempre l'individuazione di una legge scientifica di copertura [47]. C'è addirittura chi, in dottrina, avanza l'ipotesi che in talune situazioni – in cui il sistema di i.a. agisce in maniera del tutto *estranea* rispetto all'addestramento ricevuto – il *decision making* algoritmico possa interrompere il nesso di causalità, ai sensi dell'art. 41, co. 2, c.p. [48].

Ma è soprattutto sul piano della colpa che si colgono le principali difficoltà nell'accertamento di una responsabilità penale del produttore. Anche a voler tacere delle complessità connesse all'identificazione del soggetto

---

[45] Si vedano, per tutti, F. Stella, *Giustizia e modernità*, cit.; C. Piergallini, *Danno da prodotto e responsabilità penale*, cit.; A. Di Martino, *Danno e rischio da prodotti. Appunti per la rilettura critica di un'esperienza giurisprudenziale italiana*, in R. Bartoli (a cura di), *Responsabilità penale e rischio nelle attività mediche e d'impresa: un dialogo con la giurisprudenza*, Firenze University press, Firenze, 2010, p. 437 ss.

[46] Così, ad esempio, afferma Pedro Domingos, uno dei più noti ricercatori in materia di *machine learning*: «*developing successful machine learning applications requires a substantial amount of 'black art' that is difficult to find in textbooks*», v. P. Domingos, *A Few Useful Things to Know about Machine Learning*, cit., p. 78.

[47] Cass., sez. un., 11 luglio 2022, n. 30328, Franzese. Sull'"eredità" della sentenza Franzese – e sulle questioni ancora aperte nell'accertamento della causalità in materia penale – v. i contributi contenuti nel *focus* dedicato ai "Vent'anni dalla sentenza Franzese", in *Riv. it. med. leg.*, n. 4, 2022, p. 963 ss.

[48] S. Preziosi, *La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati*, cit., p. 716 ss.; *contra* C. Piergallini, *Intelligenza artificiale*, cit., pp. 1761-1762; M. Lanzi, Self-driving cars *e responsabilità penale*, cit., p. 238 ss.; S. Gless, E. Silverman, T. Weigend, *If robots cause harm, who is to blame?*, cit., pp. 432-433.

personalmente responsabile all'interno delle strutture societarie[49], residuerebbe comunque la questione di fondo circa la possibilità di rimproverare al produttore un *evento lesivo concretamente imprevedibile*[50]. Nello specifico, particolarmente tortuoso potrebbe risultare l'accertamento del duplice nesso tra colpa ed evento.

Prendiamo, ad esempio, il caso di un incidente mortale che abbia coinvolto un veicolo a guida autonoma, in relazione al quale si riesca a stabilire che a determinare il malfunzionamento del veicolo sia stato il mancato riconoscimento di un segnale di stop da parte del sistema di *image recognition*. In un'ipotesi del genere, il giudice – aderendo ad una concezione estensiva del giudizio di prevedibilità dell'evento, ormai invalsa in giurisprudenza, sebbene non esente da criticità[51] – potrebbe, in taluni casi, considerare provato il *nesso di concretizzazione del rischio* tra regola cautelare violata ed evento lesivo. Tale nesso, per restare al nostro esempio, potrebbe essere ritenuto sussistente laddove il produttore abbia violato gli *standard* in materia di addestramento del sistema di i.a., predisposti al fine di garantire l'accuratezza delle predizioni algoritmiche – e, dunque, volendo portare alle estreme conseguenze il processo di "astrazione" nella ridescrizione dell'evento[52], di evitare incidenti stradali.

Ma ancora più problematica pare la questione dell'accertamento dell'efficacia impeditiva della condotta alternativa conforme al dovere. Entrano qui in gioco l'*imprevedibilità* e l'*opacità* dei sistemi di *machine learning*: ri-

---

[49] Sul c.d. *many hands problem*, nel settore dell'intelligenza artificiale, v. M. Lanzi, Self-driving cars *e responsabilità penale*, cit., p. 256 ss.; A. Giannini, J. Kwik, *Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of Ai and Autonomous Vehicles*, in *Criminal Law Forum*, vol. 34, 2023, p. 58; più in generale, sulla difficoltà di individuare il soggetto personalmente responsabile all'interno delle organizzazioni complesse v., per tutti, A. Gargani, *Posizioni di garanzia nelle organizzazioni complesse: problemi e prospettive*, in *Riv. trim. dir. pen. econ.*, 2017, n. 3-4, p. 515 ss.

[50] Così A. Cappellini, *Reati colposi e tecnologie dell'intelligenza artificiale*, in G. Balbi e al. (a cura di), *Diritto penale e intelligenza artificiale. "Nuovi scenari"*, Giappichelli, Torino, 2022, p. 26. Sul concetto di *prevedibilità genericamente prevedibile* v. anche C. Piergallini, *Intelligenza artificiale*, cit., p. 1762; v. anche S. Beck, *Intelligent agents and criminal law – Negligence, diffusion of liability and electronic personhood*, in *Robotics and Autonomous Systems*, vol. 86, 2016, cit., p. 139; M. Lanzi, Self-driving cars *e responsabilità penale*, cit., *passim*, spec. p. 170.

[51] Per una ricostruzione critica di tale orientamento giurisprudenziale si v., per tutti, G. Civello, (voce) *Prevedibilità e reato colposo*, in M. Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 1017; C. Piergallini, *Il paradigma della colpa nell'età del rischio*, cit., p. 1692.

[52] In questo senso, G. Civello, (voce) *Prevedibilità e reato colposo*, cit., p. 1017; sulla patologica ridescrizione dell'evento *ex post* v. anche S. Zirulia, *Esposizione a sostanze tossiche e responsabilità penale*, cit., p. 322; C. Piergallini, *Il paradigma della colpa nell'età del rischio*, cit., p. 1692.

manendo al nostro esempio, come si potrà dimostrare, infatti, che un *training* conforme agli *standard* del settore avrebbe potuto impedire l'incidente stradale, se gli stessi ricercatori faticano a comprendere e a spiegare il *decision making* algoritmico?[53] Trattandosi di modelli di tipo probabilistico, e non deterministico, sarà difficile affermare che l'errore algoritmico (e quindi, in ipotesi, l'evento lesivo) si sia verificato *a causa* della violazione della regola cautelare e che esso non rientri, invece, nell'ordinario tasso di errore dell'algoritmo. Il problema si porrà soprattutto quando la deviazione rispetto alla norma cautelare positivizzata è minima (si pensi, per restare al nostro esempio, all'addestramento di un sistema di i.a. con un *dataset* leggermente diverso – in termini quantitativi o qualitativi – rispetto a quanto richiesto dagli *standard* di settore).

Di fronte alla paralizzazione dei meccanismi imputativi del diritto penale d'evento, insomma, emerge quell'interrogativo che è ben noto agli studiosi che si occupano del diritto penale nell'era della complessità: dobbiamo rassegnarci ad un certo grado di ineffettività del diritto penale[54] – inadatto ad attagliarsi ai fenomeni lesivi derivanti dal progresso tecnologico – oppure è auspicabile un "adattamento" del diritto penale al "Mondo Nuovo", nel tentativo di conservarne la funzione general-preventiva[55]? Quest'ultima prospettiva, se attuata in relazione ai reati di evento, potrebbe sfociare, come già accaduto in passato, in un'estrema flessibilizzazione della categoria colposa, con progressivo affievolimento del rigore metodologico che il principio di legalità richiederebbe per la dimostrazione della responsabilità penale.

---

[53] Solleva tale problema, nello specifico settore delle auto a guida autonoma, M. Lanzi, Self-driving cars *e responsabilità penale*, cit., p. 247.

[54] Sull'effettività nel diritto penale imprescindibile è il riferimento a C.E. Paliero, *Il principio di effettività del diritto penale*, in *Riv. it. dir. proc. pen.*, 1990, p. 430.

[55] Tale interrogativo accompagna, talvolta esplicitamente, talvolta sullo sfondo, molte delle riflessioni sul c.d. "diritto penale del rischio", v. L. Stortoni, *Angoscia tecnologica ed esorcismo penale*, cit., p. 80; A. Perin, *La crisi del "modello nomologico" fra spiegazione e prevedibilità dell'evento nel diritto penale. Note introduttive e questioni preliminari sul fatto tipico colposo*, in *Riv. it. dir. proc. pen.*, 2014, fasc. 3, p. 1387; L.A. Zapatero, *Introduzione*, in L. Stortoni, L. Foffani (a cura di), *Critica e giustificazione del diritto penale nel cambio di secolo*, cit., p. 17: «Si pongono due opzioni di fondo: o scavare trincee contro il "moderno" diritto penale, o tentare di montare e cavalcare questo cavallo bizzoso dei moderni fenomeni materiali che premono sul sistema penale».

5. Dal disvalore di evento al disvalore di azione? Sulla negligente gestione del rischio da intelligenza artificiale

È troppo presto, ci pare, per stabilire se l'imprevedibilità e l'opacità dei sistemi intelligenti costituiranno un *limite strutturale* al riconoscimento di una responsabilità penale del produttore, per il verificarsi di un evento lesivo algoritmico, o se invece il problema sarà piuttosto *probatorio*, attinente alla difficoltà di provare al di là di ogni ragionevole dubbio gli elementi del reato, in un contesto di sostanziale inesplicabilità dei modelli di *machine learning* [56], oltre che di frammentazione dei centri decisionali all'interno delle imprese.

Quale che sia la ragione, dovremmo comunque prepararci all'eventualità che, anche a fronte di condotte macroscopicamente negligenti da parte delle società produttrici, l'affermazione di una responsabilità penale personale per le offese derivanti dall'attivazione dei sistemi di i.a. costituirà – a meno di non voler rinunciare ai principi cardine del diritto e del processo penale – una rarità.

Un congedo integrale dal diritto penale, tuttavia, veicolerebbe un preoccupante messaggio deresponsabilizzante alle società produttrici, anche in settori in cui potrebbero emergere rischi non trascurabili per beni giuridici di primaria importanza [57]. La sfida, per il diritto penale, sarà allora duplice: da un lato, bisognerà evitare che istanze solidaristiche e di rassicurazione sociale conducano alla ricerca di un responsabile ad ogni costo, un "capro espiatorio", in violazione delle garanzie fondamentali del sistema penale; dall'altro lato, tuttavia, la tendenziale imprevedibilità dei sistemi di i.a. non potrà essere considerata una giustificazione per prassi sciatte e dinamiche di "disimpegno morale" [58].

In quest'ottica, ci pare che una prima soluzione possa venire, *de jure condendo*, dalla predisposizione di forme anticipate di tutela penale, che, quantomeno in determinati settori di sviluppo dell'intelligenza artificiale, sanzionino:

---

[56] In particolare, è già molto accesso il dibattito sui criteri di ammissibilità, nei processi penali, della *machine evidence*, v. S. Gless, *AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials*, in *Georgetown Journal of International Law*, vol. 51, No. 2, 2020, p. 195 ss.; P.W. Nutter, *Machine learning evidence: Admissibility and weight*, in *Journal of Constitutional Law*, vol. 23, No. 3, 2019, p. 919 ss.

[57] Sulla necessità di mantenere un presidio penalistico v. L. Picotti (ed.), *Resolution on Traditional Criminal Law Categories And AI*, cit., §§ 9 ss.

[58] A. Bandura, *Disimpegno morale. Come facciamo del male continuando a vivere bene*, Erickson, Trento, 2017.

(i) la creazione o il mantenimento di rischi illeciti a beni giuridici di primaria importanza, quali la vita e l'integrità fisica [59];
(ii) la mancata comunicazione di informazioni rilevanti per la gestione del rischio [60];
(iii) l'inottemperanza alle ingiunzioni dell'autorità pubblica [61].

È noto come la dottrina sia tradizionalmente restìa al ricorso agli schemi del pericolo astratto [62]; nei contesti di incertezza scientifica, poi, più che di illeciti di pericolo si dovrebbe forse parlare di "illeciti di rischio", stante l'assenza di conoscenze specifiche circa l'idoneità lesiva dei prodotti [63]. In relazione al rischio da intelligenza artificiale, tuttavia, l'anticipazione della tutela penale – come modalità principe di "gestione mediante pena dell'organizzazione sociale" [64] – parrebbe giustificata dalla già evidenziata peculiarità del fenomeno: se, infatti, l'evento lesivo algoritmico non è mai pienamente preventivabile (né evitabile) dal produttore, il nucleo del rimprovero non consiste tanto nel verificarsi dell'offesa, quanto nell'aver commercializzato un prodotto intelligente – di cui ben si conoscono le potenzialità lesive – in assenza delle dovute cautele.

Tale considerazione è tra l'altro da collegarsi a quella posizione che, in dottrina, suggerisce un complessivo ripensamento della responsabilità col-

---

[59] N. Amore, *L'effetto della robotica e dell'ia nell'imputazione giuridica degli eventi infausti*, cit., p. 249 ss.; F. Consulich, Flash offenders, cit., p. 1053; L. Romanò, *La responsabilità penale al tempo di ChatGPT*, cit., p. 84.

[60] È la tesi di G. Forti, *"Accesso" alle informazioni sul rischio e responsabilità: una lettura del principio di precauzione*, cit., *passim*; per un "aggiornamento" di tale proposta al contesto produttivo dell'intelligenza artificiale v. C. Piergallini, *Intelligenza artificiale*, cit., p. 1773.

[61] C. Piergallini, *Intelligenza artificiale*, cit., p. 1773; M. Lanzi, Self-driving cars *e responsabilità penale*, cit., p. 224.

[62] V. per tutti F. Stella, *Giustizia e modernità*, cit.; F. D'Alessandro, *Pericolo astratto e limiti soglia. Le promesse non mantenute del diritto penale*, Giuffrè, Milano, 2012, p. 255 ss.

[63] Sulla proposta di introdurre, *de jure condendo*, il c.d. *illecito di rischio*, ossia un tipo di illecito basato sulla "propensione al rischio" di una determinata condotta, v. C. Piergallini, *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali*, cit., p. 534 ss. In generale, sul rapporto tra i concetti di *rischio* e *pericolo*, nella riflessione penalistica, v. C. Perini, *Il concetto di rischio nel diritto penale moderno*, Giuffrè, Milano, 2010, *passim*, spec. p. 367 ss.

[64] A. Vallini, *Antiche e nuove tensioni tra colpevolezza e diritto penale artificiale*, Giappichelli, Torino, 2003, p. 98; sulla distinzione tra fattispecie penalistiche che tutelano beni giuridici e fattispecie che invece tutelano funzioni (*lato sensu*) amministrative, v. T. Padovani, *Tutela dei beni e tutela di funzioni nella scelta fra delitto, contravvenzione e illecito amministrativo*, in *Cass. pen.*, 1987, p. 670 ss.

posa, attraverso la valorizzazione – da sempre negletta nel nostro Paese [65] – del *disvalore di azione*, rispetto al solo *disvalore di evento* [66]. Le critiche che possono essere mosse all'idea della responsabilità colposa per l'evento lesivo sono ampiamente note: se l'evento costituisce il discrimine tra la punizione e la non punizione di una condotta, la responsabilità finisce per dipendere dal puro caso, dal momento che la verificazione dell'evento, oltre che dall'azione o omissione difformi dallo *standard* cautelare, dipende soprattutto da circostanze esterne e, per l'appunto, sostanzialmente aleatorie [67]. D'altra parte, lo stesso trattamento sanzionatorio viene calibrato, in maniera del tutto irrazionale, sulla portata lesiva dell'evento stesso, secondo un'ottica di fatto retribuzionista, che prescinde completamente da qualsiasi considerazione relativa all'efficacia preventiva delle norme penali: i soggetti che risponderanno dell'evento lesivo, infatti, sono semplicemente quelli *più sfortunati*, e non necessariamente quelli più negligenti [68].

Tali osservazioni – che ordinariamente non sono tali da scardinare la centralità, nel nostro ordinamento, delle fattispecie colpose d'evento – acquisiscono una maggiore pregnanza nel settore della produzione dei sistemi di i.a., considerato che l'evento lesivo algoritmico costituisce l'esito finale di una lunga serie di "decisioni" imponderabili ed imprevedibili [69].

---

[65] Per le ragioni culturali che, nel nostro ordinamento, hanno favorito l'affermarsi del primato del disvalore dell'evento – sostanzialmente legate ad un'interpretazione "forte" del principio di offensività – si rinvia a M. Mantovani, *Contributo ad uno studio sul disvalore di azione nel sistema penale vigente*, Bononia University Press, Bologna, 2014, *passim*, spec. 8 ss.; v. anche N. Mazzacuva, *Il disvalore di evento nell'illecito penale*, Giuffrè, Milano, 1983, che, tuttavia, nel suo lavoro, si propone di mettere in luce proprio una tendenza evolutiva al ridimensionamento del *disvalore dell'evento*. Nello specifico settore dei rapporti tra attività produttiva e principio di precauzione v. C. Pongiluppi, *Principio di precauzione e reati alimentari. Riflessioni sul rapporto "a distanza" tra disvalore d'azione e disvalore d'evento*, in *Riv. trim. dir. pen. econ.*, 2010, *passim*, spec. p. 240 ss.

[66] L. Cornacchia, *Responsabilità colposa: irrazionalità e prospettive di riforma*, in *Arch. pen.*, 2022, n. 2, *passim*; L. Eusebi, (voce) *Sistema sanzionatorio e reati colposi*, in M. Donini (diretto da) *Reato colposo*, *Enc. dir. – I Tematici*, cit., p. 1212 ss.

[67] Su questi aspetti v. L. Cornacchia, *Responsabilità colposa: irrazionalità e prospettive di riforma*, cit., p. 1 ss.; M.C. Del Re, *Per un riesame della responsabilità colposa*, in *Indice penale*, 1985, p. 31 ss., nell'ambito di una critica complessiva alla colpa quale elemento soggettivo rimproverabile all'agente; cfr. D. Castronuovo, *La colpa penale*, cit., p. 105 e ss., per una ricca ricostruzione, anche nella dottrina straniera, di queste posizioni. In argomento v. anche G. Marinucci, *La colpa per inosservanza di leggi*, cit., p. 121 ss.

[68] L. Cornacchia, *Responsabilità colposa: irrazionalità e prospettive di riforma*, cit., p. 2; v. anche L. Eusebi, (voce) *Sistema sanzionatorio e reati colposi*, cit., p. 1201 ss.

[69] F. Consulich, Flash offenders, cit., p. 1053.

Uno spunto per l'introduzione di una tutela penale anticipata sembrerebbe venire, d'altra parte, dal già citato Regolamento europeo sull'intelligenza artificiale, che obbliga gli Stati membri ad introdurre sanzioni «effettive, proporzionate e dissuasive» per il caso di mancato rispetto della disciplina ivi stabilita (v. art. 99, *AI Act*). Sebbene l'approccio del legislatore europeo sembri mostrare un *favor* per le sanzioni pecuniarie amministrative, non è detto che gli ordinamenti interni non possano optare per l'introduzione di apposite norme incriminatrici. Ovviamente, l'opera di individuazione delle condotte la cui violazione determini l'applicazione di una sanzione penale costituirà operazione delicata, da improntarsi ai criteri di precisione e di *extrema ratio*[70], e possibilmente da svolgersi proprio in collaborazione con le società produttrici, spesso dotate di un'*expertise* tecnologica e di una conoscenza delle fonti di pericolo che non ha paragoni nel settore pubblico[71].

Infine, un'efficace prevenzione del rischio da intelligenza artificiale non potrà non passare da una più approfondita riflessione sulla *colpa di organizzazione*. È questo, ci pare, l'elefante nella stanza: la discrepanza tra carattere intrinsecamente *plurisoggettivo* della criminalità d'impresa e conformazione *personalistica* della responsabilità penale si acuisce, infatti, nel contesto della produzione di sistemi intelligenti, in cui l'evento lesivo sembra fuoriuscire definitivamente dalla sfera di controllo del singolo partecipante alle decisioni d'impresa. La colpa che potrà essere imputata al produttore è infatti intrinsecamente una colpa da *disorganizzazione*, da *mancata gestione* o *scorretta gestione del rischio*[72]. In questa prospettiva, potrebbero allora amplificarsi quelle istanze, provenienti da una parte della dottrina, che sostengono la necessità di favorire una maggiore indipendenza dei criteri ascrittivi della responsabilità dell'ente rispetto a quelli della persona fisica, attraverso la valorizzazione del principio di "autonomia delle

---

[70] In questo senso, N. AMORE, *L'effetto della robotica e dell'ia nell'imputazione giuridica degli eventi infausti*, cit., p. 249.

[71] C. PIERGALLINI, *La responsabilità del produttore: una nuova frontiera del diritto penale?*, cit., p. 1128; G. FORTI, *"Accesso" alle informazioni*, cit., 192 ss.

[72] Concordano sulla necessità di individuare un modello di responsabilità amministrativa degli enti applicabile al settore dello sviluppo di intelligenza artificiale v. L. PICOTTI (ed.), *Resolution on Traditional Criminal Law Categories And AI*, cit., § 19.b.; B. PANATTONI, *Intelligenza artificiale*, cit., p. 362 ss.; L. ROMANÒ, *La responsabilità penale al tempo di ChatGPT*, cit., p. 85; N. AMORE, *L'effetto della robotica e dell'ia nell'imputazione giuridica degli eventi infausti*, cit., p. 250; per un'ampia disamina delle tecniche normative utilizzabili v. V. MONGILLO, *Corporate criminal liability for AI-related crimes: possible legal techniques and obstacles*, in L. PICOTTI, B. PANATTONI (ed.), *Traditional Criminal Law Categories and AI*, cit., p. 77 ss.

responsabilità dell'ente" (art. 8, d.lgs. 231 del 2001)[73], o attraverso l'introduzione di nuovi meccanismi diretti di responsabilizzazione (para) penale dell'ente[74].

Una funzione general-preventiva, d'altra parte, potrebbe essere svolta anche dalla predisposizione di sanzioni amministrative rivolte alla compagine societaria[75]: una soluzione che avrebbe il pregio di colpire il destinatario diretto delle regole cautelari in materia di produzione e sviluppo di intelligenza artificiale, senza che il filtro della responsabilità penale individuale si trasformi in una "caccia al colpevole" all'interno dei contesti organizzati[76].

---

[73] C. PIERGALLINI, *Intelligenza artificiale*, cit., pp. 1756-1756; F. CONSULICH, *Il nastro di Möbius*, cit., p. 195 ss.

[74] A. GARGANI, *Profili della responsabilità collettiva da reato colposo*, in *Riv. trim. dir. pen. econ*., 2022, n. 1-2, p. 59 ss.; con specifico riferimento ai reati ambientali L. MALDONATO, *Il crimine ambientale come crimine delle corporations: cooperazione pubblico-privato e responsabilità indipendente dell'ente*, in *Riv. trim. dir. pen. econ*., 2021, n. 3-4, p. 527 ss.; in relazione, invece, ai reati alimentari v. E. BIRRITTERI, *Salute pubblica, affidamento dei consumatori e diritto penale*, Giappichelli, Torino, 2022, p. 352 ss.

[75] Nello specifico settore delle auto a guida autonoma v. M. LANZI, Self-driving cars *e responsabilità penale*, cit., p. 228 ss., che ritiene che tali sanzioni possano essere più efficaci – sul piano general-preventivo – rispetto all'estensione dell'ambito applicativo del d.lgs. 231/2001. Va tra l'altro rilevato, sebbene in maniera inevitabilmente cursoria, come illeciti formalmente amministrativi potrebbero comunque rientrare nella nozione "sostanziale" di *matière pénale* propugnata dalla Corte edu a partire dalla sentenza C. edu, GC, *Engel e a. c. Paesi Bassi*, 8 giugno 1976, caso n. 5100/1971; in argomento v., per tutti, F. MAZZACUVA, *Le pene nascoste*, Giappichelli, Torino, 2017; L. MASERA, *La nozione costituzionale di materia penale*, Giappichelli, Torino, 2018. Una finalità afflittiva e deterrente parrebbe, *prima facie*, emergere già dalle indicazioni contenute nel già citato art. 99 *AI Act*, che, ad esempio, in merito alla violazione «del divieto delle pratiche di IA di cui all'articolo 5» prevede «sanzioni amministrative pecuniarie fino a 35 000 000 EUR o, se l'autore del reato è un'impresa, fino al 7 % del fatturato mondiale totale annuo dell'esercizio precedente, se superiore» (art. 99, § 3, *AI Act*).

[76] M. CATINO, *Trovare il colpevole. La costruzione del capro espiatorio nelle organizzazioni*, Il Mulino, Bologna, 2022.

INTELLIGENZA ARTIFICIALE E INVESTIGAZIONE PENALE PREDITTIVA*

di Lucio Camaldo

1. *L'intelligenza artificiale nell'attività investigativa predittiva*

Nel racconto intitolato «*Rapporto di minoranza*» [1], pubblicato nel 2002, dal quale è stato tratto il noto film «*The Minority Report*» di Steven Spielberg, l'autore Philip K. Dick prefigura un futuro in cui i delitti di omicidio nella città di Washington saranno completamente scomparsi, grazie a un sistema chiamato *Precrimine*. Con tale strumento, che si fonda sulle premonizioni di tre individui dotati di poteri extrasensoriali di precognizione amplificati (detti *Precog*), la polizia può, infatti, impedire gli omicidi prima che essi avvengano e arrestare i potenziali "colpevoli". In questo modo, non viene punito il fatto (che non avviene), bensì l'intenzione di compierlo e chi potrebbe realizzarlo.

Dalla fantascienza alla realtà il passo è stato piuttosto breve: alla fine del

---

* Il presente contributo è stato pubblicato sul fascicolo 1/2024 della *Rivista italiana di diritto e procedura penale*. Rispetto alla versione contenuta nella *Rivista*, il testo qui riportato è stato opportunamente aggiornato, tenendo in considerazione, in particolare, l'approvazione e la pubblicazione del testo definitivo del Regolamento dell'Unione europea sull'intelligenza artificiale (Reg. 2024/1689).

[1] Cfr. P.K. Dick, *Rapporto di minoranza e altri racconti*, trad. P. Prezzavento, Roma, 2002.

2009, l'*Office of Justice Assistance*, in collaborazione con il *Los Angeles Police Department*, ha organizzato un *Symposium* dedicato alla *predictive policing*. Si tratta della prima occasione in cui ricercatori, criminologi, sociologi, specialisti del diritto, funzionari di diversi dipartimenti di polizia si sono incontrati per esaminare e discutere l'impatto della polizia predittiva nell'ambito della giustizia penale, nonché le ricadute in relazione alla tutela dei diritti fondamentali e soprattutto sulla sfera della *privacy* e sul trattamento dei dati personali [2].

L'investigazione predittiva, come affermato in dottrina, consiste nelle attività rivolte allo studio e all'applicazione di metodi statistici con l'obiettivo di "predire" chi potrà commettere un reato oppure dove e quando potrà essere commesso un delitto, al fine di prevenirne la commissione [3].

Questa "predizione" si basa sulla rielaborazione 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 in cui si svolgono le azioni criminali, al periodo dell'anno e alle condizioni atmosferiche maggiormente connesse alla commissione di determinati delitti [4]. Sono raccolte, altresì, informazioni relative all'origine etnica, al livello di scolarizzazione, alle condizioni economiche, alle caratteristiche somatiche riconducibili a soggetti appartenenti a determinate categorie criminologiche, come, ad

---

[2] Sul tema, v. W.L. Perry, M. McInnis, C.C. Price, S. Smith, J.S. Hollywood, *Predictive Policing: The Role of Crime Forecasting,* in *Law Enforcement Operations*, Santa Monica, 2013.

[3] In questi termini, v. F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in G. Balbi, F. De Simone, A. Esposito, S. Manacorda (a cura di), *Diritto penale e intelligenza artificiale. Nuovi scenari*, Torino, 2022, p. 6. Sull'argomento, v., altresì, R.E. Kostoris, *Intelligenza artificiale, strumenti predittivi e processo penale*, in *Cass. pen*., 2024, n. 5, p. 1642 ss.; A. Balsamo, *L'impatto dell'intelligenza artificiale nel settore della giustizia*, in *Sist. pen*., 22 maggio 2024; G. Barone, *Giustizia predittiva e certezza del diritto*, Pisa, 2024; G. Ubertis, *Intelligenza artificiale e giustizia predittiva*, in *Sist. pen*., 16 ottobre 2023; L. Luparia Donati, G. Fiorelli, *Diritto probatorio e giudizi criminali ai tempi dell'intelligenza artificiale*, in *Dir. pen. cont. – Riv. trim*., 2022, n. 2, p. 34; P.P. Paulesu, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Arch. pen*., 2022, n. 1, p. 13; D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dell'intelligenza artificiale*, in *Arch. pen*., 2020, n. 3, p. 6; S. Signorato, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc*., 2020, n. 2, p. 605 ss.; Ead., *Il diritto a decisioni penali non basate esclusivamente su trattamenti automatizzati: un nuovo diritto derivante dal rispetto della dignità umana*, *ivi*, 2021, p. 101 ss.; G. Canzio, L. Luparia Donati, *Prova scientifica e processo penale*, II ed., Milano, 2022; S. Lorusso, *La sfida dell'intelligenza artificiale al processo penale nell'era digitale*, in *Sist. pen*., 28 marzo 2024.

[4] Cfr. F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, cit., p. 6.

esempio, potenziali terroristi. L'intelligenza artificiale provvede all'incrocio, mediante algoritmi, dei dati immagazzinati attraverso fonti diverse (banche dati della polizia, *databrokers*, *social networks*, *internet*, impianti di videosorveglianza, etc.).

Il procedimento di *predictive policing* si articola in tre fasi: la prima consiste nell'inserimento dei dati (di una o più tipologie) nel sistema; segue, poi, l'analisi dei dati inseriti attraverso un metodo algoritmico, allo scopo di elaborare la specifica previsione cui il sistema è finalizzato; infine, tale previsione viene utilizzata da parte degli operatori di polizia per adottare le decisioni strategiche e le tattiche sul campo.

La cifra che contraddistingue la polizia predittiva è proprio il mutato paradigma alla base delle strategie di *crime management* che passano dall'essere informate da un approccio esclusivamente "reattivo", secondo cui le forze di polizia intervengono a fronte della notizia circa la realizzazione di un reato ai fini dell'individuazione del relativo autore, a uno di tipo "proattivo", dove l'intervento della polizia precede e prescinde dall'attività criminale, al fine di prevenirla [5].

L'impiego di *software* basati sull'intelligenza artificiale, pertanto, ha rivoluzionato le attività di *law enforcement*, in quanto la previsione del crimine è diventata una priorità [6].

Gli strumenti di polizia predittiva possono essere suddivisi principalmente in due categorie: da un lato, vi sono i *place-based systems*, che consentono di prevedere il compimento di reati tramite la loro localizzazione; dall'altro, si rinvengono i *person-based systems*, preordinati a elaborare profili criminali individuali [7].

---

[5] Sul tema, cfr. V. Nicolì, *La predizione nell'attività di polizia*, in AA.VV., *Giurisdizione penale, intelligenza artificiale ed etica del giudizio*, Atti del convegno di studio "Enrico De Nicola", tenutosi *online* in data 15 ottobre 2020, Milano, 2021, p. 45; R. Pelliccia, *Polizia predittiva: il futuro della prevenzione criminale?*, in *cyberlaws.it*, 9 maggio 2019; L. Bennett Moses, J. Chan, *Algorithmic prediction in policing: assumptions, evaluation, and accountability*, in *Policing and society*, 2018, p. 806 ss.

[6] In argomento, v. E. Carpanelli, *Il ricorso all'intelligenza artificiale nel contesto di attività di law enforcement e di operazioni militari: brevi riflessioni nella prospettiva del diritto internazionale*, in *DPCE online*, 2022, n. 1.

[7] Cfr. L. Algeri, *Intelligenza artificiale e polizia predittiva*, in *Dir. pen. proc.*, 2021, n. 6, p. 729; F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, cit., p. 7; D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dell'intelligenza artificiale*, cit., pp. 6-7.

2. *I sistemi di polizia predittiva basati sulla localizzazione dei reati*

Con i *software* di polizia predittiva basati sulla localizzazione dei reati (*place-based systems*) vengono individuate le «zone calde» (*hotspots*), ossia i luoghi che costituiscono il possibile futuro scenario di un'eventuale commissione di determinati reati. Questi *software* attingono alla criminologia ambientale, che studia come i *target* criminali si muovono nello spazio e nel tempo, riservando particolare attenzione alla distribuzione geografica del crimine e al ritmo delle attività giornaliere [8]. L'idea di fondo consiste, infatti, nella considerazione secondo cui gli eventi criminali hanno luogo al ricorrere di determinati fattori spazio-temporali ovvero alla convergenza di delinquenti, di vittime o di obiettivi in contesti specifici in un tempo e in uno spazio definiti.

Dati e informazioni sugli eventi criminali vengono raccolti e catalogati in banche dati integrate e successivamente inseriti in un *software* che li analizza con il fine ultimo di trasformare tali informazioni dapprima in conoscenza sul dove e quando sia possibile che avvenga un crimine e poi in una guida per la prevenzione. Attraverso l'analisi dei dati relativi ai luoghi di maggiore concentrazione dei crimini avvenuti in passato e l'utilizzo di modelli predittivi, le forze dell'ordine possono organizzare in maniera più efficace le risorse a propria disposizione, distribuendosi in modo più mirato sul territorio cittadino e possono essere presenti nelle zone in cui è previsto che si verifichi un reato quel giorno e in quella fascia oraria con interventi specifici.

Un interessante esempio di *software* di questo tipo è rappresentato da *XLAW* che è stato ideato dall'ispettore di polizia Elia Lombardo presso la Questura di Napoli [9]. Tale strumento si basa su un algoritmo capace di elaborare una mole enorme di dati estrapolati dalle denunce inoltrate alla polizia, nonché dalle banche dati e dai *social network*, facendo 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 [10].

---

[8] A tal proposito, v. S. Vezzadini, *Profilo geografico e crime mapping. Il contributo della criminologia ambientale allo studio del delitto*, in R. Bisi (a cura di), *Scena del crimine e profili investigativi: quale tutela per le vittime?*, Milano, 2006, p. 83 ss.; L. Algeri, *Intelligenza artificiale e polizia predittiva*, cit., p. 730; F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, cit., p. 7.

[9] V. M. Iaselli, *XLAW: la polizia predittiva è realtà*, in *Altalex.com*, 28 novembre 2018. Cfr. anche L. Algeri, *Intelligenza artificiale e polizia predittiva*, cit., p. 723.

[10] Cfr. E. Lombardo, *Sicurezza 4P. Lo studio alla base del software XLAW per prevedere e prevenire i crimini*, Venezia, 2019; C. Morelli, *XLAW, il brevetto italiano di polizia predittiva*, in www.xlaw.it.

Il sistema traccia una mappa del territorio, dove vengono evidenziate le zone a più alto rischio (le "zone calde") 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 [11].

Poiché rapine, borseggi e furti hanno caratteristiche di ciclicità e sono tendenzialmente stanziali, in quanto messi in atto da soggetti deviati e modestamente organizzati, che usano questi espedienti per costruire un profitto in un arco temporale relativamente breve, è possibile individuare delle vere e proprie «riserve di caccia» [12].

Il sistema si basa anche sulle caratteristiche del sospettato, come genere, altezza, cittadinanza, segni distintivi e altri aspetti biometrici.

Gli operatori di polizia ricevono degli *alert*, ad esempio, riguardo a un potenziale furto in modo da sorvegliare in anticipo una determinata zona. Il programma, grazie a un sistema geografico informativo, fornisce, infatti, agli agenti una mappa di rischio che raffigura ogni trenta minuti i luoghi e gli orari precisi in cui si potrà consumare un crimine con un anticipo anche di due ore, descrivendo il tipo di reato, il *modus operandi* dell'autore, il tipo di preda e di *target*.

Basandosi su controlli selettivi e sequenziali, in risposta agli allarmi predittivi elaborati tramite intelligenza artificiale, il *software* in esame è in grado di trasformare il tradizionale metodo di pronto intervento dei funzionari di pubblica sicurezza.

Il sistema è stato utilizzato, oltre che a Napoli, anche a Prato, Salerno, Venezia, Modena, Parma, per testarlo in varie città italiane, con l'obiettivo di migliorare la prevenzione dei delitti nelle aree urbane. Secondo le dichiarazioni rese dall'ideatore di *XLAW*, nel corso di un'intervista televisiva del 29 dicembre 2018, il sistema ha contribuito ad abbattere il tasso di criminalità, assestandosi su una percentuale del -22% nella città di Napoli e del -39% in quella di Prato.

L'utilizzo del sistema *XLAW* ha prodotto risultati positivi sotto svariati profili: in termini di efficacia operativa, si è registrata una diminuzione significativa di reati come furti e rapine; sotto il profilo della valorizzazione del capitale umano, l'adozione del *software* ha comportato il miglioramento della motivazione e della capacità decisionale strategica degli operatori

---

[11] Cfr. F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. uomo*, 2019, n. 10, p. 11.

[12] V. C. Morelli, *Furti e rapine: a sventarli ci pensa l'intelligenza artificiale!*, in *Altalex.com*, 6 maggio 2019.

di controllo del territorio. Si è, inoltre, rilevata una diminuzione dei costi per la collettività: *XLAW* ha determinato, infatti, la razionalizzazione degli interventi per garantire la sicurezza pubblica, la riduzione dei chilometri percorsi dalle pattuglie di polizia, con un considerevole risparmio di carburante, e la riduzione dello *stress* emotivo degli agenti. Infine, si evidenzia il miglioramento della percezione di sicurezza e della fiducia nelle istituzioni da parte dei cittadini, un'evoluzione positiva nella reputazione professionale degli operatori, nonché una migliore rappresentazione delle azioni di contrasto alla criminalità da parte dei *mass media* [13].

Nella medesima tipologia dello strumento appena analizzato, rientra anche *Pelta*, che è un *software* ideato per garantire la sicurezza urbana e presentato con questo slogan: «Vedere per prevedere, prevedere per provvedere» [14].

Come emerge dalla *brochure* informativa, si tratta, anche in questo caso, di un sistema al servizio della polizia locale per prevenire fenomeni illeciti e di degrado nelle città e risparmiare sui costi di gestione della sicurezza. Sulla base della più evoluta analisi di rischio, *Pelta* permette di organizzare interventi, risorse, dotazioni strumentali, stabilendo le reali esigenze in base al grado, alla collocazione e all'evoluzione del rischio nel tempo e nello spazio. Il *software* nasce da uno studio accreditato da più centri di ricerca sui fenomeni di insicurezza urbana, quali furti, scippi, rapine, borseggi, spaccio di stupefacenti, abusi di ogni genere, prostituzione e incidenti stradali, soprattutto quelli che hanno una correlazione con comportamenti illeciti come, ad esempio, la guida in stato di ebrezza o l'uso di sostanze stupefacenti oppure quelli generati da insidie presenti sul manto stradale.

Secondo i produttori di *Pelta*, il modello alla base della soluzione, frutto di anni di studio multidisciplinare, si fonda su principi euristici e su un evoluto procedimento di analisi, mai applicato prima, per analizzare il rischio criminale.

Il suo impiego sposta il costrutto strategico dell'azione di controllo da una visione riparatoria del danno ad una visione probabilistica del rischio; quindi, da una logica di rincorsa dei problemi e degli effetti che essi generano, tipica della permanente emergenza, ad una che lavora sugli schemi della prevenzione.

---

[13] Su questo tema, v. G. Suffia, A. Lavorgna, S. Icardi, *Polizia "smart" tra paure e realtà: un'analisi esplorativa sulla rappresentazione mediatica dello smart policing in Italia*, in *Studi sulla questione criminale*, 2022, n. 3, p. 95.

[14] V. sito web www.pelta.it.

Sperimentato per lungo tempo in più contesti, come città [15], centri commerciali o aree private, la soluzione tecnologica e metodologica alla base di *Pelta* ha ottenuto prestigiose validazioni, tra cui il premio innovazione Smau 2015 e numerosi altri riconoscimenti.

3. *Gli strumenti tecnologici fondati sulle serialità criminali individuali*

Diversamente dalle tecniche sinora esaminate, i *person-based systems* seguono le serialità criminali (*crime linking*) di determinati soggetti per prevedere dove, come e quando i medesimi commetteranno il prossimo reato.

Mentre la *hotspot analysis* è volta a segnalare aree con alta incidenza di reati, con la conseguenza di criminalizzare le zone stesse senza risolvere il problema, i sistemi di *crime linking* puntano alla ricerca di comportamenti ripetuti che possano condurre ai responsabili. Il presupposto su cui si basano è il seguente: soltanto una piccola quota della popolazione è responsabile di episodi di violenza, sicché l'individuazione di tali soggetti e il conseguente intervento mirato nei loro confronti si rivelano essenziali per ridurre il tasso di criminalità.

Questi strumenti richiedono la compilazione di elenchi di persone ritenute «a rischio», nonché una *social network analysis* (elementi che, nella prassi, risultano spesso combinati tra di loro).

Si è correttamente osservato che i *software* di polizia predittiva rientranti nella categoria in esame, ossia quelli che si fondano su algoritmi in grado di prevedere se e quando uno specifico soggetto porrà in essere un dato reato, «sono in grado di fornire maggiori profili di interazione con il procedimento penale, con particolare riferimento alle fasi di raccolta di elementi probatori su cui fondare l'eventuale responsabilità di un imputato» [16].

Il sistema più conosciuto è *KeyCrime*, creato nel 2008 da Mario Venturini, dirigente della Polizia di Stato presso la Questura di Milano [17]. Que-

---

[15] Cfr. L. Bariella, *Polizia predittiva: al via la sperimentazione a Caorle*, in *Altalex.com*, 24 maggio 2021.

[16] V. D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dell'intelligenza artificiale*, cit., p. 7.

[17] A tal riguardo, v. L. Algeri, *Intelligenza artificiale e polizia predittiva*, cit., p. 731; A.D. Signorelli, *Il software italiano che ha cambiato il mondo della polizia predittiva*, in *Wired.it*, 18 maggio 2019; G. Santucci, *Milano. Il programma anti rapine diventa una startup della sicurezza*, in www.corriere.it, 18 aprile 2019; C. Morabito, *La chiave del crimine*, in www.poliziadistato.it, luglio 2015; M. Serra, *Rapinatore seriale catturato grazie al software "KeyCrime"*, in www.lastampa.it, 5 gennaio 2018.

sto *software* si basa sull'analisi dei dati di indagine acquisiti in relazione a precedenti reati (*input*) per fornire un'indicazione probabilistica delle future serie criminali (*output*).

L'ispirazione è scaturita dall'analisi degli elementi presenti in un atto criminale e dalla convinzione crescente che un adeguato esame di tali elementi, supportato da un potente elaboratore con capacità di calcolo e valutazione, avrebbe consentito di comprendere in dettaglio le dinamiche legate ai reati seriali. Attraverso l'esame delle variabili di una moltitudine di episodi già accaduti, l'algoritmo, infatti, provvede a segnalare le serie criminali effettuate dagli stessi soggetti e può prevedere dove potranno consumarsi le prossime azioni criminali.

Le informazioni e i dati raccolti in base alle dichiarazioni delle persone offese vengono inseriti nel *database* unitamente a tutti gli altri elementi oggettivi relativi al fatto (immagini da telecamere, tracce biologiche rinvenute, accertamenti diretti della polizia intervenuta sul posto dopo il fatto).

Il *software* si concentra sulla creazione di profili dettagliati dei potenziali autori di reati, analizzando le caratteristiche e le abitudini dei criminali, raccogliendo e incrociando dati investigativi per identificare modelli comportamentali che possono essere utilizzati per prevedere la futura attività delinquenziale di un individuo.

KeyCrime è stato applicato, per la prima volta, nella città di Milano e poi anche in tutta la provincia, al fine di ottenere una soluzione avanzata nell'analisi dei crimini, concentrandosi specificamente sulla prevenzione delle rapine, che, per la loro gravità e impatto sociale, costituiscono una minaccia rilevante.

In virtù del suo modello predittivo, l'utilizzo di questo sistema ha consentito di pianificare pattugliamenti efficaci, trasferendo agli organi di polizia le informazioni necessarie per operare in condizioni di massima sicurezza, nonché di organizzare servizi preventivi mirati in specifiche aree cittadine, prevedendo l'insorgenza di attività criminali. In diversi casi, tale strumento è servito a evitare la commissione di reati, consentendo alle forze dell'ordine di intervenire prima che questi avvenissero, arrestando i responsabili.

Le caratteristiche del sistema in esame offrono numerosi vantaggi, sia a livello organizzativo, ottimizzando il tempo e il metodo nella gestione della sicurezza pubblica, sia a livello di attività investigativa: *KeyCrime* facilita, infatti, il collegamento di procedimenti tra loro connessi a titolo oggettivo, soggettivo o funzionale, con impatti positivi sulle decisioni e sui tempi di conclusione delle indagini.

Sulla base delle sperimentazioni effettuate dalla Questura di Milano con

KeyCrime, è stato creato, più recentemente, un ulteriore sistema di analisi automatizzata in ausilio alle attività di polizia, denominato *Giove*, che agisce in due sensi: prevenzione e repressione [18].

Il progetto informatico è stato ideato nel 2020 dal Dipartimento di pubblica sicurezza del Ministero dell'Interno, che intende metterlo a disposizione di tutte le Questure in Italia.

Il nuovo sistema si avvale di un *set* di domande da porre alla vittima in fase di denuncia con la possibilità di inserire *file* multimediali audio o video e altra documentazione, in modo da consentire di rilevare gli elementi ricorrenti utili alle forze di polizia per le loro attività investigative.

Mentre il sistema *KeyCrime* è stato ideato per contrastare soprattutto le rapine in ambito commerciale, *Giove* potrebbe essere usato anche per molestie e violenze sessuali, furti in abitazione, truffe e raggiri.

L'uso di un sistema del genere potrebbe però comportare una violazione del diritto alla *privacy* e una lesione delle libertà personali degli individui.

Il Garante per la protezione dei dati personali, pertanto, dovrà esprimere il suo parere sul tema, senza il quale *Giove* non potrà essere utilizzato. Occorre comprendere, in particolare, quali banche dati e quali informazioni vengono usate per addestrare l'algoritmo, se le vittime di reato sono obbligate o meno a rispondere alle domande utilizzate per l'addestramento del *software* e chi è il responsabile del trattamento dei dati.

È opportuno segnalare, altresì, che è stata recentemente presentata un'interrogazione parlamentare [19], volta a ottenere chiarimenti in ordine ai seguenti quesiti: a) quali interventi il Ministero dell'Interno intenda mettere in atto per introdurre il sistema *Giove* in Italia e se esistano altri *software* di questo tipo già in uso o dei quali si prospetta l'utilizzo; b) quali aziende siano state coinvolte nella definizione di questa tecnologia, della sua implementazione e del suo sviluppo; c) quale sia lo stato dell'arte dell'interlocuzione con il Garante per la protezione dei dati personali in ordine a una valutazione di impatto che l'introduzione di questo sistema comporterebbe; d) quale tipo di dati e quali *batch* si intenda utilizzare per andare a comporre la memoria operativa del sistema; e) che livello di individuazione sia possibile e ottenibile senza violare la *privacy* dei soggetti;

---

[18] V. F. Ongaro, B. Simonini, *Software italiano Giove per la polizia predittiva, pro e contro*, in *Agenda digitale*, 2023; V. Nicolì, *La predizione nell'attività di polizia*, cit., pp. 47-48.

[19] Cfr. Senato della Repubblica, Atto di Sindacato Ispettivo n. 3-00499, con carattere d'urgenza, pubblicato il 13 giugno 2023, nella seduta n. 76, interrogazione parlamentare, presentata dall'On. Sensi (primo firmatario), al Ministro dell'Interno.

f) quali siano gli effetti anche sull'urbanistica delle città a fronte di una capacità così penetrante e intrusiva di profilazione delle persone e dei comportamenti, alla luce di un dibattito europeo ed internazionale molto negativo verso l'utilizzo di simili tecnologie così invasive e lesive dei diritti delle persone e nelle more di una decisione europea che regolerà in maniera cogente il suddetto utilizzo, vietando esplicitamente la possibilità di una polizia predittiva.

Per completare il quadro relativo ai sistemi di investigazione basati sulle nuove tecnologie, bisogna ricordare gli strumenti elaborati dal Reparto indagini telematiche dell'Arma dei Carabinieri [20]. In particolare, occorre segnalare il "Sistema di indagine" (SDI), che costituisce una banca dati interforze, sorta con finalità operative, contenente informazioni su reati, eventi con l'autore, vittime e oggetti censiti, come documenti, banconote, armi, veicoli. Vi è, inoltre, il "Sistema di controllo del territorio" (SICOTE), volto ad assicurare un efficace supporto alle attività di prevenzione generale e di controllo territoriale. Infine, il sistema ODINO (*Operational device for information, networking and observation*) è in grado di incrementare qualità e quantità dei controlli su strada, tramite un navigatore satellitare integrato e radiolocalizzato alla centrale operativa su cui sono installate applicazioni per l'accesso al sistema interforze SDI e alle banche dati a valenza info-investigativa, al fine di trasmettere alla centrale messaggi, immagini o video acquisiti.

4. *Le questioni problematiche della* predictive policing *e la regolamentazione normativa*

Nonostante i risultati positivi derivanti dall'applicazione dei *software* di polizia predittiva, basati sull'intelligenza artificiale, tali sistemi devono essere oggetto di un'attenta valutazione critica e necessitano di una specifica regolamentazione, a causa delle preoccupazioni riguardanti, come già anticipato, la tutela della *privacy*, la salvaguardia dei diritti individuali e le possibili ricadute discriminatorie etniche, razziali, religiose, sociali o di altro genere, che possono derivare dall'utilizzo dei dati raccolti da questi strumenti

[20] Per un approfondimento, v. C. Pistilli, *L'utilizzo dell'intelligenza artificiale nel campo delle attività investigative delle forze dell'ordine: tra prospettive di sviluppo ed esigenze di coordinamento*, in G. Balbi, F. De Simone, A. Esposito, S. Manacorda (a cura di), *Diritto penale e intelligenza artificiale. Nuovi scenari*, cit., p. 145 ss.

e dalla localizzazione dei fenomeni criminosi oggetto di predizione [21].

Il problema, in termini di salvaguardia della *privacy*, deriva dalla considerazione che queste applicazioni possono acquisire una grande quantità di dati in relazione alla vita, anche privata, dei cittadini e tali dati «potrebbero essere manipolati abusivamente, sottratti, deformati, con grave pregiudizio per le persone cui essi si riferiscono» [22].

Non si può nemmeno trascurare l'assenza in capo a questi dispositivi di doti tipicamente umane (intuito, senso comune, capacità di improvvisazione), che, invece, sono presenti ovviamente negli operatori della polizia [23].

Con riferimento ai *place-based systems*, come evidenziato dalla dottrina, bisogna evitare il fenomeno delle «profezie che si autoavverano» (*self-fulfilling prophecies*), ossia il circolo vizioso per cui i quartieri considerati a rischio («zone calde») attirano maggiore attenzione da parte della polizia, che rileva più criminalità con una eccessiva sorveglianza sulle comunità che vi abitano, lasciando, invece, privi di sorveglianza e di intervento altri quartieri («zone fredde») [24].

Inoltre, i sistemi di *predictive policing* sollecitano una prevenzione dei reati attraverso l'intervento attivo della polizia che si traduce in 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.) [25].

Si deve, peraltro, rilevare che gli strumenti in esame possono fornire

---

[21] Cfr. B. Pietrocarlo, *Predictive policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali*, in *Sist. pen*., 28 settembre 2023; C. Parodi, V. Sellaroli, *Sistema penale e intelligenza artificiale: molte speranze e qualche equivoco*, in *Dir. pen. cont*., 2019, n. 6, pp. 55-59; V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Discrimen*, 15 maggio 2020; P.P. Paulesu, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, cit., p. 10 ss.

[22] Sull'argomento, 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; B. Perego, *Predictive policing: trasparenza degli algoritmi, impatto sulla privacy e risvolti discriminatori*, in *BioLaw Journal*, 2020, n. 2, p. 447 ss.; A. Moretti, *L'intelligenza Artificiale nel dettato costituzionale: opportunità, incertezze e tutela dei dati personali*, in *BioLaw Journal*, 2020, n. 3, p. 365 ss.

[23] Al tal proposito, v. M. B. Magro, *Biorobotica, robotica e diritto penale*, in D. Provolo, S. Riondato, F. Yenisey (a cura di), *Genetics, Robotics, Law, Punishment*, Padova, 2014, p. 512.

[24] In tal senso, v. F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, cit., p. 8; L. Algeri, *Intelligenza artificiale e polizia predittiva*, cit., p. 733.

[25] V. ancora F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, cit., p. 9.

adeguate previsioni soltanto in relazione a determinate categorie di reati, quali, ad esempio, quelli attinenti alla criminalità urbana, come furti, rapine e spaccio di sostanze stupefacenti, non necessariamente quelli più pericolosi per la democrazia e per la libertà democratica.

Questa "anticipazione analitica" rischia, d'altro canto, di spostare eccessivamente l'attività delle forze dell'ordine, che si concentrano su quello che potrebbe accadere, anziché su ciò che è accaduto [26].

Un ulteriore aspetto che merita particolare attenzione è che la maggior parte di questi *software* sono coperti da brevetti depositati da aziende private, le quali, a buon diritto, sono gelose 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 [27].

Si tenga conto, poi, che le capacità tecniche dimostrate dall'intelligenza artificiale nell'analisi dei dati e nella formulazione di modelli predittivi consentono alle società private, oltre che di conoscere i comportamenti delle persone, altresì di influenzarne le decisioni, senza che esse siano consapevoli dell'influenza su di loro esercitata e, quindi, in assenza di un loro esplicito consenso.

Nell'ambito delle iniziative volte a regolamentare queste nuove tecniche investigative, occorre considerare, anzitutto, la Direttiva 2016/680/UE del 27 aprile 2016, il cui art. 11, par. 1 impone agli Stati membri di introdurre il divieto di decisioni basate «unicamente» su un trattamento automatizzato di dati, compresa la profilazione, che producano effetti giuridici negativi o incidano significativamente sui diritti fondamentali dell'interessato [28]. Si deve, inoltre, limitare, per quanto possibile, il margine di errore in cui può

---

[26] Sul punto, v. M. Martorana, L. Pinelli, *Polizia e giustizia predittive: cosa sono e come vengono applicate in Italia*, in *Agenda Digitale*, 2021.

[27] Così F. Basile, *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, cit., p. 9.

[28] Cfr. Direttiva 2016/680/UE del Parlamento europeo e del Consiglio del 27 aprile 2016 relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio, in *G.U.U.E.*, 4 maggio 2016, L 119/89. A tal proposito, v. G. Baccari, *Il trattamento (anche elettronico) dei dati personali per finalità di accertamento dei reati*, in A. Cadoppi, S. Canestrari, A. Manna, M. Papa (diretto da), *Cybercrime*, Torino, 2019, p. 1611 ss.; D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dell'intelligenza artificiale*, cit., p. 36 ss.

incorrere l'algoritmo, prevedendo «l'intervento umano» di un "controllore" dotato delle competenze necessarie per rilevare e correggere il vizio del risultato. Si stabilisce, altresì, che non è ammesso il trattamento automatizzato di dati sensibili o che comunque esso impone specifiche garanzie per l'interessato, in modo da minimizzare il rischio di trattamento discriminatorio ed effetti distorsivi (art. 11, par. 2 e par. 3).

Altre previsioni, relative alle valutazioni d'impatto, agli obblighi di adozione di specifiche misure da parte del titolare del trattamento dei dati personali a tutela dei diritti dell'interessato (artt. 19-28), contribuiscono a ridimensionare la portata degli effetti negativi delle tecniche di polizia predittiva, in quanto consentono, da un lato, di limitare i rischi per la *data protection*, intervenendo preventivamente, e, dall'altro, di assicurare l'informazione nei confronti delle persone coinvolte circa i loro diritti e le modalità per esercitarli concretamente (artt. 12-28).

In attuazione dell'atto normativo europeo, l'art. 16 d.lgs. n. 51/2018 [29] stabilisce che il titolare del trattamento dei dati deve mettere in atto misure tecniche e organizzative adeguate con riferimento alla protezione dei diritti degli interessati [30], nonché garantire che, per impostazione predefinita, non siano resi accessibili dati personali a un numero indefinito di persone fisiche.

Va menzionato pure il D.P.R. n. 15/2018, che disciplina, nello specifico, il trattamento dei dati personali da parte delle forze di polizia nell'esercizio dei compiti di prevenzione dei reati, tutela dell'ordine e della sicurezza pubblica, nonché di polizia giudiziaria [31].

Tra gli atti di *soft law*, oltre alla "Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi", elaborata, in data 4 dicembre 2018, dalla *European Commission for the Effi-*

---

[29] Cfr. D.lgs. 18 maggio 2018, n. 51, Attuazione della direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio.

[30] Si fa riferimento alla c.d. pseudonimizzazione, ossia una modalità particolare di trattamento intesa ad impedire che i dati personali possano essere riferiti ad un determinato soggetto senza l'utilizzo di informazioni aggiuntive a loro volta separatamente gestite.

[31] V. D.P.R. 15 gennaio 2018, n. 15, Regolamento a norma dell'articolo 57 del decreto legislativo 30 giugno 2003, n. 196, recante l'individuazione delle modalità di attuazione dei principi del Codice in materia di protezione dei dati personali relativamente al trattamento dei dati effettuato, per le finalità di polizia, da organi, uffici e comandi di polizia.

ciency of Justice (CEPEJ) [32], è opportuno ricordare la Comunicazione della Commissione europea del 25 aprile 2018 [33], con cui è stata delineata una "strategia europea" sull'intelligenza artificiale, che si concentra su tre linee guida principali: promuovere la ricerca, lo sviluppo tecnologico e le applicazioni industriali legate alle nuove tecnologie; sostenere l'impatto socio-economico derivante da tali tecnologie; sviluppare un quadro etico e giuridico coerente con i valori dell'Unione europea e allineato con le disposizioni della Carta dei diritti fondamentali.

La Commissione europea, poi, ha costituito un "Gruppo di esperti di alto livello" che ha pubblicato, nell'aprile 2019, le "*Ethical guidelines for trustworthy artificial intelligence*", dove vengono indicati gli elementi chiave per un'intelligenza artificiale che possa essere ritenuta "affidabile": legalità, eticità, robustezza tecnica e sociale. Nel febbraio 2020, è stato pubblicato il "Libro bianco sull'intelligenza artificiale" [34], con il quale sono state approfondite e discusse le strategie precedentemente definite nel 2018 per una *artificial intelligence* "*made in Europe*", incentrata sull'eccellenza europea nella ricerca e nell'industria. Questo documento ha affrontato anche questioni di rilevanza giuridica, ponendo l'accento su diritti fondamentali, *governance* dei dati e responsabilità.

Si è giunti così alla "Proposta di regolamento sull'intelligenza artificiale" [35], pubblicata in data 21 aprile 2021, la quale mira a creare un quadro normativo generale sull'intelligenza artificiale in linea con i valori europei e con la protezione dei diritti fondamentali. La proposta classifica i sistemi d'intelligenza artificiale, stabilisce i requisiti per la loro creazione, commer-

---

[32] Cfr. Commissione europea per l'efficienza della giustizia (CEPEJ), Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi, Strasburgo, 3-4 dicembre 2018. In argomento, 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 *Legisl. pen. online*, 18 dicembre 2018, p. 4; D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dell'intelligenza artificiale*, cit., p. 30 ss.

[33] V. Comunicazione della Commissione al Parlamento europeo, al Consiglio europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni sull'intelligenza artificiale per l'Europa, Bruxelles, 25 aprile 2018, COM(2018) 237 final.

[34] Cfr. Commissione Europea, Libro bianco sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia, Bruxelles, 19 febbraio 2020, COM(2020) 65 final.

[35] V. Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione, Bruxelles, 21 aprile 2021, COM(2021) 206 final. In seguito, v. Risoluzione legislativa del Parlamento europeo del 13 marzo 2024 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione.

cializzazione e utilizzo, e delinea gli strumenti di controllo e le sanzioni.

Preme evidenziare che la disciplina elaborata dalle istituzioni dell'Unione europea adotta un approccio basato sulla "piramide di rischio", classificando le applicazioni di intelligenza artificiale in tre categorie: sistemi vietati, sistemi ad alto rischio e sistemi a rischio medio/basso.

Gli strumenti di intelligenza artificiale sono vietati quando rappresentano una minaccia inaccettabile per le persone, e ciò comprende numerose situazioni, tra le quali, ad esempio, la manipolazione comportamentale di persone vulnerabili, la profilazione delle persone in base a comportamenti, *status* socio-economico o altre caratteristiche personali, l'identificazione biometrica in spazi pubblici mediante il riconoscimento facciale, quando venga effettuata in tempo reale e in modo invasivo. È vietato, altresì, l'utilizzo di sistemi che categorizzano gli individui sulla base di genere, razza, etnia, cittadinanza, religione o orientamento politico.

L'intelligenza artificiale "generativa" è soggetta, inoltre, a requisiti di trasparenza volti a garantire l'origine dei sistemi che si fondano su tale tecnologia, affinché sia possibile distinguere tra i contenuti generati artificialmente e quelli reali, contribuendo a prevenire la diffusione di informazioni fuorvianti o *deepfake*. Occorre, altresì, che i sistemi di intelligenza artificiale siano progettati in modo da prevenire la generazione di contenuti illegali e da garantire la protezione del diritto d'autore.

Nella prima versione della Proposta di regolamento sopra citata, le attività di polizia predittiva ricadevano all'interno dei sistemi di intelligenza artificiale "ad alto rischio", sicché, sebbene soggette a particolari prescrizioni, erano comunque ammesse (v. art. 6 ss. e Allegato III).

Secondo la successiva versione della Proposta, datata 14 giugno 2023, che poi è confluita nel Regolamento (UE) 2024/1689 del 13 giugno 2024 (*Artificial Intelligence Act – AI Act*) [36], i *person-based systems* (v., *supra*, § 3) potrebbero essere ricondotti, in linea generale, alle attività vietate (art. 5), in relazione alle quali il livello di rischio è considerato inaccettabile, poiché contrastante con i valori fondamentali dell'Unione [37]. Il medesimo destino

[36] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), in *G.U.U.E.*, 17 luglio 2024, L 144. Sul tema, v. G. Canzio, *AI act e processo penale: sfide e opportunità*, in *Sist. pen.*, 14 ottobre 2024; M. Torre, *Il Regolamento europeo sull'intelligenza artificiale: profili processuali*, in *Proc. pen. giust.*, 2024, n. 6, p. 1534 ss.

[37] In tal senso, v. B. Pietrocarlo, *Predictive policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali*, cit., p. 54 ss.

sembra prospettarsi anche per i *place-based systems* (v., *supra*, § 2). Nell'ambito delle pratiche vietate, infatti, sono inclusi i sistemi utilizzati per predire il verificarsi o il ripetersi di un reato basati sulla profilazione di una persona fisica o sulla valutazione dei tratti e delle caratteristiche della personalità, compresa l'ubicazione della persona, o di comportamenti criminali passati di persone fisiche o gruppi di persone fisiche.

Tale divieto, tuttavia, non si applica ai sistemi di intelligenza artificiale «utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa» (art. 5, par. 1, lett. d). In questo senso, dunque, i sistemi di *crime linking* in precedenza analizzati – come *KeyCrime* o *Giove* – non dovrebbero rientrare nel divieto di cui all'art. 5 del Regolamento [38].

Occorre, tuttavia, evidenziare che, a livello nazionale, l'utilizzabilità in sede processuale delle risultanze derivanti dalle indagini predittive deve essere ricondotta, in ogni caso, ai criteri normativi di cui all'art. 191 c.p.p., nonché alle previsioni dell'art. 526 c.p.p. A tal proposito, «appare necessario delimitare l'ampiezza della rilevanza probatoria dei risultati degli algoritmi in esame, nella misura in cui essi – lungi dal divenire una sorta di prova privilegiata – devono, all'evidenza, non solo essere rimessi alla libera valutazione del giudice – che non può aderire passivamente a quanto indicato (*rectius*, deciso) dalla "macchina" – ma anche essere sottoposti ai canoni prescritti dal codice di rito rispetto alle prove indiziarie di cui all'art. 192 c.p.p.» [39], che esigono, come è noto, riscontri e conferme da rinvenirsi negli ulteriori elementi di prova raccolti in sede processuale.

5. *L'identificazione biometrica e i programmi di riconoscimento facciale*

È opportuno dedicare speciale attenzione agli strumenti di identificazione biometrica [40]: ci si riferisce a un gruppo eterogeneo di *tools*, tesi ad automatizzare le procedure di verifica dell'identità, attraverso la valutazio-

[38] Così B. Pietrocarlo, *Predictive policing: criticità e prospettive dei sistemi di identificazione dei potenziali criminali*, cit., p. 58.

[39] Cfr. D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dell'intelligenza artificiale*, cit., p. 11.

[40] Sul tema, v. E. Sacchetto, *Spunti per una riflessione sul rapporto fra biometria e processo penale*, *in Dir. pen. cont.*, 2019, n. 2, p 467; T. Alesci, *Il corpo umano fonte di prova*, Padova, 2017, p. 89.

ne di caratteristiche fisiologiche della persona, quali le impronte facciali o digitali e la forma della mano, oppure comportamentali, come voce, firma, andatura, e si possono distinguere in sistemi che operano "in tempo reale" o, invece, "a posteriori" [41]. La prima tipologia implica l'uso istantaneo di materiale rilevato dal vivo, mentre i programmi del secondo tipo trattano i dati in differita, cioè successivamente alla raccolta.

Il Parlamento europeo ha assunto una posizione rigorosa sul riconoscimento facciale: nella risoluzione del 6 ottobre 2021 sull'intelligenza artificiale e il diritto penale [42] ha raccomandato esplicitamente di vietare l'utilizzo di questi strumenti negli spazi pubblici [43].

In aderenza con questa indicazione, il legislatore nazionale [44], per primo tra gli Stati membri dell'Unione europea, ha previsto tale divieto, salvo che si tratti di sistemi utilizzati per la prevenzione e la repressione dei reati o nell'esecuzione delle pene.

Ai sensi dell'art. 5, par. 1, lett. h), punto iii), del Regolamento (UE) 2024/1689, sopra citato, l'impiego di strumenti di riconoscimento facciale basati sull'intelligenza artificiale ai fini dello svolgimento di attività investigativa, dell'esercizio dell'azione penale o dell'esecuzione di una sanzione detentiva, è consentito limitatamente all'individuazione o alla localizzazione di soggetti nei confronti dei quali sussista il sospetto di reità.

L'Autorità giudiziaria procedente è autorizzata ad avvalersi di tali risorse esclusivamente per perseguire illeciti di particolare gravità, astrattamente punibili, secondo la legge dello Stato membro interessato, con una pena

---

[41] Cfr. J. Della Torre, *Quale spazio per i tools di riconoscimento facciale nella giustizia penale?*, in G. Di Paolo, L. Pressacco (a cura di), *Intelligenza artificiale e processo penale*, Napoli, 2022, p. 7 ss.; M. Colacurci, *Riconoscimento facciale e rischi per i diritti fondamentali alla luce delle dinamiche di relazione tra poteri pubblici, imprese e cittadini*, in G. Balbi, F. De Simone, A. Esposito, S. Manacorda (a cura di), *Diritto penale e intelligenza artificiale. Nuovi scenari*, cit., p. 119 ss.

[42] Cfr. Risoluzione del Parlamento europeo sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziaria in ambito penale, Bruxelles, 6 ottobre 2021. A tal proposito, v. G. Barone, *Intelligenza artificiale e processo penale: la linea dura del Parlamento europeo. Considerazioni a margine della risoluzione del Parlamento europeo del 6 ottobre 2021*, in *Cass. pen*., 2022, n. 3, p. 1180 ss.; A. Pingen, *EP Resolution on AI in Criminal Law and Policing*, in *Eucrim.eu*, 17 novembre 2021.

[43] V. G. Borgia, *Profili sistematici delle tecnologie di riconoscimento facciale automatizzato, anche alla luce dei futuribili sviluppi normativi sul fronte eurounitario*, in *Legisl. pen*., 2021, n. 4, p. 206 ss.

[44] V. d.l. n. 139/2021, Disposizioni urgenti per l'accesso alle attività culturali, sportive e ricreative, nonché per l'organizzazione di pubbliche amministrazioni e in materia di protezione dei dati personali, convertito dalla l. n. 205/2021.

detentiva non inferiore nel massimo a quattro anni o in relazione ai quali è prevista l'applicazione di una misura di sicurezza privativa della libertà personale della medesima durata minima.

Il ricorso, in zone pubbliche o aperte al pubblico, a sistemi di identificazione biometrica che operano in modalità *realtime* è consentito solo per l'accertamento e la repressione degli illeciti specificamente indicati nell'Allegato II, previa autorizzazione, dall'efficacia vincolante, di un'Autorità amministrativa o giudiziaria indipendente dello Stato membro in cui l'applicativo di intelligenza artificiale deve essere utilizzato, a condizione che siano rispettati i limiti e gli obiettivi di cui al par. 1, primo comma, lett. h) del già citato art. 5.

L'Allegato II contiene un elenco dettagliato di reati dall'elevato allarme sociale [45], per il contrasto dei quali il legislatore europeo ha ritenuto opportuno un bilanciamento tra l'esigenza di tutela della pubblica sicurezza e la protezione delle persone fisiche con riguardo al trattamento dei dati personali.

L'uso dei sistemi di identificazione remota delle persone fisiche operanti «in tempo reale» in spazi accessibili al pubblico è, infatti, particolarmente invasivo e comporta un'inevitabile compressione dei diritti e delle libertà fondamentali dei soggetti coinvolti dall'analisi biometrica.

Relativamente all'impiego di tali strumenti come strategia di contrasto alla criminalità, il duplice limite applicativo di un indice tassativo di reati presupposto e della previsione di una soglia di pena minima astrattamente applicabile rappresenta un presidio significativo all'inviolabilità dei diritti umani, sancita dall'art. 2 TUE e garantita, mediante il rinvio di cui all'art. 6 TUE, dalla Carta dei diritti fondamentali dell'Unione europea e dalla Convenzione europea dei diritti dell'uomo.

L'attuale assetto normativo suscita, tuttavia, alcune perplessità [46]. Anzitutto, la disciplina che vieta tali pratiche, ma che contestualmente prevede

---

[45] Si tratta, in particolare, delle seguenti fattispecie di reato: terrorismo, tratta di esseri umani, sfruttamento sessuale di minori e pornografia minorile, traffico illecito di stupefacenti o sostanze psicotrope, traffico illecito di armi, munizioni ed esplosivi, omicidio volontario, lesioni gravi, traffico illecito di organi e tessuti umani, traffico illecito di materie nucleari e radioattive, sequestro, detenzione illegale e presa di ostaggi, reati che rientrano nella competenza giurisdizionale della Corte penale internazionale, illecita cattura di aeromobile o nave, violenza sessuale, reato ambientale, rapina organizzata o a mano armata, sabotaggio, partecipazione a un'organizzazione criminale coinvolta in uno o più dei reati appena elencati.

[46] Per un maggiore approfondimento, v. M. Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa*, in *Dir. pen. cont.*, 29 maggio 2019, p. 10.

numerose eccezioni, genera dubbi sulla sua effettività. In secondo luogo, la mancanza di restrizioni relative alla vendita di queste tecnologie al di fuori dell'Unione europea solleva notevoli preoccupazioni, poiché le imprese che producono tali strumenti in ambito europeo possono liberamente venderli a Paesi terzi, i quali potrebbero utilizzarli per scopi repressivi o invasivi.

Infine, non si può trascurare la disparità di trattamento tra le due tipologie di sistemi in oggetto: nonostante il controllo dei dati biometrici "a posteriori" possa offrire maggiori garanzie rispetto alla scansione "in tempo reale" (*realtime*), ci si chiede se il ritardo temporale sia sufficiente a giustificare tale differenza di regolamentazione. Invero, anche i programmi di identificazione "a posteriori" rappresentano una grave minaccia per la *privacy* e potrebbero facilmente trasformarsi in strumenti potenti per effettuare una sorveglianza di massa.

6. *Il sistema automatico di riconoscimento delle immagini: alcuni rilievi critici*

Il sistema automatico di riconoscimento delle immagini, noto con l'acronimo S.A.R.I., è stato creato nel 2017 e costituisce una delle più efficaci dotazioni tecnologiche a disposizione della polizia per la sorveglianza ai fini di sicurezza, nonché un valido supporto per l'attività investigativa [47].

Tale strumento, che utilizza le "*faceprint*", ossia le caratteristiche principali di un viso, registrate ed elaborate a monte dal sistema, permette di confrontare l'identità ignota di un volto raffigurato in un'immagine con quelle di milioni di soggetti già presenti nel *database* contenente immagini segnaletiche, dati anagrafici e biometrici di soggetti schedati.

L'utilizzo di questo programma ha consentito una notevole accelerazione nel confronto dei dettagli identificativi di soggetti sospettati, rispetto al metodo classico, che richiedeva agli operatori di inserire manualmente tali elementi nelle maschere di ricerca della banca dati e ciò richiedeva parecchio dispendio di tempo e notevole precisione.

Il sistema di riconoscimento automatico delle immagini rientra sia nella

---

[47] Sul sistema S.A.R.I., v. L. Saponaro, *Le nuove frontiere tecnologiche dell'individuazione personale*, in *Arch. pen*., 2022, n. 1 p. 4; E. Currao, *Il riconoscimento facciale e i diritti fondamentali: quale equilibrio?,* in *Dir. pen. uomo*, 2021, n. 5, p. 16 ss.; E. Sacchetto, *Face to face: il complesso rapporto tra automated facial recognition technology e processo penale*, in *Legisl. pen*., 2020, p. 1 ss.; R. Lopez, *La rappresentazione facciale tramite software*, in A. Scalfati (a cura di), *Le indagini atipiche*, Torino, 2019, p. 239 ss.; R.V.O. Valli, *Sull'utilizzabilità processuale del Sari: il confronto automatizzato di volti rappresentati in immagini*, in *Il penalista*, 16 gennaio 2019.

categoria delle pre-investigazioni, ossia le attività realizzate quando la *notitia criminis* non si è manifestata in alcuni o in tutti i suoi contorni costitutivi, sia nella categoria delle indagini atipiche, poiché tale sistema viene utilizzato per identificazioni personali in sostituzione degli strumenti tradizionali, dopo l'iscrizione della notizia di reato nel registro di cui all'art. 335 c.p.p.

Il sistema S.A.R.I. conosce due differenti modalità di funzionamento. Anzitutto, la funzionalità "*Enterprise*" che, a seguito dell'inserimento da parte dell'operatore di un'immagine, fornisce agli agenti di polizia un meccanismo automatizzato per verificare l'identità di un individuo tramite l'analisi di tale immagine. L'identificazione è effettuata all'interno dell'archivio AFIS (*Automated Fingerprint Identification System*) contenente dati di grandi dimensioni che possono essere selezionati dall'utente a seconda delle necessità. Nello specifico, ci sono tre tipi di ricerca disponibili: la prima è basata sull'immagine facciale (modalità di ricerca del volto); la seconda si fonda su informazioni anagrafiche o descrittive collegate alle immagini presenti nell'archivio fotografico (modalità di ricerca anagrafica/descrittiva); la terza combina le due precedenti modalità. Al termine della fase di ricerca, il sistema genera una lista di candidati (c.d. "*candidate list*"), cioè una serie di profili ordinati in base a un punteggio di similarità con l'immagine che è stata inserita nel programma. Una volta interrogato l'algoritmo e acquisito l'*output*, è richiesto l'intervento di un operatore, il cui ruolo consiste nel valutare e convalidare i risultati ottenuti dal *software*. Tale valutazione è basata su procedure di confronto facciale, per le quali il personale della polizia scientifica riceve una formazione specifica.

Una diversa modalità di funzionamento del sistema S.A.R.I. è denominata "*Real Time*", la quale costituisce una soluzione completa per il riconoscimento facciale "in tempo reale" e in modalità "diretta" su più flussi video provenienti da telecamere posizionate in determinati luoghi di interesse. I volti presenti nei *frame* dei flussi video sono sottoposti ad analisi e confronto, mediante un algoritmo di riconoscimento, con quelli presenti in un elenco di controllo (contenente un numero considerevole di immagini, circa un migliaio). Qualora si verifichi una corrispondenza positiva (*match*), il sistema genera un avviso (*alert*), in grado di attirare l'attenzione degli operatori preposti, ai quali spetterà il compito finale di confermare il riconoscimento e provvedere agli interventi necessari.

La necessità di garantire elevati livelli di affidabilità e precisione negli algoritmi di riconoscimento facciale, basati sulla biometria, è emersa come tema centrale, soprattutto in considerazione delle gravi implicazioni che gli errori del *software* potrebbero determinare nel contesto del procedimento penale.

L'origine dei dati e i soggetti coinvolti nel processo di programmazione sono considerati come elementi determinanti per la formazione e l'evoluzione degli algoritmi, non potendosi trascurare il rischio che eventuali pregiudizi discriminatori possano contaminare i risultati.

Sebbene questi principi siano rilevanti in ogni contesto applicativo dell'intelligenza artificiale, si rivelano particolarmente cruciali quando vengono utilizzati gli strumenti di riconoscimento facciale.

L'innovativo sistema sopra descritto ha suscitato, infatti, l'attenzione del Garante per la protezione dei dati personali, con riferimento alle possibili implicazioni in tema di *privacy* e diritti fondamentali. In particolare, sono state avviate due istruttorie concernenti i differenti metodi di funzionamento del *software*.

A proposito del S.A.R.I. *Enterprise*, il Garante ha espresso una valutazione favorevole, poiché tale modalità non costituisce una nuova forma di trattamento dei dati personali, bensì un'automatizzazione di operazioni precedentemente eseguite manualmente [48]. Pertanto, la base normativa, già esistente per il precedente sistema AFIS [49], è considerata sufficiente per autorizzare l'attività di trattamento dei dati biometrici anche con il nuovo strumento.

Diametralmente opposto è, invece, il parere espresso dal Garante per la protezione dei dati personali con riferimento allo scenario operativo di S.A.R.I. *Real Time* [50]. Nello specifico, è stato evidenziato come tale metodo rappresenti una nuova modalità di trattamento dei dati biometrici nel contesto della sicurezza pubblica e delle indagini penali: si è passati dal tracciamento mirato di individui specifici alla possibilità di una "sorveglianza di massa" con *screening* costante degli individui in una determinata area [51].

---

[48] Cfr. Garante per la protezione dei dati personali, Parere sul sistema *Sari Enterprise*, Registro dei provvedimenti n. 440 del 26 luglio 2018. In tale provvedimento, si osserva, infatti, che la funzione in esame rappresenta «un mero ausilio all'agire umano, avente lo scopo di velocizzare l'identificazione, da parte dell'operatore di polizia, di un soggetto ricercato della cui immagine facciale si disponga, ferma restando l'esigenza dell'intervento dell'operatore per verificare l'attendibilità dei risultati prodotti dal sistema automatizzato».

[49] V. d.lgs. 18 maggio 2018, n. 51, cit., nonché Decreto del Ministro dell'Interno del 24 maggio 2017, recante l'individuazione dei trattamenti di dati personali effettuati dal Centro elaborazione dati del Dipartimento della pubblica sicurezza o da forze di polizia effettuati con strumenti elettronici, in attuazione dell'art. 53, comma 3, d.lgs. n. 196/2003.

[50] V. Garante per la protezione dei dati personali, Parere sul sistema *Sari Real Time*, Registro dei provvedimenti n. 127 del 25 marzo 2021. Per un commento sui provvedimenti del Garante, v. G. Mobilio, *Tecnologie di riconoscimento facciale. Rischi per i diritti fondamentali e sfide regolative*, Napoli, 2021.

[51] Cfr. A. Fonsi, *Prevenzione dei reati e riconoscimento facciale: il parere sfavorevole del ga-*

Non è rinvenibile, nella legislazione vigente, una disciplina adeguata che possa garantire la conformità dell'applicativo alla normativa in materia di trattamento dei dati personali.

Trattandosi di una sorta di "videoripresa algoritmica", non finalizzata a captare un comportamento comunicativo, è chiaro come, secondo l'orientamento delle Sezioni unite [52], la stessa non possa trovare ingresso neppure tramite un'applicazione analogica della disciplina relativa alle intercettazioni [53].

D'altro canto, l'utilizzo a fini investigativi e/o probatori di tale modalità del *tool* non può essere ricondotto nemmeno nell'alveo delle individuazioni o delle ricognizioni atipiche, poiché non è possibile accertare l'idoneità del programma ad assicurare il corretto accertamento dei fatti, come richiede l'art. 189 c.p.p., dato che non sono ancora disponibili le informazioni essenziali per comprendere il suo effettivo funzionamento. Non si può neanche escludere un'eventuale lesione del diritto alla libertà morale dell'individuo, in quanto i sistemi di riconoscimento automatico delle immagini, oltre a determinare un considerevole impatto sulla garanzia dell'*habeas data*, sono in grado di realizzare un marcato *chilling effect* che induce le persone, proprio per il timore di essere soggetti al trattamento biometrico da parte dei *tools* di intelligenza artificiale, senza un loro esplicito consenso oppure in luoghi dove non si aspettano di essere sorvegliate, a modificare le proprie abitudini, autolimitandosi nell'esercizio dei propri diritti fondamentali, come, ad esempio, la libertà di riunione o di associazione, al fine di evitare le possibili conseguenze negative che potrebbero derivarne.

Come già rilevato, la normativa contenuta nel Regolamento (UE) 2024/1689 prevede alcune limitazioni in relazione all'utilizzo dei sistemi di identificazione biometrica in modalità *realtime* a fini di *law enforcement* in luoghi pubblici o aperti al pubblico (v., *supra*, § 5). L'uso di tali sistemi è consentito solo se i *software* superano una valutazione preventiva di conformità da parte di un ente certificatore terzo e soltanto in situazioni di stretta necessità. A tal riguardo, si ricordano le seguenti ipotesi: l'individuazione di specifiche vittime di reato, la prevenzione di minacce imminenti alla vita o

---

rante privacy sul sistema Sari Real Time, in *Penale Diritto e procedura*, 4 maggio 2021; F. Paolucci, *Riconoscimento facciale e diritti fondamentali: è la sorveglianza un giusto prezzo da pagare?*, in *MediaLaws*, 2021, p. 213 ss.

[52] V. Cass., Sez. Un., 28 marzo 2006, n. 26795, Prisco, in *Riv. it. dir. proc. pen*., 2006, p. 1537.

[53] Cfr. G. Borgia, *Profili sistematici delle tecnologie di riconoscimento facciale automatizzato, anche alla luce dei futuribili sviluppi normativi sul fronte eurounitario*, cit., p. 64.

all'incolumità fisica delle persone, l'individuazione di autori di reati particolarmente gravi, tra i quali la partecipazione a un'organizzazione criminale, il terrorismo, la frode e il riciclaggio, il traffico di droga e di armi, la corruzione e la criminalità informatica (art. 5, par. 1, lett. h).

Si stabilisce, inoltre, che il ricorso a sistemi di *facial recognition* deve essere calibrato in relazione alla situazione specifica e alle ricadute per i diritti e le libertà delle persone coinvolte, considerando la gravità, la probabilità e l'entità del danno causato o delle conseguenze, con eventuali limitazioni temporali, geografiche o personali, necessarie e proporzionate all'uso di tali strumenti (art. 5, par. 2).

Ogni utilizzo di queste tecnologie, infine, deve essere sottoposto all'autorizzazione di un'autorità giudiziaria o amministrativa indipendente dello Stato membro, a seguito di richiesta motivata e nel rispetto delle regole nazionali (art. 5, par. 3). Questa autorizzazione è concessa sulla base di prove oggettive e ove sussistano chiare indicazioni sulla necessità e sulla proporzionalità dell'uso dei sistemi in esame, ad eccezione delle situazioni d'urgenza, che richiedono comunque una successiva convalida.

Si può osservare, in conclusione, come sia poco realistico immaginare uno scenario in cui sia totalmente precluso l'utilizzo delle tecniche di indagine basate sull'intelligenza artificiale che sono state precedentemente descritte, le quali possono consentire agli organi inquirenti di sfruttare le straordinarie capacità analitiche degli algoritmi[54]. Occorre, quindi, accogliere favorevolmente l'innovazione tecnologica, con l'auspicio, però, che la recente regolamentazione normativa si riveli in grado di assicurare un adeguato bilanciamento tra efficienza investigativa e salvaguardia dei diritti fondamentali.

[54] In proposito, v. *Strategia europea in materia di giustizia elettronica 2024-2028* (C/2025/437), in *G.U.U.E.*, 16 gennaio 2025, C/1.

L'UTILIZZO DELL'INTELLIGENZA ARTIFICIALE NELLA FORMAZIONE DELLA DECISIONE PENALE

di Letizia Mantovani

1. *Giustizia penale e predizione algoritmica: cenni introduttivi*

Se è indubbio che, sin dall'origine della società civile, il diritto ha perseguito l'ambizioso obiettivo di plasmare i rapporti umani mediante l'individuazione e il bilanciamento di interessi contrapposti, poiché "*ubi ius ibi societas*, *ubi societas ibi ius*", è altrettanto innegabile che, in anni recenti, l'innovazione tecnologica si è dimostrata un supporto tutt'altro che marginale nello svolgimento di tale compito. Non solo, l'ingerenza sempre più pervasiva dei sistemi di intelligenza artificiale nelle dinamiche relazionali ha consolidato il legame tra coscienza individuale e ragionamento automatico, al punto da renderlo quasi inscindibile, con inevitabili conseguenze sull'organizzazione della vita collettiva e, di riflesso, sulla struttura dell'ordinamento giuridico che ne regola lo svolgimento. Nell'evoluzione simbiotica di diritto e umanità, centrale è il ruolo del sistema giudiziario, il cui grado di efficienza rispecchia, segnatamente, la qualità dell'esistenza dei singoli [1]: dinnanzi all'incessante sviluppo della conoscenza e delle sue concrete applicazioni, il rinnovamento dell'amministrazione della giustizia si prospetta, allora, come inevitabile.

---

[1] G. Pascuzzi, *Il diritto dell'era digitale*, V ed., Il Mulino, Bologna, 2020, p. 17; S. Rodotà, *Tecnologie e diritti*, Il Mulino, Bologna, 1995. determinati

Tutt'altro che marginale è, poi, l'affinità tra IA e diritto in punto di indeterminatezza definitoria: compito arduo appare, difatti, l'attribuzione a tali concetti di un significato univoco e universalmente accolto. Nella risoluzione P.E. 16/02/2017, recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica, l'intelligenza artificiale è stata definita come una tecnologia «in grado di apprendere e prendere decisioni in modo autonomo e indipendente dall'uomo»; nella "Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi" è, invece, descritta come l'«insieme di metodi scientifici, teorie e tecniche finalizzate a riprodurre mediante le macchine le capacità cognitive degli esseri umani», ulteriormente categorizzabile a seconda della sua capacità di contestualizzare i problemi che le vengono sottoposti e del grado di autonomia con cui individua la migliore strategia per risolverli [2].

L'approvazione, da parte del Parlamento europeo, dell'*Artificial Intelligence Act* [3], ha, da ultimo, delineato una definizione di "sistema di intelligenza artificiale" conforme ai valori sanciti nell'articolo 2 TUE, oltre che ai diritti e alle libertà fondamentali promulgati dai Trattati e dalla Carta di Nizza: l'art. 3, infatti, riconosce l'efficacia vincolante del Regolamento in relazione ad applicativi automatizzati, programmati per operare con variabili livelli di autonomia, che, al fine di raggiungere determinati obiettivi (espliciti o impliciti), esercitano la propria capacità inferenziale per dedurre dall'*input* ricevuto specifici *output*, quali previsioni, contenuti, raccomandazioni o decisioni in grado di influenzare ambienti fisici o virtuali. Ai sensi dell'art. 6 e del contenuto dell'Allegato III, inoltre, specifiche forme di utilizzo di processi algoritmici sono state categorizzate come "ad alto rischio", con conseguente intensificazione degli obblighi di controllo gravanti sui produttori e previsione di stringenti limiti applicativi. Rientrano in tale classificazione anche i software adottati da autorità pubbliche per valutare l'affidabilità degli elementi probatori nel corso delle indagini o del perseguimento di reati, ovvero, i programmi che assistono l'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto e nell'applicazione della legge, posta, comunque, la necessità, per le parti coinvolte nell'applicazione del Regolamento, di dare priorità alla protezione dell'integrità del processo penale. *Ex* art. 6 par. 3, si prevede, inoltre, che, in dero-

[2] Si ricorda, in merito, che l'I.A. "forte" si caratterizza, idealmente, per la sua idoneità a eguagliare, se non addirittura superare, le capacità umane e che quella "debole", di converso, studia modelli in grado di emulare il meccanismo di funzionamento della mente umana.

[3] Regolamento (UE) 1689 del 13 giugno 2024.

ga a quanto prescritto dal paragrafo 2, il sistema di IA non è considerato ad alto rischio se programmato per rilevare schemi decisionali o deviazioni da schemi decisionali precedenti, ovvero, non sia comunque finalizzato a sostituire o influenzare la valutazione del soggetto deputato a decidere, in assenza di un'adeguata revisione umana.

Alla luce dell'assetto normativo delineatosi a livello europeo, la possibilità di ricorrere a strumenti basati sulla processazione algoritmica delle informazioni incrementa nel giurista la fiducia in un modello conoscitivo automatizzato, apparentemente idoneo a fornire soluzioni caratterizzate da un elevato grado di precisione [4]. Per effetto della, seppur ancora parziale, regolamentazione operata dal diritto comunitario, tale tecnologia si presenta, infatti, quale efficace presidio della certezza e della calcolabilità giuridica, da intendersi nel senso di tendenziale uniformità e prevedibilità delle decisioni, nonché di parità di trattamento dei cittadini di fronte alla legge [5].

Con specifico riguardo alla giustizia penale, lo spazio applicativo idealmente riservabile a *tools* predittivi si estende dalla fase delle indagini preliminari fino all'esecuzione della pena. Nel contesto investigativo, di particolare rilevanza è la c.d. "*predictive policing*", che consente alle forze dell'ordine di ricorrere alla computazione algoritmica per la prevenzione e il contrasto delle attività criminose [6]. Nelle fasi processuali che seguono l'esercizio dell'azione penale, le potenzialità d'impiego dell'intelligenza artificiale predittiva si concentrano, invece, sull'acquisizione e sulla valutazione degli elementi di prova, nonché sulla formazione della sentenza: si giustifica, in questo senso, lo sviluppo di programmi destinati alla ricostruzione del fatto storico sulla base del quadro indiziario disponibile [7], nonché alla previsione di spiegazioni alternative del comportamento dell'imputato ri-

---

[4] Sul punto, v. A. Garapon, J. Lassègue, *Justice digitale. Revolution graphique et rupture anthropologique*, in *International journal of law in context,* vol. 15, *Presses Universitaires de* France, 2018, p. 536 ss.

[5] G. Canzio, *Legalità penale, processi decisionali e nomofilachia*, in *Sist. Pen.*, fasc. n. 12, 2022, p. 49 e ss.

[6] Sul tema, v. F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Dir. pen. e uomo*, 29 settembre 2019, p. 27; V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in U. Ruffolo (a cura di), *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Giuffrè, Milano, 2020, p. 547 ss.; P. Severino, *Intelligenza artificiale e diritto penale*, U. Ruffolo (a cura di), in *Intelligenza artificiale – Il diritto, i diritti, l'etica*, Giuffrè, Milano, 2020, p. 531 ss.

[7] E. Nissan, *Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement*, in *AI & Society*, 2015, p. 11 e ss.

spetto all'evento delittuoso [8]. Non meno degni di interesse sono, poi, i software progettati per la verifica, mediante il raffronto con una selezionata casistica antecedente, della corretta concatenazione delle inferenze del ragionamento probatorio [9], nonché gli strumenti di analisi computazionale delle decisioni giudiziarie, in grado di formulare previsioni circa il probabile esito della controversia [10].

Alla luce di tali inediti scenari, dischiusi dall'IA nella fase istruttoria, appare difficile resistere al fascino dei benefici prospettabili in termini di modernizzazione e incremento dell'efficienza e della qualità della giustizia penale, soprattutto nel momento di formazione della decisione: il ricorso al calcolo algoritmico, infatti, potrebbe restituire obiettività, coerenza e neutralità alla funzione giurisdizionale [11], la quale attraversa, ormai da tempo, un periodo di profonda crisi [12]. Occorre, tuttavia, non sottovalutare la presenza di limiti epistemici invalicabili, che precludono alle variegate forme di giustizia predittiva la possibilità di confluire nel patrimonio cognitivo e valutativo del giudice [13].

Numerosi sono, dunque, gli spunti per una riflessione sull'effettiva compatibilità dei sistemi predittivi con lo statuto epistemologico che permea il modello accusatorio, nonché con le garanzie che l'ordinamento pone a presidio del giusto processo [14], con particolare interesse per la potenziale incidenza di questa tecnologia sull'individuazione delle argomentazioni che fondano il ragionamento giuridico sotteso alla redazione del

---

[8] E. Nissan, *Legal evidence, Police Intelligence, Crime Analysis or Detection, Forensic Testing, and Argumentation: An Overview of Computer Tools or Techniques*, in *Int'l J.L. & Info. Tech.*, vol. 17, 2009, p. 1 e ss.

[9] H. Prakken, *Modelling Reasoning about Evidence in Legal Procedure*, in *ICAIL*, 2001, p. 119 e ss.

[10] G. Ubertis, *Intelligenza Artificiale e giustizia predittiva*, in *Sist. Pen.*, 16 ottobre 2023.

[11] G. Canzio, *Intelligenza artificiale e processo penale*, in *Cass. pen.*, fasc. n. 3, 2021, p. 798; C. Burchard, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, fasc. n. 4, 2019, p. 1926.

[12] A.M. Maugeri, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra evidence-based practices e tutela dei diritti fondamentali*, fasc. n. 73, in *Arch. pen.*, 2021, p. 3 e ss.

[13] L. Luparia Donati, *Artificial Intelligence in Criminal Courts. Opportunity or Threat*, in A. Mercedez Lopez Rodriguez, M.D. Green, M.L. Kubica (a cura di), *Legal Challenges in the New Digital Age*, 2021, p. 160 ss.

[14] L. Luparia Donati, G. Fiorelli, *Diritto probatorio e giudizi criminali ai tempi dell'Intelligenza Artificiale*, in *Dir. pen. cont. – Riv. Trim.*, fasc. 2, 2022, p. 38.

provvedimento finale [15]. In altri termini, occorre comprendere se, e con che limiti, la ricostruzione algoritmica del *thema decidendum* risulti compatibile «con una materia del diritto che si nutre di principi e regole che "umanizzano" il giudizio penale in ogni settore e forma, costituendo la procedura penale la disciplina dei modi dell'accertamento e della decisione finale» [16].

L'introduzione di tali applicativi nel processo penale, in assenza di adeguate e stringenti limitazioni, rischia, infatti, di provocare un'irreversibile metamorfosi delle categorie giuridiche codificate dal legislatore del 1988, a partire dalla ricostruzione dogmatica alla stregua della quale è stato finora inquadrato il tema della verità nel processo penale [17].

2. *Predittività della decisione: verso un diritto penale certo e calcolabile?*

Allo scopo di individuare i potenziali benefici conseguenti all'impiego di sistemi predittivi nella formazione della decisione penale, risulta opportuno interrogarsi sull'origine dell'esigenza, avvertita con impellenza dai giuristi nazionali, e non solo, di assicurare una maggiore prevedibilità degli esiti processuali. Considerato uno dei pilastri fondanti dello Stato democratico, il principio di certezza del diritto [18] impone all'ordinamento di porre ogni soggetto, dotato di capacità giuridica, nella condizione di prevedere e valutare le conseguenze delle proprie condotte. Da un punto di vista giuspositivistico, tale concetto si declina, soprattutto, come «possibilità diffusa di prevedere la gamma delle conseguenze giuridiche effettivamente suscettibili di essere spontaneamente o coattivamente ricondotte ad

---

[15] Si permetta il rimando a L. Mantovani, *Algoritmi predittivi a supporto della decisione penale: sull'opportunità di uno ius dicere "calcolabile"*, in *Ciberspazioediritto*, fasc. n. 2, 2024, 389.

[16] G. Riccio, *Ragionando su intelligenza artificiale e processo penale*, in *Arch. pen.*, fasc. n. 3, 2019, p. 2.

[17] Sul tema, v. O. Mazza, *Tradimenti di un codice. La Procedura penale a trent'anni dalla grande riforma*, Giappichelli, Torino, 2020.

[18] Per un approfondimento sul tema, v. *ex multis*, G. Alpa, *La certezza del diritto nell'età dell'incertezza*, in G. Alpa, G. Iudica (a cura di), *Costituzione europea e interpretazione della Costituzione italiana*, E.S.I., Napoli, 2006; A.M. Campanale, *La decisione tra certezza e giustezza*, in *Arch. pen.*, fasc. n. 3, 2018; F. Carnelutti, *La certezza del diritto*, in *Riv. dir. civ.*, fasc. n. 20, 1943, p. 81 ss.; M. Corsale, *Certezza del diritto*, in *Enciclopedia giuridica*, Istituto della enciclopedia italiana, 1988.

atti o fatti, nonché l'ambito temporale in cui tali conseguenze giuridiche verranno in essere»[19]. *Ex* art. 25 Cost. l'incontrovertibilità del dato normativo consente, allora, al cittadino di inquadrare con sicurezza il valore, o il disvalore, di determinate azioni e, contestualmente, dota il giudice di un parametro oggettivo su cui basare la sua valutazione circa i fatti oggetto di giudizio[20]: se il cittadino deve «poter sapere in anticipo quali sono le leggi che reggono le sue azioni e che distinguono il lecito dall'illecito»[21], significativamente limitato dovrebbe, allora, risultare lo spazio riservato all'attività ermeneutica dell'organo giudicante, in quanto operazione tecnica totalmente neutrale, che identifica il ruolo del giudice di mera *bouche de la loi*[22].

Alla luce di tale interpretazione, il tema della calcolabilità della decisione giudiziale troverebbe, quindi, fertile terreno applicativo all'interno di una logica circolare dove «decidere la controversia, giudicare torto e ragione, applicare la legge, coincidono appieno e l'uno sta per l'altro»[23]. Secondo Max Weber, la fattispecie, quale paradigma unificante della dogmatica razionale, consente di perseguire la certezza del diritto mediante meccanismi razionali di tipo sillogistico: il giudice deciderebbe in modo meccanico, quasi automatico, rendendo il diritto prevedibile, calcolabile e applicabile nella genericità di casi simili e uguali[24].

Se il provvedimento che definisce il processo si fonda su una valutazione risultante dal confronto tra il dato normativo astratto e la fattispecie concreta per cui si procede, si desume che proprio dalla corretta applicazione della legge dipende il contenuto della decisione medesima. L'equazione, di matrice illuministica, secondo la quale certezza del diritto e certezza della norma costituiscono l'una condizione necessaria del-

---

[19] G. Gometz, *Indici di certezza del diritto*, in *Diritto & Questioni Pubbliche*, fasc. n. 12, 2012, p. 308 e ss.

[20] A. Cadoppi, *Il valore del precedente nel diritto penale. Uno studio sulla dimensione in action della legalità*, II ed., Giappichelli, Torino, 2014, p. 49 e ss.

[21] A. Cadoppi, *Il valore del precedente nel diritto penale. Uno studio sulla dimensione in action della legalità*, cit., p. 51; v. anche C. Beccaria, *Dei delitti e delle pene*, a cura di P. Gonnella, S. Marietta, ed. I, Giappichelli, Torino, 2022.

[22] V. Maiello, *Legge e interpretazione nel 'sistema' di dei delitti e delle pene*, in *Discrimen*, 18 novembre 2020, p. 24.

[23] N. Irti, *Un diritto incalcolabile*, in *Riv. dir. civ.*, 2015, p. 11 ss.; v. anche N. Irti, *La crisi della fattispecie*, in *Riv. dir. proc.*, 2014, p. 36; G. Canzio, *Calcolo giuridico e nomofilachia*, in AA.VV., *Calcolabilità giuridica*, a cura di A. Carleo, Il Mulino, Bologna, 2017, p. 169 ss.

[24] G. Canzio, *Nomofiliachia e diritto giurisprudenziale*, in *Dir. pubbl.*, n. 1/2017, p. 21 ss.

l'altra, tuttavia, si confronta con una realtà dai confini mutevoli e in continua evoluzione, che, molto spesso, si sottrae al "sillogismo perfetto" prima teorizzato.

L'attività ermeneutica deve essere compiuta, in primo luogo, nel rispetto dei principi costituzionali e di diritto internazionale: il contenuto della singola disposizione legislativa si confronta, così, con un parametro interposto che attribuisce alla norma nuove sfumature, espressione di un preciso valore che assume la funzione di canone interpretativo[25]. Un ulteriore fattore determinante della crisi in cui versa il principio della certezza del diritto si identifica nella formulazione della legge: il precetto normativo dovrebbe essere articolato in modo tassativo, chiaro e comprensibile sia per i destinatari, che possono regolare di conseguenza la propria condotta, sia per le pubbliche autorità nelle loro valutazioni e decisioni.

Pur nel rispetto del principio di determinatezza e delle esigenze dallo stesso tutelate[26], l'esistenza di un consolidato indirizzo giurisprudenziale può, comunque, rappresentare un'ulteriore e utile conferma «della possibilità di identificare, sulla scorta d'un ordinario percorso ermeneutico, la più puntuale valenza di un'espressione normativa in sé ambigua, generica o polisensa»[27]. L'interpretazione normativa è imprescindibile, poiché l'assoluta certezza e chiarezza dell'espressione normativa non è da sola sufficiente a perfezionare il giudizio sul fatto storico e il suo potenziale disvalore. Il contenuto semantico della regola giuridica non è un dato precostituito, unicamente desumibile dal tenore letterale della norma, bensì il frutto di una valutazione del giudice, anche alla luce delle peculiarità del caso di specie. L'ordinamento giuridico deve, quindi, dotarsi di nuovi strumenti e prassi metodologiche in grado di ridurre efficacemente i contrasti interpretativi, assicurando un'applicazione della legge il più possibile uniforme e certa, al netto delle variabili finora considerate.

La costante ricerca di stabilità e sicurezza indirizza il moderno Stato di diritto verso l'uso della tecnologia e di modelli statistici di previsione dei giudizi, come forma di tutela contro i rischi di arbitrio e di soggettivismo che possono inficiare l'attività dei magistrati. In questa prospettiva, la prevedibilità delle decisioni potrebbe garantire quel grado di certezza a cui la

---

[25] G. Barone, *Giustizia predittiva e certezza del diritto*, Pacini giuridica, Pisa, 2024, p. 105.

[26] Ossia: garantire la concentrazione nel potere legislativo della produzione della *regula iuris* e assicurare al destinatario del precetto penale la conoscenza preventiva di ciò che è lecito e di ciò che è vietato. A riguardo, v. Scevi, *La crisi della legalità nel diritto penale. Progressiva dissoluzione o transizione verso una prospettiva di crescita?*, in *Arch. pen*., fasc. n. 3, 2017, p. 14.

[27] Corte cost., n. 327 del 2008, par. 6, in *Giur. cost.*, 2008, 3534.

scienza giuridica, per sua natura, dovrebbe tendere [28].

La questione, tutt'altro che pacifica, si pone al centro di un dibattito tanto vivace quanto, per ora, ancora irrisolto. Da un lato, vi è chi sottolinea come l'impiego di software predittivi possa, entro certi limiti, favorire la verifica dei criteri adottati dal giudice nell'esercizio delle sue funzioni e, conseguentemente, la comprensione della *ratio decidendi* sottesa al provvedimento che definisce il giudizio [29]. Dall'altro, vi è, invece, chi suggerisce di rivalutare la dimensione intrinsecamente umana dell'attività giudiziaria, alla quale appartengono anche elementi "incalcolabili", come il linguaggio visivo, le emozioni o l'esperienza derivante dalla vita di relazione [30].

Occorre, pertanto, indagare non solo le possibili modalità di interazione tra amministrazione della giustizia e computazione algoritmica, ma altresì, e soprattutto, i potenziali effetti derivanti dall'impiego di quest'ultima sulla qualità del percorso argomentativo su cui si basa l'esercizio dello *ius dicere*. L'idea di fondo è che l'attività ermeneutica che caratterizza la funzione giurisdizionale sia dotata di un *animus* strettamente antropocentrico, destinato, se non a contrastare, almeno a influenzare sensibilmente i modelli di apprendimento automatico che definiscono il fenomeno della giustizia predittiva [31].

3. *La funzione aletica del processo penale messa alla prova dall'intelligenza artificiale*

Con particolare riferimento al sistema della giustizia penale, gli interventi legislativi finalizzati a prevenire la diffusione di orientamenti giurisprudenziali antitetici e contraddittori non risultano, attualmente, sufficienti a rivitalizzare la funzione nomofilattica della Corte di cassazione, il cui scopo, *ex* art. 65 R.D. 30 gennaio 1941 n. 12, è proprio di garantire l'esatta osservanza e l'uniforme interpretazione della legge e l'unità del diritto oggettivo nazionale. Pur non essendo espressamente codificata nel codice di rito, la

---

[28] F. Scamardella, M. Vestoso, *Modelli predittivi a supporto della decisione giudiziaria. Alcuni spunti di riflessione*, in *Rivista di filosofia del diritto*, fasc. n. 1, 2023, p. 137.

[29] M. Luciani, *La decisione giudiziaria robotica." Rivista Associazione italiana dei costituzionalisti*, in *Rivista Associazione italiana dei costituzionalisti*, fasc. n. 3, 2018, p. 872 ss.

[30] N. Irti, *Un diritto incalcolabile*, cit., p. 36 ss.

[31] Per approfondire, v. A. Garapon, J. Lassègue, *La giustizia digitale. Determinismo tecnologico e libertà*, Il Mulino, Bologna, 2021.

rilevanza del precedente, mutuata dal principio di *stare decisis*, che regola i sistemi di *common law*, trova un riscontro normativo negli artt. 610, 618, 627, comma 3, 628, comma 2, c.p.p. e 172-173 disp. att. c.p.p.

In particolare, la formulazione del principio di diritto da parte delle Sezioni Unite può tradursi in una fondamentale direttiva ermeneutica, in funzione della prevedibilità delle decisioni future, della coerenza del sistema e di un auspicato effetto deflattivo rispetto all'insorgere di procedimenti non necessari[32]. Il criterio che orienta la decisione sulla fattispecie concreta si ricava da una regola giuridica destinata ad essere applicata non solo in situazioni uguali ma anche in quelle simili o assimilabili: il principio di diritto, individuando una specifica modalità applicativa di una norma generale e astratta, di fatto, "universalizza" il contenuto di una singola sentenza[33].

Al fine di preservare il valore del precedente e, conseguentemente, tentare di rivitalizzare il ruolo della certezza del diritto nella giustizia penale, parte della dottrina[34] si è, quindi, interrogata sulla possibilità di fare ricorso a strumenti di giustizia predittiva. L'applicazione della giurimetria, intesa come lo studio qualitativo e quantitativo delle decisioni giudiziarie, potrebbe contribuire, ad esempio, a ridurre l'imprevedibilità degli esiti processuali, mediante l'elaborazione di modelli decisionali implementati dall'intelligenza artificiale.

Il recupero e l'organizzazione dei precedenti giurisprudenziali presenti all'interno delle principali banche dati giuridiche, finalizzato all'immediata individuazione di elementi, fattuali e normativi, comuni, consentirebbe, infatti, di raggiungere in tempi brevi una soluzione del caso concreto coerente con quanto già deciso in relazione a questioni analoghe o simili. Si attiverebbe, così, un irreversibile processo di oggettivizzazione delle decisioni che scandiscono le fasi del giudizio: «la possibilità che un preciso calcolo matematico pervenga a determinare il livello di pericolosità sociale di un individuo, riesca a ponderare il rischio di recidiva e sia ragionevolmente in grado di sostituirsi al giudice nel fondare una sentenza di condanna, prima che un ideale illuministico, sembra oggi essere divenuta una realtà storica»[35].

---

[32] G. Gorla, *voce Precedente giudiziale*, in *Enc. giur. Treccani*, 1990, p. 11.

[33] G. De Amicis, *La formulazione del principio di diritto e i rapporti tra sezioni semplici e Sezioni Unite penali della Corte di cassazione*, in *Sist. Pen*., 2020, p. 113 ss.

[34] R. Bichi, *Intelligenza artificiale, giurimetria, giustizia predittiva e algoritmo decisorio. Machina Sapiens e il controllo sulla giurisdizione*, in U. Ruffolo (a cura i) *Intelligenza artificiale. Il diritto, i diritti e l'etica*, Giuffrè, Milano, 2020, p. 429.

[35] B. Occhiuzzi, *Algoritmi predittivi: alcune premesse metodologiche*, in *Dir. pen. cont. – Riv. Trim*., 2019, n. 2, p. 393.

In questa prospettiva, però, il concetto di "giustizia predittiva" assume le sembianze di una contraddizione in termini, ai limiti dell'ossimoro[36]. L'algoritmo, replicando un'operazione tipicamente umana, seleziona e apprende informazioni, finalizza la propria ricerca e identifica il risultato più probabile, al netto delle operazioni di calcolo realizzate dal sistema operativo, con un'efficienza in larga misura superiore a qualsiasi metodo alternativo di analisi di informazioni. L'individuazione del dato normativo idoneo a regolare il singolo caso e la sua interpretazione sono, tuttavia, necessariamente correlate all'accertamento giudiziale di un fatto concreto, verificatosi nel passato. È la sola condotta umana, infatti, che, una volta realizzata, produce effetti giuridicamente rilevanti, compreso il perfezionamento di un fatto illecito di rilevanza penale.

L'attività predittiva, pertanto, non attiene né alla *regiudicanda* né all'opportunità di esercitare la pretesa punitiva nei confronti dell'autore di reato, ma incide sullo sviluppo del ragionamento giuridico su cui si fonda la decisione finale. In questo senso, «chi sostiene che in ambito giudiziario l'intelligenza artificiale serve a predire il futuro dimentica il suo utilizzo nel processo anche per conoscere il passato»[37]. Il significato di "predizione" differisce, quindi, da quello di "previsione", in quanto il primo lemma identifica l'atto di annunciare anticipatamente avvenimenti futuri, mentre il secondo attiene al risultato dell'osservazione di un insieme di dati al fine di prospettare una possibile, se non probabile al limite della certezza, situazione futura.

Occorre, poi, sottolineare che il processo penale, nella sua duplice dimensione, cognitiva e decisoria, osserva i paradigmi propri della tradizione razionalista occidentale, su cui si fonda il sapere scientifico e, conseguentemente, anche la scienza giuridica[38]. Nel modello accusatorio, il ricorso a inferenze di natura probabilistica consente, in primo luogo, l'individuazione dell'ipotesi più verosimile e, successivamente, contribuisce alla sua conferma o confutazione "al di là di ogni ragionevole dubbio"[39].

Al ragionamento giuridico si accompagna, in questo senso, un'imprescindibile retrospezione: non potendo essere oggetto di percezione diretta,

---

[36] G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Dir. pen. cont. – Riv. Trim*, 2020, n. 4, p. 76 e ss.

[37] G. Ubertis, *Intelligenza Artificiale e giustizia predittiva*, cit., p. 2.

[38] G. Canzio, *La "dike" degli antichi e la "giustizia" dei moderni: "Edipo re" e "Antigone"*, in *Diritto Penale Contemporaneo*, n. 1, 2018, p. 4 e ss.

[39] G. Canzio, *Intelligenza artificiale, algoritmi e giustizia penale*, in *Sist. Pen.*, 2021.

le cause che hanno determinato un evento devono essere necessariamente dedotte alla luce degli effetti e delle conseguenze rinvenibili nell'esperienza attuale. Il giudice, conseguentemente, è gravato da un adempimento tanto impossibile quanto indispensabile, poiché l'esigenza di concretizzare la pretesa punitiva mediante la repressione di un illecito è mitigata dal limite conoscitivo che attiene al fatto storico irripetibile [40].

Il percorso logico-argomentativo della decisione si basa su una ricostruzione che è, di per sé, idonea a limitare, ma non a elidere completamente, lo scarto tra verità fattuale e processuale. Il sapere giudiziale assolve, così, una funzione euristica, diretta alla progressiva elaborazione del provvedimento finale, mediante l'applicazione di modelli conoscitivi all'interno di un protocollo formale e ritualizzato, massima espressione dell'equilibrio legislativo tra regole e garanzie [41]. La primaria funzione dello *ius dicere* è, dunque, salvaguardata dal processo penale, che mette a servizio del giudizio «regole, modelli e discernimento» [42], finalizzati a giustificare, sotto il profilo razionale, il contenuto della sentenza.

L'impiego di massime o regole scientifiche, così come il ricorso a inferenze logico-probabilistiche esibiscono, tuttavia, margini più o meno ampi di incertezza che impongono al giudicante, in sede di motivazione, di riaffermare la supremazia della propria conoscenza [43]. Occorre, infatti, ricordare che l'individuazione e l'applicazione del precetto penale al caso concreto non possono prescindere dal mutevole, quanto pregnante, influsso del contesto sociale di riferimento né, tantomeno, dalla personale interpretazione del giudicante.

Se l'esperienza giuridica è necessariamente inquadrata all'interno di coordinate valutative soggettive, pur sorrette da una solida rete di regole epistemiche e procedurali, il magistrato è, allora, tenuto a un continuo aggiornamento del proprio sapere, così da poter far fronte con prontezza all'evoluzione della società: l'impiego di strumenti predittivi basati sull'intelligenza artificiale, quale massima espressione della rivoluzione tecnologica che sta rimodellando la civiltà del XXI secolo, appare un'allettante e tempestiva risposta a tale necessità.

D'altro canto, il modello legale di motivazione, quale momento «di un

---

[40] G. Giostra, *Prima lezione sulla giustizia penale*, Laterza, Bari, 2020, p. 3 e ss.

[41] P. Felicioni, *L'attività valutativa del giudice tra ragione ed emozione*, in G.M. Baccari, P. Felicioni, *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, Giuffrè, Milano, 2023, p. 5.

[42] F. Cordero, *Procedura penale*, Giuffrè, Milano, 2012, p. 9.

[43] E. Amodio, *Mille e una toga. Il penalista tra cronaca e favola*, Giuffrè, Milano, 2010, p. 170 ss.

complesso itinerario della ragione» [44], rischia, così, di comprimersi a favore di una mera validazione dei risultati ottenuti dalla computazione algoritmica, ritenuti affidabili proprio perché frutto di un'attività non umana e, pertanto, meno fallibile. La funzione aletica del processo penale cederebbe il passo a un approccio estremamente razionale, quasi matematico, che, pur consentendo di schermare parzialmente il *decisum* da polemiche e critiche in ordine al suo contenuto, per altro verso, condurrebbe inevitabilmente a «un giudizio senza decisione» [45].

Il fattore umano, infatti, eleva inevitabilmente la sentenza da un insieme di dati e informazioni a massima espressione dell'*ars iudicandi*. Proprio a questo fine tende l'obbligo di motivazione, sancito *ex* art. 111 c. 6 Cost. e art. 546 c.p.p.: il giudice è tenuto a giustificare personalmente la *ratio* della propria scelta in merito ai fatti oggetto del giudizio, esponendo il percorso logico su cui si è formata la sua valutazione. In questo senso, l'azione di decidere coincide con quella di motivare, poiché «l'interpretazione, pur imbrigliata da canoni e regole dettate dal legislatore, dai precedenti, dalla dottrina, non si risolve in una pura catena di passaggi deduttivi: vi sono spesso alternative logicamente equivalenti e il giudice deve scegliere e spiegare le ragioni della scelta. [...] Il libero convincimento del giudice è criterio che non si risolve in un soggettivismo arbitrario ma non può essere sostituto dall'apparente oggettività di una intelligenza artificiale» [46].

4. *I potenziali vantaggi della (parziale) automazione decisoria: le euristiche del giudice*

Come già anticipato, l'assenza di autonomia cognitiva induce l'applicativo di intelligenza artificiale a selezionare dati astratti, estrapolati dal sapere scientifico e dalla giurisprudenza, al fine di individuare e raffrontare argomentazioni giurisprudenziali il più possibile afferenti alla fattispecie concreta, oggetto di decisione. L'algoritmo predittivo s' incentra, dunque, su un giudizio di mera probabilità statistica, il cui limite è insito nello stesso concetto di eventualità.

---

[44] G. Canzio, *La "dike" degli antichi e la "giustizia" dei moderni: "Edipo re" e "Antigone"*, cit., p. 5.

[45] F.R. Dinacci, *Intelligenza artificiale tra quantistica matematica e razionalismo critico: la necessaria tutela di approdi euristici,* in *Processo penale e giustizia*, n. 6, 2022, p. 1627 ss.

[46] R. Bichi, *Intelligenza artificiale tra 'calcolabilità' del diritto e tutela dei diritti*, in *Giurisprudenza italiana*., 2019, p. 1778.

Imprescindibile è, quindi, il rimando al teorema di Bayes, secondo il quale il risultato di un'inferenza statistica corrisponde a una variabile aleatoria tra gli infiniti numeri compresi tra 0 e 1. In base alla logica bayesiana, infatti, l'attendibilità della tesi accusatoria appare direttamente proporzionale al "quoziente di verosimiglianza" [47], quale indice su cui fondare e quantificare la credenza in un'ipotesi alla luce di una prova: il concetto di probabilità si declina, dunque, nel livello di fiducia sul verificarsi o meno di un determinato evento. Il ricorso alle c.d. "*naked statistics*", tuttavia, impone la tendenziale uniformità dei casi osservati in precedenza a eventi nuovi e non ricompresi nel suddetto calcolo frequenziale.

Inoltre, secondo il filosofo statunitense John Searle, non esistono processi computazionali indipendenti dall'interpretazione umana, poiché l'attività della macchina non è altro che un processo matematico astratto che esiste solo in relazione a interpreti coscienti. La processazione algoritmica, non presupponendo stati mentali causalmente incidenti sull'attività di calcolo, impedisce all'intelligenza artificiale di dare un senso ai simboli linguistici che elabora, caratteristica, quest'ultima, tipicamente umana [48].

La simulazione di un processo cognitivo non produce, dunque, i medesimi risultati che conseguono a una reazione neurofisiologica, come accade nella mente di un soggetto pensante [49]. Conseguentemente, pur avvalendosi di canoni di certezza logico probabilistica e di un ragionamento di tipo inferenziale, gli algoritmi predittivi non sarebbero in grado di ricostruire tutte le sfumature proprie della scienza giuridica, pervenendo esclusivamente a una valutazione della ripetibilità di una condotta giuridicamente rilevante, nota ma non necessariamente reiterabile in futuro.

Pur apparendo, allora, imprescindibile, alla luce dei limiti appena osservati, un apporto significativo della mente umana nell'*iter* decisorio, il ricorso a *tools* basati sull'IA potrebbe comunque contribuire a incrementare sensibilmente l'efficienza dell'organo giudicante. Nonostante, infatti, l'attività dei giudici si basi, soprattutto, sull'interpretazione del diritto vigente, l'influenza delle ragioni ideologiche sull'esito del processo è altrettanto rilevante [50]: a decisioni eque e imparziali, basate su un rigido e ossequioso rispetto della norma, all'uopo individuata in relazione al caso con-

---

[47] C. Costanzi, *La matematica del processo: oltre le colonne d'Ercole della giustizia penale*, in *Questione giustizia*, n. 4, 2018, p. 166 ss.

[48] J.R. Searle, *La mente*, Raffaello Cortina, Milano, 2006, p. 115.

[49] G. Pasceri, *La predittività delle decisioni*, Giuffrè, Milano, 2022, p. 42 ss.

[50] R.A. Posner, *Como deciden los jueces*, Marcial Pons, Madrid, 2011, p. 71 ss.

creto, si sostituiscono, spesso, ragionamenti giuridici influenzati da convinzioni personali, informazioni superflue o erronee, che alterano inevitabilmente il contenuto e la qualità del giudizio [51].

Nell'*iter* decisionale si inserisce, così, il fenomeno delle euristiche, scorciatoie cognitive emblematiche delle aporie e dei limiti tipici del ragionamento giudiziale. Si tratta di schemi decisionali semplificati che assolvono una funzione di economia di pensiero, con cui l'essere umano accelera il processo di scelta, reiterando comportamenti passati, con il rischio di commettere errori sistematici di valutazione (*bias*) [52]. I meccanismi su cui si fonda il pensiero razionale sono, pertanto, soggetti a deviazioni sistematiche da un modello di ragionamento ritenuto a priori "corretto" e, in quanto tali, possono essere anticipati mediante valutazioni di tipo probabilistico: determinante appare, in questo senso, il ruolo degli algoritmi predittivi.

Secondo l'euristica "della rappresentatività", nel momento in cui deve stabilire la priorità di un determinato evento al fine di determinare la propria scelta, l'agente tende a basarsi su una stima di quanto tale situazione sia significativa rispetto alla gamma di scenari prospettabili, attinenti all'oggetto della decisione. Declinato in una prospettiva giuridica, tale fenomeno si manifesta, ad esempio, nella ricerca del precedente giurisprudenziale: mediante il raffronto con casi simili, avvenuti nel passato, l'interprete individua elementi, fattuali e di diritto, su cui fondare le proprie determinazioni in merito alla questione sottoposta alla sua attenzione. In un'ottica di miglioramento, quantitativo e qualitativo, di tale funzionalità cognitiva, è stato prospettato l'intervento dell'intelligenza artificiale al fine di ridurre la percentuale di errore, propria dell'euristica umana, trasformandola in un calcolo statistico: «si può quindi immaginare un'applicazione che legga gli scritti delle parti e identifichi l'argomento della questione da decidere, e che, conseguentemente, cerchi la giurisprudenza applicabile per determinare una o più possibili soluzioni alternative, fornendo anche dati percentuali del rapporto di frequenza di ciascuna di tali alternative. Si tratterebbe di un programma che, pur non imponendo al giudice la sentenza, sgraverebbe il giudice stesso di una parte importante del proprio lavoro» [53].

Mediante l'euristica "dell'accessibilità", invece, l'individuo valuta la probabilità che un evento si verifichi a seconda della maggiore o minore facilità

---

[51] G. Cevolani, V. Crupi, *Come ragionano i giudici: razionalità, euristiche e illusioni cognitive*, in *Criminalia – Annuario di scienze penalistiche*, 2017, p. 202.

[52] J. Nieva-Fenoll, *Intelligenza artificiale e processo*, Giappichelli, Torino, 2019, p. 33.

[53] *Ibidem*.

con cui tende a ricordarlo. Nel corso del procedimento penale, ad esempio, la percezione di un pericolo potenzialmente imminente, tale da giustificare l'adozione di una misura cautelare detentiva, sulla base delle esigenze codificate dall'art. 274 c.p.p., potrebbe influenzare il giudicante anche nell'accertamento della effettiva responsabilità penale dell'imputato, determinando un'illegittima correlazione tra applicazione della custodia cautelare e successiva condanna di un soggetto già *in vinculis*. Tale processo mentale discende dal pregiudizio della c.d. "correlazione illusoria", in base al quale una persona viene ricordata e rappresentata dal giudicante esclusivamente o prevalentemente per una determinata caratteristica o in relazione a un determinato fatto. Un software basato sull'apprendimento automatico potrebbe costituire un valido ausilio per contrastare tali variabili emotive, contribuendo, almeno apparentemente, al raggiungimento di un giudizio più imparziale.

L'euristica "dell'ancoraggio e dell'aggiustamento" si basa sull'assunto secondo cui, solitamente, l'essere umano, per affrontare un problema, tende a semplificare la realtà e a farsi un'idea della migliore soluzione ipotizzabile, sin dall'inizio. Tale propensione, per quanto contrastata dalla conoscenza tecnica e dalla legge, che impone un *iter* procedurale specifico e predefinito, nonché dei precisi limiti allo sviluppo di una personale regola di giudizio, è propria dell'attività del giudicare. In altri termini, per quanto possa assumere informazioni e prove che contrastano con l'ipotesi inizialmente formulata, il giudice tenderà a reinterpretare tali informazioni a sostegno della propria tesi, dando origine a un *bias* di conferma. In questo caso, però, l'intervento dell'intelligenza artificiale si prospetta tutt'altro che salvifico: anche una macchina tende all'ancoraggio e all'aggiustamento, perché si adegua sempre a quanto indicato dall'algoritmo sulla base del quale è stata programmata. La processazione computazionale, infatti, è altamente sensibile all'introduzione di nuovi dati di ingresso, ma non anche all'inserimento di elementi che suggeriscono riconsiderazioni del risultato finale senza modificare gli *input* iniziali[54].

L'euristica "affettiva", infine, dimostra come l'essere umano sia condizionato dalle variabili emotive indotte dal linguaggio o dall'apparenza: se, da una parte, tale condizione risulta una componente imprescindibile dell'*ars iudicandi*, dall'altro, comporta un'inevitabile imprevedibilità del contenuto della decisione. Sotto questo profilo, l'agente artificiale, appare, invece, estraneo a tali condizionamenti e potenzialmente idoneo a ridurre sensibilmente il margine di aleatorietà che caratterizza l'*iter* logico seguito dal giudice[55].

---

[54] G. Pasceri, *La predittività delle decisioni*, cit., p. 56 ss.

[55] J. Nieva-Fenoll, *Intelligenza artificiale e processo*, cit., p. 41.

5. *L'effetto* **black box** *e la neutralità (solo) presunta dell'intelligenza artificiale: i limiti della decisione algoritmica*

Pur ipotizzando un futuribile impiego di software predittivi a supporto della decisione penale, occorre, comunque, considerare il rischio che l'inconscia fiducia riposta dall'uomo «nelle tecnologie, ritenute oggettive e meritevoli di fiducia per il solo fatto di essere tecnologie»[56], dissimuli l'inevitabile influenza delle dinamiche sociali, che ne hanno reso possibile il funzionamento. L'operatività di sistemi basati sull'apprendimento automatico è necessariamente influenzata da interventi esterni, che incidono sulla formazione e sulla scelta dei dati elaborati dall'algoritmo[57], il quale «è ontologicamente condizionato dal sistema di valori e dalle intenzioni di chi ne commissiona la creazione e/o di chi lo crea»[58].

Nei processi di apprendimento basati sul *deep learning* – a differenza dei sistemi computazionali di tipo deterministico, nei quali, a partire da un dato *input*, si ottiene sempre il medesimo *output* – aumenta sensibilmente il rischio di opacità per quanto riguarda le tecniche di elaborazione degli *input* da parte dell'algoritmo. Molti applicativi di I.A. di ultima generazione sono, infatti, in grado di riprogrammarsi autonomamente durante la processazione delle informazioni, al fine di perseguire più efficacemente l'obiettivo per cui sono stati progettati: l'operatore umano non è in grado di conoscere né lo stato computazionale della macchina in un preciso momento né l'esatta sequenza procedurale che ha portato l'agente artificiale a ottenere un determinato risultato, nonostante sia conosciuta la base dati di partenza.

Il verificarsi di tale circostanza, denominata "effetto *black box*", ostacola fortemente, sotto il profilo applicativo, l'inserimento di algoritmi predittivi nel sistema giudiziario penale. Se le modalità di analisi dei dati e le correlazioni funzionali al raggiungimento dell'*output* finale non sono riconoscibili, anche il percorso logico seguito dalla macchina, la struttura e persino il contenuto dell'attività computazionale rimangono, di fatto, ignoti. All'impossibilità di verificare l'attendibilità delle previsioni realizzate dal software consegue, quindi, l'inutilizzabilità delle stesse in sede processuale. Si deve, poi, considerare che i risultati conseguiti dalla macchina sono soggetti all'inter-

---

[56] P. Comoglio, *Prefazione*, in J. Nieva-Fenoll, *Intelligenza artificiale e processo*, Giappichelli, Torino, 2019, pp. X-XV.

[57] O. Di Giovine, *Il judge-bot e le sequenze giuridiche in materia penale (intelligenza artificiale e stabilizzazione giurisprudenziale)*, in *Cassazione penale*, pp. 951-965.

[58] S. Signorato, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo"*, in *Rivista italiana di diritto e procedura penale*, pp. 605-616.

pretazione personale di chi ne fa uso e che le informazioni sottoposte a processazione algoritmica discendono, comunque, da un'imprescindibile categorizzazione dei dati di partenza, operata dai programmatori e indotta dalle esigenze degli stessi utilizzatori: gli *input* così elaborati, allora, «perdono il margine di flessibilità interpretativa che caratterizza la comprensione degli accadimenti, specialmente se riferiti all'esperienza umana»[59].

Inoltre, con particolare riferimento alla decisione penale e al relativo obbligo di motivazione, la valorizzazione del percepito soggettivo del giudice non sembra sostituibile da valutazioni puramente algoritmiche: la scienza argomentativa, *ars boni et aequi*, implica il necessario apporto di prospettive personali, che esaltano la consapevolezza e la sensibilità del giudicante. La parte motiva del provvedimento finale, perderebbe, altrimenti, di significato.

È pur vero che elementi riconducibili a una logica "matematica" dell'*iter* processuale[60] sono rinvenibili, astrattamente, nel dettato normativo di cui all'art. 533, c. 1, c.p.p., nella misura in cui si richiede, ai fini della condanna, una netta preponderanza della probabilità di colpevolezza rispetto a quella di innocenza. In questo senso «se nessun giudizio storico è tale che sia assolutamente impossibile predicare il contrario, il concetto di verità processuale si può ottenere soltanto a prezzo di una determinazione quantitativa sulle probabilità contrarie»[61].

Per effetto della regola di giudizio prevista dal codice di rito grava sulla pubblica accusa l'incognita dell'esistenza del fatto di reato, con conseguente quantificazione del relativo onere probatorio nella misura dell'oltre ogni ragionevole dubbio[62]. Il positivo superamento di tale canone da parte del giudice, alla luce delle risultanze processuali e dell'accertamento in concreto basato sulle evidenze disponibili al momento della decisione, indica la soluzione di un quesito che ha sì a che fare con la logica, ma ancora di più con l'esperienza e l'intuizione dell'uomo. Giudicare è, prima di tutto, un atto morale, cui consegue il dovere del magistrato di giustificare le proprie decisioni, risaltando, mediante ragione, ciò che è giusto.

L'idea che al giudice sia riconosciuta la possibilità di conformare le proprie determinazioni alle previsioni di un software basato sull'intelligenza artificiale, oltre a inquadrare l'attività decisoria in una rigida forma di de-

---

[59] Cfr. G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, cit., p. 76 ss.

[60] F. Carnelutti, *Matematica e diritto*, in *Riv. Dir. Proc.*, 1951, fasc. n. 1, p. 201.

[61] F. Cordero, *Note sul procedimento probatorio*, in *Jus*, fasc. n. 1-2, 1963, p. 45

[62] E.M. Catalano, *Ragionevole dubbio e logica della decisione*, Giuffrè, Milano, 2016, p. 4 ss.

terminismo, comporterebbe un inaccettabile svilimento dell'indipendenza dell'autorità giudiziaria, costituzionalmente sancita *ex* art. 101, comma 2 Cost., nonché del requisito della naturalità dell'organo giudicante, previsto dall'art. 25, comma 1, Cost. Il magistrato rischierebbe di deresponsabilizzare il proprio ruolo, delegando parte della funzione giudicante a un'entità esterna ed estranea alla legge, l'algoritmo.

Il ricorso a software predittivi, basati su un approccio *case-based*[63], innesterebbe un processo di auto-avveramento suscettibile di produrre un duplice effetto negativo. Da una parte, il circolo virtuoso della giurisprudenza si arresterebbe, a favore di un appiattimento dell'attività decisoria, non più in grado di mutuare la costante trasformazione del contesto sociale di riferimento nell'interpretazione delle norme giuridiche. D'altro canto, si verificherebbe, in via mediata, un'implicita egemonia della regola giurisprudenziale, essendo la decisione dell'applicativo di intelligenza artificiale improntata sul canone, proprio dei sistemi di *common law*, dello "*stare decisis*": da un sistema improntato sul principio di legalità si passerebbe, così, a un ordinamento fondato sul precedente, in evidente violazione degli artt. 25, comma 2, 101, comma 2 e 111, comma 1, Cost.

In conclusione, l'accertamento della responsabilità penale non può fondarsi unicamente su risultanze algoritmiche che, per quanto apparentemente oggettive, sottendono un procedimento di elaborazione non verificabile *ex post* e, di conseguenza, non del tutto attendibile. In tal senso, sia la legislazione europea, con gli articoli 22 reg. n. 2016/679 e 11 dir. n. 2016/680/UE, sia l'ordinamento italiano, mediante l'art. 8 d.lgs. n. 51 del 2018, vietano, in generale, decisioni fondate esclusivamente sul trattamento automatizzato dei dati, prescrivendo un tassativo intervento umano nel processo decisionale, così da garantire il controllo, la validazione o la smentita della decisione automatica. La predominanza del "fattore umano" si pone, dunque, quale irrinunciabile garanzia, al consapevole utilizzo del dato conoscitivo da parte dell'organo giudicante, nel rispetto della funzione cognitiva propria del razionalismo critico e delle regole epistemologiche poste a sua tutela dal codice di rito[64].

---

[63] K.D. Ashley, *Artificial Intelligence and Legal Analytics*, Cambridge University Press, Cambridge, 2017; A. Santosuosso, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Mondadori, Milano, 2020; E. Bassoli, *Algoritmica giuridica. Intelligenza artificiale e diritto*, Amon Edizioni, Milano, 2021.

[64] Cfr. F.R. Dinacci, *Intelligenza artificiale tra quantistica matematica e razionalismo critico: la necessaria tutela di approdi euristici*, in *Processo penale e giustizia*, cit., p. 1637.

INTELLIGENZA ARTIFICIALE E SISTEMA PENITENZIARIO[*]

di Gaia Caneschi

1. *Una premessa necessaria*

È ormai diffusa la consapevolezza che l'inarrestabile avvento dell'intelligenza artificiale potrebbe comportare un vero e proprio «mutamento di paradigma» nella giustizia penale [1]. In effetti, se le innovazioni che fino a ieri sembravano appartenere ad un futuro remoto, quasi fantascientifico, sono entrate a far parte della realtà quotidiana, tanto da avere condotto taluno a parlare di un «nuovo capitolo della storia umana» [2], non sembra pensabile che il sistema penale possa restare immune a questa rivoluzione.

---

[*] Il presente contributo è già stato pubblicato sul fascicolo 1/2024 della *Rivista italiana di diritto e procedura penale*. Rispetto alla versione contenuta nella *Rivista*, il testo qui riportato è stato opportunamente aggiornato, tenendo in considerazione, in particolare, l'approvazione e la pubblicazione del testo definitivo del Regolamento dell'Unione europea sull'Intelligenza artificiale (Reg. 2024/1689).

[1] È l'espressione utilizzata da G. Canzio, *Il dubbio e la legge*, in *Dir. pen. cont.*, 20 luglio 2018, p. 4. A. Garapon, J. Lasségue, *Justice digitale. Révolution graphique et rupture anthropologique*, Parigi, 2018, *passim*, per descrivere il fenomeno dell'ingresso dell'intelligenza artificiale in un ambito – quello della giustizia penale – pensato e strutturato sull'uomo, parlano icasticamente di una «frattura antropologica».

[2] Così L. Floridi, *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Milano, 2022, p. 11.

Sebbene i tempi di reazione dell'ordinamento nazionale nell'ammodernamento della giustizia suggeriscano il contrario, almeno a livello sovranazionale si assiste a sempre più frequenti interventi tesi a normare questo innovativo settore: basti pensare alla "*Digital agenda*" dell'Unione europea [3] e al recente «*Artificial Intelligence Act*» [4].

I rischi e le opportunità derivanti dall'utilizzo delle tecnologie fondate sull'intelligenza artificiale nel processo penale sono già oggetto di un ampio dibattito scientifico [5] e non è certo un caso che, in questo dibattito, a restare in ombra sia stato, finora, il piano dell'esecuzione della pena detentiva e dell'ordinamento penitenziario. Un settore, quest'ultimo, regolamentato da una fonte principale che risale al 1975 [6] e che non è più stato oggetto di una (pur più volte invocata) riforma organica [7], ma solo di interventi

---

[3] Cfr. Commissione europea [COM (2018) 237 final], *L'intelligenza artificiale per l'Europa*, 25 aprile 2018.

[4] *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)*, in *G.U.U.E.*, 12 luglio 2024, L 1689, sul quale v. S. Quattrocolo, *Intelligenza artificiale e processo penale: le novità dell'AI Act*, in *Diritto di difesa*, 16 gennaio 2025. L'accordo prodromico al Regolamento, raggiunto il 9 dicembre 2023, trae origine da Commissione europea [COM (2021) 206 final], *Proposta di regolamento che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione*, 21 aprile 2021, sulla quale v. G. Contissa, F. Galli, F. Godano, G. Sartor, *La nuova Proposta di Regolamento europeo sull'intelligenza artificiale: questioni giuridiche e approcci regolatori*, in *Nuove questioni di informatica forense*, a cura di R. Brighi, Roma, 2021, p. 387.

[5] Nell'ambito di una letteratura sempre più vasta, cfr. C. Buchard, *L'intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società*, in *Riv. it. dir. proc. pen.*, 2019, p. 1908 ss.; M. Gialuz, *Quando il processo penale incontra l'intelligenza artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, in *Dir. pen. cont.*, 29 maggio 2019; R.E. Kostoris, *Predizione decisoria*, diversion *processuale e archiviazione*, in *Dir. pen. cont. – Riv. trim.*, 2021, n. 2, p. 42 ss.; L. Lupária, *Artificial Intelligence in Criminal Courts. Opportunity or Threat?*, in *Legal Challenges in the New Digital Age*, a cura di A.M. Lopez Rodriguez, M.D. Green, M.K. Kubica, Leiden, 2021, p. 160 ss.; V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *Intelligenza artificiale. Il diritto, i diritti, l'etica*, a cura di U. Ruffolo, Milano, 2022, p. 547 ss.; S. Quattrocolo, *Artificial intelligence, Computational Modelling and Criminal Proceedings. A Framework for a European Legal Discussion*, New York, 2020; G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, in *Dir. pen. cont. – Riv. trim.*, 2020, n. 4, p. 75 ss.

[6] Si tratta della l. 26 luglio 1975, n. 354, recante «Norme sull'ordinamento penitenziario e sull'esecuzione delle misure privative e limitative della libertà».

[7] Uno strutturato tentativo di riforma è stato quello elaborato dalla Commissione istituita con d.m. 19 luglio 2017 e presieduta dal Prof. G. Giostra, i cui lavori sono solo in parte confluì-

rapsodici sensibilmente intensificatisi, fino a diventare quasi incessanti, negli ultimi anni [8].

Sembra quasi paradossale, pertanto, parlare di futuro, di progresso tecnologico, di possibili impieghi dell'intelligenza artificiale nel sistema penitenziario, in un momento storico – quello attuale – nel quale, secondo l'ultima rilevazione del Ministero della giustizia, negli istituti di pena nazionali sono presenti 62.281 persone a fronte di una capienza regolamentare di 51.283 [9]. Un livello di sovraffollamento disumano, che riporta l'ordinamento nazionale indietro di oltre un decennio, all'epoca cioè della nota pronuncia della Corte di Strasburgo nel caso Torreggiani c. Italia [10], che sancì l'incompatibilità delle condizioni carcerarie nazionali con l'art. 3 Cedu ed ebbe l'effetto di destare una certa attenzione sul tema, tanto da condurre all'introduzione di alcune modifiche normative, peraltro rivelatesi non risolutive [11].

---

ti nei d.lgs. 2 ottobre 2018, n. 123 e 124. In argomento, v. P. Bronzo, F. Siracusano, D. Vicoli, *La riforma penitenziaria: novità e omissioni del nuovo "garantismo penitenziario"*, Torino, 2019.

[8] A partire dai già citati d.lgs. n. 123 e 124 del 2018; v. anche la l. 9 gennaio 2019, n. 3 (c.d. "spazzacorrotti"), che ha esteso ai condannati e agli internati per delitti contro la pubblica amministrazione la disciplina prevista per i reati c.d. "ostativi", salve poi le modifiche introdotte sul punto dal d.l. 31 ottobre 2022, n. 162 conv. dalla l. 30 dicembre 2022, n. 199, intervenuto in maniera organica sul testo dell'art. 4-*bis* ord. penit. (sul quale si spenderanno considerazioni più specifiche *infra*, n. § 3.1.); senza trascurare il d.lgs. 10 ottobre 2022, n. 150 (c.d. "riforma Cartabia"), che ha introdotto nell'ordinamento la disciplina della giustizia riparativa e ha notevolmente modificato le pene sostitutive, con l'introduzione di nuove tipologie e di diverse modalità di esecuzione.

[9] Cfr. *Detenuti presenti e capienza regolamentare degli istituti penitenziari per regione di detenzione. Situazione al 31 marzo 2025*, in *www.ministerodellagiustizia.it*; per un'analisi dei dati, si rinvia allo studio del Garante nazionale dei diritti delle persone private della libertà personale sull'indice di sovraffollamento della popolazione detenuta, *Analisi storica 2020-2024 sul sovraffollamento negli Istituti penitenziari*, a cura di E. Cappelli e G. Suriano, in *Sist. pen*., 22 gennaio 2024.

[10] Corte edu, 8 gennaio 2013, Torreggiani c. Italia, in *Dir. pen. cont*., 9 gennaio 2013, con nota di F. Viganò, *Sentenza pilota della Corte EDU sul sovraffollamento delle carceri italiane: il nostro Paese chiamato all'adozione di rimedi strutturali entro il termine di un anno*. Sulla sentenza, v. anche G. Della Morte, *La situazione carceraria italiana viola strutturalmente gli standard sui diritti umani (a margine della sentenza Torreggiani c. Italia)*, in *Dir. umani e dir. internaz*., 2013, p. 147 ss.; M. Dova, *Torreggiani c. Italia, un barlume di speranza nella cronaca del sistema sanzionatorio*, in *Riv. it. dir. proc. pen.*, 2013, p. 948 ss.; G. Tamburino, *La sentenza Torreggiani e altri della Corte di Strasburgo*, in *Cass. pen*., 2013, p. 11 ss.

[11] Sul tema, v. A. Della Bella, *Il carcere oggi: tra diritti negati e promesse di rieducazione*, in *Dir. pen. cont. – Riv. trim*., 2017, n. 4, p. 42 ss.

2. *L'adeguamento tecnologico del sistema penitenziario*

Senza concrete prospettive di soluzione della situazione appena descritta, sembra davvero difficile portare a compimento quella (invero) ineludibile opera di ammodernamento di cui il sistema penitenziario necessita.

Nel 2021, tuttavia, tra le proposte formulate dalla Commissione per l'innovazione del sistema penitenziario con l'obiettivo di migliorare la quotidianità penitenziaria [12], si è avuta quella di procedere al suo adeguamento tecnologico, seguendo una duplice linea d'azione: da una parte si è puntato a potenziare l'ordine e la sicurezza all'interno del carcere attraverso l'uso di strumenti avanzati e, dall'altra parte, si è proposto di introdurre delle innovazioni inerenti al trattamento penitenziario con lo scopo di migliorare la qualità della vita detentiva [13].

Sospinte dall'intento di riallineare la vita carceraria alla matrice costituzionale e alle indicazioni sovranazionali, le proposte della Commissione rivolte ad una migliore gestione dell'ordine e della sicurezza sono concepite in un'ottica funzionale alle necessità trattamentali e non più solo nella dimensione della corretta gestione del carcere. Come è stato condivisibilmente osservato [14], se è vero che solo un ambiente sicuro può garantire una corretta erogazione del trattamento, è del pari evidente che assicurare l'ordine e la sicurezza con forme che comprimono le esigenze trattamentali significa svilire il primario scopo rieducativo della detenzione.

A partire da un rimarcato divieto di ricorrere alla forza fisica e psichica (se non in casi eccezionali) [15], un necessario passo in avanti sul fronte della sicurezza è stato individuato nell'implementazione (o nel rafforzamento nel

---

[12] Commissione per l'innovazione del sistema penitenziario, istituita con d.m. 13 settembre 2021 e presieduta dal prof. M. Ruotolo. Per approfondire i lavori della Commissione, cfr. *Innovazione del sistema penitenziario: la relazione finale della Commissione Ruotolo*, in *Sist. pen*., 11 gennaio 2022; nonché F. Siracusano, *Verso un carcere più umano e solidale: brevi riflessioni a margine delle proposte della Commissione Ruotolo*, in *Riv. it. dir. proc. pen.*, 2022, p. 849 ss.

[13] Le proposte di intervento sono illustrate da M. Ruotolo, *Il sistema penitenziario e le esigenze della sua innovazione*, in *Biolaw Journal*, 2022, n. 4, p. 31 ss. Come viene segnalato nella *Relazione finale*, cit., p. 13, si tratta preliminarmente di un problema di risorse, «ancorché alcune delle innovazioni proposte possano essere realizzate con costi contenuti (anche perché determinanti la riduzione contestuale di altre spese), e un'attenta valutazione dei fabbisogni, nonché la piena disponibilità alla standardizzazione delle buone pratiche già esistenti».

[14] Così F. Siracusano, *Verso un carcere più umano e solidale: brevi riflessioni a margine delle proposte della Commissione Ruotolo*, cit., p. 858.

[15] Si segnalano al riguardo le proposte di intervento sugli artt. 2 e 82 d.p.r. 30 giugno 2000, n. 230 («Regolamento recante norme sull'ordinamento penitenziario e sulle misure privative e limitative della libertà»).

caso di strutture in cui fossero già presenti impianti tecnologici idonei) di sistemi anti-drone, di *metal detector* fissi o di *body scanner*, per impedire l'accesso di oggetti la cui disponibilità non è consentita alle persone detenute, senza dover ricorrere a più invasivi sistemi di controllo [16].

In una direzione analoga va la proposta di semplificare le operazioni di accesso dei visitatori per i colloqui, introducendo sistemi di controllo biometrico in grado di agevolare la procedura di identificazione. La ricognizione automatica del dato permetterebbe infatti alle persone già registrate nel sistema di entrare e di uscire dall'istituto nella giornata del colloquio senza l'articolata procedura di registrazione cartacea, con un evidente risparmio di tempo e di risorse in termini organizzativi.

Una convinzione piuttosto radicata all'interno della Commissione ministeriale, e ribadita più volte nella *Relazione finale*, è infatti quella secondo cui dall'investimento di risorse nella tecnologia deriverebbe un innalzamento in termini qualitativi della sicurezza, ma anche una semplificazione del lavoro del personale penitenziario [17].

È per questo motivo che, con un'altra proposta, si è suggerito di standardizzare il progetto "*Move*", già adottato presso la Casa Circondariale di Roma Rebibbia Nuovo Complesso (e successivamente esportato alla Casa Circondariale di Lecce, nella quale ha assunto il nome di "*Free man*"). Si tratta sostanzialmente di un sistema automatizzato per la gestione della circolazione delle persone detenute attraverso i vari reparti dell'istituto, in base al quale gli spostamenti avvengono senza l'accompagnamento della polizia penitenziaria ma utilizzando un *pass* digitale dotato di codice a barre, e imponendo alcune prescrizioni quali il divieto di fermarsi durante il percorso e di partecipare ad assembramenti non autorizzati negli spazi comuni. Il risultato, anche grazie all'utilizzo di impianti di video-sorveglianza, è che le movimentazioni sono più veloci e l'attività del personale di vigilanza è alleggerita [18].

---

[16] Oltre ad integrare una grave infrazione di tipo disciplinare (art. 77, comma 1 n. 8, 9 e 16 reg. esec.), l'accesso indebito di dispositivi idonei alla comunicazione da parte di soggetti detenuti, con la novità introdotta dall'art. 9 d.l. 21 ottobre 2020, n. 130, conv. in l. 18 dicembre 2020, n. 173, è ora oggetto della fattispecie incriminatrice di cui all'art. 391-*ter* c.p.

[17] Così C. Cantone, *L'uso della tecnologia per la gestione della sicurezza interna*, in *Pena e nuove tecnologie. Tra "trattamento" e "sicurezza"*, a cura di G. Fiorelli, P. Gonnella, A. Massaro, A. Riccardi, M. Ruotolo, S. Talini, Napoli, 2022, pp. 254-255.

[18] Così Circ. D.a.p. del 18 novembre 2022, n. 0442486U, la quale, nel diffondere sul territorio una serie di iniziative mutuate dai lavori della Commissione Ruotolo, ha suggerito di verificare la possibilità di implementare il programma "*Move*" nei vari istituti di pena.

Il ricorso alla tecnologia si rivela nondimeno fondamentale anche sul versante del trattamento.

Nell'ottica del mantenimento delle relazioni affettive, centrale nell'opera di trattamento (artt. 15 e 16 ord. penit.), in seno alla Commissione ministeriale è emersa una chiara consapevolezza dell'importanza del potenziamento delle comunicazioni a distanza, le quali, già durante l'emergenza pandemica, si erano rivelate un indispensabile strumento di contatto dei detenuti con i propri familiari: in proposito, l'art. 221, comma 10, d.l. 10 maggio 2020, n. 34, conv. l. 17 luglio 2020, n. 77, aveva stabilito che, su richiesta dell'interessato o quando la misura risulti indispensabile per la salvaguardia della salute, i colloqui con i congiunti o con altre persone cui hanno diritto i condannati, gli internati e gli imputati, possono essere svolti a distanza [19].

È utile ricordare che né la legge sull'ordinamento penitenziario, né il relativo regolamento di esecuzione (che ancora oggi disciplina l'uso del telegrafo e del fax), contengono una specifica normativa con riguardo alle comunicazioni in videochiamata, bensì solo un riferimento generico, come quello dell'art. 18, comma 9, l. 354/75, agli «altri tipi di comunicazione». Nonostante tale lacuna, l'amministrazione penitenziaria sembra avere inquadrato l'opportunità dello strumento: fin dal 2019, è stato avviato, in via sperimentale e per il solo circuito di media sicurezza, un programma per l'utilizzo delle videochiamate a distanza tramite *Skype for business* [20]. Con una successiva Circolare [21], si è poi stabilito che, oltre alle tre modalità "tradizionali" previste per garantire il mantenimento dei contatti con l'esterno (colloqui visiti, conversazioni telefoniche e corrispondenza epistolare), fossero confermate ed estese a tutti i circuiti penitenziari anche le videochiamate come modalità alternative di fruizione dei colloqui visivi.

Benché i numerosi aspetti positivi dell'apertura alle videochiamate siano evidenti, non si può fare a meno di osservare che la Circolare ha comunque prodotto il paradossale effetto di configurare il colloquio telefonico – benché in modalità di videochiamata – come una sostituzione del colloquio in

[19] La piattaforma principalmente utilizzata è quella di *Skype for business*, anche se, in alcuni casi, è stato autorizzato lo svolgimento di videochiamate tramite l'applicativo *WhatsApp*, attivato su apparecchi telefonici acquistati dall'amministrazione penitenziaria. Per uno sguardo alle esperienze straniere v. M.C. Locchi, N. Pettinari, *L'utilizzo di Skype in carcere al fine del mantenimento e del rafforzamento dei rapporti dei detenuti con il mondo esterno*, in *Arch. pen*., 2020, n. 1, p. 12 ss.

[20] Cfr. Circ. D.a.p. del 30 gennaio 2019, n. 0031246U.

[21] Si tratta della Circ. D.a.p. del 26 settembre 2022, n. 3696/6146, p. 4.

presenza. È la stessa Circolare infatti ad ammettere che il ricorso a tale modalità di contatto rende «non necessarie le lunghe e defatiganti operazioni di perquisizione dei soggetti che fanno ingresso in carcere in occasione dei colloqui "in presenza", consentendo il contenimento del rischio che, in tale frangente, possano essere introdotti, dall'esterno, oggetti non consentiti» [22]. Al contrario: il mezzo di comunicazione innovativo della videochiamata non deve sostituire il colloquio in presenza dei familiari, trattandosi di uno strumento da utilizzare in aggiunta alle visite, solo qualora la presenza fisica sia impossibile o particolarmente complicata dalla distanza geografica o dalle condizioni di salute del familiare.

Risponde alla logica di un concreto miglioramento della quotidianità penitenziaria anche la realizzazione di *totem touch* per la più rapida gestione delle richieste dei detenuti: terminali multimediali, fruibili in diverse lingue, che consentirebbero di sostituire le richieste cartacee (es. le c.d. "domandine" mod. 393, gli ordini di sopravvitto mod. 72, ma anche le istanze indirizzate alla magistratura di sorveglianza). Se effettivamente implementata, questa innovazione, oltre che razionalizzare un procedimento a dir poco arcaico, avrebbe delle ricadute positive in termini di effettività della tutela dei diritti dei detenuti, che si fonda sull'imprescindibile presupposto di rendere tracciabili le richieste.

Un'altra area di intervento in cui il supporto della tecnologia potrebbe essere decisivo per migliorare le condizioni detentive è quello della tutela della salute. Se introdotti, digitalizzazione del fascicolo sanitario e telemedicina contribuirebbero ad ottenere cure più celeri ed un complessivo miglioramento degli *standard* sanitari. Senza andare a sostituire le prestazioni mediche in presenza quando necessarie, un primo consulto in videoconferenza potrebbe costituire un presidio diagnostico efficace per la definizione di un percorso assistenziale completo.

Le soluzioni indicate dalla Commissione Ruotolo per perseguire il miglioramento delle condizioni detentive, come si vede, non comportano l'introduzione di strumentazioni d'avanguardia, bensì si tratta del mero, ma non procrastinabile, adeguamento tecnologico di un sistema fortemente retrogrado.

---

[22] Cfr. Circ. D.a.p., n. 3696/6146, cit., p. 8, la quale, nel descrivere i vantaggi delle videochiamate, prosegue aggiungendo che «la circostanza che il colloquio "a distanza" possa essere interrotto in ogni caso di condotte inappropriate, consente al contempo di soddisfare le sempre essenziali e imprescindibili esigenze di sicurezza».

3. L'impiego dell'intelligenza artificiale in carcere

Prima di addentrarsi nell'esame dei suoi possibili utilizzi nel sistema penitenziario, conviene soffermarsi brevemente sulla definizione di intelligenza artificiale [23]. Al riguardo, considerato che non esiste una nozione condivisa [24], conviene accogliere quella contenuta nella Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi [25], che si riferisce ad essa come «l'insieme di metodi scientifici, teorie e tecniche finalizzate a riprodurre mediante le macchine le capacità cognitive degli esseri umani» [26].

Con la pubblicazione nella Gazzetta Ufficiale dell'Unione europea, si è recentemente concluso il percorso di approvazione del Regolamento europeo (il c.d. «*Artificial Intelligence Act*») [27], il quale, all'art. 3, par. 1, defini-

---

[23] A partire dagli studi di A.M. Turing che, negli anni '50 del secolo scorso, sulla base di un *test* (il c.d. "*The imitation game*"), si interrogava sulla possibilità che le macchine potessero essere considerate "pensanti": cfr. A.M. Turing, *Computing, machinery and intelligence*, in *Mind*, 1950, pp. 433-460. L'espressione "*Artificial Intelligence*" sarebbe stata invece coniata da John McCarthy nel 1955, come ricorda F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi d'indagine*, in *Dir. pen. uomo*, 29 settembre 2019, p. 4.

[24] Così M. Ienca, *Intelligenza2. Per un'unione di intelligenza naturale e artificiale*, Torino, 2019, p. 13. Sulle difficoltà di definire un fenomeno in continua evoluzione, v. G. Padua, *Intelligenza artificiale e giudizio penale: scenari, limiti, prospettive*, in *Proc. pen. giust*., 2021, 6, p. 1481.

[25] Questa è la definizione riportata nell'Appendice III della *Carta etica per l'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente*, adottata dalla Commissione europea per l'efficienza della giustizia (CEPEJ), nel corso della XXXI Riunione plenaria, Strasburgo, 3 dicembre 2018, [CEPEJ (2018) 14], per un commento alla quale 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 *Leg. pen*., 18 dicembre 2018.

[26] La stessa Carta etica ulteriormente distingue «tra intelligenze artificiali 'forti' (capaci di contestualizzare problemi specializzati di varia natura in maniera completamente autonoma) e intelligenze artificiali 'deboli' o 'moderate' (alte prestazioni nel loro ambito di addestramento)». Cfr. G. Ubertis, *Intelligenza artificiale, giustizia penale, controllo umano significativo*, cit., p. 78.

[27] Si è così superata la definizione dell'art. 3, par. 1, della Proposta di Regolamento europeo, secondo cui il "sistema di intelligenza artificiale" è un «*software* sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato 1 che può, per una determinata serie di obiettivi definiti dall'uomo, generare *output*, quali contenuti, previsioni, raccomandazioni o decisioni»: così, Commissione europea [COM (2021) 206 final], *Proposta di regolamento che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione*, cit., p. 43. Al riguardo, osserva L. Maldonato, *Risk and need assessment tools e riforma del sistema sanzionatorio*, in *Intelligenza artificiale e processo penale. Indagini, prove e giudizio*, a cura di G. Di Paolo, L. Pressacco, Napoli, 2022, p. 143, che «i *software* di intelligenza artificiale simulano i *pattern* del ragionamento umano mirando, però, a ottenere un miglioramento esponenziale delle relative *performances*».

sce il "sistema di intelligenza artificiale" quale «sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'*input* che riceve come generare *output* quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali».

È evidente che le potenzialità operative di strumentazioni del genere nel processo penale, e nella fase di esecuzione della pena, sono enormi: accanto al possibile sviluppo delle banche-dati e al miglioramento in termini organizzativi delle strutture materiali e delle risorse di personale, la vastissima latitudine dell'uso degli algoritmi si estende dalla prevenzione della consumazione di reati [28] e l'identificazione dei responsabili, all'adozione di *tools* di giustizia predittiva a supporto dei giudici nell'attività decisionale [29] e nella commisurazione della pena, fino alla profilazione della persona per predirne le future condotte e calcolare le probabilità di recidiva.

3.1. Big data *per la verifica del superamento della c.d. ostatività penitenziaria*

Con l'espressione *Big data* ci si riferisce ad una «grande quantità di tipi diversi di dati prodotti con un'alta velocità da un grande numero di fonti di diverso tipo» [30], la cui gestione richiede oggi nuovi strumenti e metodi, come le tecnologie di apprendimento automatico (*machine learning*) basate su sistemi di intelligenza artificiale.

Un'interessante prospettiva di impiego di questa formidabile capacità computazionale risiede nella possibilità di indicizzare gli elementi conoscitivi che sono nella disponibilità della magistratura di sorveglianza ai fini decisori, per razionalizzare la procedura contenendone tempi e costi e, soprattutto, ottenere una tendenziale uniformità decisionale [31].

---

[28] V., in proposito, L. Algeri, *Intelligenza artificiale e polizia predittiva*, in *Dir. pen. proc.*, 2021, p. 724 ss.; nonché S. Quattrocolo, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale*, in *Cass. pen.*, 2019, p. 1750.

[29] Come nota P.P. Paulesu, *Intelligenza artificiale e giustizia penale. Una lettura attraverso i principi*, in *Arch. pen.*, 2022, n. 1, p. 1, gli strumenti di intelligenza artificiale possono «contribuire a risolvere problemi investigativi, problemi cautelari, problemi probatori, problemi decisori».

[30] Questa è la definizione della Commissione europea [COM (2014) 442 final], *Verso una florida economia basata sui dati*, 2 luglio 2014, 5. Sul tema, v. U. Ruffolo, *Intelligenza artificiale, machine learning e responsabilità da algoritmo*, in *Giur. it.*, 2019, f. 7, p. 1689.

[31] Valuta la possibile indicizzazione degli elementi con riferimento a vari ambiti del proce-

Uno degli ambiti della fase esecutiva in cui l'istruttoria è particolarmente complessa è quello che attiene alla verifica del superamento della c.d. ostatività penitenziaria, ai sensi dell'art. 4-*bis* ord. penit. come recentemente modificato [32].

Non potendo analizzare in questa sede i dettagli della riforma [33], basti ricordare che il rinnovato regime di accesso ai benefici penitenziari e alle misure alternative sostituisce alla presunzione assoluta di incompatibilità per i condannati per i delitti indicati nel c. 1 dell'art. 4-*bis* ord. penit. non collaboranti un sistema fondato su condizioni e presupposti che l'istante deve allegare e documentare, volti ad escludere sia l'attualità sia il pericolo di ripristino di collegamenti con l'associazione criminale di appartenenza. Per un secondo gruppo di reati ostativi, non connotati dalla matrice mafiosa o terroristica, il nuovo c. 1-*bis* dell'art. 4-*bis* ord. penit. stabilisce che i benefici possano essere concessi, anche in assenza di collaborazione, allegando elementi che consentano di escludere l'attualità – ma non il pericolo di ripristino – di collegamenti con il contesto criminale nel quale il reato è stato commesso.

Sul piano processuale, il comma 2 del rinnovato art. 4-*bis* ord. penit. introduce alcune modifiche alla disciplina procedurale cui la magistratura di sorveglianza deve attenersi per la valutazione dell'istanza: trattasi in particolare di dettagliate informazioni rese dal comitato provinciale per l'ordine e la sicurezza pubblica competente in relazione al luogo di detenzione del condannato. Inoltre, «anche al fine di verificare la fondatezza degli elementi offerti dall'istante», il giudice deve acquisire il parere del pubblico ministero presso il giudice che ha emesso la sentenza di primo grado ovvero, se si tratta di condanne per reati di cui all'art. 51, commi 3-*bis* e 3-

dimento istruttorio in esecuzione C. Fiorio, *Predizione algoritmica e giurisdizione di sorveglianza*, in *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, a cura di G.M. Baccari, P. Felicioni, Milano, 2023, p. 260 ss.

[32] Per un'analisi delle modifiche introdotte nella disciplina dell'art. 4-*bis* ord. penit. dall'art. 1 d.l. 31 ottobre 2022, n. 162, conv. dalla l. 30 dicembre 2022, n. 199, v. E. Dolcini, *L'ergastolo ostativo riformato* in articulo mortis, in *Sist. pen*., 7 novembre 2022. La riforma fa seguito alla decisiva pronuncia con la quale la Corte costituzionale ha dichiarato l'illegittimità costituzionale dell'art. 4-*bis* ord. penit. nella parte in cui non prevedeva la possibilità per i condannati per reati c.d. "ostativi", che non fossero collaboranti, di accedere al beneficio dei permessi premio, qualora fossero stati acquisiti «elementi tali da escludere sia l'attualità dei collegamenti con la criminalità organizzata, sia il pericolo di ripristino di tali collegamenti»: così Corte cost., sent. 4 dicembre 2019, n. 253, in *Giur. cost*., 2019, p. 3120.

[33] Per un approfondimento sul sistema delle preclusioni ostative ai benefici penitenziari e alle misure alternative alla detenzione, v. S. Lonati, *Verso il tramonto dell'ostatività penitenziaria: un'attesa lunga trent'anni*, in *Arch. pen*., 2022, n. 2.

quater, c.p.p., del pubblico ministero presso il tribunale del capoluogo del distretto ove è stata pronunciata la sentenza di primo grado, e del Procuratore nazionale antimafia e antiterrorismo; devono essere acquisite informazioni dalla direzione dell'istituto di pena; devono essere disposti accertamenti circa le condizioni reddituali e patrimoniali dell'istante e del suo nucleo familiare, oltre che sulla pendenza o definitività di misure di prevenzione personali o patrimoniali; devono essere raccolti anche gli eventuali elementi di prova contraria, «quando dall'istruttoria svolta emergano indizi dell'attuale sussistenza di collegamenti con la criminalità organizzata, terroristica ed eversiva o con il contesto nel quale il reato è stato commesso, ovvero del pericolo di ripristino di tali collegamenti» [34].

L'ampliamento delle fonti di conoscenza cui la magistratura di sorveglianza deve ricorrere per valutare le domande di ammissione ai benefici penitenziari spiega il significato del recente Protocollo del Procuratore nazionale antimafia (c.d. "Protocollo Melillo" del 29 dicembre 2022) [35], che ha lo scopo principale di adottare misure organizzative avanzate per agevolare la formazione e la condivisione di informazioni. Allo scopo è stata creata, all'interno del S.i.d.n.a. [36] , una "Area sicura di condivisione informativa", denominata "Procedure 4-*bis* ord. pen.", che consiste in un'architettura informatica uniforme, accessibile a tutti gli utenti abilitati, che è in grado di accogliere una mole significativa di dati aggregati per cartelle nominative corrispondenti alle persone detenute per i reati di cui all'art. 51, commi 3-*bis* e 3-*quater*, c.p.p.

Benché la banca-dati non sia governata da un algoritmo capace di elaborare le informazioni e fornire un risultato immediatamente fruibile dal magistrato di sorveglianza ai fini della decisione sulla richiesta di uno dei benefici penitenziari previsti dalla legge, essa rappresenta un valido supporto all'attività giurisdizionale, quanto meno in relazione a quell'ambito dell'istruttoria che attiene a dati "oggettivi" e che non riguarda quei criteri metagiuridici (come la dissociazione, la rivisitazione critica del proprio

[34] Le criticità di un'istruttoria altamente complessa come quella che risulta dal nuovo testo dell'art. 4-*bis* ord. penit. sono messe in luce da M. Passione, *A proposito del d.l. 162/2022: rilievi costituzionali e proposte di modifica, con particolare riferimento alla disciplina in materia di* 4-bis *o.p.*, in *Sist. pen*., 28 novembre 2022.

[35] Protocollo d'intesa del 29 dicembre 2022, in *Sist. pen*., 5 maggio 2023, sul quale v. G. Filocamo, *Nuovo art. 4 ord. penit.: il Protocollo operativo della DNAA e i risvolti applicativi del "regime probatorio rafforzato" per i condannati non collaboranti*, in *Sist. pen*., 2023, n. 5, p. 213.

[36] Si tratta del Sistema informativo direzione nazionale antimafia che attua la disposizione di cui all'art. 371-*bis*, comma 3, lett. c), c.p.p., che attribuisce al Procuratore nazionale antimafia il compito di acquisire ed elaborare notizie, informazioni e dati sulla criminalità organizzata.

operato, la collaborazione) che pure accompagnano la valutazione della magistratura di sorveglianza. In futuro, per rendere più funzionale lo strumento, potrebbe essere utile uniformare le tecniche di redazione delle informative demandate ai vari organi (comitati provinciali per l'ordine e la sicurezza pubblica, questure, procure territoriali, distrettuali e nazionali) e quelle delle relazioni di sintesi, rideterminandone i criteri [37].

3.2. *Algoritmi predittivi della pericolosità sociale*

Benché l'impiego più noto avvenga in ambito di misure di sicurezza (è qui infatti che all'art. 203 c.p. il concetto trova una definizione legislativa) [38], la pericolosità sociale è una categoria polifunzionale, che rileva in plurime sedi dell'ordinamento, tant'è che è del tutto lecito chiedersi se di essa esista una «nozione ontologicamente unitaria» [39].

Nella fase dell'esecuzione della pena, la pericolosità sociale costituisce in primo luogo un parametro per la differenziazione del regime penitenziario. Basti pensare all'accesso alle misure alternative alla detenzione: l'affidamento in prova al servizio sociale è possibile solo quando «assicuri la prevenzione del pericolo che egli [il condannato] commetta altri reati» (art. 47 ord. penit.), anche nell'ipotesi di affidamento in prova in casi particolari, previsto per condannati tossicodipendenti o alcoldipendenti (art. 94 D.p.r. 9 ottobre 1990, n. 309); alla detenzione domiciliare si può accedere solo se la misura è «idonea ad evitare il pericolo che il condannato commetta altri reati» (art. 47-*ter*, c. 1-*bis*, ord. penit.); la detenzione domiciliare speciale può essere concessa solo «se non sussiste un concreto pericolo di commissione di ulteriori delitti» (art. 47-*quinquies* ord. penit.). La valutazione circa la pericolosità sociale è significativa – ancorché inespressa – anche con riferimento alla semilibertà, alla quale il condannato può essere ammesso «quando vi sono le condizioni per un graduale reinserimento (...)

---

[37] Condivisibilmente C. Fiorio, *Predizione algoritmica e giurisdizione di sorveglianza*, cit., p. 274, propone al riguardo un'opera di «rimeditazione normativa» calibrata su una concezione maggiormente laica della pena.

[38] Per approfondimenti, cfr. R. Bartoli, *Pericolosità sociale, esecuzione differenziata della pena, carcere (appunti "sistematici" per una riforma "mirata" della pericolosità sociale)*, in *Riv. it. dir. proc. pen*., 2013, p. 717 ss.; M. Pelissero, *Pericolosità sociale e doppio binario. Vecchi e nuovi modelli di incapacitazione*, Torino, 2008, p. 107 ss.

[39] Così 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, p. 645 ss.

nella società» (art. 50 ord. penit.). La pericolosità sociale rileva poi anche per la concessione dei permessi premio (art. 30-*ter* ord. penit.) perché, accanto al requisito della regolare condotta, ad essere valutata è proprio l'assenza di pericolosità sociale del condannato.

Nel tentativo di semplificare un'analisi complessa e che in larga misura trascende i limiti del presente lavoro, si può dire che il giudizio sulla pericolosità sociale consiste in una valutazione circa la probabilità di commissione di ulteriori reati [40].

La qualità di «persona socialmente pericolosa» è accertata dal magistrato di sorveglianza sulla base dei criteri di cui all'art. 133 c.p.p. (gli stessi, cioè, su cui si fonda la commisurazione della pena): la personalità del reo, la sua condotta di vita, il contesto sociale e familiare di provenienza, le modalità e le caratteristiche del fatto commesso e le sue conseguenze. In particolare, la valutazione verte sui dati anamnestici, sulle perizie psichiatriche eventualmente effettuate nel corso del processo e su quelle criminologiche [41], sull'esistenza di precedenti penali e di procedimenti pendenti, sugli elementi desumibili dalla sentenza di condanna, sul comportamento *post delictum* e sull'andamento del percorso trattamentale.

Occorre domandarsi se le peculiarità di tale giudizio possano essere affidate a macchine dotate di intelligenza artificiale, come già avviene negli Stati Uniti [42], in cui diffusa applicazione hanno trovato gli strumenti di giustizia predittiva. Con riferimento a questa ampia categoria, una fondamentale distinzione va immediatamente compiuta tra strumenti predittivi dell'intelligenza artificiale a fini decisori e predizioni decisorie: mentre i primi forniscono elementi su cui basare una decisione, le seconde mirano a determinare la prevedibilità delle future decisioni su casi simili [43].

---

[40] Sul tema, v. A. Cabiale, *L'accertamento giudiziale della pericolosità sociale fra presente e futuro*, in *Arch. pen*., 2022, n. 2; nonché, per approfondimenti sulle metodologie di *risk assessment*, G. Zara, *Valutare il rischio in ambito criminologico. Procedure e strumenti per l'*assessment *psicologico*, Bologna, 2016.

[41] Il magistrato di sorveglianza può ricorrere alla perizia criminologica, ammessa nella fase esecutiva e vietata nella fase di cognizione secondo quanto previsto dall'art. 220, c. 2, c.p.p.; può altresì essere d'ausilio la perizia psichiatrica, sempre consentita. Quanto agli apporti conoscitivi specialistici sulla personalità dell'imputato si rinvia a M. Montagna, *I confini dell'indagine personologica nel processo penale*, Roma, 2013. Rileva il contrasto tra il divieto di perizia criminologica in fase di commisurazione della pena e i criteri di cui all'art. 133 c.p. S. Quattrocolo, *Artificial Intelligence, Computational Modelling and Criminal Proceedings. A framework for A European Legal Discussion*, cit., p. 174.

[42] Per riferimenti più specifici all'ordinamento statunitense si rinvia a B.L. Garrett, J. Monahan, *Judging risk*, in *California Law Review*, 2020, p. 439 ss.

[43] Sul punto, v. R.E. Kostoris, *Predizione decisoria*, diversion *processuale e archiviazione*, cit.,

Un celebre *software* di predizione decisoria, il cui uso ha dato adito a molte discussioni, è COMPAS [44], che opera processando le informazioni inerenti a precedenti giudiziari, a dati statistici e alle risposte fornite dall'imputato stesso ad un questionario di 137 domande, divise in cinque categorie: «*criminal involvement, relationships/lifestyles, personality/attitudes, family and social exclusion*». Il modello computazionale, elaborando i dati immessi in relazione ad un campione statistico di popolazione rilevante, prevede un rischio di recidiva, senza spiegare quello che, con riferimento alla motivazione di qualsiasi provvedimento giurisdizionale, chiameremmo il "percorso logico-argomentativo".

Tra chi ne sostiene le potenzialità e chi ne mette in luce gli aspetti critici, nell'analisi circa l'impiego dello strumento questi ultimi sembrano essere prevalenti e sono concentrati principalmente nell'arbitraria scelta dei fattori di rischio che, immessi nella macchina, consentono di formulare una prognosi circa la possibilità di recidiva e di esprimere un giudizio in ordine al collocamento, alla supervisione e alla gestione degli autori di reato. In particolare, alcuni studi hanno sollevato dubbi in ordine all'effettiva capacità predittiva del *software* ed hanno notato l'esistenza di possibili distorsioni dovute a discriminazioni [45], che determinano la sua non imparzia-

---

p. 44. Il versatile impiego dei c.d. *risk assessments tools* in funzione dell'emissione di provvedimenti giurisdizionali si presta tanto alla fase cautelare quanto a quella decisionale: cfr. G. Contissa, G. Lasagni, G. Sartor, *Quando a decidere in materia penale sono (anche) algoritmi e IA: alla ricerca di un rimedio effettivo*, in *Dir. internet*, 2019, n. 4, p. 619 ss.; con riferimento alla fase cautelare, invece, v. E. Guido, *Intelligenza artificiale e procedimento penale: ragionando di valutazione del rischio* de libertate, in *Arch. pen*., 2023, n. 1. Con riferimento all'utilizzo dei c.d. *risk assessment tools* nella fase esecutiva, v. S. Quattrocolo, *Equo processo penale e sfide della società algoritmica*, in *BioLaw Journal*, 2019, n. 1, p. 142, che sottolinea come, nel sistema statunitense, tali strumenti vengano utilizzati anche per le decisioni di *bail* e/o di *sentencing*. Sul punto, v. anche L. D'Agostino, *Gli algoritmi predittivi per la commisurazione della pena*, in *Dir. pen. cont. – Riv. trim*., 2019, n. 2, p. 356.

[44] Si tratta dell'acronimo di *Correctional Offender Management Profiling for Alternative Sanction*, uno strumento che opera attraverso l'elaborazione di un questionario composto da 137 domande, dati statistici e precedenti giudiziari. Al riguardo v. F. Basile, *Intelligenza artificiale e diritto penale: quattro possibili percorsi d'indagine*, cit., p. 19 ss. Un altro esempio di algoritmo predittivo è costituito da *Hart Assessment Risk Tool* (HART), in uso nel Regno Unito. A proposito dell'evoluzione degli strumenti di *risk assessment*, G. Zara, D.P. Farrington, *Criminal Recidivism: explanation, prediction and prevention*, Oxon, 2016, p. 148 ss.

[45] Come ricorda L. D'Agostino, *Gli algoritmi predittivi per la commisurazione della pena*, cit., p. 364, per l'evidente effetto discriminatorio prodotto da alcune variabili (quali quelle socioeconomiche) «alcuni autori suggeriscono di espungerle dai parametri di *risk-assessment*, limitando l'analisi ai soli precedenti penali del reo, all'età del primo arresto, e alle caratteristiche del crimine commesso». Sottolinea i limiti in punto di affidabilità scientifica di uno strumento ine-

lità. Anche la natura proprietaria del *software*, che preclude la conoscenza del suo funzionamento e pertanto impedisce la controllabilità dei risultati, come è stato giustamente sottolineato [46], costituisce un aspetto problematico non trascurabile.

Nel noto caso Loomis, tuttavia, la Corte Suprema del Wisconsin ha negato che l'utilizzo della macchina in fase di commisurazione della pena (*sentencing*) avesse leso il diritto ad un giusto processo dell'imputato: in relazione alla scarsa trasparenza del *tool*, che avrebbe impedito di valutarne l'attendibilità scientifica, la Corte ha ritenuto che, sulla base del manuale d'uso, fosse comunque possibile confrontare gli *input* (ossia i dati individuali inseriti) e gli *output* prodotti (ossia le valutazioni compiute). Tra gli argomenti della decisione, più convincente è quello che punta sul c.d. controllo umano significativo, secondo il quale le valutazioni compiute dall'algoritmo devono configurarsi come mero supporto al giudice nell'attività di commisurazione della pena [47].

Del resto, che una decisione non possa derivare esclusivamente dai risultati forniti da una macchina è un principio consolidato anche all'interno della cornice delle fonti continentale: l'art. 11 dir. 2016/680/UE [48] e l'art. 8 d.lgs. 18 maggio 2018, n. 51 vietano decisioni che producano effetti giuri-

---

vitabilmente esposto a forme di discriminazione A.M. Maugeri, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra* evidence based practices *e tutela dei diritti fondamentali*, in *Arch. pen*., 2021, n. 1, p. 7. Nello stesso senso, S. Signorato, *Giustizia penale e intelligenza artificiale. Considerazioni in tema di algoritmo predittivo*, in *Riv. dir. proc.*, 2020, n. 2, p. 614.

[46] Cfr. D. Kehl, P. Guo, S. Kessler, *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiative*, in *Berkman Klein Center for Internet & Society*, *Harvard Law School*, 2017, p. 11; S. Carrera, V. Mitsilegas, M. Stefan, *Criminal Justice, Fundamental Rights and the Rule of Law in the Digital Age. Report of CEPS and QMUL Task Force*, 2021, p. 60.

[47] State v. Loomis, 881 NW 2d 749 (Wis 2016). Per un commento alla decisione, v. *Criminal Law – Sentencing Guidelines – Wisconsin Supreme Court Requires Warnings before Use of Algorithmic Risk Assessment in Sentencing – State v. Loomis*, in *Harvard Law Review*, 2017, p. 1530 ss. Nel caso di specie, sulla base del *risk assessment*, la corte locale aveva irrogato una pena detentiva pari a sei anni (senza parole) e cinque anni di *extended supervision*, pena elevata in rapporto ai fatti per cui l'imputato si era dichiarato colpevole. Nell'impugnazione, venivano contestati diversi profili di violazione del principio del *due process*, evidenziando criticità legate all'uso in fase deliberativa della pena dello strumento di *risk assessment*.

[48] Dir. 2016/280/UE del Parlamento europeo e del Consiglio del 27 aprile 2016, dedicata alla protezione del trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, sulla quale v. S. Signorato, *Le indagini digitali. Profili strutturali di una metamorfosi investigativa*, Torino, 2018, p. 87 ss.

dici negativi o che incidano in modo significativo sulla sfera dell'interessato basate unicamente su trattamenti automatizzati. A ciò bisogna però aggiungere che il divieto esce piuttosto ridimensionato dalla previsione di una facoltà di deroga, condizionata dall'esistenza di garanzie adeguate, tra le quali deve essere riconosciuta la possibilità del titolare del trattamento dei dati di ottenere l'intervento umano [49].

Non bisogna poi trascurare che l'art. 101 Cost., stabilendo il principio di soggezione del giudice alla sola legge, è presidio valido per ritenere che gli strumenti di predizione decisoria non possano vincolare il libero convincimento del giudice. Sempre la Costituzione (art. 25, c. 2. Cost.) fornisce al sistema gli anticorpi verso quella che alcuni hanno preconizzato come una possibile traslazione verso un intollerabile diritto penale d'autore, nel cui ambito la pericolosità sociale verrebbe desunta esclusivamente dagli schemi comportamentali che il *software* ha tracciato come ricorrenti nel campione di riferimento [50].

Quelle che provengono da oltreoceano sono e restano mere suggestioni: il processo statunitense è caratterizzato da snodi procedimentali (*pretrial process*, *sentencing*, *parole*) del tutto sconosciuti al nostro sistema, per cui ragionare in termini di mera "importazione" risulterebbe fuorviante.

Tuttavia, non è seriamente pensabile che l'ingresso di algoritmi predittivi nell'ordinamento nazionale possa essere escluso, tanto più con riferimento alla fase di esecuzione della pena in cui il divieto di esprimere giudizi personologici sul condannato è decaduto.

Una recente circolare del Dipartimento dell'amministrazione penitenziaria (D.a.p.) [51] ha in effetti riconosciuto che per la valutazione prognostica della recidiva occorre adottare un «approccio polifonico» che, accanto alle metodologie consolidate, non potrà in futuro prescindere dagli strumenti di *risk assessment*. Pertanto, in una rinnovata prospettiva tecnologica, nella dinamica penitenziaria un ipotetico *report* prodotto dalla macchina potrebbe essere configurato come una sorta di "perizia 2.0" (purché

---

[49] Al riguardo, v. ancora S. Signorato, *Le indagini digitali. Profili strutturali di una metamorfosi investigativa*, cit., pp. 101-103, che sostiene la necessità di riconoscere l'esistenza di un nuovo diritto, denominato «diritto a decisioni non basate esclusivamente su trattamenti automatizzati».

[50] In questo senso, M. Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale: luci ed ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, cit., p. 21, che mette in luce come ciò sarebbe contrario anche al principio di individualizzazione del trattamento sanzionatorio desumibile dall'art. 27, c. 1 e 3, Cost.

[51] Circ. D.a.p. del 5 ottobre 2023, n. 0062758.U, *L'esecuzione penale esterna quale sistema di* probation. *Linee di indirizzo e indicazioni operative*, p. 7.

fondata su un metodo scientifico *evidence-based* riconosciuto e validato dalla comunità di esperti di riferimento) [52], di cui il magistrato di sorveglianza potrebbe avvalersi ai fini dell'ammissione alle misure alternative o della concessione di benefici penitenziari, ma anche per strutturare i contenuti e gli obiettivi del programma di trattamento: momenti valutativi accomunati dall'utilità dell'apporto scientifico nella prospettiva della rieducazione del condannato [53]. Senza dimenticare, tuttavia, che la qualità del risultato e la sua attendibilità sarebbero strettamente dipendenti dal controllo operato sulla tipologia di *input* immessi nella macchina e dalla centralità del ruolo valutativo comunque affidato al giudice [54].

Esclusa pertanto la sostituzione del giudice umano con gli strumenti di predizione, la prospettiva da prediligere sembra essere quella delineata dalla Carta etica europea: un utilizzo consapevole della tecnologia basata sull'intelligenza artificiale nel rispetto dei diritti umani fondamentali.

3.3. *Forme avanzate di sorveglianza e controllo*

A voler prescindere per un momento dal piano dell'allarme per la disumanità delle condizioni detentive, un secondario – ma non del tutto trascurabile – ordine di conseguenze dell'incessante crescita del numero dei detenuti da analizzare sarebbe quello inerente alla capacità del sistema di gestire la loro sorveglianza. È possibile ritenere che dai potenziali futuri utilizzi dell'intelligenza artificiale in carcere discendano benefici anche in ordine alla gestione della sicurezza?

Sistemi di monitoraggio robotico, telecamere tridimensionali, sensori

---

[52] Nel procedimento probatorio *in executivis*, sebbene debba riconoscersi il ruolo principale alla prova documentale, grazie al richiamo operato dall'art. 185 disp. att. c.p.p. alla testimonianza e alla perizia, è possibile ritenere ammissibili tutti i mezzi di prova la cui assunzione non risulti incompatibile con il procedimento esecutivo: in questi termini, C. Fiorio, *Predizione algoritmica e giurisdizione di sorveglianza*, in *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, cit., p. 262.

[53] Il possibile utilizzo degli strumenti di *risk assessment* per la predisposizione del trattamento penitenziario è oggetto dell'interessante studio di G. Zara, *Tra il probabile e il certo. La valutazione del rischio di violenza e di recidiva criminale*, in *Dir. pen. cont.*, 20 maggio 2016.

[54] Sul rapporto tra il giudice e la macchina è stato messo in luce da G. Canzio, *Il dubbio e la legge*, in *Dir. pen. cont.*, 20 luglio 2018, p. 3 ss., il rischio di «una ridotta responsabilità del giudicante, con l'inaccettabile effetto di conformismo e sclerotizzazione del formante giurisprudenziale». Tuttavia, nota S. Lorusso, *La sfida dell'intelligenza artificiale al processo penale nell'era digitale*, in *Sist. pen.*, 28 marzo 2024, p. 9, l'intelligenza artificiale potrebbe fare da antidoto al c.d. "giudice emotivo", le cui distorsioni possono condizionare la decisione.

per identificare comportamenti irregolari o inappropriati sono in effetti già oggetto di sperimentazioni negli istituti penitenziari della Corea del Sud e di Hong Kong. Nel carcere cinese di Yancheng, un esempio di c.d. "*smart jail*", esiste un sistema di sorveglianza continua dei detenuti, attuata attraverso una rete di sensori e di telecamere collegate ad un "cervello" centrale che è in grado di monitorare ininterrottamente ogni persona e allertare il personale in caso di condotte sospette [55]. Anche il sistema penitenziario di Singapore persegue un progressivo processo di realizzazione di carceri "senza contatto", nel cui ambito, attraverso l'utilizzo di strumenti denominati "Avatar" e "Vadar" [56], la sicurezza è gestita da automi.

Guardando ai sistemi continentali, è interessante riportare l'esperienza dei Paesi Bassi, nel cui sistema penitenziario sono stati introdotti dei braccialetti identificativi a radio-frequenza (RFID), come parte di una struttura integrata di gestione della sicurezza che poggia su congegni automatizzati e sensori di monitoraggio dei movimenti. Benché queste tecnologie possano dirsi più efficaci rispetto alla tradizionale sorveglianza umana sul piano del controllo, uno studio condotto negli Stati Uniti ha rilevato che RFID ha un impatto molto scarso in termini di deterrenza delle condotte vietate [57].

Gli argomenti di quanti sostengono le potenzialità del progresso tecnologico al servizio dell'ordine e della sicurezza poggiano sulla convinzione che una sorveglianza continua consentirebbe di prevenire le condotte violente e le rivolte attraverso la captazione dei comportamenti sospetti, di intervenire tempestivamente in caso di atti di autolesionismo e di ridurre il fenomeno dell'ingresso e del contrabbando di beni il cui possesso è vietato.

Altrettanto chiare sono tuttavia le controindicazioni di un modello di iper-sicurezza [58]. Occorre infatti tener presente che l'inviolabilità dei diritti

---

[55] Per un approfondimento sull'utilizzo di sistemi di sorveglianza digitale nel mondo, v. C. McKay, *The carceral automation: Digital prisons and technologies of detention*, in *International Journal for Crime, Justice and Social Democracy*, 2022, 11, pp. 104-105.

[56] Acronimi rispettivamente di *Advanced Video Analytics To detect Aggression* e di *Video Analytics to Detect Abnormal Behaviour*.

[57] In argomento, v. R.L. Halberstadt, N.G. La Vigne, *Evaluating the use of radio frequency identification device (RFID) technology to prevent and investigate sexual assault in a correctional setting*, in *Prison Journal*, 2011, p. 227 ss.: attraverso un *chip* incorporato in un braccialetto o in una cavigliera antimanomissione, la tecnologia RFID monitora i detenuti e il rispetto da parte loro delle regole penitenziarie, oltre a fornire la prova di eventuali infrazioni.

[58] Cfr. P. Puolakka, S. Van De Steene, *Artificial intelligence in prison in 2030: an exploration of the future of AI in prison*, in *Advancing Corrections Journal*, 2021, p. 134, i quali sottolineano condivisibilmente che, nell'introduzione di tecnologie di sorveglianza avanzate, è necessario delimitare dove gli strumenti possano essere posizionati (nelle celle, nel corpo della persona dete-

fondamentali esige che la loro compressione, nello stato di detenzione, debba essere comunque proporzionale alle effettive necessità del trattamento penitenziario e che, soprattutto, sia compatibile con la dignità, intesa quale nucleo minimo degli stessi diritti fondamentali [59].

Un'interessante opportunità di impiego della tecnologia di sorveglianza potrebbe semmai aversi nel contesto delle misure alternative alla detenzione. In un recente rapporto, il Gruppo di lavoro delle Nazioni Unite sulla detenzione arbitraria ha accolto con favore l'adozione di braccialetti elettronici di ultima generazione collegati ad *internet*, che permettono di limitare la necessità di ricorrere al carcere [60].

Si tratta di una prospettiva indubitabilmente affascinante, ma di difficile traslazione nell'ordinamento interno. Ben nota è infatti la deludente esperienza nazionale del braccialetto elettronico "tradizionale" [61], la cui perdurante indisponibilità ha prodotto soluzioni giurisprudenziali tutt'altro che eque [62] e non gli effetti deflativi sperati.

---

nuta, nelle aree comuni), precisare la durata temporale del loro utilizzo, cosa essi misurino (posizione, movimenti, frequenza cardiaca, respiro), come debba avvenire la misurazione (attraverso segnali come suono, video, foto, segnaletica digitale) e per quanto tempo i dati vadano conservati (controllo *real-time* con cancellazione immediata, anonimizzazione oppure custodia per un certo arco temporale per future analisi).

[59] Con la Raccomandazione R (2006) 2-rev (consultabile nel sito del Consiglio d'Europa all'indirizzo: *https://www.coe.int*), 1° luglio 2020, il Consiglio d'Europa ha aggiornato le Regole penitenziarie europee, che, alla Regola 49, prevedono che «il buon ordine in carcere deve essere mantenuto tenendo conto delle esigenze di sicurezza, incolumità e disciplina, garantendo allo stesso tempo ai detenuti condizioni di vita rispettose della dignità umana e offrendo loro un programma completo di attività in conformità con la Regola 25».

[60] Consiglio dei diritti umani delle Nazioni Unite, *Racial discrimination and emerging digital technologies: a human rights analysis. Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*, UN Doc. A/HRC/44/57, 2020, par. 48. Cfr. A. Riccardi, *Opportunità e rischi delle nuove tecnologie nel sistema penale e penitenziario nella prospettiva dei diritti umani*, in *Pena e nuove tecnologie. Tra "trattamento" e "sicurezza"*, cit., p. 232.

[61] Sul quale v. L. Cesaris, *Dal* panopticon *alla sorveglianza elettronica*, in *Il decreto "antiscarcerazioni"*, a cura di M. Bargis, Torino, 2001, p. 51 ss.

[62] Poiché la legge prevede che il braccialetto elettronico venga applicato quando il giudice abbia verificato la disponibilità del dispositivo presso la polizia giudiziaria (art. 275-*bis* c.p.p.), ci si è chiesti cosa si dovesse fare nel caso in cui tale disponibilità non vi fosse stata. In proposito, le Sezioni Unite hanno affermato che in simili evenienze il giudice debba reiterare il vaglio di adeguatezza della misura cautelare e scegliere di nuovo tra arresti domiciliari e custodia carceraria, dovendosi respingere qualsivoglia automatismo nell'applicazione delle restrizioni della libertà personale: Cass., sez. un., 28 aprile 2016, Lovisi, in CED Cass., n. 266651, sulla quale v. I. Guerini, *Più braccialetti (ma non necessariamente meno carcere): le Sezioni Unite e la portata ap-*

3.4. *Nuovi elementi del trattamento rieducativo*

La messa a disposizione dei detenuti di alcuni strumenti tecnologici è controversa. Ci si chiede, in particolare, se il diritto di accesso ad *internet* sia un diritto autonomo o se esso rilevi solo in quanto mezzo per godere di altri diritti umani (come il diritto all'informazione, all'istruzione, al lavoro e, più in generale, il diritto al pieno sviluppo della personalità umana) [63].

Alcuni sostengono che ammettere l'utilizzo di *computer* in carcere, consentire un (limitato) accesso ad *internet*, permettere l'uso di *devices* personali per le comunicazioni, oltre a contrastare con la visione "punitiva" della detenzione, porrebbe seri problemi in ordine alla sicurezza (basti pensare al rischio che il condannato continui a perpetrare l'attività criminosa dal carcere) [64]. Altri, invece, mettono in luce le opportunità che derivano dall'accesso a *internet*, considerandolo come la naturale evoluzione di diritti che sono già riconosciuti alle persone private della libertà personale (come quello di mantenere i legami affettivi attraverso le comunicazioni con il mondo "esterno", ma anche avvalersi della tecnologia nel campo del lavoro e dell'istruzione, per fruire della c.d. didattica a distanza) [65].

Grazie alle importanti aggiunte operate dal d.lgs. 2 ottobre 2018, n. 123, tramite le quali sono stati inseriti nell'art. 18 ord. penit. i riferimenti agli «strumenti di comunicazione disponibili e previsti dal regolamento», nonché ai «siti informativi» e, soprattutto, agli «altri tipi di comunicazione», attualmente la legge riconosce al detenuto la possibilità di avvalersi di

---

plicativa degli arresti domiciliari con la procedura di controllo del braccialetto elettronico, in *Dir. pen. cont.*, 24 giugno 2016.

[63] Sul punto, v. O. Pollicino, *The Right to Internet Access: Quid Iuris?*, in *The Cambridge Handbook of New Human Rights*, a cura di A. Von Arnauld e al., Cambridge, 2020, p. 263 ss. La Corte di Strasburgo ha affermato che, nonostante l'assenza di un obbligo in capo agli Stati di garantire ai detenuti l'accesso ad *internet*, laddove uno Stato offra tale possibilità questa non può essere arbitrariamente limitata, a pena di interferire con il godimento di alcuni diritti fondamentali, quali quelli sanciti dagli artt. 8 e 10 Cedu: cfr. Corte edu, sez. II, 19 gennaio 2016, Kalda c. Estonia.

[64] Cfr., v. C. McKay, *Video links from prison: Permeability and the carceral world*, in *International Journal for Crime, Justice and Social Democracy*, 2016, p. 21 ss.

[65] Sul tema, v. P. Scharff Smith, *Imprisonment and internet-access: Human rights, the principle of normalization and the question of prisoners access to digital communication technology*, in *Nordic Journal of Human Rights*, 2012, p. 454. È d'obbligo il richiamo alle *The United Nations Standard Minimum Rules for the Treatment of Prisoners* (c.d. Mandela rules) che, alla regola 58, riconoscono il diritto dei detenuti ai contatti con il mondo esterno anche «mediante corrispondenza per iscritto e utilizzando, ove disponibile, la telecomunicazione, gli strumenti elettronici, digitali e qualsiasi altro mezzo».

internet. Tuttavia, le modalità di accesso avrebbero dovuto essere disciplinate da un regolamento che non è stato ancora emanato. Il tema, al centro di alcune circolari del D.a.p. che sembravano aver preso atto dell'importanza delle tecnologie informatiche per numerose iniziative di natura trattamentale [66], non figura tra i punti più impellenti dell'agenda legislativa: troppo drammatiche le condizioni detentive nelle carceri italiane e troppo pressante l'esigenza di risolvere una situazione di sovraffollamento ormai intollerabile (che invero costituisce una precondizione rispetto a qualunque altra iniziativa volta a migliorare la qualità della vita detentiva), di talché non stupisce che la disuguaglianza prodotta dal divario digitale sia oggi relegata al limbo dei molti problemi irrisolti del sistema penitenziario.

Eppure, risultati decisamente incoraggianti sul piano rieducativo provengono da quelle sperimentazioni che hanno recentemente introdotto percorsi professionalizzanti volti all'alfabetizzazione digitale: formare specialisti in consulenza informatica, esperti di robotica, *cybersecurity* o *internet of things* significa avere la lungimirante visione di offrire al mondo del lavoro specializzazioni non ancora così diffuse, ma certamente imprescindibili nel prossimo futuro. Significa, in altre parole, offrire un'opportunità professionale in linea con il reingresso in società come prescritto dall'art. 20, c. 3, ord. penit., a mente del quale «l'organizzazione e i metodi del lavoro penitenziario devono riflettere quelli del lavoro nella società libera al fine di far acquisire ai soggetti una preparazione professionale adeguata alle normali condizioni lavorative per agevolarne il reinserimento sociale».

Questi programmi sembrano passi in avanti minimali rispetto all'esperienza finlandese, nella quale è in corso di sperimentazione il c.d. "*Smart prison project*". Dal 2021, nell'istituto penitenziario femminile di Hämeenlinna, ogni cella è dotata di un *computer* collegato alla rete, che può essere utilizzato dalle detenute per inviare messaggi, richieste, fare videochiamate con gli operatori penitenziari e con i propri familiari. I dispositivi sono dotati di un accesso limitato ad *internet* che consente di studiare, di fare acquisti *online* e di gestire alcuni aspetti della quotidianità, oltre a rendere possibile ricevere assistenza sanitaria da remoto. Sia gli operatori penitenziari che le detenute hanno ricevuto un *training* di preparazione all'utilizzo

[66] Circ. D.a.p. 2 novembre 2015, n. 366755, sulla possibilità di accesso ad internet da parte dei detenuti. Sull'argomento v. D. GALLIANI, *Internet e la funzione costituzionale rieducativa della pena*, in *Dir. pen. cont.*, 2 maggio 2017, p. 18; v. anche Circ. D.a.p., 5 dicembre 2018, n. 0381497; Circ. D.a.p., 12 marzo 2020, n. 0084702.

della tecnologia ed è incentivata la partecipazione a programmi di studio di elementi di intelligenza artificiale [67].

Nonostante i promotori del progetto sostengano che questo approccio favorirà il reingresso delle persone in società e ridurrà il tasso di recidiva, sono state sollevate numerose preoccupazioni, in larga misura condivisibili, in ordine al rischio di "deumanizzazione" che potrebbe discendere da una concezione detentiva di questo tipo [68]: un conto è fornire gli strumenti digitali per colmare il *gap* con il progresso tecnologico del mondo esterno e favorire – attraverso percorsi di studi e lavoro specializzati – il reinserimento, altro è incrementare l'alienazione da isolamento che rischia di prodursi riducendo ulteriormente i contatti umani del detenuto. Resta in effetti imprescindibile, come già sopra osservato, che l'accesso alle tecnologie informatiche nel contesto carcerario rappresenti un'opportunità aggiuntiva, e non sostitutiva, rispetto al diritto di mantenere anche relazioni umane ed affettive in presenza, seppur coi limiti dettati dal regime di detenzione.

4. *I prossimi passi*

Nel complesso quadro sin qui ripercorso, un'interessante visuale sul futuro è dischiusa dai lavori del Consiglio europeo per la Cooperazione Penologica (*Council for Penological Co-operation* "PC-CP"), organismo istituito in seno al Consiglio d'Europa e dedicato alla definizione di *standard* e principi nell'ambito dell'esecuzione delle sanzioni penali. Consapevole dell'ineluttabile avvento dell'intelligenza artificiale, il Consiglio ha elaborato una proposta di raccomandazione [69] sugli aspetti etici ed organizzativi dell'uso dell'intelligenza artificiale e delle tecnologie digitali correlate da parte dei servizi penitenziari.

Nel documento, si osserva come il ricorso all'intelligenza artificiale potrebbe da un lato consentire di creare modalità più avanzate di raccolta ed

---

[67] V. B. Lindström, P. Puolakka, *Smart prison: The preliminary development process of digital self-services in Finnish prisons*, in *International Corrections & Prisons Association*, 28 luglio 2020.

[68] R. Johnson, K. Hail-Jares, *Prison and technology: General lessons from the American context*, in *Handbook on Prisons*, a cura di Y. Jewkes, J. Bennett, B. Crewe, London, 2016, p. 284 ss., nel mettere in luce come questi strumenti potrebbero finire per sostituire il contatto umano, utilizzano la significativa immagine di «*baby-sitter* elettroniche».

[69] Consiglio europeo per la Cooperazione Penologica, *Ethical, Strategic and Operational Guidance on the Use of Artificial Intelligence in Prison and Probation Services and the Private Companies acting on their Behalf*, 10 settembre 2021.

elaborazione dei dati sulle persone detenute e, dall'altro lato, potenziare le tecniche di sorveglianza al fine di rendere più penetrante il controllo (*prison security*).

Quanto alle valutazioni automatizzate circa il pericolo di recidiva – che, come si è visto, si prestano oggi ad essere gestite da sistemi di intelligenza artificiale –, il PC-CP mette in luce come affidare decisioni in via esclusiva alle macchine, anziché al giudice, sia una scelta molto imprudente, che non terrebbe conto delle incertezze attuali sul loro funzionamento e della loro fallibilità [70].

Il Consiglio rileva ancora che non sembra – allo stato attuale delle conoscenze tecnologiche – che dal ricorso all'intelligenza artificiale possano dischiudersi delle prospettive di riduzione del ricorso al carcere o che le tecnologie offrano rimedi per il problema del sovraffollamento. Concentrandosi allora sulle implicazioni all'interno degli istituti penitenziari, si rileva che le *smart prisons* (quelle già esistenti, specie in realtà extra-europee) sono caratterizzate da tecnologie di sorveglianza remota e da formidabili capacità automatizzate di fornitura dei servizi ai detenuti; tuttavia, viene espresso l'auspicio che, anziché tendere ad emulare questo modello e quindi limitarne l'utilizzo allo scopo di incrementare la sicurezza, gli strumenti di intelligenza artificiale vengano piuttosto utilizzati nei percorsi riabilitativi.

Sarebbe prematuro dire oggi quale approccio debba essere adottato, ed il tema dovrà essere di certo ritrattato – il Consiglio stima nel 2032 – alla luce dei futuri sviluppi dell'intelligenza artificiale. Le suggestioni provenienti dai lavori del Consiglio appaiono comunque apprezzabili, nel senso di connotarsi per quello che pare dovere, in effetti, necessariamente essere l'equilibrio di fondo a cui dovrà ispirarsi l'utilizzo delle tecnologie di intelligenza artificiale nel settore penitenziario, mantenendo sempre una salda consapevolezza sia delle opportunità che esse potrebbero rappresentare nell'ausilio dell'attività giurisdizionale, sia delle insidie ai diritti fondamentali dei detenuti a cui un loro incontrollato affermarsi nel contesto carcerario potrebbe condurre.

---

[70] Secondo il documento, lo sviluppo tecnologico potrebbe essere senz'altro più utile nei sistemi di *probation*, attraverso cioè l'implementazione di nuove forme di monitoraggio elettronico, che utilizza gli *smartphone* sia come dispositivi di localizzazione, sia come una sorta di "*probation with app*", un sistema di sorveglianza che consente di monitorare in tempo reale le azioni del condannato. Tale *app* «*can detect (and possibly predict) potentially risky behavior. Based on the nature of the event, an AI could autonomously take a number of different actions to address the risk. Those might include, alerting the supervising officer or a mentor, or initiating a chat-bot system through an offender's mobile device that is trained to de-escalate situations*». Lo strumento potrebbe rivelarsi particolarmente utile per il trattamento di cura della tossicodipendenza e della salute mentale.

DIRITTO DELL'UNIONE EUROPEA E INTELLIGENZA ARTIFICIALE. RIFLESSI SUL PROCEDIMENTO PENALE*

di VALENTINA VASTA

1. *Premessa*

Il diffondersi dei sistemi d'intelligenza artificiale (IA), che consentono, in sempre più numerosi ambiti della vita quotidiana, lo svolgimento di operazioni e compiti complessi in precedenza svolti dall'essere umano, ha reso necessario l'intervento del diritto in diversi settori dell'ordinamento.

Il fenomeno ha coinvolto inevitabilmente l'ambito giudiziario, con la possibilità di molteplici applicazioni dell'intelligenza artificiale anche nel procedimento penale. Qui, rispetto ai tradizionali strumenti, scientifici e tecnologici, a disposizione delle autorità giudiziarie e di polizia, l'intelligenza artificiale si impone con un carattere inedito, dato dalla capacità di effettuare previsioni probabilistiche, sulla base dell'elaborazione di dati e parametri, e di riprodurre capacità cognitive proprie dell'essere umano.

Il ricorso a sistemi di IA, oltre che per finalità preventive, nel procedimento penale, come strumento d'indagine, per scopi probatori e nel mo-

* Il presente contributo è già stato pubblicato sul fascicolo 1/2024 della *Rivista italiana di diritto e procedura penale*. Rispetto alla versione contenuta nella *Rivista*, il testo qui riportato è stato opportunamente aggiornato, tenendo in considerazione, in particolare, l'approvazione e la pubblicazione del testo definitivo del Regolamento dell'Unione europea sull'Intelligenza artificiale (Reg. 2024/1689).

mento decisionale, potrebbe consentire un più efficace esercizio della giurisdizione. Si assicurerebbero sia l'efficientamento delle operazioni di *law enforcement*, con risvolti positivi in termini di completezza e celerità delle indagini penali, sia la prevedibilità delle decisioni, con la conseguente riduzione degli errori giudiziari e della c.d. *sentencing disparity*.

L'ingresso dell'intelligenza artificiale nelle attività investigative, cognitive e decisionali della giustizia penale, però, può incidere negativamente sulle garanzie fondamentali: dal *vulnus* al principio di eguaglianza, considerando che «in effetti, l'algoritmo – per antonomasia – è antiegualitario, perché considera alcuni fattori di rischio e non altri» [1], alla compromissione del diritto di difesa e dei principi del "giusto processo": è evidente, infatti, che una decisione fondata – ed è così "appiattita" – sull'esito dell'algoritmo mette a rischio i principi di imparzialità e libero convincimento del giudice, nonché il diritto al ricorso contro una decisione ingiusta.

Vantaggi e opportunità dell'impiego dell'intelligenza artificiale impongono, in sostanza, «una parallela consapevolezza dei pericoli che essa dischiude in ragione della sua opacità, della parziale e crescente sottrazione al controllo umano, della possibilità di errori ed esiti discriminatori, potenzialmente lesivi dei diritti fondamentali» [2].

Per questo, nell'Unione europea l'elaborazione normativa in tema di intelligenza artificiale ha iniziato un cammino, che è rapidamente proseguito fino alla predisposizione di una disciplina organica che cerca un punto di equilibrio tra efficienza e tutela dei diritti fondamentali nello sviluppo e nell'utilizzazione di sistemi di IA. Peraltro, va detto che nel contesto europeo esistono già da tempo regole e principi entro i quali scorgere un impianto normativo in relazione all'impiego dell'intelligenza artificiale nel procedimento penale.

2. *Il cammino dell'Unione europea nella regolamentazione dell'uso dei sistemi di intelligenza artificiale*

L'Unione europea ha approvato il *Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale e*

---

[1] V. Manes, *L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia*, in *DisCrimen*, 15 maggio 2020, p. 12.

[2] C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *BioLaw Journal-Rivista di BioDiritto*, 2021, n. 3, p. 416.

modifica alcune leggi dell'Unione[3], c.d. *AI Act*, al termine di un percorso costellato da diverse iniziative delle istituzioni europee finalizzate a predisporre forme di *governance* dell'intelligenza artificiale, strumentali a uno sviluppo del settore in conformità ai valori consolidati nell'ordinamento europeo.

La Commissione, dapprima con la pubblicazione delle *Draft Ethics Guidelines for Trustworthy AI*, elaborate da un gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale[4] e poi con il Libro bianco[5], ha fin da subito rimarcato l'esigenza che l'intelligenza artificiale sia affidabile, etica e antropocentrica e non possa prescindere dal controllo umano.

Il Regolamento (UE) sull'intelligenza artificiale è anche il risultato di un dibattito tra le istituzioni dell'Unione europea, che ha visto la Commissione, in un primo momento, propendere per interventi di modifica e integrazione degli atti già in vigore[6], contrapporsi al Parlamento, invece, favorevole ad adottare un nuovo atto normativo che regoli in modo generale e organico la materia dell'intelligenza artificiale[7]. Con la Proposta di Rego-

---

[3] *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)*, in *G.U.U.E.*, 12 luglio 2024, L 1689.

[4] *The European Commission's high-level expert group on artificial intelligence*, *Draft Ethics Guidelines for Trustworthy AI*, in www.ec.europa.eu; gli orientamenti del gruppo di esperti sono stati condivisi dalla Commissione con la Comunicazione *Creare fiducia nell'intelligenza artificiale antropocentrica*, 8 aprile 2019, COM(2019), 168 final, nella quale accoglie con favore l'individuazione dei sette requisiti fondamentali dell'IA: intervento e sorveglianza umani; robustezza tecnica e sicurezza; riservatezza e *governance* dei dati; trasparenza; diversità; non discriminazione ed equità; benessere sociale e ambientale; *accountability*.

[5] *Libro bianco sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia*, 19 febbraio 2020, COM(2020), 65 final.

[6] Nel *Libro bianco sull'intelligenza artificiale*, cit., p. 15, «la Commissione ha ritenuto che il quadro legislativo possa essere migliorato per affrontare le situazioni e i rischi individuati come derivanti dall'impiego dell'intelligenza artificiale»; anche il *position paper* dal titolo *Innovative and Trustworthy AI: two sides of the same coin*, dell'8 ottobre 2020, firmato da 14 Stati membri dell'Unione europea, critica la previsione di porre prescrizioni e adempimenti obbligatori in materia di intelligenza artificiale, suggerendo, invece, di intervenire con un regime di *soft law*, capace di assicurare così la necessaria flessibilità a una tecnologia in rapida e continua evoluzione.

[7] V. per tutte la *Risoluzione del Parlamento europeo del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale (2020/2016(INI))*, in *G.U.U.E.*, 23 marzo 2022, C 132/17, che invita «la Commissione a valutare se sia necessario adottare una specifica azione legislativa volta a precisare ulteriormente

lamento (UE) del 21 aprile 2021[8] la Commissione, però, si è espressa per questa seconda soluzione, motivata dalla necessità di introdurre una disciplina uniforme che contribuisca «all'obiettivo dell'Unione di essere un leader mondiale nello sviluppo di un'intelligenza artificiale sicura, affidabile ed etica» (punto 5 delle premesse) e sia funzionale a evitare che «normative nazionali divergenti possano determinare una frammentazione del mercato interno e diminuire la certezza del diritto per gli operatori che sviluppano o utilizzano sistemi di IA» (punto 2 delle premesse).

L'Unione europea, quindi, ha scelto di operare attraverso una «legislazione rigida a carattere orizzontale»[9], che mira a garantire lo sviluppo e il funzionamento di un mercato unico digitale e a tutelare i diritti e i valori fondamentali contro i rischi derivanti dall'uso di *software* di intelligenza artificiale.

L'originaria proposta della Commissione ha subìto, poi, gli emendamenti del Parlamento, approvati il 14 giugno 2023, che però hanno lasciato immutata la logica *risk-oriented* della disciplina chiaramente mutuata dal GDPR (*General Data Protection Regulation*)[10], il Regolamento generale sulla protezione dei dati[11].

Tale Regolamento (UE), in primo luogo, indica una serie di divieti **di utilizzo dell'intelligenza artificiale**[12], perché fonte di rischi inaccettabili

---

i criteri e le condizioni per lo sviluppo, l'uso e la diffusione di applicazioni e soluzioni di IA da parte delle autorità di contrasto e giudiziarie» (par. 35).

[8] *Proposta di Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativo dell'Unione*, 21 aprile 2021, COM(2021), 206 final.

[9] F. Donati, *Diritti fondamentali e algoritmi nella proposta di regolamento sull'intelligenza artificiale*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. I, *Diritti fondamentali, dati personali e regolazione*, Il Mulino, Bologna, 2022, p. 121.

[10] *Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati)*, in *G.U.U.E.*, 4 maggio 2016, L 119/1.

[11] Secondo L. Floridi, *The European Legislation on AI: a Brief Analysis of its Philosophical Approach*, in *Philosophy & Technology*, 2022, p. 218, infatti, «*from an ethical perspective, the AIA inherits the same foundational approach seen in the GDPR: it is based on protecting human dignity and fundamental rights*».

[12] L'elenco è stato ampliato dal Parlamento, includendo anche i sistemi di categorizzazione biometrica che utilizzano caratteristiche sensibili (ad esempio genere, razza, etnia, stato di cittadinanza, religione, orientamento politico), i sistemi di polizia predittiva (basati su profilazione, posizione o precedenti penali), i sistemi di riconoscimento delle emozioni nelle forze dell'ordine, alle frontiere, sul posto di lavoro e nelle scuole e quelli di identificazione biometrica

per la salute, la sicurezza e i diritti fondamentali della persona. In secondo luogo, prevede regole di trasparenza armonizzate e requisiti specifici e stringenti per gli utilizzi dell'intelligenza artificiale in ambiti ritenuti "ad alto rischio" [13], che impongono valutazioni di conformità, certificazioni, obblighi di registrazione e monitoraggio del funzionamento dei relativi sistemi, nel cui ambito si innesta pure la valutazione d'impatto sui diritti fondamentali [14]. Solo un obbligo di segnalazione agli utenti è invece imposto per i sistemi "a rischio limitato". Nessuna cautela è invece prevista per quelli "a rischio minimo".

Per i sistemi definiti "ad alto rischio" [15], la strategia dell'Unione europea è stata orientata sin da subito a spostare il presidio delle garanzie "a monte", principalmente su *deployers* e *providers* [16]. In ogni caso, ai sensi del Regolamento (UE), tali sistemi devono essere progettati e sviluppati in modo tale da consentire un'efficace supervisione da parte delle persone fisiche durante il loro utilizzo [17]. Ciò in quanto afferiscono ad attività che richie-

---

remota in tempo reale ed *ex post* in spazi pubblici, pur con la previsione di alcune eccezioni per le forze dell'ordine.

[13] Il relativo elenco è riportato in un apposito allegato del Regolamento (UE): *ANNEX III*.

[14] Il *Fundamental Rights Impact Assessment* (FRIA) è stato inserito dal Parlamento nella prima bozza del Regolamento (UE), all'art. 29(a), rendendolo obbligatorio per gli utilizzatori di sistemi di IA "ad alto rischio", i quali, in base ai relativi risultati, devono sviluppare piani per mitigarne eventuali impatti negativi sui diritti fondamentali e, laddove ciò non sia possibile, devono cessarne l'utilizzo, informandone i fornitori e le autorità nazionali; in tema, v. C. Novelli, *L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*, in www.federalismi.it, 2024, n. 2, pp. 110-111. L'A., in primo luogo, nota come «il bilanciamento dei diritti, operazione complessa persino per i giuristi qualificati delle Corti, spingerebbe i *deployers* ad assumere consulenti esperti legali o esternalizzare questa valutazione. Anche assumendo che tutto ciò abbia un costo ragionevole per i *deployers*, questi ultimi si farebbero carico in definitiva di decisioni particolarmente sensibili»; in secondo luogo, rileva la carenza di omogeneità nelle valutazioni FRIA.

[15] Per un'analisi dettagliata dei sistemi "ad alto rischio" nell'ambito della Proposta di regolamento si rinvia a G. Finocchiaro, *La proposta di regolamento sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in *Dir. inf.*, 2022, pp. 313-320.

[16] M. Veale, F. Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act. Analysing the good, the bad, and the unclear elements of the proposed approach*, in *Computer Law Review International*, 2021, n. 4, pp. 102-103, notano che «*the vast majority of all obligations fall on the 'provider': in short, person or body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark*».

[17] L'art. 14 (*Human oversight*), il cui testo è rimasto inalterato nel corso dei negoziati, prevede infatti che «i sistemi di IA "ad alto rischio" sono progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente su-

dono una componente e una responsabilità tipicamente umane, non delegabili alla macchina [18].

Tra i sistemi "ad alto rischio", già la Commissione nella proposta di Regolamento (UE) aveva indicato espressamente l'opportunità di classificare «alcuni sistemi di IA destinati all'amministrazione della giustizia e ai processi democratici […] in considerazione del loro impatto potenzialmente significativo sulla democrazia, sullo Stato di diritto, sulle libertà individuali e sul diritto a un ricorso effettivo e a un giudice imparziale». Fra questi vanno necessariamente inclusi «i sistemi di IA destinati ad assistere le autorità giudiziarie nelle attività di ricerca e interpretazione dei fatti e del diritto e nell'applicazione della legge a una serie concreta di fatti» (considerando n. 40).

In ambito penalistico, diversi sono i contesti in cui i sistemi di intelligenza artificiale possono trovare applicazione – dagli strumenti d'indagine, come i noti *software* di riconoscimento facciale, all'attività predittiva, fino alla determinazione del contenuto di una decisione giudiziaria – su cui la regolamentazione europea è destinata a impattare [19].

La Proposta di Regolamento (UE) è stata oggetto di intensi negoziati tra le istituzioni europee (i c.d. triloghi) conclusisi con l'accordo provvisorio dell'8 dicembre 2023, in base al quale anche il Comitato dei Rappresentanti permanenti (Coreper) con il voto del 2 febbraio 2024 ha adottato la posizione del Consiglio dell'UE sull'*AI Act*, il quale, in seguito all'approvazione del Parlamento e al voto del Consiglio, è diventata la prima forma di regolamentazione organica dell'intelligenza artificiale a livello mondiale, con la pubblicazione in gazzetta ufficiale del 12 luglio 2024.

3. *Le regole preesistenti*

Le regole *ad hoc* sull'intelligenza artificiale del nuovo Regolamento (UE) andranno a integrare altri atti, sia di *soft law*, sia vincolanti, già vigenti nel

---

pervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso»; per un'analisi della disposizione e dei relativi profili critici si rinvia a L. Enqvist, *'Human oversight' in the EU artificial intelligence act: what, when and by whom?*, in *Law, Innovation and Technology*, 2023, vol. 15, n. 2, p. 508 ss.

[18] C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, cit., p. 429, mettono in luce come la regola «dimostr[i] un apprezzabile realismo e, al contempo, lanc[i] una sfida culturale impegnativa».

[19] Si rinvia nel merito al par. 4.

diritto dell'Unione europea e rilevanti ai fini dell'impiego di sistemi di IA [20].

In particolare, il legislatore eurounitario ha da tempo dedicato specifica attenzione al tema della tutela del diritto alla *privacy*, declinato nella Carta di Nizza nella sua duplice accezione di diritto al «rispetto della vita privata e della vita familiare» (art. 7) e alla «protezione dei dati di carattere personale» (art. 8). Infatti, nel c.d. *Pacchetto privacy* del 2016 [21] si ritrovano alcune regole fondamentali per l'utilizzo dei sistemi di intelligenza artificiale, i quali hanno come presupposto tecnico essenziale l'elaborazione dei c.d. *big data*. Nel settore della giustizia penale, di particolare rilievo è la direttiva (UE) 2016/680 (c.d. direttiva *Law enforcement*), relativa al trattamento dei dati personali da parte delle autorità competenti a fini di «prevenzione, indagine, accertamento e perseguimento dei reati».

3.1. *L'automazione nell'assunzione delle decisioni penali*

La questione se affidare all'intelligenza artificiale l'attività "normalmente umana" del decidere – intesa come capacità di formulare giudizi e determinare il contenuto di una decisione – assume un'importanza fondamentale se si tratta di una decisione giudiziaria, soprattutto in materia penale. La giurisdizione penale vive, infatti, di principi – si pensi alla soggezione del giudice solo alla legge, all'obbligo di motivazione dei provvedimenti, al diritto a un ricorso effettivo – che rischiano di indebolirsi fino a perdere significato in

[20] V. considerando 9, *Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)*, cit., per cui «le norme armonizzate stabilite nel presente regolamento dovrebbero applicarsi in tutti i settori e, in linea con il nuovo quadro legislativo, non dovrebbero pregiudicare il vigente diritto dell'Unione, in particolare in materia di protezione dei dati, tutela dei consumatori, diritti fondamentali, occupazione e protezione dei lavoratori e sicurezza dei prodotti, al quale il presente regolamento è complementare».

[21] Esso si compone: del *Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati)*, cit.; della *Direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio*, in *G.U.U.E.*, 4 maggio 2016, L 119/89; della *Direttiva (UE) 2016/681 del Parlamento europeo e del Consiglio del 27 aprile 2016 sull'uso dei dati del codice di prenotazione (PNR) a fini di prevenzione, accertamento, indagine e azione penale nei confronti dei reati di terrorismo e dei reati gravi*, in *G.U.U.E.*, 4 maggio 2016, L 119/132.

rapporto a una decisione automatizzata, in quanto assunta da una macchina e non dal giudice penale libero nel suo convincimento.

Ci si interroga, allora, sulla possibilità di avvalersi dell'utilizzo di *software* durante l'intero procedimento, che possano acquisire ed elaborare tutte le informazioni rilevanti per il caso concreto e, sulla base di decisioni preimpostate assunte in casi simili, giungere alla singola decisione.

In questi casi, il diritto dell'Unione europea offre una tutela inerente alla protezione dei dati personali: l'art. 11 della direttiva (UE) 2016/680 prevede il divieto di decisioni basate «unicamente su un trattamento automatizzato, compresa la profilazione, che produca[no] effetti giuridici negativi o incida[no] significativamente sull'interessato», a meno che le stesse siano autorizzate dal diritto dell'UE o dello Stato membro al quale è soggetto il titolare del trattamento e siano previste «garanzie adeguate per i diritti e le libertà dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento»[22]. Dalla disposizione, quindi, è possibile enucleare un diritto a non essere destinatari di decisioni che si basino in via esclusiva su trattamenti automatizzati, che non prevedono alcun coinvolgimento dell'essere umano nel processo decisionale[23].

Per quanto attiene all'intensità della tutela, la direttiva (UE) 2016/680, in realtà, pur replicando quanto previsto anche al di fuori della materia penale dall'art. 22 del Regolamento (UE) 2016/679, in ordine a una decisione basata unicamente su un trattamento automatizzato[24], pare più incisiva rispetto a quella stabilita dalla disciplina generale[25]. I destinatari di

---

[22] Per S. Signorato, *Il diritto a decisioni penali non basate esclusivamente su trattamenti automatizzati: un nuovo diritto derivante dal rispetto della dignità umana*, in *Riv. dir. proc.*, 2019, p. 106, il divieto in questione è solo apparente; dello stesso avviso è L. Pressacco, *Intelligenza artificiale e ragionamento probatorio nel processo penale*, in G. Di Paolo, L. Pressacco (a cura di), *Intelligenza artificiale e processo penale. Indagini, prove, giudizio*, Editoriale scientifica, Napoli, 2022, p. 104.

[23] Si trova già esplicita menzione nel considerando n. 38 del diritto dell'interessato a «non essere oggetto di una decisione che valuta aspetti personali che lo concernono basata esclusivamente su un trattamento automatizzato».

[24] La disposizione, al par. 1, stabilisce che «l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona»; anche in questo contesto, l'interessato gode del diritto, riconosciutogli al par. 3, «di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestarne la decisione».

[25] Sottolinea infatti M. Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, in *Dir. pen. cont.*, 29 maggio 2019, p. 16, che «la direttiva 2016/680/UE costituisce una *lex specialis* rispetto al regolamento».

una decisione automatizzata assunta per finalità di prevenzione e repressione dei reati, infatti, non possono rinunciare preventivamente al diritto a invocare il controllo umano, come invece risulterebbe consentito dalla norma del GDPR, nel caso in cui la decisione «si basi sul consenso esplicito dell'interessato» (par. 2, lett. c)[26].

Il testo del primo paragrafo dell'art. 22 del Regolamento (UE) 2016/679 lo si trova, peraltro, letteralmente trasfuso nell'art. 56 del Regolamento (UE) 2017/1939[27], che riconosce all'interessato «il diritto di non essere sottoposto a una decisione dell'EPPO basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona».

In relazione all'estensione applicativa del divieto, occorre intendersi sul significato da attribuire all'espressione «decisione basata unicamente su un trattamento automatizzato». Non v'è dubbio che il diritto UE vieti, al di fuori di una specifica previsione normativa – nazionale o dell'Unione – e in assenza delle garanzie previste a tutela dell'interessato, le decisioni giurisdizionali adottate solo attraverso *software* di intelligenza artificiale, senza alcun coinvolgimento dell'intelligenza umana sul risultato decisionale.

La formulazione letterale delle disposizioni sembrerebbe avallare, però, un'interpretazione più ampia del divieto, esteso anche all'assunzione di decisioni fondate unicamente su *output* di un meccanismo automatizzato, vale a dire su elementi di prova provenienti da tecnologie di intelligenza artificiale[28]. D'altronde, i sistemi basati su algoritmi sono estremamente varie-

---

[26] S. Signorato, *Le indagini digitali. Profili strutturali di una metamorfosi investigativa*, Giappichelli, Torino, 2018, p. 101.

[27] *Regolamento (UE) 2017/1939 del Consiglio del 12 ottobre 2017 relativo all'attuazione di una cooperazione rafforzata sull'istituzione della Procura europea («EPPO»)*, in *G.U.U.E.*, 31 ottobre 2017, L 283/1.

[28] Sul punto, M. Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, cit., p. 17, propone di valutare gli elementi di prova ottenuti mediante strumenti di intelligenza artificiale alla stregua degli indizi, che necessitano di essere corroborati da altri elementi di prova. Per l'A., infatti, «accanto all'obbligo di un intervento umano andrebbe ritenuta sussistente quella che, nel lessico processualpenalistico, chiameremmo regola di valutazione»; tale interpretazione è condivisa pure da L. Pressacco, *Intelligenza artificiale e ragionamento probatorio nel processo penale*, cit., p. 106, il quale, pur riconoscendovi una certa «rigidità», ritiene «plausibile affermare che tali informazioni sarebbero idonee a dimostrare la sussistenza di determinati fatti soltanto con il concorso di ulteriori elementi di prova, in grado di bilanciarne – rispettivamente – l'eccessiva genericità o specificità», che sono «caratteristiche piuttosto comuni tra gli elementi generati o raccolti mediante dispositivi di IA» (così a p. 121); v. anche K. La Regina, *I.A. e ragionamento giuridico: la*

gati e, a seconda dello scopo per cui sono programmati, suscettibili di differenti applicazioni, potendo pertanto impattare in maniera e misura differente sulla decisione penale. Così, l'IA con finalità analitico-computazionali – ad esempio finalizzata al calcolo della pena o dei termini di prescrizione o alla raccolta e alla catalogazione di pronunce giurisprudenziali – si affianca a quella con finalità predittiva – ad esempio in relazione alla reiterazione di un reato o alla pericolosità sociale – che elabora dati reali per formulare una risposta in termini probabilistici, «cioè una decisione dal punto di vista logico-cognitivo»[29].

In sostanza, può dirsi che il legislatore dell'UE escluda la possibilità di una «delega totale all'automazione»[30] nell'assunzione delle decisioni giudiziarie in materia penale[31], e richieda il controllo umano *ex post* sulle valutazioni emesse dai sistemi di IA, così da mantenere la necessaria interazione tra macchina ed essere umano (c.d. "*human in the loop*")[32].

In altri termini, come emerge dalle *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*, «se un essere umano riesamina il risultato del processo automatizzato e tiene conto di altri fattori nel prendere la decisione finale, tale decisione non sarà "basata unicamente" sul trattamento automatizzato»[33],

---

giustizia prevedibile, in G.M. Baccari, P. Felicioni (a cura di), *La decisione penale tra intelligenza emotiva e intelligenza artificiale*, Giuffrè, Milano, 2023, p. 182; secondo G. Ubertis, *Intelligenza artificiale e giustizia predittiva*, in *Sist. pen.*, 16 ottobre 2023, p. 7, «disconoscere ciò, significherebbe trascurare che la "percezione soggettiva" è ineludibile per la comprensione delle circostanze emergenti nel processo e relative alla complessità degli elementi oggettivo e soggettivo delle fattispecie penali, mentre mancherebbe la possibilità di conformare l'esito processuale alle imprevedibili cadenze della dialettica che si svolge in sede processuale».

[29] V. G. Padua, *Intelligenza artificiale e giudizio penale*, in *Proc. pen. giust.*, 2021, p. 1490.

[30] L'espressione è di G. Lasagni, *Difendersi dall'intelligenza artificiale o difendersi con l'intelligenza artificiale? Verso un cambio di paradigma*, in G. Di Paolo, L. Pressacco (a cura di), *Intelligenza artificiale e processo penale. Indagini, prove, giudizio*, Editoriale scientifica, Napoli, 2022, p. 70.

[31] A tal proposito si parla di «principio non esclusività»; in tema v. L. Luparia, *Diritto probatorio e giudizi criminali ai tempi dell'intelligenza artificiale*, in R. Giordano, A. Panzarola, A. Police, S. Preziosi, M. Proto (a cura di), *Il diritto nell'era digitale. Persona, Mercato, Amministrazione, Giustizia*, Giuffrè, Milano, 2022, p. 782; ugualmente, in relazione all'art. 22 GDPR, v. S. Maranella, *La protezione dei dati personali contro un uso distopico dell'AI*, in *op. cit.*, p. 63.

[32] In termini generali, v. anche *Libro bianco sull'intelligenza artificiale*, cit., p. 23, il quale in tema di prescrizioni nell'impiego dell'intelligenza artificiale "ad alto rischio" annovera «la sorveglianza umana», che «aiuta a garantire che un sistema di IA non comprometta l'autonomia umana o provochi altri effetti negativi».

[33] *Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profi-*

purché il controllo non rappresenti un puro adempimento formale, ma sia tale da poter incidere significativamente sulla decisione[34]. Inoltre, l'art. 11 della direttiva (UE) 2016/680 sembra circoscrivere, in materia penale, il divieto di decisioni interamente automatizzate ai soli provvedimenti negativi nei confronti dell'interessato, risultando sempre legittimi se producono invece effetti favorevoli.

Sul tema, ha preso posizione il Parlamento europeo con un atto d'indirizzo politico – la risoluzione del 6 ottobre 2021 *sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*[35] – che, in modo radicale, ha raccomandato la previsione di un «divieto dell'uso dell'intelligenza artificiale e delle relative tecnologie per l'emanazione delle decisioni giudiziarie» (par. 16).

Secondo l'Europarlamento, infatti, «gli esseri umani che fanno affidamento unicamente sui dati, i profili e le raccomandazioni generati dalle macchine, non saranno in grado di condurre una valutazione indipendente» (par. 15); il che impone di affidare le decisioni con effetti giuridici «sempre» a un essere umano, che «possa essere ritenuto responsabile per le decisioni adottate» (par. 16).

A ben guardare, la garanzia fondamentale rappresentata dalla centralità dell'intervento umano successivo sulle decisioni automatizzate, già prevista dall'art. 11 della direttiva (UE) 2016/680, più che la possibilità di mettere in discussione il merito del risultato tecnico fornito dagli strumenti di intelligenza artificiale, assicura l'individuazione del titolare del potere di emissione della decisione, ossia il giudice. Spetta, cioè, sempre a quest'ultimo il compito di valutare il risultato statistico o predittivo dall'algoritmo, consi-

---

lazione ai fini del regolamento 2016/679, adottate il 3 ottobre 2017, nella versione emendata e adottata in data 6 febbraio 2018 dal Gruppo di lavoro articolo 29 per la protezione dei dati.

[34] Altrimenti, afferma M. Brakan, *Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond*, in *International journal of law and information technology*, 2019, vol. 27, n. 2, p. 103, «*a formalistic interpretation, involving the human only as a necessary part of procedure but ultimately leaving the decision power to the machine, would not ensure a sufficiently high level of data protection of the data subject*»; in senso conforme v. anche G.N. La Diega, *Against the Dehumanisation of decision-Making. Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information*, in *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 2018, vol. 9, n. 2, p. 19.

[35] *Risoluzione del Parlamento europeo del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale (2020/2016(INI))*, cit.; a commento v. G. Barone, *Intelligenza artificiale e processo penale: la linea dura del Parlamento europeo. Considerazioni a margine della Risoluzione del Parlamento europeo del 6 ottobre 2021*, in *Cass. pen*., 2022, p. 1180.

derandone il peso dimostrativo in relazione all'oggetto del giudizio, mediante un ragionamento critico che valorizzi le specificità del caso concreto.

3.2. *L'impiego di* software *che utilizzano dati biometrici*

Pur di fronte alle forti potenzialità in termini di efficienza nella prevenzione e nella repressione dei reati, da tempo l'Unione europea mostra un atteggiamento restrittivo rispetto all'impiego dell'intelligenza artificiale che sfrutta dati biometrici, tra cui assume rilievo il funzionamento dei *tool* di riconoscimento facciale.

Innanzitutto, una precisa definizione di dati biometrici è contenuta all'art. 4, punto 14 del GDPR, ripresa anche da altre fonti [36], da intendersi quali «dati personali ottenuti da un trattamento tecnico specifico relativi alle caratteristiche fisiche, fisiologiche o comportamentali di una persona fisica che ne consentono o confermano l'identificazione univoca, quali l'immagine facciale o i dati dattiloscopici» [37].

Trattandosi di dati particolarmente sensibili, in quanto idonei a identificare in modo univoco un individuo, l'art. 10 della direttiva (UE) 2016/680 ne assoggetta il trattamento da parte dalle autorità di *law enforcement* a criteri piuttosto stringenti. Esso è «autorizzato solo se strettamente necessario [38], soggetto a garanzie adeguate per i diritti e le libertà dell'interessato

---

[36] Art. 3, par. 18, *Regolamento (UE) 2018/1725 del Parlamento europeo e del Consiglio del 23 ottobre 2018 sulla tutela delle persone fisiche in relazione al trattamento dei dati personali da parte delle istituzioni, degli organi e degli organismi dell'Unione e sulla libera circolazione di tali dati, e che abroga il regolamento (CE) n. 45/2001 e la decisione n. 1247/2002/CE*, in *G.U.U.E.*, 21 novembre 2018, L 295/39; art. 3, par. 14, *Direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio*, cit.

[37] L'esigenza dell'*AI Act* di coordinamento con la disciplina già in vigore riguarda anche le definizioni già contenute in altri atti dell'Unione europea; per quanto attiene alla definizione di «dati biometrici», il considerando n. 14 avverte che quella utilizzata nel Regolamento (UE) («i dati biometrici possono consentire l'autenticazione, l'identificazione o la categorizzazione delle persone fisiche e il riconoscimento delle emozioni delle persone fisiche») debba essere interpretata «alla luce di» quella fornita dal GDPR e dalle direttive di settore.

[38] Per CGUE, Grande Sezione, 30 gennaio 2024, C-118/22, *NG*, par. 48, l'espressione «definisce una condizione rafforzata di liceità del trattamento di tali dati e implica, segnatamente, un controllo particolarmente rigoroso del rispetto del principio della "minimizzazione dei dati", quale derivante dall'articolo 4, paragrafo 1, lettera c), della direttiva 2016/680, del quale tale requisito costituisce un'applicazione specifica a detti dati sensibili».

e soltanto: a) se autorizzato dal diritto dell'Unione o dello Stato membro; b) per salvaguardare un interesse vitale dell'interessato o di un'altra persona fisica; c) se il suddetto trattamento riguarda dati resi manifestamente pubblici dall'interessato».

Il trattamento, anche automatizzato di dati biometrici intesi a identificare in modo univoco una persona è consentito a Europol dall'art. 30, par. 2, del Regolamento (UE) 2016/794 [39], come recentemente modificato dal Regolamento (UE) 2022/991 [40], con la «finalità di prevenire o combattere forme di criminalità organizzata nell'ambito degli obiettivi di Europol», ma pur sempre nel rispetto dei diritti e delle libertà dell'interessato attraverso le garanzie previste dallo stesso Regolamento. Tale disposizione è stata introdotta in occasione della generale investitura conferita a Europol quale promotore dello sviluppo e della diffusione di un'intelligenza artificiale «etica, affidabile e antropocentrica, che è soggetta a solide garanzie in termini di protezione, sicurezza, trasparenza, spiegabilità e diritti» (considerando n. 49).

Non deve trascurarsi che qualunque trattamento di dati biometrici finalizzato all'identificazione univoca di una persona fisica, siccome rappresenta una compressione dell'esercizio dei diritti alla *privacy* riconosciuti dagli artt. 7 e 8 della Carta dei diritti fondamentali dell'Unione europea, è in ogni caso soggetto al rispetto del principio di proporzionalità previsto dall'art. 52, per cui possono essere apportate limitazioni ai diritti riconosciuti dalla Carta «solo laddove siano necessarie e rispondano effettivamente a finalità di interesse generale riconosciute dall'Unione o all'esigenza di proteggere i diritti e le libertà altrui».

Il Parlamento europeo, sempre con la risoluzione del 6 ottobre 2021, ha espresso, però, forte preoccupazione rispetto all'utilizzo di tali *software*, sia per «la controversa validità scientifica della tecnologia di riconoscimento utilizzata», sia, e in special modo, perché «l'uso dei dati biometrici è correlato in senso ampio al principio di dignità umana, che è la base di tutti i diritti fondamentali garantiti dalla Carta», osservando che «l'uso dell'iden-

---

[39] *Regolamento (UE) 2016/794 del Parlamento europeo e del Consiglio dell'11 maggio 2016 che istituisce l'Agenzia dell'Unione europea per la cooperazione nell'attività di contrasto (Europol) e sostituisce e abroga le decisioni del Consiglio 2009/371/GAI, 2009/934/GAI, 2009/935/GAI, 2009/936/GAI e 2009/968/GAI*, in *G.U.U.E.*, 24 maggio 2016, L 135/53.

[40] *Regolamento (UE) 2022/991 del Parlamento europeo e del Consiglio dell'8 giugno 2022 che modifica il regolamento (UE) 2016/794 per quanto riguarda la cooperazione di Europol con le parti private, il trattamento dei dati personali da parte di Europol a sostegno di indagini penali, e il ruolo di Europol in materia di ricerca e innovazione*, in *G.U.U.E.*, 27 giugno 2022, L 169/1.

tificazione biometrica nel contesto delle attività di contrasto e giudiziarie dovrebbe sempre essere considerato ad alto rischio» (par. 30) [41].

3.2.1. *Il riconoscimento facciale automatico*

Tra le tipologie di tecniche biometriche il cui impiego, specie in ambito penale, è visto con particolare diffidenza vi sono i *software* di riconoscimento facciale automatico [42]. Dalla fotografia del volto è possibile estrarre una serie di caratteristiche – come, ad esempio, la radice del naso, il punto di attacco dei capelli sulla fronte, il punto più sporgente dello zigomo (c.d. *faceprints*) – che consentono di creare il modello elettronico (*template*) di quello specifico tratto biometrico [43], una sorta di "codice a barre" dell'individuo, che può essere usato con finalità identificative in diversi contesti, comprese le attività di prevenzione e perseguimento dei reati. Il risultato fornito è comunque di tipo probabilistico, in quanto il sistema compara il singolo volto con tutti gli altri disponibili in una banca dati, assegnando un valore che indica il grado di probabilità che le due immagini rappresentino la stessa persona, con la possibilità anche di riscontrare una perfetta corrispondenza.

Tale metodo, rispetto ad altri di riconoscimento biometrico, se, da un lato, ha il vantaggio di non essere invasivo per il soggetto passivo, dall'altro lato, non è da questi controllabile, tanto da poter consentire l'identificazione senza la necessità di alcun atteggiamento collaborativo. Di qui, può cogliersi l'elevata efficienza di tale tecnologia, che permette la cattura dell'immagine a distanza senza che il soggetto possa accorgersene, minata, però, da un riconosciuto margine di errore piuttosto elevato, dovuto a fattori esterni che ne influenzano la *performance* – si pensi alla distanza del soggetto dal sensore, alla qualità delle immagini o alla tipologia di luce, oltre al fatto che le stesse caratteristiche del volto sono soggette a mutamenti

---

[41] La conclusione era stata peraltro condivisa dalla Commissione nel *Libro bianco sull'intelligenza artificiale*, cit., p. 20.

[42] A tale riguardo, basta ricordare come in un caso in materia di passaporto biometrico CGUE, Sez. IV, 17 ottobre 2013, C-291/12, *Micheal Schwarz*, par. 48, abbia riconosciuto come, laddove venga raccolta l'immagine facciale, la fotografia del volto possa provocare imbarazzo fisico e psichico per l'interessato.

[43] Così E. SACCHETTO, *Brevi riflessioni sui fondamenti e limiti del rapporto fra automated faced-based human recognition technology e procedimento penale*, in A. PAJNO, F. DONATI, A. PERRUCCI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, vol. II, *Amministrazione, responsabilità, giurisdizione*, Il Mulino, Bologna, 2022, p. 541 ss., cui si rinvia per una sintetica ed efficace spiegazione del funzionamento del riconoscimento facciale tramite *software*.

– o anche alla presenza di veri e propri *bias* nello sviluppo dell'algoritmo dovuti alla qualità dei dati con cui si alimenta, con il serio rischio che il *software* possa prestarsi a controlli discriminatori e oppressivi da parte delle autorità pubbliche [44].

L'elevata probabilità di riconoscere falsamente un determinato soggetto, anche legata a pregiudizi che influenzano la progettazione dell'algoritmo [45], fa sì che l'impiego dei programmi di riconoscimento facciale – e di identificazione biometrica in generale – possa creare frizioni con la presunzione d'innocenza (art. 48 della Carta dei diritti fondamentali dell'Unione europea) [46], tanto nelle attività di polizia predittiva finalizzate a prevedere il momento e il luogo di commissione di un reato (c.d. *hotspots*) o il suo potenziale autore, quanto in fase investigativa nelle operazioni di identificazione dei responsabili di un reato già commesso.

Non stupisce, allora, che la regolamentazione di tale tecnica di controllo sia stata controversa nonché, come si dirà, oggetto di ripensamenti nel corso della stesura dell'*AI Act*.

Le istituzioni europee, tuttavia, anche al di fuori dell'*iter* legislativo dell'*hard law* avevano già assunto una linea tesa a porre un freno all'uso dei *software* di riconoscimento facciale nelle attività di contrasto. Tra queste, l'Agenzia europea per i diritti fondamentali (FRA) ha ritenuto opportuno che l'uso nelle attività di polizia sia «*strictly determined*» e limitato, oltre

---

[44] Tale aspetto è stato messo in luce anche dal documento della *European Union Agency for Fundamental Rights* (FRA), *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, 2020, p. 27, consultabile in www.fra.europa.eu, per cui «*an important cause of discrimination is the quality of data used to develop algorithms and software. […]. Yet to date, facial images used to develop algorithms in the Western world often over-represent white men, with lower numbers of women and/or individuals of other ethnic backgrounds. As a result, facial recognition systems worked well for white men, but not for black women*».

[45] Secondo uno studio condotto dal *National Institute of Standards and Technology* degli Stati Uniti, l'utilizzo di molti *software* di riconoscimento facciale riporta molti più errori di identificazione nei confronti delle persone afroamericane, soprattutto se donne, e asiatiche rispetto alle persone caucasiche (v. P. Grother, M. Ngan, K. Hanaoka, *Face Recognition Vendor Test (FVRT). Part 3: Demographic Effects*, in www.nvlpubs.nist.gov); su tale studio e in generale sul tema v. J. Della Torre, *Quale spazio per i tools di riconoscimento facciale nella giustizia penale?*, in G. Di Paolo, L. Pressacco (a cura di), *Intelligenza artificiale e processo penale. Indagini, prove, giudizio*, Editoriale scientifica, Napoli, 2022, p. 20 ss., il quale spiega che all'origine di tali discriminazioni vi sia «il pericolo che i creatori degli algoritmi, non selezionando in modo sufficientemente attento i volti con cui "allenare" gli applicativi, finiscano per "trasmettere" al sistema pregiudizi umani».

[46] G. Barone, *Intelligenza artificiale e processo penale: la linea dura del Parlamento europeo. Considerazioni a margine della Risoluzione del Parlamento europeo del 6 ottobre 2021*, cit., p. 1188.

che all'identificazione delle persone scomparse e delle vittime di reato, compresi i bambini, alla lotta al terrorismo e altri gravi reati, ricalcando gli ormai consolidati limiti previsti per l'accesso delle forze dell'ordine ai *database* dell'UE su larga scala [47].

Un approccio ancora più rigido è quello adottato dal Parlamento, che, con la più volte richiamata risoluzione del 2021, ha chiesto una moratoria dell'impiego degli strumenti di riconoscimento facciale nelle attività di contrasto, «finché le norme tecniche non possano essere considerate pienamente conformi con i diritti fondamentali, i risultati ottenuti siano privi di distorsioni e non discriminatori, il quadro giuridico fornisca salvaguardie rigorose contro l'utilizzo improprio e un attento controllo democratico e adeguata vigilanza, e vi sia la prova empirica della necessità e proporzionalità della diffusione di tali tecnologie»; a esclusione dei casi in cui tali strumenti siano usati per l'identificazione delle vittime (par. 27).

4. *Gli approdi dell'***AI Act** *per la giustizia penale*

Rispetto ai profili fino a ora analizzati e ferme le disposizioni del GDPR e della direttiva (UE) 2016/680, il Regolamento (UE) sull'intelligenza artificiale è destinato a incidere con livelli di intensità diversa.

L'*AI Act* ha ampliato il catalogo degli usi vietati dei sistemi di intelligenza artificiale, in quanto portatori di un livello di rischio "inaccettabile", in ragione della potenziale violazione di diritti fondamentali, ricomprendendo, fra l'altro, lo *scraping* non mirato di immagini da internet o da telecamere a circuito chiuso per creare o espandere *database* di riconoscimento facciale (art. 5, par. 1, lett. e), i sistemi di categorizzazione biometrica che utilizzano caratteristiche sensibili, come la razza, le opinioni politiche, le convinzioni religiose o filosofiche o l'orientamento sessuale (art. 5, par. 1, lett. g) o il c.d. *social scoring* (art. 5, par. 1, lett. c).

Rimane vietata la polizia predittiva mediante l'uso di sistemi di intelligenza artificiale per effettuare valutazioni sul rischio di commissione di un reato o di recidiva di una persona fisica o di un gruppo, sulla base della profilazione o della valutazione dei tratti e delle caratteristiche della personalità (art. 5.1, par. 1, lett. d). In questo caso, il rischio che l'Unione europea ha ritenuto "inaccettabile" è la lesione della presunzione d'innocenza,

[47] *European Union Agency for Fundamental Rights (FRA)*, *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, cit., p. 25.

per la quale le persone fisiche nell'UE dovrebbero essere sempre giudicate in base al fatto commesso (considerando 42).

Di gran rilievo nell'ambito della prevenzione e del contrasto al crimine è, inoltre, il divieto di sistemi di identificazione biometrica remota *real-time* in spazi accessibili al pubblico. Tuttavia, l'originario divieto totale da parte del Parlamento è stato temperato, in fase di negoziazione, da un numero limitato di eccezioni, cui si accompagnano una serie di cautele di salvaguardia [48]. Nel dettaglio, l'uso del sistema di IA è consentito nella misura in cui sia strettamente necessario per: *i*) la ricerca mirata di potenziali vittime specifiche di sequestro di persona, tratta di essere umani, sfruttamento della prostituzione, nonché di persone scomparse; *ii*) la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone o di una reale minaccia attuale o potenziale di attacco terroristico; *iii*) la localizzazione o l'identificazione di una persona sospettata di aver commesso un reato, al fine di indagarlo, processarlo o sottoporlo all'esecuzione della pena di specifici reati [49], puniti nello Stato membro interessato con una pena o una misura di sicurezza privativa della libertà personale della durata, nel massimo, di almeno quattro anni, come stabilito dalla legge di tale Stato membro (art. 5, par. 1, lett. h).

Se, da un lato, il testo normativo restringe ai casi previsti l'uso dei sistemi di identificazione biometrica in tempo reale, dall'altro lato, la formulazione letterale delle eccezioni al divieto, talora generica (v. in particolare l'eccezione di cui al punto *ii*), presta il fianco a rischi di disomogeneità applicativa nei diversi Stati membri. È innegabile, inoltre, che tale rischio si presenti anche rispetto alla possibilità di utilizzare il sistema per la ricerca degli autori di determinati reati individuati con riferimento al limite massimo edittale della pena, che muta a seconda delle legislazioni nazionali.

---

[48] Sulla necessità di non imporre un divieto assoluto delle *facial recognition technology*, ma di puntare piuttosto su «*certain bans, limitations, regulations and monitoring*» v. V.L. Raposo, *Ex machina: preliminary critical assessment of the European Draft Act on artificial intelligence*, in *International Journal of Law and Information Technology*, 2022, n. 30, p. 96.

[49] L'elenco è riportato all'allegato II-Elenco dei reati di cui all'articolo 5, paragrafo 1, primo comma, lettera h), punto iii): «terrorismo, tratta di esseri umani, sfruttamento sessuale di minori e pornografia minorile, traffico illecito di stupefacenti o sostanze psicotrope, traffico illecito di armi, munizioni ed esplosivi, omicidio volontario, lesioni gravi, traffico illecito di organi e tessuti umani, traffico illecito di materie nucleari e radioattive, sequestro, detenzione illegale e presa di ostaggi, reati che rientrano nella competenza giurisdizionale della Corte penale internazionale, illecita cattura di aeromobile o nave, violenza sessuale, reato ambientale, rapina organizzata o a mano armata, sabotaggio, partecipazione a un'organizzazione criminale coinvolta in uno o più dei reati elencati sopra».

L'utilizzo da parte delle forze dell'ordine dei sistemi di identificazione biometrica remota in tempo reale richiede comunque l'autorizzazione preventiva dell'autorità giudiziaria o di un'autorità amministrativa indipendente dello Stato membro, fatta salva, in casi di urgenza, la possibilità di una convalida *ex post* entro le ventiquattro ore successive all'uso del sistema, dovendosi, in mancanza, procedere immediatamente all'eliminazione e alla cancellazione di dati e risultati (art. 5, par. 3). L'attività, però, può essere autorizzata dall'autorità competente solo all'esito di un positivo vaglio di necessità e proporzionalità dell'uso del sistema di identificazione biometrica per il conseguimento di uno degli obiettivi per cui è consentita dal Regolamento e solo se l'autorità di polizia abbia eseguito il *Fundamental Rights Impact Assessment* (FRIA) (art. 5, par. 2)[50].

È possibile, invece, il ricorso a sistemi di identificazione biometrica remota a posteriori, se si tratta di attività "ad alto rischio", ma solo al fine di ricercare una persona condannata o sospettata di aver commesso un grave reato. Tale uso deve comunque essere limitato a quanto strettamente necessario per l'indagine su uno specifico reato ed è comunque soggetto all'autorizzazione dell'autorità competente, la cui decisione è vincolante e soggetta al controllo giurisdizionale, per l'uso di tale sistema, tranne quando è utilizzato per l'identificazione iniziale di un potenziale sospettato sulla base di fatti oggettivi e verificabili direttamente connessi al reato (art. 26, par. 10).

Il Regolamento (UE) ha, infine, inserito anche una regola di utilizzabilità degli *output* dei sistemi di identificazione biometrica, escludendo che possano da soli fondare una decisione che produca effetti giuridici negativi in capo a una persona (artt. 5, par. 3; 26, par. 10).

Un'ampia gamma di sistemi di IA "ad alto rischio" è autorizzata nell'ambito delle attività di prevenzione e contrasto al crimine rimanendo soggetta a una serie di requisiti e obblighi per la diffusione e l'uso all'interno dell'UE. Tuttavia, in vista del potenziale di tali strumenti in uso alle forze di polizia, è prevista una "procedura d'emergenza" che consente alle autorità di contrasto, in situazioni di urgenza, dovute a eccezionali motivi di ordine pubblico o di specifico, concreto e imminente pericolo per la vita o l'integrità fisica delle persone, di utilizzare un sistema di intelligenza artificiale "ad alto rischio" che non abbia superato la procedura di valutazione della conformità (art. 46, par. 2).

Diversamente, l'*AI Act* non sembra aver recepito l'invito del Parlamen-

[50] Si rinvia a nt. 14.

to a vietare l'impiego di sistemi di IA finalizzato all'emissione di decisioni giurisdizionali in materia penale[51].

Nell'ambito dell'amministrazione della giustizia, sono infatti classificati "ad alto rischio" i sistemi di intelligenza artificiale destinati a essere utilizzati da un'autorità giudiziaria nella ricerca e nell'interpretazione dei fatti e del diritto, nonché nell'applicazione della legge a casi simili (allegato III, par. 8, lett. a). Rientra in tale categoria l'impiego di metodi computazionali nell'ambito delle decisioni penali, che rimane così assoggettato al solo controllo di conformità affidato alle stesse aziende sviluppatrici[52].

Alla normativa dell'Unione europea sull'intelligenza artificiale, pur a fronte del frammentario e incompleto sviluppo, può di sicuro riconoscersi un primo apprezzabile risultato: il riconoscimento dell'imprescindibilità dell'intervento umano, che mantenendo il controllo sull'algoritmo, consente l'utilizzo di un nuovo metodo sinergico uomo-macchina idoneo a incrementare efficienza e qualità delle decisioni giurisdizionali.

---

[51] *Risoluzione del Parlamento europeo del 6 ottobre 2021 sull'intelligenza artificiale nel diritto penale e il suo utilizzo da parte delle autorità di polizia e giudiziarie in ambito penale*, cit., sulla quale si rinvia al par. 3.1.

[52] Per S.F. Schwemer, L. Tomada, T. Pasini, *Legal AI Sistems in the EU's proposed Artificial Intelligence Act*, in *Proceedings of the Second International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace*, 2021, LegalAIIA, «*it is noteworthy that the proposal does not follow a rights-based approach, which would, e.g., introduce new rights for individuals that are subject to decisions made by AI systems. Instead, it focuses on regulating providers and users of AI systems in a product regulation-akin manner*».

PROCESSO PENALE E DECISIONI ALGORITMICHE: GLI STRUMENTI DI VALUTAZIONE DEL RISCHIO

di Alessia Di Domenico

1. *Strumenti di* risk assessment *innovativi e prospettive d'oltreoceano*

Nell'era dell'intelligenza artificiale (di seguito, anche "IA"), assistiamo alla diffusione ormai inarrestabile di algoritmi predittivi che promettono di guidarci, anche nel delicato contesto dell'accertamento penale, verso l'orizzonte di una decisione più "oggettiva" e "giusta" [1].

Tra le possibili applicazioni dell'IA nell'ambito della giustizia penale, interessanti prospettive sembrerebbero arrivare dagli Stati Uniti con gli *algorithmic risk assessment tools* (che chiameremo anche "strumenti algoritmici di valutazione del rischio") [2], sistemi attuariali [3] che valutano il rischio

[1] Sul tema, cfr. diffusamente F. Basile, I*ntelligenza artificiale e diritto penale: quattro possibili percorsi di indagine*, in *Diritto Penale e Uomo*, 29.09.2019, disponibile su www.dirittopenaleuomo.org; D. Polidoro, *Tecnologie informatiche e procedimento penale: la giustizia penale "messa alla prova" dall'intelligenza artificiale*, in *Arch. pen.*, *(Web)*, 3, 2020, disponibile su www.archiviopenale.it; M. Amisano, *Prevedere – e non predire attraverso gli algoritmi e le loro insidie*, in *Arch. Pen. (Web)*, 2, 2022, disponibile su www.archiviopenale.it; G. Canzio, *Intelligenza artificiale e processo penale*, in *Cass. pen.*, 3, 2021, p. 797 ss.; M. Caterini, *Il giudice penale robot*, in *Leg. pen.*, 19.12.2020, disponibile su www.lalegislazionepenale.eu.

[2] Per una prima analisi, cfr. M. Hamilton, *Evaluating Algorithmic Risk Assessment*, in *New Crim. L. Rev.*, 2, 2021, p. 156 ss.; R.F. Lowden, *Risk Assessment Algorithms: The Answer to an*

di ricaduta criminale ed individuano i bisogni criminogeni del reo, e che potrebbero dunque essere un utile supporto all'attività giurisdizionale.

In particolare, l'incredibile diffusione di questi strumenti nel sistema d'oltreoceano ci pone di fronte alla necessità di considerarne punti di forza e criticità, per comprendere se simili *tools* potrebbero trovare spazi di operatività nel nostro ordinamento, considerando eventuali rischi di collisione con le garanzie del giusto processo e con la tutela dei diritti fondamentali dell'imputato.

Di fatto anche in Italia, con la rapida evoluzione che ha attraversato l'IA negli ultimi anni, si discute ormai spesso dell'opportunità di utilizzare sistemi di *machine learning* nell'ambito della valutazione prognostica del rischio di recidiva e di pericolosità sociale. Più precisamente, il dibattito sulle prospettive degli strumenti di *risk assessment* è sorto a seguito nel famoso caso *Loomis* [4], deciso dalla Corte Suprema del Wisconsin nel 2016. Tale vicenda, che discuteva dell'applicazione di *risk assessment tools* nell'ambito della misura di *probation* e della relativa possibile violazione del diritto dell'imputato ad un giusto processo [5], è diventata particolarmente nota anche in Italia, ed ha alimentato un nuovo interesse, da parte della dottrina [6]

Inequitable Bail System?, in *NCJOLT*, 19, 2018, p. 221 ss.; B.L. Garrett, J. Monahan, *Judging Risk*, in *Calif. Law Rev.*, 108, 2020, p. 439 ss.; A. Christin, *Risk-Assessment Tools in the U.S. Criminal Justice System: Construction, diffusion, and uses*, in *Data & Civil Rights*: *A New Era of Policing and Justice, Washington D.C.*, 27.10.2015; R.A. Berk, *Artificial Intelligence, Predictive Policing, and Risk Assessment for Law Enforcement*, in *Annu. Rev. Crimino.*, 37, 2021, p. 209 ss.

[3] Si tratta di sistemi "attuariali" in quanto, allineandosi ai metodi della matematica attuariale, calcolano la probabilità di eventi incerti utilizzando regole statistiche. Per un approfondimento su questa tipologia di strumenti nell'ambito della giustizia penale, cfr. B.E. Harcourt, *Against prediction: Profiling, policing, and punishing in an actuarial age*, University of Chicago Press, Chicago, 2007, p. 39 ss.

[4] *State v. Loomis*, 881 N.W.2d 749 (2016).

[5] In particolare, secondo la difesa di Loomis, l'uso dello strumento di valutazione del rischio COMPAS nel giudizio di determinazione della pena aveva violato il diritto all'equo processo sotto tre profili: 1) il diritto ad essere condannato ad una determinata pena sulla base di informazioni accurate, delle quali non si poteva disporre in quanto coperte da diritti di proprietà industriale; 2) il diritto di essere condannato ad una pena individualizzata; 3) l'uso improprio del dato di genere nella determinazione della pena. La Corte Suprema ha però ritenuto che l'uso di COMPAS non avesse violato il diritto di *Loomis* all'equo processo, ritenendo COMPAS un mezzo di calcolo affidabile e osservando soprattutto come gli fosse stato attribuito un peso minimo nella decisione.

[6] V., a titolo esemplificativo, M. Gialuz, *Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei* risk assessment tools *tra Stati Uniti ed Europa*, in *Dir. pen. cont.*, 29.05.2019, disponibile su www.archiviodpc.dirittopenaleuomo.org; S. Quattrocolo, *Quesiti nuovi e soluzioni antiche? Consolidati paradigmi normativi vs rischi e paure della giustizia digitale*

e dell'opinione pubblica, verso le opportunità e i pericoli connessi a tali tecnologie.

Nei successivi passaggi, si cercheranno quindi di individuare sinteticamente le principali caratteristiche di questi strumenti e le modalità con cui questi vengono utilizzati nel sistema statunitense, al fine di valutare possibili scenari che potrebbero aprirsi anche nel nostro ordinamento e delineare alcuni principi che si ritengono irrinunciabili per il loro utilizzo.

2. *La "giustizia attuariale" nel sistema statunitense*

Le ragioni di un crescente interesse verso gli strumenti di IA in esame ci sembrano evidenti: tali algoritmi predittivi promettono infatti di effettuare previsioni molto accurate, grazie ad indagini statistiche che godono della disponibilità di grandi quantità di dati [7].

Come si anticipava, il tema della giustizia predittiva assume una nuova, inedita, dimensione con i rapidi progressi che hanno attraversato le scienze computazionali negli ultimi anni; eppure, è bene precisare che l'utilizzo di previsioni algoritmiche e metodi statistici non è una vera e propria novità nel settore della giustizia penale, o quantomeno non lo è in alcuni ordinamenti come quello statunitense. Di fatto, negli Stati Uniti la c.d. "giustizia attuariale" ha radici ben più profonde, e da oltre un secolo si ricorre a metodi statistici di valutazione del rischio, in grado di prevedere l'avvenimento di eventi incerti come la reiterazione del reato [8].

I *risk assessment tools*, nel sistema statunitense, hanno dunque attraver-

---

"predittiva", in *Cass. pen*., 4, 2019, p. 1756 ss.; L. Maldonato, Risk and need assessment tools *e riforma del sistema sanzionatorio: strategie collaborative e nuove prospettive*, in G. Di Paolo, L. Presacco (a cura di), *Intelligenza artificiale e processo penale: indagini, prove, giudizio*, Editoriale Scientifica, Trento, 2022, p. 137 ss.; D. Zingales, Risk assessment*: una nuova sfida per la giustizia penale?*, in *Diritto penale e uomo*, 9.12.2021, disponibile su www.dirittopenaleuomo.org; A.M. Maugeri, *L'uso di algoritmi predittivi per accertare la pericolosità sociale: una sfida tra* evidence based practices *e tutela dei diritti fondamentali*, in *Arch. pen. (Web)*, 1, 2021, disponibile su www.archiviopenale.it.

[7] In questi termini, cfr. D. Polidoro, *Tecnologie informatiche e procedimento penale*, cit., p. 4; più in generale, sul tema dell'elaborazione di dati da parte di sistemi di IA, cfr. C. Snjiders, U. Matzat, U.D. Repis, *Big Data: Big Gaps of Knowledge in the Field of Internet Science*, in *Int. J. Inf. Sci.*, 7, 2012, p. 1.

[8] Per un approfondimento sui primi importanti sviluppi della giustizia attuariale nel sistema statunitense, v. H. Kemshall, *Risk, Actuarialism, and Punishment*, in *Oxford Criminology and Criminal. Justice*, 31.08.2017, disponibile su www.oxfordre.com, pp. 2-3; B.E. Harcourt, *Against prediction*, cit., p. 39.

sato diverse "generazioni", passando dall'utilizzo di procedure più informali e soggettive all'applicazione di metodologie attuariali innovative, che sfruttano algoritmi di *simple machine learning*[9].

Nell'impossibilità, in questa sede, di ripercorrere in maniera esaustiva l'evoluzione degli strumenti di valutazione del rischio negli Stati Uniti, si cercherà di soffermarsi sulle principali caratteristiche degli strumenti di ultima generazione, cercando di comprendere, anche sul piano tecnico, come venga concretamente valutato il rischio di recidiva.

3. *Costruzione dell'algoritmo e possibili spazi applicativi*

È possibile distinguere alcuni passaggi comuni nella costruzione di uno strumento di valutazione del rischio di natura attuariale, che si riassumeranno brevemente di seguito[10].

La prima fase è costituita dalla raccolta di dati che alimenteranno l'algoritmo. In particolare, la "popolazione di base" scelta nell'ambito dei *risk assessment tools*, e dunque l'*input* di tali strumenti, è costituita da dati di individui che nel passato sono stati arrestati o condannati (e dunque supervisionati in carcere o nell'ambito delle *corrective measures*).

Un ulteriore passaggio preliminare è l'individuazione dei c.d. "fattori di rischio" o "fattori predittivi". In particolare, dovranno individuarsi quali fattori, tra i dati raccolti, potranno contribuire ad aumentare o diminuire il rischio di recidiva, stabilendo il grado di correlazione di ogni variabile con il rischio da valutare. Ogni strumento di valutazione del rischio, a seconda degli obiettivi da raggiungere e delle scelte del programmatore, può prendere in considerazione fattori di rischio diversi: tra quelli più comuni, vengono annoverati il sesso, l'età, il livello di scolarizzazione, i precedenti penali ed i precedenti contatti con le forze dell'ordine, la presenza di un certo tasso di criminalità nella rete di conoscenze, l'eventuale abuso di sostanze stupefacenti o alcoliche[11].

---

[9] Sulle diverse generazioni di *risk assessment*, ed in particolare sulla distinzione tra approccio clinico ed attuariale, cfr. A. Giannini, *Lombroso 2.0: on AI and Predictions of Dangerousness in Criminal Justice*, in *Reveu Internationale de Droit Penal*, 92, 2021, p. 184; A. Andrews, J. Bonta, J. Stephen Wormith, *The recent past and near future of risk and/or need assessment*, in *Crime Delinq.*, 52, 1, 2006, p. 7.

[10] Per ulteriori approfondimenti, cfr. J.M. Eaglin, *Costructing Recividism Risk*, in *Emory Law J.*, 59, 2017, p. 72.

[11] Sulla scelta dei fattori di rischio, cfr. anche B.L. Garrett, *Judging Risk*, cit., p. 448 ss.; S. Bornstein, *Antidiscriminatory Algorithms*, in *Ala. L, Rev.*, 70, 2018, p. 519 ss.

Una volta terminato tale momento preliminare, lo strumento potrà calcolare in termini probabilistici il rischio di reiterazione del reato e valutare la pericolosità sociale del soggetto esaminato, confrontando i dati messi a disposizione nel caso concreto con quelli di individui con caratteristiche simili. Il risultato ottenuto verrà poi tradotto in una "classe di rischio" (es. basso, medio, alto), in cui verrà collocato l'individuo valutato dallo strumento [12].

Nel sistema statunitense, i modelli algoritmici in esame sono stati implementati in tre principali ambiti e fasi procedimentali [13]: 1) nella fase *pretrial*, in sostituzione della misura del *cash bail*; 2) nel *sentencing*, per le valutazioni relative alla determinazione della pena (nello specifico, il risultato algoritmico viene inserito nel *presentence investigation report* redatto dal *probation officer*); 3) nella fase esecutiva, per l'elaborazione del trattamento penitenziario e per valutare l'applicazione delle misure di *probation* e di *parole*.

4. *Le ragioni di una crescente diffusione e le criticità dell'algoritmo predittivo*

Interrogandoci sulle principali ragioni che hanno portato alla diffusione dei *risk assessment tools* nel sistema statunitense, possono individuarsi, anzitutto, esigenze di maggiore obiettività e neutralità [14] della decisione penale [15] in contrasto alle pratiche discriminatorie e ai *bias* che inevitabilmente possono influenzare le valutazioni del giudice.

Allo stesso tempo, anche il progredire delle scienze tecniche e dei metodi statistici ha contribuito ad alimentare il "movimento attuariale" e l'evoluzione delle tecniche di *risk assessment*, conducendo ad una nuova fiducia verso la possibilità di conoscere "scientificamente" il criminale e di prevedere il suo comportamento [16].

---

[12] Su quest'ultima fase, v. inoltre M. Henry, *Risk assessment: explained*, in *The Appeal*, 25.5.2019, www.theappeal.org.

[13] A. Christin, *Risk-Assessment Tools in the U.S. Criminal Justice System*, cit., p. 4.

[14] Sul tema, v. anche M. Bagaric, J. Svilar, M. Bull, D. Hunter, N. Stobbs, *The solution to the pervasive bias and discrimination in the criminal justice system: transparent and fair artificial intelligence*, in *Am. Crim. L. Rev.*, 59, 2021, p. 96 ss.

[15] Simili tecniche si sono diffuse anche nell'ambito della c.d. "*predictive policing*", per contrastare le discriminazioni nell'operato delle forze dell'ordine. Sul punto, cfr. E. Pietrocarlo, Predictive Policing; *criticità e prospettive dei sistemi di identificazione dei potenziali criminali*, in *Sist. pen.*, 28.09.2023, disponibile su www.sistemapenale.it, p. 18.

[16] Come evidenziato anche da B.E. Harcourt, *Against prediction*, cit., p. 174.

Da ultimo, la profonda crisi delle carceri statunitensi (si fa riferimento, in particolare, al problema del sovraffollamento carcerario) ha alimentato un crescente interesse verso tecnologie "razionali ed efficienti", in grado di sopperire alla mancanza di personale e di migliorare la gestione della popolazione carceraria [17].

Tuttavia, sono molteplici le perplessità che emergono dall'utilizzo degli strumenti predittivi in esame; tali criticità sono state più volte messe in luce dalla dottrina nordamericana, e riguardano l'effettiva imparzialità di tali sistemi e l'attendibilità degli *outputs* (suscettibili di errori e spesso basati su *inputs* discriminatori) [18] nonché la loro compatibilità con i principi del giusto processo sanciti nel Quinto e nel Quattordicesimo Emendamento della Costituzione americana [19].

Tali rischi emergono chiaramente dall'analisi di diverse recenti pronunce giurisprudenziali americane. Di seguito si fornirà qualche esempio, utile a comprendere in che modo la giurisprudenza abbia risposto all'implementazione di *risk assessment tools* nelle diverse fasi del procedimento penale.

5. *Alcuni recenti pronunce americane: l'opacità dell'algoritmo e la discrezionalità del giudice*

Al di là dell'emblematico caso *Loomis*, che si è reso già protagonista di molti dibattiti e sul quale quindi non ci soffermeremo, vi sono numerose pronunce che affrontano direttamente i principali problemi connessi all'utilizzo di strumenti di valutazione del rischio nel processo penale.

In particolare, una delle principali preoccupazioni discusse dalla giuri-

---

[17] In questo senso, cfr. N. James, *Risk and Needs Assessment in the Criminal Justice System*, in *Congressional Research Service*, 24.07.2015, disponibile su www.everycrsreport.com; K. Hannah-Moffat, *Actuarial Sentencing: An Unsettled Proposition*, in *Justice Q.*, 30, 2, 2013, p. 273.

[18] L. Rodriguez, *All data is not credit data: closing the gap between the fair housing act and algorithmic decisionmaking in the lending industry*, in *Colum. L. Rev.*, 120, 7, 2020, p. 1844 ss.; T. Douglas, J. Pugh, I. Singh, *Risk Assessment Tools in Criminal Justice and Forensic Psychiatry: The Need for Better Data*, in *Eur. Psychiatry*, 42, 2017, p. 134 ss.; S. Bornstein, *Antidiscriminatory Algorithms*, cit., p. 519; in dottrina italiana, sul punto, cfr. V. Manes, *L'oracolo algoritmico e la giustizia penale, al bivio tra tecnologia e tecnocrazia*, in U. Ruffolo (a cura di), *Intelligenza artificiale, Il diritto, i diritti, l'etica*, Giuffrè, Milano, 2020, p. 12.

[19] In argomento, cfr. M. Brenner, J.S. Gersen, M. Haley, M. Lin, A. Merchant, R.J. Millett, S.K. Sarkan, D. Wegner, *Constitutional Dimensions of Predictive Algorithms in Criminal Justice*, in *Harvard Civil Rights – Civil Liberties Law Review*, 55, 2020, p. 267; A.Z. Huq, *Constitutional Rights in the Machine-Learning State*, in *Cornell Law Rev.*, 105, 2020, p. 1875.

sprudenza riguarda la mancanza di trasparenza dello strumento, dovuta principalmente alla tutela dei diritti di proprietà intellettuale sull'algoritmo. L'opacità del *risk assessment tool* rischia infatti di collidere con il diritto dell'imputato ad un giusto processo, come emerge ad esempio nella pronuncia *People v. Canedo*[20]. Nella *concurring opinion*, infatti, si afferma che l'imputato ha diritto ad essere giudicato sulla base di «*accurate information*», e di contestare le informazioni contenute nel *Presentence Investigation Report* (compreso il risultato fornito dallo strumento di valutazione del rischio). Fatta questa premessa, «*it is unclear (…) what it might mean to measure the accuracy of a prediction about an individual's future conduct and how that prediction might be challenged without knowing how it was formulated*». Nella pronuncia si ritiene che la natura proprietaria dell'algoritmo possa inevitabilmente celare discriminazioni e pregiudizi perpetrati dallo strumento: «*without knowing what the algorithm is, it is difficult to know whether and how race, class, and other personal factors influence a potentially biased score*».

L'evidente problema dell'opacità dell'algoritmo, dovuto ad esigenze di protezione dei diritti di proprietà industriale, non appare tuttavia di facile risoluzione al cospetto di molteplici algoritmi predittivi che vengono attualmente utilizzati negli Stati Uniti, tra cui il noto *risk assessment tool* COMPAS, di proprietà della società Equivant. A titolo esemplificativo, la sentenza *Rayner v. New York State Department of Corrections and Community Supervision*[21], decisa dalla Corte Suprema della Contea di Albany (New York) nel 2023, riconosce la problematicità della mancanza di trasparenza del *tool*, ma non accoglie comunque la richiesta di *disclosure* sullo strumento da parte della difesa. Secondo la Corte, infatti, «*the value of the risk and needs assessment technology embodied in COMPAS-NY, both to Equivant and to potential competitors, is self-evident and its disclosure would almost assuredly complicate, if not compromise equivant's competitive position in New York (and perhaps in other states)*».

Oltre al problema dell'opacità del sistema di IA, si discute molto, a livello giurisprudenziale, sul peso effettivo che dovrebbe avere l'algoritmo predittivo nella decisione, e dunque sul difficile rapporto tra utilizzo di algoritmi e discrezionalità giudiziaria. A tal proposito, si ritiene utile segnalare la pronuncia *DeWess v. State*[22], emanata dalla Corte Suprema dell'In-

[20] *People v. Canedo*, 507 Mich. 1029 (2021).

[21] *Rayner v. New York State Department of Corrections and Community Supervision*, 81 Misc.3d 281 (2023).

[22] *DeWees v. State*, 180 N.E.3d 261 (2022).

diana nel 2022. Nella sentenza, si chiarisce che l'utilizzo di *risk assessment tool* a supporto della decisione giudiziaria (nel caso di specie, nelle decisioni relative alla misura del *bail*) non deve risultare in un «*change of [the] judicial flexibility*». Di fatto, l'unico obbligo che viene imposto ai giudici è quello di considerare il risultato dello strumento, e di conseguenza la decisione finale potrà sempre discostarsi dall'*output*. Secondo la Corte Suprema, «*while highly useful and important for trial courts to consider as a broad statistic tool, an evidence-based assessment like IRAS is no substitute for a judicial determination of bail but is merely supplemental to all other evidence informing the trial court's decision*». In altre parole, all'algoritmo viene conferito un peso non determinante, e in nessun modo vincolante per il giudice, che dovrà valutare l'*output* fornito insieme ad altri elementi di prova e potrà liberamente allontanarsi dal "suggerimento algoritmico".

6. *Brevi conclusioni: alcuni principi-guida per un "giusto" utilizzo degli strumenti di valutazione del rischio*

Cercando di trarre delle brevi conclusioni, anche alla luce delle problematiche e delle preoccupazioni sollevate dalla giurisprudenza negli Stati Uniti, sembrerebbe opportuno individuare alcuni principi-guida, che dovrebbero assicurare, senza dover rinunciare aprioristicamente all'utilizzo degli strumenti in esame, il rispetto dei diritti costituzionali dell'imputato e dei principi del giusto processo.

Anzitutto, sembra opportuno aprire la strada a strumenti algoritmici *trasparenti*, in grado di assicurare l'"*accessibilità*" delle informazioni elaborate e la "*verificabilità*" del risultato algoritmico[23]. Il diritto della difesa di conoscere tutti gli elementi che hanno portato alla decisione va dunque

[23] Sul requisito della "trasparenza" dei sistemi di IA si è recentemente soffermato, a livello europeo, anche il Regolamento europeo sull'intelligenza artificiale (c.d. "AI Act"): in particolare, all'art. 13 («*Trasparenza e fornitura di informazioni ai deployer*»), si prevede che i sistemi di IA ad alto rischio siano «progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire ai *deployer* di interpretare l'*output* del sistema e utilizzarlo adeguatamente». Più in generale, può dirsi che il problema della mancanza di trasparenza e di accessibilità del sistema di IA è una delle questioni più discusse dalla dottrina, anche con specifico riferimento ai *risk assessment tools*: cfr., a titolo esemplificativo, K.J. Strandburg, *Rulemaking and Inscrutable Automated Decision Tools*, in *Colum. L. Rev.*, 119, 2019, p. 1851; M. Hamilton, *Risk-Needs Assessment*, cit., p. 240 ss.; S. Quattrocolo, *Quesiti nuovi e soluzioni antiche?*, cit., p. 1761; A. Maugeri, *L'uso di algoritmi predittivi per accertare la pericolosità sociale*, cit., p. 20.

considerato prevalente rispetto alla tutela dei diritti di proprietà intellettuale sull'algoritmo. L'opacità del *tool* rischierebbe di incidere gravemente sui principi del contraddittorio, in quanto renderebbe impossibile alla difesa contestare la validità e l'attendibilità dei metodi alla base dello strumento, nonché di evidenziare eventuali errori dell'algoritmo e possibili discriminazioni attuate, anche indirettamente. Tutte le caratteristiche e le informazioni elaborate dallo strumento devono invece essere "cristallizzate" e messe a disposizione del difensore. Naturalmente, il problema dell'opacità dell'algoritmo è di più facile risoluzione di fronte ad algoritmi di *machine learning* relativamente semplici, come può essere lo strumento COMPAS e tutti i principali strumenti di valutazione del rischio utilizzati attualmente nel sistema statunitense: la problematica potrebbe infittirsi, nei prossimi anni, al cospetto di algoritmi di *machine learning* più sofisticati, la cui mancanza di trasparenza potrà non essere dovuta solo alla natura proprietaria dell'algoritmo, ma soprattutto alla complessità dello strumento.

Un ulteriore principio che viene evidenziato dalla dottrina americana è quello della "*coerenza*" [24] delle valutazioni operate dal *risk assessment tool*: deve quindi impedirsi che il dinamismo del sistema di IA ed il suo apprendimento automatico possa condurre lo strumento a fornire, a parità di *input* e di elementi considerati dallo strumento, risultati diversi in momenti diversi. Soggetti con caratteristiche simili potrebbero quindi essere valutati diversamente dallo strumento a seconda del momento in cui il calcolo viene effettuato, alla luce dell'esperienza nel frattempo maturata dal *tool*.

Da ultimo, vi sono i principi di "*non vincolabilità*" e "*non esclusività*" [25]. In particolare, il risultato dell'algoritmo non può considerarsi vincolante per il giudice, e non deve in alcun modo limitare il suo libero convincimento; al tempo stesso, l'*output* non può ritenersi da solo sufficiente a fondare la decisione giudiziaria, ma deve essere sempre corroborato da ulteriori

---

[24] In questo senso, v. J. Villasenor, V. Foggo, *Artificial Intelligence, Due Process, Criminal Sentencing*, in *Mich. St. L. Rev.*, 2, 2020, p. 347; v., inoltre, D. Zingales, *Risk assessment*, cit., p. 18.

[25] Tale principio si allinea a quanto affermato, a livello europeo, dall'art. 11 della Direttiva 2016/680/UE, che ravvisa il divieto di decisioni basate unicamente su un trattamento automatizzato (compresa la profilazione) «che producano effetti giuridici negativi o incidano significativamente sull'interessato, salvo che siano autorizzate dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento e che prevedano garanzie adeguate per i diritti e le libertà dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento». Sul tema, cfr. A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *BioLaw – Rivista di BioDiritto*, 1, 2019, p. 80.

elementi di prova[26]. Ciò assicurerà l'indipendenza del giudice di fronte all'algoritmo predittivo.

I principi appena enucleati dovranno tenersi in considerazione nel valutare l'opportunità di implementare strumenti di valutazione del rischio anche nell'ordinamento italiano, immaginandone possibili applicazioni nella fase cautelare, nella determinazione della pena ed in fase esecutiva (similmente a quanto è avvenuto negli Stati Uniti).

L'esperienza statunitense, così come brevemente accennata in questa sede, ci consente quindi di essere consapevoli dei possibili scenari che potrebbero profilarsi anche nel nostro ordinamento. I principi sopra delineati, alla luce delle questioni emerse dalla giurisprudenza americana, si ritengono così irrinunciabili anche all'orizzonte di un'esplosione della "giustizia attuariale" anche in Europa, e della possibile diffusione di strumenti di valutazione del rischio nel nostro sistema penale.

[26] In questi termini, v. M. GIALUZ, *Quando la giustizia penale incontra l'intelligenza artificiale*, cit., p. 21.

INTELLIGENZA ARTIFICIALE E FUTURO

PROBLEM SOLVING CREATIVO E INTELLIGENZA ARTIFICIALE

di Daniela Grieco e Gary Charness

1. *Introduzione*

La creatività è una caratteristica fondamentale dell'intelligenza umana e rappresenta un'abilità di vitale importanza non solo quando si devono completare task complessi in ambito lavorativo, ma anche quando la vita quotidiana ci pone davanti sfide che richiedono soluzioni innovative. La creatività si fonda su un insieme articolato di capacità che comprende «l'associazione di idee, il ricordo, la percezione, il pensiero analogico, la ricerca di uno spazio-problema strutturato [...]»[1]. Perché un risultato sia creativo, si richiede che vada oltre ciò che è considerato "standard".

L'intelligenza artificiale generativa è una delle innovazioni recenti più significative nel campo della tecnologia. È quella branca dell'informatica che simula i processi che l'intelligenza umana usa per la risoluzione dei problemi e l'apprendimento, in modo che i programmi per computer possano eseguire questi processi al posto degli esseri umani. ChatGPT 3.5, lanciato nel novembre 2022, e le sue versioni successive, attingono a un'enorme quantità di dati e ricorrono al "machine learning" per rispondere a domande, produrre disegni o animazioni, conversare, scrivere saggi su qualsiasi argomento possibile, classificare oggetti mai visti prima, persi-

[1] Si veda a questo proposito M.A. Boden, *Creativity and artificial intelligence*, in *Artificial Intelligence*, 1998, 103, pp. 347-356.

no fare le olimpiadi di matematica o la peer-review di articoli scientifici[2]. ChatGPT si basa su un tipo di intelligenza artificiale generativa chiamata "Large Language Models" (LLM), che sfrutta enormi quantità di dati, implementa algoritmi in grado di apprendere da questi dati, e può contare su una potenza computazionale così elevata da consentire di portare a termine l'intero processo in pochi istanti.

Di fronte alle sorprendenti performance di ChatGPT, il timore che l'intelligenza artificiale possa sostituire le persone in tutte o in alcune delle loro mansioni si è diffuso rapidamente. Se, da un lato, è già emerso chiaramente che l'intelligenza artificiale sia in grado di fare alcune cose in maniera più efficiente e rapida rispetto alle persone (ad esempio sintetizzare testi e scrivere semplici codici di programmazione), tuttavia immaginare e quantificare quale potrebbe essere il suo rendimento nei task creativi è difficile. L'intelligenza artificiale è infatti programmata per elaborare le informazioni secondo una procedura specifica e ottenere un determinato risultato, e per sua caratteristica strutturale non può discostarsi dalle istruzioni ricevute: i LLM sono di fatto un esercizio statistico in cui le parole della richiesta fatta a ChatGPT vengono convertite in numeri, raggruppate con altre parole con significato simile e collegate ad altre parole secondo la frequenza "appresa" di queste associazioni, dove l'apprendimento avviene sulla base della probabilità osservata nelle associazioni di parole precedentemente osservate nei dataset che il modello ha a disposizione. Per questa ragione, i risultati che i LLM generano non possono essere spontanei o imprevedibili come ci si aspetta, invece, dalla creatività umana. Poiché l'associazione tra parole si basa su una certa probabilità, il risultato dei LLM non è deterministico e quindi non del tutto prevedibile. Tale prevedibilità del testo generato dipende dall'impostazione dei parametri, in particolare della c.d. "temperatura", che regola il grado di originalità, ma allo stesso tempo anche la casualità, del testo generato.

La creatività consiste nel trovare nuove soluzioni a problemi che altri potrebbero non aver ancora preso in esame, sovvertendo le regole, guardando una questione da prospettive diverse e "pensando fuori dagli schemi". I programmi per computer che utilizzano l'intelligenza artificiale sono codificati per raggiungere i risultati (esatti) che viene chiesto loro di ottenere. Questo è il motivo per cui sono generalmente utilizzati in attività in

[2] Per un'interessante discussione dei task in cui ChatGPT ha ottenuto ottimi risultati, si veda W. Dai, J. Lin, H. Jin, T. Li, Y.S. Tsai, D. Gašević, E.G. Chen, *Can large language models provide feedback to students? A case study on ChatGPT*, 2023. In Proceedings of the 23rd IEEE international conference on advanced learning technologies (pp. 323-325).

cui è necessario un elevato livello di precisione. La creatività è invece difficilmente riconducibile a un insieme di istruzioni o a una formula matematica, e rappresenta un processo umano aperto a molteplici interpretazioni, contributi e punti di vista.

Lo studio che presentiamo in questo articolo ha lo scopo di testare se l'intelligenza artificiale sia in grado di eguagliare o addirittura superare la performance degli esseri umani in task creativi con un diverso grado di "apertura". Come illustrato in un nostro precedente lavoro[3], «un problema è definibile come perfettamente chiuso quando il metodo per risolverlo è noto e replicabile […]», come avviene nei c.d. task "algoritmici"[4], si pensi per esempio a una equazione. Al contrario, un task si dice aperto quando al partecipante viene richiesto di inventare, trovare o scoprire qualcosa di totalmente nuovo"[5]. Questa classificazione ovviamente non esaurisce tutte le possibili dimensioni della creatività nei molteplici contesti in cui il problem solving creativo si applica. Tuttavia, il nostro contributo fornisce alcuni risultati preliminari su questo dibattito così attuale.

Lo studio si basa sulla stessa metodologia impiegata nei nostri articoli precedenti[6] nei quali, tramite una serie di esperimenti di laboratorio, si riproduce un determinato setting decisionale in un ambiente controllato e in modo semplificato, ma sufficientemente ricco da presentarne le caratteristiche fondamentali. In particolare, questo esperimento prevede che alcuni soggetti valutino il livello di creatività delle risposte a un task creativo precedentemente prodotte da altri esseri umani, oppure generate tramite ChatGPT.

La sezione 2 di questo articolo presenta l'esperimento e i suoi risultati. Nella sezione 3 sintetizziamo le ipotesi principali di un modello che interpreta tali risultati. La sezione 4 presenta un ulteriore esercizio in cui è l'intelligenza artificiale, e non le persone, a valutare la performance creati-

---

[3] Ci riferiamo a G. Charness, D. Grieco, *Creativity and incentives*, in *Journal of the European Economic Association*, 2019, 17(2), pp. 454-496.

[4] Per un approfondimento su questa tipologia di task e le loro peculiarità, si veda M.A. Collins, T.M. Amabile, *Intrinsic motivation and artistic creativity: The effects of naturally-occurring interest, affect, and involvement*, in *Annual Meeting of the Eastern Psychological Association*, Boston, 1992.

[5] Una prima classificazione che ha distinto tra task creativi aperti e chiusi si deve a K. Unsworth, *Unpacking creativity*, in *Academy of Management Review*, 2001, 26(2), pp. 289-297.

[6] Oltre all'articolo già citato nella nota 3, si vedano G. Charness, D. Grieco (2022). *Creativity and ambiguity tolerance*, in *Economics Letters*, 218, p. 110720, e G. Charness, D. Grieco, *Creativity and corporate culture*, in *The Economic Journal*, 2023, p. uead012.

va di individui e ChatGPT. Nella sezione 5, offriamo alcuni spunti di discussione e concludiamo.

2. *Descrizione dell'esperimento e risultati*

Come precedentemente illustrato, i task possono essere classificati a seconda del loro livello di apertura, collocandosi idealmente lungo un continuum che inizia con i task perfettamente chiusi, o "algoritmici" – dove l'accuratezza nel seguire le istruzioni consente di raggiungere l'unica soluzione possibile – e task perfettamente aperti, dove le soluzioni possono essere molteplici, variegate fra loro, e persino difficili da confrontare. Inoltre, nel caso dei task aperti, non è sufficiente seguire le istruzioni per assicurarsi che il risultato raggiunto sia effettivamente considerato creativo. Un maggiore grado di apertura del task implica quindi una minore probabilità di risolvere con successo il problema seguendo unicamente le istruzioni, e quindi comporta una maggiore varietà e originalità nelle possibili risposte. Ciò potrebbe influenzare il giudizio dei valutatori, poiché – in quanto chiamati a giudicare un prodotto creativo – è probabile che essi diano un peso notevole all'originalità del contenuto e che premino le risposte che si differenziano maggiormente dalle altre.

L'esperimento ha un design 2×2, con due trattamenti ("Chiuso" vs. "Aperto"), che differiscono nel grado di apertura del task creativo, e due agenti ("Umani" vs. Intelligenza Artificiale) che producono risposte al task, con quattro condizioni in totale: IA-Chiuso (IA-C), IA-Aperto (IA-A), Umani-Chiuso (U-C), Umani-Aperto (U-A). Il campione sperimentale è costituito da 251 individui, reclutati sulla piattaforma online Prolific, che mette in contatto ricercatori e potenziali partecipanti a studi e sondaggi. Ciascun individuo valuta sei risposte (tre prodotte dagli esseri umani e tre prodotte dall'intelligenza artificiale) estratte casualmente da un insieme di 160 risposte e presentate a loro volta in ordine casuale. In totale, abbiamo raccolto e analizzato 1506 valutazioni, e ciascuna risposta ha ricevuto almeno nove valutazioni indipendenti. I paragrafi seguenti descrivono il protocollo sperimentale in maniera più dettagliata.

Trattamento Chiuso. Nel Trattamento Chiuso, i soggetti valutano le risposte che 40 partecipanti hanno dato nell'esperimento di laboratorio al seguente task: "Scrivi una storia creativa utilizzando obbligatoriamente le seguenti parole: casa, zero, perdono, curva, rilevanza, mucca, albero, pianeta, anello, inviare." Questo task è classificato come "chiuso" (in termini

relativi, a differenza del task aperto descritto di seguito) perché prevede il vincolo di utilizzare obbligatoriamente dieci parole predefinite. Tale requisito rende il task un problema da "risolvere", ma comunque lo mantiene paragonabile al task aperto descritto di seguito.

Altrettante risposte al medesimo task sono state prodotte utilizzando l'intelligenza artificiale (nello specifico, ChatGPT 3.5), variando la temperatura da 0 a 2 (valore massimo possibile) e lasciando gli altri parametri al loro valore predefinito, ad eccezione del parametro di lunghezza del testo prodotto, che è stato aumentato rispetto al valore di default. I soggetti hanno effettuato la loro valutazione – assegnando un punteggio compreso tra 1 e 10 – senza sapere se le risposte fossero state prodotte da un essere umano o da una macchina, e hanno scritto un breve testo che motivasse il voto assegnato.

Trattamento Aperto. Come nel precedente trattamento, anche nel Trattamento Aperto i soggetti valutano le risposte che 40 partecipanti hanno dato al task: "Se avessi il talento di inventare qualcosa semplicemente pensandola, cosa creeresti e perché?". Questo task è classificato come "aperto" perché i soggetti sono completamente liberi di inventare senza alcun vincolo. Anche in questo caso, lo stesso numero di risposte è stato prodotto da esseri umani e da ChatGPT, e i soggetti valutatori hanno formulato il loro giudizio senza sapere chi avesse fornito la risposta. L'impostazione dei parametri di ChatGPT è rimasta invariata rispetto al Trattamento Chiuso, in modo da rendere i due trattamenti confrontabili tra loro.

L'analisi delle performance creative dei due tipi di agenti – umani e intelligenza artificiale – nei due task – chiuso e aperto – dà spunti di riflessione molto interessanti. I risultati mostrano infatti che la prestazione media degli esseri umani nel task chiuso è significativamente peggiore di quella dell'intelligenza artificiale. Tuttavia, questo risultato è completamente ribaltato nel task aperto, dove gli esseri umani raggiungono una performance media significativamente migliore.

Nel task chiuso, se si considerano i punteggi medi di ciascun soggetto attribuiti alle risposte dell'intelligenza artificiale (da qui in poi, nella descrizione dei risultati, abbrevieremo con IA) e degli umani, il punteggio medio dell'IA è pari a 6.136 in un range di punteggio che va da 1 a 10, ed è significativamente più alto di quello degli umani, pari in media a 5.331. Nel task aperto, il punteggio medio degli esseri umani è pari a 6.266, significativamente superiore a quello dell'IA, uguale in media a 5.108. La differenza (1.953) è altamente significativa, essendo pari a oltre 10 volte l'errore standard.

È interessante notare che la performance dell'intelligenza artificiale nel task chiuso non è correlata con il parametro "temperatura" che, lo ricordiamo, regola il livello di casualità con cui le parole vengono combinate nelle risposte generate, e rappresenta pertanto un parametro che regola l'originalità dell'output creativo. Per quanto riguarda il task aperto, i risultati mostrano una correlazione leggermente positiva quando la temperatura è inferiore o uguale a 1, dove 1 è il livello di default, e una forte correlazione negativa quando la temperatura è maggiore di 1. Dal momento che una temperatura elevata aumenta la casualità dell'associazione delle parole nella risposta, l'output creativo risulta contemporaneamente più originale, ma anche incoerente e poco chiaro, e questi incide negativamente sulle valutazioni di queste risposte.

Per avere una migliore comprensione delle ragioni per cui l'intelligenza artificiale non ha una buona performance nel task aperto, abbiamo condotto un'analisi del contenuto del testo che i valutatori hanno scritto per motivare il punteggio assegnato. Una percentuale non trascurabile di risposte nel task aperto è giudicata "non sufficientemente creativa": il 25,47% quando scritte da IA, contro il 21,41% quando scritte da esseri umani, mentre le percentuali si riducono rispettivamente al 13,54% e 15,62% nel caso del task chiuso.

Un'altra frazione consistente di risposte è stata classificata come "priva di senso": tale percentuale è molto più alta per le risposte generate dall'intelligenza artificiale in entrambi i task: 14,36% contro 3,79% nel task aperto, 11,98% contro il 5,47% nel task chiuso. È interessante notare che, sebbene la temperatura sia impostata in modo identico nei due task, la presenza di vincoli riduce la portata delle associazioni casuali nel task chiuso anche quando si imposta un livello di temperatura molto elevato; questo supporta ulteriormente la nostra distinzione tra task chiusi e aperti. Infatti, la necessità di rispettare i vincoli del task chiuso (in particolare, dovere usare obbligatoriamente le dieci parole fornite) restringe la gamma dei possibili output. Ciò non vale per il task aperto e questo si riflette sull'output prodotto e, a seguire, sulla valutazione che esso riceve.

Per quanto concerne le caratteristiche demografiche dei valutatori, non si osserva alcuna differenza nelle valutazioni medie di uomini e donne nel task chiuso. Nel task aperto, invece, notiamo che gli uomini hanno assegnato punteggi significativamente più bassi sia a ChatGPT che agli esseri umani rispetto alle donne. Sebbene non vi sia alcuna differenza nel livello di creatività quando si confrontano uomini e donne in termini di performance creativa, gli uomini sembrano essere più esigenti nel valutare la performance creativa di altre persone in task che richiedono un elevato conte-

nuto di innovatività. Questo risultato è in linea con alcuni lavori precedenti[7] che mostrano che gli uomini sono più concentrati sull'output creativo, mentre le donne tendono a preoccuparsi maggiormente del processo, favorendo la novità e l'originalità invece della fluidità o del grado di elaborazione del risultato finale.

Infine, i dati mostrano che né l'appartenenza a un determinato gruppo etnico, né il tempo impiegato per completare l'esperimento hanno alcun effetto sulle valutazioni assegnate.

3. *Il modello*

Per fare luce sui fattori che possono spiegare le differenze nelle performance creative di esseri umani e intelligenza artificiale, presentiamo ora un semplice modello che ha lo scopo di sintetizzare i meccanismi con cui gli esseri umani affrontano un task creativo rispetto all'intelligenza artificiale, distinguendo tra task con differente grado di apertura. I due parametri chiave del modello saranno poi stimati strutturalmente, mettendo in evidenza in quale misura esseri umani e intelligenza artificiale sono influenzati dall'apertura del task, e calcolando il peso relativo che la capacità di giudizio, propria soltanto degli esseri umani, assume rispetto al contenuto che si sarebbe prodotto seguendo pedissequamente le istruzioni del task.

Il cuore del modello è costituito da una equazione che evidenzia come l'agente che si approccia a un determinato task – non necessariamente creativo – sperimenti un "costo psicologico" nel momento in cui l'output che produce si discosti dall'output che si potrebbe ottenere seguendo esattamente le istruzioni. Tuttavia, il ruolo di tali istruzioni diventa via via meno rilevante quanto più il task è aperto, dal momento che l'apertura del task è associata alla presenza di molteplici soluzioni e alla produzione di un output maggiormente insolito. Per l'intelligenza artificiale, inoltre, variare il parametro della temperatura assegnando un valore più elevato amplifica questa distanza dalle istruzioni, poiché una temperatura maggiore aumenta la variabilità nell'output. In un task perfettamente chiuso, come un'equazione, è sufficiente seguire le istruzioni per completare perfettamente il task e raggiungere l'obiettivo. In un task perfettamente aperto, la performance creativa dell'intelligenza artificiale può invece arrivare a essere de-

[7] Si veda in particolare G.T. HILL, *Sex and gender differences in humour, creativity and their correlations*, in *Dissertations Abstracts International. Section A: Humanities and Social Sciences* 62 (2A), 2001, p. 473.

terminata dal valore della temperatura, che governa il grado di casualità delle associazioni tra parole. Tuttavia, come già emerso, aumentare la temperatura significa aumentare contemporaneamente originalità e "stranezza": una temperatura bassa dà combinazioni più consuete e meno creative, mentre l'aumento della temperatura induce combinazioni molto originali che potrebbero non avere senso, e di conseguenza non essere riconosciute come creative da chi è chiamato a valutarle. Mentre l'intelligenza artificiale non è in grado di capire fino a che punto l'originalità delle combinazioni possa essere compresa dagli altri, gli esseri umani riescono invece a farlo.

Per tenere conto di questa importante abilità, il modello include una seconda determinante della performance creativa, che cattura il ruolo del giudizio o della supervisione umana: gli esseri umani sono capaci di adeguare la loro performance creativa al contesto, andando oltre rispetto a quanto prescritto dalle istruzioni ma riuscendo al contempo a rimanere comprensibili per gli altri esseri umani, e operando quindi una sorta di "controllo di qualità". L'output creativo ottenuto è quindi una media ponderata tra l'output che si sarebbe prodotto semplicemente seguendo le istruzioni e il giudizio su ciò che è creativo, e che soprattutto può essere percepito come tale dalle altre persone.

La stima strutturale dei parametri del modello evidenzia, coerentemente con i risultati già menzionati sopra, che il giudizio umano ha rilevanza significativa solo nei task sufficientemente aperti. Ciò significa che, per task relativamente più chiusi, seguire le istruzioni è sufficiente e non sono necessari ulteriori aggiustamenti. Nel task aperto, invece, in base ai nostri calcoli il ruolo relativo del giudizio umano rispetto alle istruzioni pesa per circa il 35% del punteggio creativo medio degli esseri umani. Tale rilevanza aumenta restringendo l'analisi alle risposte dell'intelligenza artificiale prodotte settando la temperatura a un livello elevato, e confrontandole con le prestazioni umane: sebbene il punteggio medio sia molto simile, quando il livello di casualità delle associazioni generate dall'intelligenza artificiale aumenta, la capacità di giudizio degli esseri umani risulta ancora più rilevante, arrivando a contribuire per il 45% del loro punteggio creativo. La crescente casualità nelle risposte generate dall'intelligenza artificiale determina un aumento della diversità delle risposte ma, affinché la diversità diventi creatività, è necessario un aggiustamento che solo gli esseri umani al momento sono capaci di compiere.

In sintesi, le stime strutturali mostrano che, nel task aperto, il giudizio umano costituisce una componente rilevante della performance creativa umana, rappresentando così un ingrediente chiave del suo successo rispet-

to all'intelligenza artificiale. Questo tipo di giudizio non ha invece alcun ruolo nei task chiusi, nei quali i vincoli del task restringono già di per sé la gamma di possibili risposte senza rendere necessario alcun controllo a posteriori.

4. *Intelligenza artificiale come valutatore*

Un'ulteriore questione legata all'importanza del giudizio umano riguarda la capacità dell'intelligenza artificiale di valutare la performance creativa altrui. L'intelligenza artificiale al momento è già utilizzata per pratiche di selezione come lo screening dei curricula dei candidati, la valutazione dell'affidabilità creditizia dei clienti, la determinazione dei prezzi di alcuni beni. Il quesito generale che ci poniamo è se l'intelligenza artificiale sia davvero in grado di valutare le prestazioni umane, specie in task in cui l'intelligenza artificiale stessa non eccelle, come accade nel caso del nostro task aperto.

Per rispondere a questa domanda, abbiamo replicato l'esperimento chiedendo a ChatGPT di valutare le stesse 160 risposte precedentemente esaminate da individui. Come sopra, abbiamo due trattamenti (Chiuso vs. Aperto) e due possibili agenti (umani vs. IA) che producono risposte al task, con le stesse quattro condizioni in totale. Per mantenere il procedimento quanto più confrontabile con l'esperimento descritto nella Sezione 2, anche queste istruzioni chiedono a ChatGPT di estrarre sei risposte dalle 80 risposte presenti in ciascun insieme e di valutarle, senza sapere se siano state prodotte da un essere umano o da ChatGPT. Le estrazioni sono state ripetute fino a quando ciascuna risposta è stata valutata nove volte, come nel precedente esperimento con valutatori umani.

I risultati mostrano che i punteggi creativi assegnati dall'intelligenza artificiale agli esseri umani e all'intelligenza artificiale nelle diverse condizioni non differiscono in modo significativo tra loro, né nel task chiuso, né nel task aperto. È interessante notare che i punteggi assegnati dall'intelligenza artificiale tendono ad essere più generosi di quelli degli esseri umani e che gli errori standard – che misurano la variazione rispetto alla media, quindi la variabilità delle valutazioni – sono considerevolmente inferiori rispetto a quelli degli esseri umani. Inoltre, l'intelligenza artificiale evita i punteggi estremi (i voti 1-2 e 10 non vengono mai assegnati). Infine, non esiste alcuna correlazione significativa tra il punteggio assegnato dall'intelligenza artificiale e la temperatura, sia per il task chiuso che per il task aperto.

Come sopra, utilizziamo l'analisi del testo per comprendere le motivazioni che l'intelligenza artificiale associa ai suoi punteggi. Ciò che emerge è

che non solo l'intelligenza artificiale valuta pochissime risposte come "non sufficientemente creative", ma non descrive mai alcun contenuto della risposta come "insensato", indipendentemente dal fatto che la risposta sia stata prodotta dall'intelligenza artificiale o da un essere umano. Per quanto riguarda le risposte valutate come "non abbastanza creative", la percentuale di risposte nel task aperto è del 6,67% se scritte dall'intelligenza artificiale, contro il 4,44% se scritte da esseri umani, mentre tali percentuali si riducono ulteriormente, rispettivamente a 1.11% e 5.56%, nel caso del task chiuso. Questi risultati mostrano che non solo l'intelligenza artificiale non ha capacità di giudizio quando produce risposte creative, ma non è nemmeno in grado di rilevare questa mancanza di giudizio quando le valuta.

5. *Discussione e conclusioni*

I progressi dell'intelligenza artificiale stanno avvenendo così rapidamente da aver generato in molte categorie di lavoratori il timore che i suoi ulteriori sviluppi possano avere un impatto drasticamente negativo e permanente sulle loro condizioni lavorative. Tuttavia, nel momento in cui scriviamo è troppo presto per avere un'idea chiara del ruolo che l'ultima generazione di modelli di intelligenza artificiale avrà su lavoratori, produttività e mansioni.

I risultati del nostro studio si concentrano sul problem solving creativo e mostrano che, al momento, gli esseri umani hanno ancora una performance migliore dell'intelligenza artificiale in task creativi sufficientemente aperti, vale a dire task in cui i soggetti sono tenuti a trovare, inventare o scoprire nuove soluzioni. Questo risultato è ancora più interessante se pensiamo che i soggetti sperimentali coinvolti non si erano "auto-selezionati" rispetto al task, ossia hanno partecipato all'esperimento senza essere stati informati in anticipo sul tipo di task che avrebbero completato. Ci aspettiamo quindi che i risultati non possano che essere ancora più netti selezionando partecipanti altamente creativi, come potrebbe accadere coinvolgendo persone che vivono e lavorano in ambienti in cui la creatività è fondamentale, si pensi ad esempio a tutti coloro che lavorano nella c.d. "industria creativa".

Il nostro modello illustra il meccanismo della produzione creativa da parte tanto dell'intelligenza artificiale quanto degli esseri umani e assume che l'output creativo sia il risultato della media ponderata tra l'output che si produrrebbe semplicemente seguendo le istruzioni e l'aggiustamento che

si introduce formulando un giudizio su ciò che è creativo, e soprattutto che è percepito come creativo dalle altre persone. Le stime mostrano che la capacità delle persone di "aggiustare" l'output rispetto a quanto prescritto dalle istruzioni determina una percentuale compresa tra il 35% e il 45% del punteggio creativo degli umani nel task aperto, mentre non ha rilevanza nel task chiuso. Il giudizio umano sembra diventare ancora più rilevante quando a ChatGPT è richiesto di aumentare il livello di casualità delle associazioni tra parole, poiché ChatGPT risulta incapace di distinguere tra creatività e casualità e, di conseguenza, incapace di rilevare risposte in cui la sequenza di parole non ha senso.

Nel momento in cui scriviamo, l'intelligenza artificiale sembra richiedere ancora un intervento correttivo di supervisione umana piuttosto significativo quando si affrontano compiti altamente creativi, e questo attualmente compensa la sua maggiore efficienza e rapidità. Se è difficile pensare che l'intelligenza artificiale possa essere "addestrata" a essere creativa quanto gli esseri umani nel prossimo futuro, formare i lavoratori in modo che imparino a comprendere come essa funziona e reagisce agli stimoli potrebbe probabilmente migliorare la produzione creativa dell'intelligenza artificiale stessa.

I nostri risultati forniscono sostegno a favore delle enormi possibilità di sfruttare con successo le complementarità tra l'intuizione e il giudizio degli esseri umani, e la precisione e le capacità computazionali dell'intelligenza artificiale. Inoltre, si sottolinea l'importanza di investire nel "capitale creativo", in linea con quanto recentemente riscontrato in uno studio condotto in 16 Paesi europei[8]: per i lavoratori più esposti all'intelligenza artificiale, emerge che in media i tassi di occupazione sono *aumentati*, in particolare nel caso di categorie caratterizzate da una percentuale relativamente più elevata di lavoratori giovani e qualificati. Allo stesso modo, un gruppo di ricercatori[9] ha mostrato che le aziende statunitensi che adottano l'intelligenza artificiale sono quelle che crescono più rapidamente. Tuttavia, come integrare con successo l'intelligenza umana e l'intelligenza artificiale è una questione complessa e affascinante su cui il nostro studio dà soltanto uno spunto iniziale di riflessione, e che ci aspettiamo genererà dibattito per molti anni a venire.

---

[8] Si veda S. Albanesi, A. Dias Da Silva, J.F. Jimeno, A. Lamo, A. Wabitsch, *New technologies and jobs in Europe*, in *CEPR Discussion Paper* No. 18220, 2023.

[9] Si veda a questo proposito D. Acemoglu, G. Anderson, D. Beede, C. Buffington, E. Childress, E. Dinlersoz, E.N. Zolas, *Advanced technology adoption: Selection or causal effects?*, in *AEA Papers and Proceedings*, 2023, 113, pp. 210-14.

L'INTELLIGENZA ARTIFICIALE GENERATIVA NELLA PUBBLICA AMMINISTRAZIONE*

di Renato Ruffini

La pubblica amministrazione ha da sempre avuto rapporti complessi con gli strumenti informatici. Sintetizzando la cosa seppur con una certa generalizzazione ha sempre speso molto e ottenuto poco. Questo per vari motivi, non ultimo il fatto che le unità amministrative che necessitano di sistemi e programmi informatizzati sono molte e di dimensioni medio piccole e con limitate competenze tecniche.

Su questo ci sono state lodevoli ma limitatissime eccezioni. Come quella per esempio dell'INPS e dell'INAIL, che date le caratteristiche del loro sistema produttivo, ma probabilmente anche grazie al genio dell'allora presidente Gianni Billia, sono riuscite a coniugare in modo molto efficace informatica e sviluppo organizzativo ormai da vari lustri.

Oggi la spesa in ITC ha valori non elevatissimi (6,9 mld pari al 10% delle spese) e limitate competenze professionisti presenti all'interno delle amministrazioni. In relazione allo sviluppo dei sistemi informativi e degli applicativi quello che sembra accadere è una sostanziale fatica degli operatori (e della istituzione) nello stare al passo con la loro evoluzione. È difficile, in altri termini, che l'informatica riesca a trainare l'innovazione se mancano le competenze professionali interne per gestire i processi di sviluppo organizzativo.

Data questa situazione da cui è necessario partire è lecito domandarsi come si comporteranno le amministrazioni pubbliche, o dovrebbero comportarsi, di fronte all'intelligenza artificiale generativa (GenAI) a linguaggio naturale che costituisce, ad oggi, una reale rivoluzione sia dei comportamenti umani individuali sia dei processi produttivi nelle organizzazioni.

* Si ringrazia Luca Mari per il contributo dato al presente lavoro.

In proposito occorre in primo luogo avere ben chiaro che a monte dell'utilizzo di forme di intelligenza artificiale di qualsiasi tipologia è necessario avere elevate capacità di gestire ed archiviare i dati, così come di proteggerli e garantirne la sicurezza. Su questo aspetto attualmente ci sono grandi sforzi da parte delle amministrazioni e da parte degli organi e delle agenzie di coordinamento delle stesse.

Un'altra questione che credo sia fondamentale è capire come saranno gestite le relazioni tra individui e intelligenza artificiale ed in particolare chi decide cosa, o in altri termini chi è responsabile di cosa nelle relazioni tra persona e GenAI all'interno dell'organizzazione pubblica. Questo tema è rilevante sia per affrontare la questione della responsabilità individuale derivante dalla decisione supportata dall'intelligenza artificiali ma anche per utilizzarla in modo utile e innovativo e tale da sviluppare anche le competenze delle persone. l.

Al momento GenAI ha tre limiti:

– Non è direttamente connesso al mondo empirico.
– Opera passivamente (reagisce alle richieste ma non si attiva autonomamente).
– Non impara dalle conversazioni che fa poiché la memoria a breve termine che mantiene il contenuto delle conversazioni non viene riservata nella memoria a lungo termine per proseguire l'addestramento della rete neurale.

Nonostante tali limiti, manifesta un comportamento che spesso assimiliamo più ad un essere umano che non ad un software di cui siamo abituati e questo perché il comportamento dei sistemi di GenAI deriva da addestramento, e non da programmazione. Vi sono poi altri due limiti che mettono in evidenza l'importanza di questa loro caratteristica: non sono sempre affidabili nella correttezza delle informazioni che producono e il loro comportamento è difficilmente spiegabile in termini di singole cause specifiche, anche se ultimamente descrivono il percorso logico adottato nel dare le risposte. Questi "limiti" derivano proprio dal fatto che s'invertono i paradigmi dei sistemi software tradizionali. Un software agisce su istruzioni precise e il malfunzionamento può essere spiegato e corretto. Questo non si può fare con GenAI che produce informazioni in modo più o meno corretto in funzione di ciò che ha appreso. I suoi errori non sono "bug" ma opinioni sbagliate.

Di chi è la responsabilità del loro errore?

In un terremoto la responsabilità della casa distrutta può essere del terremoto stesso (responsabilità causale), del sismologo che non ha previsto corret-

tamente la sismicità della zona (responsabilità cognitiva) o dell'architetto che non ha adottato i criteri antisismici necessari (responsabilità morale).

Nel caso della intelligenza artificiale (per esempio guida autonoma) la responsabilità, sia essa negativa (colpa) o positiva è sempre delle persone che l'addestrano. Anche se GenAI può formulare strategie, identificare obiettivi in modo autonomo, la responsabilità dei risultati ottenuti sarà sempre di esseri umani, o gli sviluppatori, o gli addestratori o gli utilizzatori (con grandi difficoltà poi e rilevare l'effettiva responsabilità).

Questa rivoluzione tecnologica ci conduce a riscoprire il valore e la responsabilità di ciascuno di noi esseri umani, nei confronti di noi stessi e della società in cui operiamo.

Occorre di conseguenza che si attivino relazioni responsabili con questa tecnologia che da un lato consentano agli individui di relazionarsi con il mondo e di apprendere in continuazione fino a migliorare le proprie capacità di innovazione.

Questo non è stato fatto, probabilmente, con altre tecnologie più tradizionali ma "attive" come per esempio dei motori di ricerca o le piattaforme di social media o social network che di fatto negli anni hanno modificato i nostri modi di pensare, di prestare attenzione e di comunicare in forma scritta e parlata, senza che ce ne rendessimo conto, ancorché noi stessi abbiamo contribuito a questi cambiamenti. Un po' provocatoriamente potremmo dire che sono state le tecnologie ad addestrarci e non viceversa.

Al momento la GenAI nelle organizzazioni pubbliche ha avuto vari utilizzi:

1. Automazione e Risposte a Domande

- Chatbot e assistenti virtuali: Utilizzati per rispondere automaticamente alle domande dei cittadini, riducendo il carico di lavoro sui dipendenti pubblici. Questi sistemi possono fornire informazioni su procedure amministrative, scadenze e modulistica.
- FAQ e supporto automatico: L'IA generativa può creare e aggiornare le FAQ dinamicamente in base alle domande più frequenti e alle nuove esigenze.

2. Creazione di Documentazione e Reportistica

- Generazione automatica di documenti. L'IA può creare bozze di documenti ufficiali, report e relazioni in modo rapido, basandosi su dati e modelli predefiniti.
- Redazione assistita. Aiuta i funzionari nella preparazione di discorsi, lettere e comunicati, migliorando la qualità e riducendo i tempi di stesura.

3. Analisi dei Dati e Previsioni

- Elaborazione dei dati. L'IA generativa è utile per analizzare grandi volumi di dati pubblici e trasformarli in report e grafici comprensibili.
- Simulazioni e previsioni. Genera scenari ipotetici per supportare la pianificazione e la decisione, ad esempio per gestire crisi o valutare politiche pubbliche.

4. Formazione e Aggiornamento del Personale

- Corsi di formazione. L'IA può creare contenuti didattici personalizzati per formare il personale, includendo simulazioni e spiegazioni adattive.
- Manuali e guide aggiornabili. I sistemi generativi possono produrre manuali che si aggiornano automaticamente con le modifiche legislative.

5. Traduzione e Accessibilità

- Traduzioni automatiche. Facilitano la comprensione di documenti in lingue diverse, migliorando l'accessibilità per comunità multilingue.
- Sintesi e trascrizione. Converte automaticamente discorsi o riunioni in trascrizioni leggibili, semplificando la distribuzione delle informazioni.

6. Innovazione e Creatività nei Servizi

- Proposte di miglioramento. L'IA può suggerire nuove idee per l'ottimizzazione dei processi amministrativi e la creazione di progetti pilota innovativi.
- Design di interfacce ed esperienze utente. Supporta nella creazione di portali web più intuitivi e applicazioni user-friendly.

Tutte queste applicazioni ed altre possibili appaiono estremamente rilevanti; tuttavia, è necessario interrogarsi quali sono le conseguenze pratiche che tutto questo ha nel funzionamento delle organizzazioni e nelle competenze degli individui.

Per capire come GenAI si inserisce nel mondo delle organizzazioni e i suoi possibili effetti possiamo operare sia in modo induttivo che deduttivo.

Nel primo caso possiamo chiederci come GenAI modifica il modo di lavorare in specifici contesti e un campo di indagine privilegiato può essere quello degli sviluppatori di software che già da tempo usano l'intelligenza artificiale nel loro lavoro. In proposito è stata fatta una breve indagine e alcuni focus group presso una società italiana.

Ci si può pertanto domandare:

– come cambiano le persone e le loro competenze,
– come cambiano l'organizzazione, i ruoli e le responsabilità.

In primo luogo, come sa chi ha interagito con i sistemi di GenAI questi non sono concepiti secondo il classico paradigma procedurale: tali sistemi si pongono come degli interlocutori, degli sparring partner che possono essere chiamati a svolgere compiti complessi e poco strutturati, magari chiariti progressivamente nel corso dell'interazione, come suggerire idee e identificare possibili miglioramenti a testi dati. E questo, evidentemente, come risultato non di una programmazione, ma di un addestramento. Non è infatti immaginabile che il comportamento di sistemi di GenAI come i chatbot si possa ottenere dall'esecuzione di algoritmi scritti esplicitamente da programmatori. È invece il risultato del calcolo di funzioni parametriche notevolmente complesse – le reti neurali artificiali che costituiscono il nucleo di questi sistemi – che sono state adattate alle applicazioni richieste assegnando opportunamente i valori ai tanti parametri delle funzioni, in un processo appunto di addestramento. Addestramento invece di programmazione, dunque.

I nuovi sistemi addestrati potrebbero affiancare, più che sostituire, i tradizionali sistemi programmati. Dal punto di vista degli sviluppatori di software, questo prefigura perciò due scenari complementari: sistemi di GenAI da sviluppare e sistemi di GenAI con cui sviluppare.

Nel primo caso gli sviluppatori saranno appunto meno programmatori e più istruttori di reti neurali.

La seconda idea è che cambiano le condizioni e le modalità dell'interazione tra uomo e macchina, con interfacce non più solo statiche/procedurali ma anche dinamiche/conversazionali. I sistemi di GenAI ci stanno facendo intravvedere opzioni di interazione meno strutturate e più flessibili, analogamente a quanto avviene in una conversazione, in cui i problemi vengono analizzati e risolti tramite uno scambio progressivo e non preimpostato di informazioni, di cui è un esempio quella tecnica di interazione che non per nulla si chiama oggi *chain-of-thought prompting* ("domande seguendo un flusso di pensiero"). Grazie alla loro capacità di conversazione e di mantenere il filo del discorso anche nel caso di dialoghi lunghi e complessi, i chatbot ci stanno facendo immaginare la possibilità di applicazioni con interfacce multimodali (testuali in forma scritta, testuali in forma parlata, per immagini, per video) completamente nuove, che richiederanno nuovi paradigmi di ergonomia, per bilanciare la facilità dell'interazione con la sua efficacia ed efficienza.

In pratica con questi sistemi si attiva una logica conversazionale, basata

su conoscenze e capacità ampie, che superano la separazione tra cultura tecnica e cultura umanistica.

Questa logica di cambiamento può essere per analogia sviluppata in altri diversi contesti organizzativi.

In generale si registra un diffuso consenso tra gli studiosi che i chatbot possono essere interpretati come consulenti virtuali a cui vengono delegati certi compiti (esecutivi o strategici), mantenendo noi stessi la gestione e la responsabilità del processo.

Imprese private traggono già beneficio dalle applicazioni di IA, ad esempio per ridisegnare i processi e le funzioni, innovare i modelli di business e i servizi offerti ai consumatori.

La ricerca scientifica sul tema si è concentrata sui benefici derivanti dall'utilizzo dell'IA per promuovere la diversità aziendale, il reclutamento, influenzare le decisioni degli HR manager.

Ma gran parte delle applicazioni concernono l'AI più che i GenAI. Ad esempio, gli algoritmi di People Analytics vengono utilizzati durante i colloqui di lavoro. Questi algoritmi consistono nell'utilizzare il Machine learning per produrre automaticamente le caratteristiche dei candidati (come le conoscenze, abilità, attitudini o tratti delle personalità) dalle loro risposte durante i colloqui. È stato dimostrato che questi strumenti sono in grado di aumentare l'efficacia dei processi di selezione del personale e ridurre i bias inconsci degli intervistatori.

In linea generale, si può ipotizzare in prima approssimazione che i sistemi organizzativi evolveranno sulla base dell'evoluzione di alcune questioni centrali:

a. Skill and education. GenAI ha la potenzialità per cambiare la meaningfulness e la natura del lavoro sviluppando la creatività delle persone. Il problema in tale contesto e su come sviluppare le conoscenze relative alle capacità di utilizzo di GenAI, come sviluppare strategie "blue sky" aperte e generative. Si creerà l'esigenza di semplificare i processi classici di natura amministrativa con automazione e delega alla line. La nuova categoria di lavoro saranno quelle connesse all'AI trainers, explainer e sustainers. Si pone il problema di come generare gli skill necessari attraverso la formazione e la pratica.
b. Business productivity. Su tale aspetto è al momento assai difficile capire l'impatto di GenAI sulla produttività del business. Occorre in questo caso porre attenzione su come vogliamo valutare la produttività. Il tema è che siamo in un contesto di servizio di difficile valutazione. Probabilmente, anche a causa della logica conversazionale di GenAI, occorre ri-

vedere il concetto di produttività superando l'ottica temporale e introducendo anche elementi di produzione di valore più complessi.

c. Etica. Tale aspetto sarà centrale nelle questioni organizzative, non a caso è stato oggetto di una lotta interna nelle governance di OpenAI. Le questioni in questo contesto sono innumerevoli. È necessario comprendere la responsabilità, i rischi reputazionali per le organizzazioni, la difficoltà di capire l'attribuzione della proprietà intellettuale, la trasparenza dei processi decisionali, etc. In tale contesto le organizzazioni, già oggi molto impegnate, dovranno essere ancora più attente a queste tematiche e presidiarli in modo efficace.
d. Metodi di ricerca e nuovi linguaggi. Un campo di immediato utilizzo di GenAI è stato proprio quello della ricerca. In questo contesto cambiano i metodi, le logiche di produzione e il linguaggio. Sono tutti elementi potenzialmente impattanti sulle imprese e sulla società che dovranno essere indagati con attenzione.

In pratica il futuro si presenta complesso ed interessante ed è fondamentale che la pubblica amministrazione, per il ruolo che le è proprio, sia protagonista dell'evoluzione dell'utilizzo di queste strumentazioni.

Nel contesto pubblico sono ad oggi innumerevoli i dibattiti e le attività formative sul tema dell'intelligenza artificiale generativa il che appare corretto e auspicabile. Anzi, crediamo sia necessaria fare passi avanti in proposito e coinvolgere la maggior parte dei dipendenti pubblici su tale questione ed attivare pratiche diffuse e controllate di utilizzo di questi strumenti.

Nessuno operatore della p.a. che utilizza queste strumentazioni dovrà operare senza una sufficiente consapevolezza del proprio ruolo e delle sue responsabilità. Gestire questi strumenti implica capacità di "addestramento" quindi competenze evolute e consapevolezza senza la quale da addestratori diventeremo soggetti addestrati. È quindi estremamente utile in questa fase diffondere la consapevolezza presso dei dipendenti pubblici della rilevanza dello strumento e di una formazione di base rispetto ad esso, all'interno di un contesto regolamentare organizzativo in cui l'utilizzo dello strumento, anche se fatto individualmente (e non con uno specifico applicativo) nello sviluppo delle proprie mansioni, sia comunque dichiarato al fine di comprendere la pervasività e la direzione di questo importante fenomeno. Si tratta in sostanza di elaborare delle guide per l'uso della intelligenza artificiale generativa in grado di offrire un quadro generale e identifichi problematiche e sfide relative all'uso della stessa e offra indirizzi di policy e buone pratiche, e un affinamento constante della analisi ed in-

terpretazione delle direzioni evolutive del fenomeno. Naturalmente su tale strada la p.a. italiana, in particolare con Agid si è già mossa, ma forse il tema è ancora considerato come una questione da tecnici e non da "persone normali" come in realtà è.

In sintesi, riteniamo che la formazione dei pubblici dipendenti su questo tema sarà cruciale. Una sfida centrale per i governi sarà data dal superamento delle deficienze esistenti inerente le competenze delle persone ancora scarse, le tecnologie insufficienti, l'inadeguatezza dei dati e dalla convinzione che l'AI sia ancora poco utile per molte aree di attività del settore pubblico. Infatti, per affrontare in modo adeguato la sfida della intelligenza artificiale generativa occorre attivare una corretta combinazione di competenze delle persone, strumenti tecnologici e dati. Tutto ciò non è ancora sviluppato. Questa sfida comporta, inoltre, non solo un rilevante investimento economico ma anche organizzativo che coinvolga in modo organico tutti i dipendenti pubblici mobilitando le persone e con esse la tecnologia. Perché oggi con l'intelligenza artificiale, persone e tecnologia, cresceranno insieme con le relazioni virtuose o meno, che riusciranno ad attivare.

Bibliografia

LONGO J., *The Transformative Potential of Artificial Intelligence for Public Sector Reform*, in *Canadian Public Administration*, 2024.

MARI L., *L'Intelligenza Artificiale di Dostoevskij – Riflessioni sul futuro, la conoscenza, la responsabilità umana*, Il Sole 24 Ore, Milano, 2024.

SIMONDON G., *Du mode d'existence des objets techniques*, Flammarion, Paris, 2024.

WIRTZ B. W.-WEYERER, J. C.-STURM B. J., *The dark sides of artificial intelligence: An integrated AI governance framework for public administration*, in *International Journal of Public Administration*, 43(9), 2020, pp. 818-829.

INTELLIGENZA ARTIFICIALE E DIRITTO DEL LAVORO: RISCHI (LAVORISTICI), OPPORTUNITÀ (OCCUPAZIONALI), SFIDE (REGOLATIVE)

di Marco Biasi

Sommario: 1. Premessa. – 2. Rischi. – 3. Opportunità. – 4. Sfide. – 5. Conclusioni (aperte).

1. *Premessa*

Come risaputo, il diritto del lavoro è destinato a seguire il cambiamento organizzativo, il quale, a sua volta, si colloca a valle del progresso tecnologico [1].

Sotto questo aspetto, l'avvento dell'intelligenza artificiale non costituisce una deviazione dal consolidato paradigma, se non fosse per la rapidità con la quale la corrente trasformazione sta investendo le modalità di lavoro e le relative tecniche protettive.

Si consideri, a titolo esemplificativo, la voce del *Digesto* dedicata al "Lavoro Digitale" e redatta da chi scrive nel tutt'altro che remoto, quanto meno in apparenza, anno 2022: l'approfondimento guardava alle due questioni, allora ritenute centrali, nonché, sotto certi versi, complementari (in quanto esempi, nell'un caso, di autonomia nella subordinazione, nell'altro caso, di subordinazione nell'autonomia), del lavoro agile e del lavoro tramite piattaforma [2].

Di converso, l'intelligenza artificiale, che pure fungeva da (efficace) ausilio per le prestazioni rese dai prestatori di lavoro operanti da remoto e da

---

[1] V., per tutti, F. Carinci, *Rivoluzione tecnologica e diritto del lavoro: il rapporto individuale*, in *Dir. Lav. Rel. Ind.*, 1985, p. 203 ss.; G. Vardaro, *Tecnica, tecnologia e ideologia della tecnica nel diritto del lavoro*, in *Pol. Dir.*, 1986, 1, p. 75 ss.; L. Nogler, *Tecnica e subordinazione nel tempo della vita*, in *Dir. Lav. Rel. Ind.*, 2015, p. 337 ss.

[2] M. Biasi, *Lavoro digitale (voce)*, in *Dig. Disc. Priv. Sez. Comm.*, Aggiornamento, IX, 2022, p. 259 ss.

(necessario) mezzo di funzionamento delle piattaforme digitali, era rimasta sullo sfondo di una discussione nella quale è invece entrata prepotentemente nell'ultimo triennio, anche per effetto dell'iniziativa legislativa europea sfociata nell'adozione, nella primavera/estate del 2024, del c.d. AI Act [3].

[3] Ci si riferisce, naturalmente, al Regolamento UE 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale), pubblicato sulla Gazzetta Ufficiale dell'Unione Europea del 12 luglio 2024. Per un commento alla versione dell'AI Act approvata dal Parlamento Europeo nella seduta del 13 marzo 2024 e ratificata dal Consiglio il 21 maggio 2024, M. Peruzzi, *Intelligenza artificiale e lavoro: l'impatto dell'AI Act nella ricostruzione del sistema regolativo Ue di tutela*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, Giuffrè, Milano, 2024, p. 115 ss.; in generale, sul rapporto tra lavoro e AI, v., senza ambizione di completezza, S. Scagliarini, I. Senatori (a cura di), *Lavoro, Impresa, e Nuove Tecnologie dopo l'AI Act*, Fondazione Marco Biagi, Modena, 2024; M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit.; M. Peruzzi, *Intelligenza artificiale e lavoro. Uno studio su poteri datoriali e tecniche di tutela*, Giappichelli, Torino, 2023; F.V. Ponte, *Intelligenza artificiale e lavoro. Organizzazione algoritmica, profili gestionali, effetti sostitutivi*, Giappichelli, Torino, 2024; A. Sartori, *Intelligenza artificiale e gestione del rapporto di lavoro. Appunti da un cantiere ancora aperto*, in *Var. Temi Dir. Lav.*, 2024, 3, p. 806 ss.; M. Corti, *Intelligenza artificiale e partecipazione dei lavoratori. Per un nuovo umanesimo del lavoro*, in *Dir. Rel. Ind.*, 2024, 3, p. 615 ss.; M. Biasi, *Problema e sistema nella regolazione lavoristica dell'intelligenza artificiale: note preliminari*, in *federalismi*, 2024, 30, p. 162 ss.; L. Zoppoli, *Il diritto del lavoro dopo l'avvento dell'intelligenza artificiale: aggiornamento o stravolgimento? Qualche (utile) appunto*, in *WP C.S.D.L.E. "Massimo D'Antona".IT*, n. 489/2024; F. Bano, *Algoritmi al lavoro. Riflessioni sul management algoritmico*, in *Lav. Dir.*, 2024, 1, p. 133 ss.; C. Cristofolini, *Navigating the impact of AI systems in the workplace: strengths and loopholes of the EU AI Act from a labour perspective*, in *Italian Labour Law e-Journal*, 2024, 17, 1, p. 75 ss.; S. Ciucciovino, *La disciplina nazionale sulla utilizzazione dell'intelligenza artificiale nel rapporto di lavoro*, in *Lav. Dir. Eur.*, 2024, 1, p. 1 ss.; A. Alaimo, *Il Regolamento sull'Intelligenza Artificiale: dalla proposta della Commissione al testo approvato dal Parlamento. Ha ancora senso il pensiero pessimistico?*, in *federalismi*, 2023, 25, p. 133 ss.; Ead., *Il Regolamento sull'Intelligenza Artificiale. Un treno al traguardo con alcuni vagoni rimasti fermi*, *ibidem*, p. 231 ss.; A. Ingrao, Hic sunt leones! *La piramide del rischio costruita dalla proposta di Regolamento sulla intelligenza artificiale (emendata)*, in *Lav. e Prev. Oggi*, 2023, 11-12, p. 778 ss.; E. Signorini, *Lavoro e tecnologia: connubio tra opportunità e rischi*, in *federalismi*, 2023, 29, p. 202 ss.; P. Lambertucci, *Intelligenza artificiale e tutela del lavoratore: prime riflessioni*, in *Arg. Dir. Lav.*, 2022, 5, p. 897 ss.; G. Zampini, *Intelligenza artificiale e decisione datoriale algoritmica. Problemi e prospettiva*, *ivi*, 2022, p. 467 ss.; P. De Petris, *Sul "potere algoritmico" dell'impresa digitale*, in A. Bellavista, R. Santucci (a cura di), *Tecnologie digitali, poteri datoriali e diritti dei lavoratori*, Giappichelli, Torino, 2022, p. 41 ss.; C. Ciccia Romito, *Intelligenza artificiale nei rapporti di lavoro: tra regolazione e sfide future*, in *Ciberspazio e diritto*, 2022, 3, p. 353 ss.; P. Tullini, *La nuova proposta europea sull'intelligenza artificiale e le relazioni di lavoro*, in *Trabajo, Persona, Derecho, Mercado*, 2022, 5, p. 99 ss.; L. Zappalà, *Informatizzazione dei processi decisionali e diritto del lavoro: algoritmi, poteri datoriali e responsabilità del prestatore nell'era dell'intelligenza arti-*

In un brevissimo lasso di tempo, l'intelligenza artificiale (d'ora innanzi, "AI") si è imposta come uno strumento tutt'altro che "di nicchia" (dunque, non ad uso e consumo delle sole piattaforme come Uber o Glovo), venendo utilizzata trasversalmente e, perciò, anche nell'ambito delle lavorazioni più classiche [4]: all'AI si ricorre oggi (invero, non ancora troppo) comunemente, sia per giungere ad una decisione umana ma basata sui dati forniti e "lavorati" dalla macchina (*workforce analytics*), sia per addivenire ad una determinazione suggerita, se non direttamente adottata, dal *software* "intelligente" (*algorithmic management*) [5]. Nell'una così come nell'altra ipotesi, l'AI si avvale di un algoritmo [6], che si nutre dei dati [7] inseriti *ex*

---

ficiale, in *Biblioteca 20 maggio*, 2021, 2, p. 98 ss.; G. Bolego, *Intelligenza artificiale e regolazione delle relazioni di lavoro: prime riflessioni*, in *Labor*, 2019, 1, p. 51 ss.; nella letteratura internazionale, v., *ex multis*, A. Ponce del Castillo (ed.), *Artificial intelligence, labour and society*, ETUI, Brussels, 2024; A. Aloisi, *Regulating Algorithmic Management at Work in the European Union: Data Protection, Non-discrimination and Collective Rights*, in *Int. Journ. Comp. Lab. Law & Ind. Rel.*, 2024, 1, p. 37 ss.; I. Ajunwa, *The Quantified Worker. Law and Technology in the Modern Workplace*, Cambridge University Press, Cambridge, 2023; O. Lobel, *The Law of AI for Good*, in *University of San Diego School of Law Research Paper* n. 23-001; 30 gennaio 2023; B. Waas, *Artificial Intelligence and Labour Law*, in *WP Hugo Sinzheimer Institute*, n. 2022/17; J. Adams-Prassl, H. Abraha, A. Kelly-Lyth, M. Silberman, S. Rakshita, *Regulating Algorithmic management: A blueprint*, in *Eur. Lab. Law Journ.*, 2023, 14, 2, p. 124 ss.; A. Aloisi, V. De Stefano, *Between risk mitigation and labour rights enforcement: Assessing the transatlantic race to govern AI-driven decision-making through a comparative lens*, *ibidem*, p. 283 ss.; J. Adams-Prassl, *What if Your Boss was an Algorithm? Economic Incentives, Legal Challenges, and the Rise of Artificial Intelligence at Work*, in *Comp. Lab. Law & Pol. Journ.*, 2019, 41, 1, p. 123 ss.; A. Aloisi, E. Gramano, *Artificial Intelligence is Watching You at Work: Digital Surveillance, Employee Monitoring, and Regulatory Issues in the EU Context*, *ibidem*, p. 95 ss.; M. Otto, *Workforce Analytics v Fundamental Rights Protection in the EU in the Age of Big Data*, *ivi*, 40, p. 389 ss.

[4] K. Crawford, *Né intelligente né artificiale. Il lato oscuro dell'IA*, il Mulino, Bologna, 2021, p. 65 ss.; S. Baiocco *et al.*, *The Algorithmic Management of work and its implications in different contexts*, in *Background Paper Series of the Joint EU-ILO Project "Building Partnerships on the Future of Work"*, n. 9/2022; C. Kellogg *et al.*, *Algorithms at Work: The New Contested Terrain of Control*, in *Academy of Management Annals*, 2020, n. 1, p. 366 ss.

[5] L. Zappalà, *Management algoritmico*, in AA.VV., *Lavoro e tecnologie. Dizionario del diritto del lavoro che cambia*, Giappichelli, Torino, 2022, p. 150 ss.; E. Signorini, *Lavoro e tecnologia: connubio tra opportunità e rischi*, in *federalismi*, 2023, 29, p. 206.

[6] L. Zappalà, *Algoritmo*, in AA.VV., *Lavoro e tecnologie. Dizionario del diritto del lavoro che cambia*, Giappichelli, Torino, 2022, p. 17 ss.

[7] L. Zappalà, *Big data*, in Aa.Vv., *Lavoro e tecnologie*, cit., p. 28; P. Tullini, *Dati*, in M. Novella, P. Tullini (a cura di), *Lavoro digitale*, Giappichelli, Torino, 2022, p. 105 ss.; cfr. anche la decisione n. 112 del 30 marzo 2023 del Garante per la Protezione dei Dati Personali, che ha temporaneamente sospeso l'operatività di Chat GPT (operata da OpenAI L.L.C.) per il pericolo di un uso incontrollato e della diffusione dei dati degli utenti del servizio.

ante e di quelli progressivamente acquisiti attraverso l'interazione con i terzi, agevolata, nei modelli più avanzati, dalle capacità di autoapprendimento della macchina stessa [8].

Per quanto non si possa revocare in dubbio che l'AI abbia già inciso sull'organizzazione del lavoro, non pare perciò solo corretto, sul piano del metodo, adagiarsi sulle narrazioni spesso affrettate e pregiudizialmente polarizzate degli apocalittici (ovvero degli scettici e dei nostalgici per vocazione) e degli integrati (o aprioristicamente entusiasti di ogni trasformazione in sé, specie se tecnologica) [9].

Piuttosto, pare preferibile assumere una postura che guardi alla tematica in esame con equilibrio e con prudenza e che miri al contemperamento, che segna la cifra dello stesso AI Act, tra la spinta all'innovazione e la dimensione antropocentrica.

Proprio per questo, nelle pagine che seguono l'accento verrà sì posto inizialmente sui rischi che le nuove tecnologie comportano per i diritti fondamentali dei lavoratori, con particolare riferimento alle tematiche della salute e sicurezza della persona e della protezione dei dati (par. 2).

In seguito, però, verranno messe in evidenza pure le opportunità derivanti dall'utilizzo dell'intelligenza artificiale, le quali spaziano dalla salvaguardia della salute e sicurezza all'apertura di nuovi spazi, come il metaverso, utili anche in chiave di formazione e di successiva certificazione, mediante la *blockchain*, delle competenze acquisitele (par. 3).

Da ultimo, l'accento verrà posto sulla necessità di una ricomposizione del sistema regolativo e di una riorganizzazione delle fonti attraverso un equilibrato dosaggio tra le tensioni universalistiche e la doverosa attenzione verso le peculiarità dei singoli settori (par. 4).

2. *Rischi*

Come rilevato in apertura, le imprese si avvalgono con maggiore frequenza dell'intelligenza artificiale per l'adozione di un ampio novero di decisioni in materia di gestione del personale, le quali possono riguardare le fasi del reclutamento, del controllo, della valutazione, della promozio-

[8] L. Zappalà, *Machine learning*, in AA.VV., *Lavoro e tecnologie. Dizionario del diritto del lavoro che cambia*, Giappichelli, Torino, 2022, p. 147 ss.

[9] Riassumono le due tensioni interne al dibattito, anche giuslavoristico, sull'AI, A. Aloisi, V. De Stefano, *Il tuo capo è un algoritmo. Contro il lavoro disumano*, Laterza, Bari, 2020, spec. p. 84 ss.

ne e persino della cessazione del rapporto di lavoro [10].

Il ricorso all'AI può senza dubbio comportare dei rischi per i diritti fondamentali dei lavoratori, oltre che degli individui in generale [11].

L'argomento proverebbe, tuttavia, troppo qualora da ciò inducesse a concludere che vi sia un'incompatibilità di fondo tra l'evoluzione dell'AI ed il rispetto dei diritti fondamentali delle persone [12].

All'opposto, proprio dall'AI Act emerge chiaramente l'idea per cui la tecnologia debba fungere da mezzo a servizio dell'uomo (e non il contrario), donde il reiterato richiamo, all'interno del testo (oltre che in altri documenti di varia natura approvati, pressoché contestualmente, al di fuori dei confini europei [13]), all'idea di un'AI antropocentrica [14].

Tale impostazione, che permea l'intero articolato normativo euro-unitario, si estende alle disposizioni che esso dedica al lavoro, pure costruite attorno al concetto di rischio [15].

---

[10] Ampiamente, I. Ajunwa, *The Quantified Worker. Law and Technology in the Modern Workplace*, cit., spec. 75 ss.; nella letteratura nazionale, S. Renzi, *Decisioni automatizzate, analisi predittive e tutela della privacy del lavoratore*, in *Lav. Dir.*, 2022, 3, p. 583 ss.; A. Pizzoferrato, *Automated decision-making in HRM*, in *Lav. Giur.*, 2022, 11, p. 1030 ss.; A. Rota, *Rapporto di lavoro e big data analytics: profili critici e risposte possibili*, in *Lab. & Law Issues*, 2017, 3, 1, p. 32 ss.

[11] Sempre di alto rischio, ma riferito alla propaganda ed alle informazioni errate (ed alla sostituzione dell'uomo, anche efficiente, con la macchina), ha parlato una importante lettera aperta del 29 marzo 2023 di influenti imprenditori del settore dell'AI (Elon Musk e Steven Wozniak su tutti), i quali invitavano allora alla sospensione delle attività di addestramento dei "*Giant AI experiments*" e alla predisposizione di protocolli condivisi di sicurezza visionati da esperti indipendenti, così da poter garantire l'accuratezza, la sicurezza, la trasparenza, la robustezza, l'allineamento, l'affidabilità e la lealtà dello strumento in parola.

[12] Per tutti, G. Ziccardi, *L'intelligenza artificiale e la regolamentazione giuridica: una relazione complessa*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 13 ss.

[13] M. Bassini, *Intelligenza artificiale e diritti fondamentali: considerazioni preliminari*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 26, ed ivi i riferimenti alla Bletchley Park Declaration del 2 novembre 2023 (https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023?) e all'*executive order* del 30 ottobre a firma dell'allora Presidente americano Joe Biden.

[14] Ciò significa pure che l'imputazione degli atti adottati su indicazione o su suggerimento dell'AI rimane in capo al datore di lavoro, il quale non viene quindi deresponsabilizzato dall'AI Act, che guarda all'intelligenza artificiale come ad un oggetto e non ad un soggetto di diritto: in termini, A. Sartori, *L'impatto dell'intelligenza artificiale sul controllo e la valutazione della prestazione, e sull'esercizio del potere disciplinare*, in *Lav. Dir. Eur.*, 2024, 3, p. 5, ed ivi ulteriori richiami.

[15] L. Zappalà, *Sistemi di IA ad alto rischio e ruolo del sindacato alla prova del* risk-based approach, in *Lab. & Law Issues*, 2024, 10, 1, p. 52 ss.; M. Barbera, *"La nave deve navigare. Rischio e responsabilità al tempo dell'impresa digitale*, in *Labour & Law Issues*, 2023, 9, 2, p. 3 ss.; P. Loi,

Da un lato, si collocano i sistemi a rischio inaccettabile (come i sistemi di riconoscimento delle emozioni sul posto di lavoro o di categorizzazione biometrica delle persone, volti a dedurre caratteristiche sensibili come la razza, le opinioni politiche, l'affiliazione sindacale, l'orientamento sessuale e il credo religioso), la cui immissione in commercio all'interno dell'Ue viene vietata, al pari del relativo utilizzo [16].

Dall'altro lato, si pongono i sistemi ad alto rischio, tra i quali rientrano i *software* utilizzati, in generale, nel campo dell'occupazione, della gestione dei lavoratori e dell'accesso al lavoro autonomo [17]. L'impiego di tali strumenti non è precluso in radice [18], ma è condizionato al rispetto di una serie obblighi che il Regolamento pone in capo, a monte, al fornitore (verifica, mappatura, istituzione di un sistema di gestione dei rischi e formazione dell'utilizzatore) e, a valle, all'utilizzatore (ossia, ai nostri fini, al datore di lavoro, chiamato a garantire la trasparenza, ad attuare la sorveglianza e ad effettuare, laddove prevista, la valutazione di impatto) [19].

Per di più, è lo stesso AI Act a puntualizzare, secondo un modello di relazione tra le fonti dell'ordinamento *multi-level* già sperimentato con il GDPR (rispetto al quale l'AI Act si pone a sua volta in un rapporto di complementarietà) [20], che al legislatore nazionale (ma anche alle parti sociali) è consentito approntare una migliore protezione dei lavoratori rispet-

---

Il rischio proporzionato nella proposta di regolamento sull'IA e i suoi effetti nel rapporto di lavoro, in *federalismi*, 8 febbraio 2023.

[16] V. Art. 5 AI Act.

[17] Cfr. il considerando 57 dell'AI Act: "Anche i sistemi di IA utilizzati nel settore dell'occupazione, nella gestione dei lavoratori e nell'accesso al lavoro autonomo, in particolare per l'assunzione e la selezione delle persone, per l'adozione di decisioni riguardanti le condizioni del rapporto di lavoro la promozione e la cessazione dei rapporti contrattuali di lavoro, per l'assegnazione dei compiti sulla base dei comportamenti individuali, dei tratti o delle caratteristiche personali e per il monitoraggio o la valutazione delle persone nei rapporti contrattuali legati al lavoro, dovrebbero essere classificati come sistemi ad alto rischio, in quanto tali sistemi possono avere un impatto significativo sul futuro di tali persone in termini di prospettive di carriera e sostentamento e di diritti dei lavoratori".

[18] In termini di "carattere permissivo" del Regolamento, I. Senatori, *Introduzione. L'AI Act: un nuovo tassello nella costruzione dell'ordinamento del lavoro digitale*, in S. Scagliarini, I. Senatori (a cura di), *Lavoro, Impresa, e Nuove Tecnologie dopo l'*AI Act, cit., p. 7.

[19] M. Peruzzi, *Intelligenza artificiale e lavoro: l'impatto dell'AI Act nella ricostruzione del sistema regolativo Ue di tutela*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 129 ss.

[20] M. Peruzzi, *L'impatto dell'IA nella selezione del personale, negli annunci di lavoro mirati a filtrare le candidature, nella determinazione della "reputazione" della persona che lavora e nell'assegnazione dei compiti*, in *Lav. Dir. Eur.*, 2024, 3, p. 13.

to a quella garantita dalla cornice europea sull'intelligenza artificiale [21].

In questo senso, assume un'indubbia rilevanza l'art. 1-*bis* d.lgs. n. 152/1997 [22], che è stato introdotto dal d.lgs. n. 104/2022 ("decreto trasparenza") e, dunque, in epoca anteriore alla pubblicazione dell'AI Act nella Gazzetta Ufficiale dell'Unione Europea e, soprattutto, alla decorrenza dei relativi effetti [23]. La disposizione nazionale già prevede un capillare obbligo di informativa a favore dei lavoratori e dei loro rappresentanti [24], operante in

---

[21] V. art. 2, par. 11, dell'AI Act.

[22] In dottrina, E. Dagnino, *Il diritto interno: i sistemi decisionali e di monitoraggio (integralmente) automatizzati tra trasparenza e coinvolgimento*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 147 ss.; L. Tebano, *I diritti di informazione nel D.Lgs. n. 104/2022. Un ponte oltre la trasparenza*, in *Lav. Dir. Eur.*, 2024, 1, p. 1 ss.; E.C. Schiavone, *Gli obblighi informativi in caso di strumenti decisionali o di monitoraggio automatizzati*, in D. Garofalo, M. Tiraboschi, V. Filì, A. Trojsi (a cura di), *Trasparenza e attività di cura nei contratti di lavoro. Commentario ai decreti legislativi n. 104 e n. 105 del 2022*, ADAPT, Modena, 2023, p. 211 ss.; M.T. Carinci, S. Giudici, P. Perri, *Obblighi di informazione e sistemi decisionali e di monitoraggio automatizzati (art. 1-bis "Decreto Trasparenza"): quali forme di controllo per i poteri datoriali algoritmici?*, in *Labor*, 2023, 1, p. 7 ss.; M. Corti, *L'intelligenza artificiale nel decreto trasparenza e nella legge tedesca sull'ordinamento aziendale*, in *federalismi*, 13 dicembre 2023; A. Allamprese, S. Borelli, *L'obbligo di trasparenza senza la prevedibilità del lavoro. Osservazioni sul decreto legislativo n. 104/2022*, in *Rass. Giur. Lav.*, 2022, I, 4, p. 677 ss.; L. D'Arcangelo, *Diritti d'informazione e dati personali nel lavoro automatizzato. Rilievi sistematici sulla normativa applicabile*, in *Dir. Merc. Lav.*, 2023, 2, p. 347 ss.; A. Donini, *Informazione sui sistemi decisionali e di monitoraggio automatizzati: poteri datoriali e assetti organizzativi*, *ivi*, 2023, 1, 85 ss.; A. Viscomi, *Per una sandbox giuslavoristica. Brevi note a partire dal "decreto trasparenza"*, in *Lab. & Law Issues*, 2023, 9, 1, p. 124 ss.; G. Peluso, *Obbligo informativo e sistemi integralmente automatizzati*, in *Lab. & Law Issues*, 2023, 9, 2, p. 99 ss.; R. Covelli, *Lavoro e intelligenza artificiale: dai principi di trasparenza algoritmica al diritto alla conoscibilità*, *ibidem*, spec. p. 90 ss.; G.A. Recchia, *Condizioni di lavoro trasparenti, prevedibili e giustiziabili: quando il diritto di informazione sui sistemi automatizzati diventa uno strumento di tutela collettiva*, *ivi*, 2023, 1, p. 34 ss.; M. Marazza, F. D'Aversa, *Dialoghi sulla fattispecie dei "sistemi decisionali o di monitoraggio automatizzati" nel rapporto di lavoro (a partire dal Decreto trasparenza)*, in *giustiziacivile.com*, 8 novembre 2022; L. Zappalà, *Appunti su linguaggio, complessità e comprensibilità del lavoro 4.0: verso una nuova procedimentalizzazione dei poteri datoriali*, in *WP C.S.D.L.E. "Massimo D'Antona".IT*, n. 464/2022, p. 26 ss.; B. Rossilli, *Gli obblighi informativi relativi all'utilizzo di sistemi decisionali e di monitoraggio automatizzati indicati nel decreto "Trasparenza"*, in *federalismi*, 5 ottobre 2022; S. Renzi, *La trasparenza dei sistemi algoritmici utilizzati nel contesto lavorativo fra legislazione europea e ordinamento interno*, in *La Nuova Giuridica*, 2022, 2, p. 72 ss.

[23] Ai sensi dell'art. 113 dell'AI Act, la generalità delle previsioni dell'AI Act si applica dal 2 agosto 2026. Tuttavia, i divieti di cui all'art. 5 del Regolamento trovano applicazione a decorrere dal 2 febbraio 2025, mentre gli obblighi connessi ai sistemi ad alto rischio di cui all'art. 6, par. 1, dell'AI Act si applicano a decorrere dal 2 agosto 2027.

[24] Trib. Palermo 3 aprile 2023, in *Arg. Dir. Lav.*, 2023, 5, p. 1004 con nota di S. Renzi, *Obblighi di trasparenza in materia di sistemi automatizzati: il Tribunale di Palermo precisa il contenuto dell'informativa ex art. 1-bis d.lgs. n. 152 del 1997*; Trib. Palermo 20 giugno 2023, in *Arg.*

tutte le ipotesi in cui i datori di lavoro (o i committenti di una prestazione coordinata e continuativa) si avvalgano di sistemi decisionali o di monitoraggio integralmente automatizzati, il cui funzionamento si basa evidentemente sull'AI (per quanto non espressamente menzionata nel corpo del testo)[25].

3. *Opportunità*

A latere dei sunteggiati rischi, molteplici risultano le opportunità derivanti per le lavoratrici e per i lavoratori dall'impiego dell'AI.

Si consideri, a titolo esemplificativo, la materia della sicurezza sul lavoro. Se, per un verso, l'utilizzo dell'AI può comportare notevoli rischi per la salute fisica e mentale dei lavoratori, per altro verso, le nuove tecnologie possono contribuire alla protezione della salute e della sicurezza degli stessi. In altri termini, il rapporto tra sicurezza sul lavoro ed AI può essere declinato nella duplice e complementare dimensione, da un lato, della sicurezza *dall'AI* (da intendersi quale protezione, da parte del *deployer*/datore di lavoro, dai rischi ingenerati dall'AI), e, dall'altro lato, della sicurezza *attraverso l'AI* (da ricondursi alla riduzione dei rischi mediante il ricorso all'AI, eventualmente assistita dalla *robotica*[26], da parte del *deployer*/datore di lavoro)[27].

Dir. Lav., 2023, 6, p. 130 con nota di P. De Petris, *L'art. 28 L. n. 300/1970 come strumento di effettività del diritto alla "trasparenza algoritmica"*, ove si evidenzia come la Dir. (UE) n. 2019/1152 consentisse l'adozione di regole di miglior favore e come dovesse perciò essere disattesa l'eccezione di incostituzionalità dell'art. 1-*bis*, D.Lgs. n. 152/1997 che la società resistente aveva chiesto al Giudice siciliano di sollevare; Trib. Torino, 5 agosto 2023, in *Arg. Dir. Lav.*, 2024, 1, p. 111, con nota di A.A. Scelsi, *L'informativa sui sistemi automatizzati, se lacunosa, integra gli estremi della condotta antisindacale*; Trib. Torino 12 marzo 2024, *Foro it.*, 2024, 4, I, c. 1282.

[25] Sottolineano, però, la differenza tra la definizione di "sistema di intelligenza artificiale" nell'art. 3, par. 1, n. 1, dell'AI Act e di "strumento decisionale o di monitoraggio integralmente automatizzato" nell'art. 1-*bis*, comma 1, d.lgs. n. 152/1997 S. Ciucciovino, *La disciplina nazionale*, cit., 10; L. Lazzeroni, *Responsabilità sociale d'impresa 2.0 e sostenibilità digitale. Una lettura giuslavoristica*, Usiena Press, Siena, 2024, p. 212 ss.

[26] E. Fiata, *Robotica e lavoro*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 611 ss.; V. Maio, *Il diritto del lavoro e le nuove sfide della rivoluzione robotica*, in *Arg. Dir. Lav*., 2018, 6, p. 1414 ss.; P.E. Pedà, *Il diritto del lavoro nell'era della robotica e dell'intelligenza artificiale*, in *Mass. Giur. Lav*., 2023, 1, spec. p. 88 ss.; L. Zappalà, *Robotizzazione*, in AA.VV., *Lavoro e tecnologie*, cit., p. 194 ss.; D. Gottardi, *Da Frankenstein ad Asimov: letteratura 'predittiva', robotica e lavoro*, in *Labour & Law Issues*, 2018, 4, 2, p. 1 ss.; R. Cingolani, D. Andresciani, *Robot, macchine intelligenti e sistemi autonomi: analisi della situazione e delle prospettive*, in G. Alpa (a cura di), *Diritto e intelligenza artificiale*, Pacini, Pisa, 2020, p. 23 ss.

[27] S. Marassi, *Intelligenza artificiale e sicurezza sul lavoro*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 207 ss., ed ivi molteplici esempi di strumenti che impiega-

Ad analoghe conclusioni si può giungere in riferimento alla c.d. discriminazione algoritmica. Da un lato, infatti, un sistema decisionale di AI, pur presentandosi in apparenza neutro ed obiettivo (in quanto automatico), può dare comunque luogo a forme (dirette e indirette) di discriminazione[28], le quali risultano vieppiù insidiose, in quanto celate dietro l'opacità dell'algoritmo[29]. Dall'altro lato, la garanzia della trasparenza, imposta, dalla normativa nazionale e sovranazionale (v. *supra*), per *ogni* decisione adottata o suggerita dall'algoritmo, può costituire un argine preventivo che distinguerebbe "in positivo" le decisioni automatizzate[30] dalle scelte "interamente umane": laddove, infatti, per queste ultime non sia previsto per legge uno specifico obbligo di motivazione (l'esempio tipico è la selezione di un/a candidato/a da parte di una/un *HR manager*), una determinazione avente una natura o effetti discriminatori potrebbe più difficilmente venire alla luce rispetto a quanto avverrebbe nel caso di uno scrutinio dei CV at-

---

no l'AI nella prospettiva di salvaguardare la salute e sicurezza dei lavoratori. In generale, sul rapporto tra AI e sicurezza sul lavoro, cfr. S. Cairoli, *Intelligenza artificiale e sicurezza sul lavoro: uno sguardo oltre la siepe*, in *Dir. Sic. Lav*., 2024, 2, p. 26 ss.; C. Romeo, *Tutela e sicurezza del lavoro nell'era della intelligenza artificiale: profili biogiuridici*, in *Lav. Giur*., 2024, 5, p. 445 ss.; A. Tardiola, *Tre quesiti sul rapporto tra sicurezza sul lavoro e AI*, in *Lav. Dir. Eur*., 2024, 3, p. 1 ss.; C. Timellini, *Verso una Fabbrica Intelligente: come l'AI invita a ripensare la tutela della salute e della sicurezza dei lavoratori*, in *Var. Temi Dir. Lav.*, 2023, 4, p. 828 ss.; E. Sena, *"Management" algoritmico e tecniche di tutela dei lavoratori tra tutela della "privacy" e sicurezza sul lavoro: quale ruolo per il sindacato?*, in *Arg. Dir. Lav*., 2023, 6, p. 1160 ss.

[28] Sotto questo aspetto, si ritiene che una fase particolarmente delicata sia quella dell'accesso al lavoro, visto che la profilazione individuale attraverso lo *screening* (virtuale e non) dei lavoratori presenta, specie se condotta o "assistita" dall'AI, notevoli rischi per i diritti fondamentali dei candidati e delle candidate: ampiamente, A. Riccobono, *Intelligenza artificiale e limiti al social media profiling nella selezione del personale*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 253 ss.

[29] V., *ex multis*, G. Gaudio, *Le discriminazioni algoritmiche*, in *Lav. Dir. Eur*., 2024, 1, p. 1 ss.; E. Falletti, *Discriminazione algoritmica. Una prospettiva comparata*, Giappichelli, Torino, 2022, spec. p. 260 ss.; M. Peruzzi, *Il diritto antidiscriminatorio al test di intelligenza artificiale*, in *Labour & Law Issues*, 2021, 7, 1, p. 48 ss.; nella letteratura internazionale, cfr., almeno, A. Kelly-Lyth, *Algorithmic discrimination at work*, in *Eur. Lab. Law Journ*., 2023, 14, p. 152 ss.; J. Ludwig, S. Mullainathan, C.R. Sustein, *Discrimination in the age of algorithms*, in *Journal of Legal Analysis*, 2018, 10, p. 113 ss.; P.T. Kim, *Data-Driven Discrimination at Work*, in *Wm. & Mary L. Rev*., 2017, 58, 3, p. 857 ss.; in giurisprudenza, Trib. Bologna 31 dicembre 2020, in *Arg. Dir. Lav*., 2021, 1, p. 771, con nota di M. Biasi, *L'algoritmo di* Deliveroo *sotto la lente del diritto antidiscriminatorio…e del relativo apparato rimediale*; Trib. Palermo 17 novembre 2023, in *Arg. Dir. Lav*., 2024, 3, p. 640, con nota di C. Pareo, *I modelli organizzativi fondati su meccanismi premiali indifferenti alla presenza di fattori protetti sono discriminatori.*

[30] P. De Petris, *La discriminazione algoritmica. Presupposti e rimedi*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 245.

traverso un'AI, sul cui meccanismo interno di funzionamento, invece, il *deployer*/datore di lavoro sarebbe tenuto, in forza della normativa euro-unitaria sull'AI, a dare una piena contezza.

Ancora, non si possono sottovalutare le opportunità derivanti dalla tecnologia in termini di creazione di nuove figure professionali. Si consideri, da una parte, l'ecosistema videoludico ed in particolare il segmento degli *esports*, i quali si presentano come lo sviluppo naturale dello sport tradizionale in un mondo digitalizzato, decorporizzato e, proprio per questo, non assoggettabile *de plano* alle regole che si applicano al lavoro sportivo "non virtuale" [31]. Dall'altro lato, si assiste alla diffusione degli *influencers* e dei *content creators*, i quali, grazie alle capacità creative, alla fedeltà del loro seguito e all'abilità nell'*engagement* del pubblico, si rivelano fondamentali per incrementare la visibilità e il successo delle imprese operanti nei settori più disparati: eppure, anche per queste figure si è ancora lontani da un chiaro statuto di protezione [32], dal momento che la relativa attività si posiziona a cavallo tra la creazione artistica, la prestazione lavorativa e, come da ultimo affermato in una recente pronuncia del Tribunale di Roma, la promozione di prodotti altrui in forza di un rapporto di agenzia [33].

4. *Sfide*

Secondo una visione oltremodo pessimistica, l'AI è destinata a sostituire progressivamente il lavoro umano, il che renderebbe opportuno avviare una riflessione sull'introduzione di un reddito minimo universale [34].

---

[31] F. Santini, *Il microcosmo videoludico: gli* esports, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 689 ss.

[32] L. Torsello, *Le figure dell'influencer e del content creator del web: lavoro o cessione del diritto d'autore?*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 711 ss.; cfr. anche l'approfondimento in *Labour & Law Issues*, 2021, 7, 2, ed ivi i contributi di A. Rota, *I creatori di contenuti digitali sono lavoratori?*, I., p. 1 ss., di P. Iervolino, *Sulla qualificazione del rapporto di lavoro degli influencers*, I., p. 26 ss., e di L. Torsello, *Il lavoro degli influencers: percorsi di tutela*, I., p. 52 ss.; V. Barzotti, I. Jerussi, *Creatori di contenuti digitali, primi passi a sostegno di un nuovo modo di lavorare*, in *Lav. Dir. Eur.*, 2022, 3, p. 1 ss.; cfr., da ultimo, C. Barnard, *The Serious Business of Having Fun: EU Legal Protection for Those Working Online in the Digital Economy*, in *Int. Journ. Comp. Lab. Law & Ind. Rel.*, 2023, 39, 2, p. 125 ss.

[33] Trib. Roma 4 marzo 2024, n. 2615, in *labor*, 16 settembre 2024, con commento di F. Capurro, *La nozione di agente di commercio e la figura dell'influencer: antichi principi, solite ambiguità, inediti equivoci*.

[34] J. Kaplan, *Le persone non servono. Lavoro e ricchezza nell'epoca dell'intelligenza artificiale*, Luiss University Press, Roma, 2016.

Senonché, si è replicato persuasivamente che un approccio neo-luddista si pone in contrasto con un assetto normativo e valoriale che esalta la libertà attraverso il lavoro e non dal lavoro [35], a tacere degli studi che evidenziano una correlazione tra l'utilizzo dell'intelligenza artificiale e l'incremento della produttività del lavoro [36].

In questa prospettiva, se si vuole che l'attuale trasformazione digitale risulti, come in passato, foriera (anche) di nuove opportunità occupazionali, è essenziale insistere sulla (o, meglio, cogliere la sfida della) formazione [37], la quale, come si è efficacemente argomentato, costituisce uno, se non il principale, degli antidoti avverso la temuta "disoccupazione tecnologica" [38].

D'altro canto, la stessa tecnologia può fungere da efficace ausilio nel processo, tanto della formazione, quanto della relativa certificazione.

Da un lato, le attività formative possono avvalersi di inediti "spazi"

---

[35] N. Irti, *Umanesimo del lavoro e civiltà tecnica*, in *Riv. Dir. Civ.*, 2023, 2, p. 203; M.W. Finkin, *Technology and Jobs: The Agony and the Ecstasy*, in *Comp. Lab. Law & Pol. Journ.*, 2019, 41, p. 221 ss.

[36] D. Calderara, *IA e voci variabili della retribuzione: i premi di risultato*, in *Lav. Dir. Eur.*, 2024, 3, p. 3 ss., ed ivi ulteriori richiami.

[37] Cfr. G. Piglialarmi, *L'impatto dell'intelligenza artificiale sul sistema della sicurezza sociale: problemi e prospettive*, in *Lav. Dir. Eur.*, 2024, 3, pp. 8-9. in generale, sulla formazione come chiave del mercato del lavoro nel tempo della digitalizzazione, v. S. Ciucciovino, *Professionalità, occupazione e tecnologia nella transizione digitale*, in *federalismi*, 2022, 8, p. 129 ss.; M. Brollo, *Tecnologie digitali e nuove professionalità*, in *Dir. Rel. Ind.*, 2019, 2, p. 468 ss.; N. De Angelis, *Il vaccino e la cura. La formazione dei lavoratori come strumento per il mercato*, in *Dir. Merc. Lav.*, 2022, 3, p. 611 ss.; G.R. Simoncini, *L'incidenza della rivoluzione digitale nella formazione dei lavoratori*, in *Lav. Giur.*, 2018, 1, p. 39 ss.

[38] F. Lamberti, *Formazione, occupabilità e certificazione delle competenze (tramite blockchain): un'alternativa alla "disoccupazione tecnologica"*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 281 ss.; cfr. Dipartimento per la Trasformazione Digitale e Agenzia Italiana per l'Italia Digitale (AGID), *Strategia Italiana per l'Intelligenza Artificiale*, 2024, 12: "nonostante i numerosi studi di settore condotti negli ultimi anni, non vi è una comune visione dell'impatto che i sistemi di Intelligenza Artificiale avranno sul mondo del lavoro. Non risultano probabilmente attendibili gli scenari catastrofici immaginati sul lungo periodo, mentre più verosimili appaiono, nel medio periodo, gli scenari in cui nuove competenze e professionalità andranno a sostituire quelle esistenti. Una strategia che spinge sull'adozione di sistemi di Intelligenza Artificiale accelererà evidentemente questo (inevitabile) processo di trasformazione. Ecco perché tale processo dovrà essere guidato e regolato, prevedendo – nelle sue azioni strategiche di maggiore impatto la dovuta attenzione al capitale umano e alle persone. Cruciali saranno quindi le iniziative dell'area strategica della Formazione, e più nello specifico i percorsi di *upskilling* e *reskilling*, prestando grande attenzione a preservare e migliorare la qualità del lavoro a valle dell'adozione di sistemi di IA e del riposizionamento del personale".

(*recte*, strumenti), come il Metaverso [39] e di nuove tecniche, come la *gamification* [40], che pure necessita di un attento monitoraggio, volto ad escludere che la dimensione – apparentemente – ludica del fenomeno non celi una forma di "dominio tecnologico" sulla forza lavoro [41].

Dall'altro lato, la certificazione delle *skills* delle lavoratrici e dei lavoratori può giovarsi del registro della *blockchain* [42], come ha dimostrato l'esperimento di Metapprendo, realizzato nell'ambito del CCNL Metalmeccanici 2021 [43]. Tralasciando i profili strettamente tecnici, tale soluzione si segnala per la valorizzazione del ruolo centrale delle parti sociali, non solo nel governo di uno strumento che opera secondo il modello della bilateralità, ma anche nel mercato del lavoro, dal momento che l'istituzione di Metapprendo ha sostanzialmente ovviato, nel settore metalmeccanico, alla mancata istituzione di fascicolo elettronico del lavoratore *ex* d.lgs. 14 settembre 2015, n. 150 [44]. D'altro canto, la disponibilità e la circolazione, sorvegliata e minimizzata, di alcuni dati e informazioni riguardanti le lavoratrici e i lavoratori potrebbe, anche al di fuori del settore metalmeccanico, apportare notevoli benefici in termini di promozione delle potenzialità occupazionali

---

[39] Cfr. già l'apertura nei confronti dei nuovi spazi, anche virtuali, per la formazione dei lavoratori da parte dell'art. 20 del d.l. n. 36/2022 e, da ultimo, l'Interpello n. 3/2024 del Ministero del Lavoro, sul quale v. A. Mariotti, *Realtà virtuale e formazione: il metaverso ha (forse) trovato il suo spazio*, in *AgendaDigitale.eu*, 9 agosto 2024. In tema, cfr., ampiamente, N. De Angelis, *La formazione dei lavoratori nel metaverso e al metaverso*, in AA.VV., *Diritto e universi paralleli*, Edizioni Scientifiche Italiane, Napoli, 2023, p. 107 ss.

[40] R. Pettinelli, *La ludicizzazione della prestazione di lavoro: la* gamification, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 663 ss.; M.A. Cherry, *The Gamification of Work*, in *Hofstra Law Review*, 2012, 40, 4, p. 851 ss.; cfr. anche L. Zappalà, *Gamification*, in AA.VV., *Lavoro e tecnologie. Dizionario del diritto del lavoro che cambia*, cit., p. 112 ss.

[41] A. Perulli, *Capitalismo delle piattaforme e diritto del lavoro. Verso un nuovo sistema di tutele?*, in A. Perulli (a cura di), *Lavoro autonomo e capitalismo delle piattaforme*, Padova, 2018, p. 117.

[42] C. Faleri, *Blockchain*, in AA.VV., *Lavoro e tecnologie. Dizionario del diritto del lavoro che cambia*, cit., p. 33; D. Garofalo, *Blockchain, Smart contract e machine learning: alla prova del diritto del lavoro*, in *Lav. Giur.*, 2019, 10, p. 869 ss.; E. De Marco, *Lavorare per un algoritmo. Blockchain e la rivoluzione copernicana del mercato del lavoro*, in A. Nuzzo (a cura di), *Blockchain e autonomia privata. Fondamenti giuridici*, Luiss University Press, Roma, 2019, p. 215 ss.

[43] M. Faioli, *Artificial Intelligence as the third element of labour relations. Lessons from the Italian reform of workers' classification system*, in A. Lo Faro (a cura di), *New Technology and Labour Law. Selected topics*, Giappichelli, Torino, 2023, p. 75 ss.

[44] S. Ciucciovino, A. Toscano, M. Faioli, *MetApprendo. Il primo caso di Social blockchain su larga scala. Formazione continua, contrattazione collettiva e aspetti di innovazione digitale*, in *Federalismi*, 8 febbraio 2023.

delle persone: infatti, mediante la creazione, la registrazione e l'aggiornamento, attraverso il sistema della *blockchain*, dell'identità professionale di ciascuno, sarebbe possibile incrociare, eventualmente con l'ausilio dell'AI, la domanda e l'offerta di lavoro più efficacemente di quanto si sia potuto fare sinora con i mezzi "tradizionali" [45], vieppiù scongiurando il pericolo della "fine del lavoro" che secondo le visioni più pessimistiche sarebbe inevitabilmente correlato all'avvento dell'intelligenza artificiale.

5. *Conclusioni (aperte)*

È difficile dubitare che il Regolamento Europeo sull'Intelligenza Artificiale (AI Act) rappresenti un punto di partenza e non di arrivo nel processo di adeguamento della normativa, lavoristica e non, all'AI.

Non per nulla, è stato lo stesso AI Act a puntualizzare che alla legge nazionale e alla contrattazione collettiva è consentito approntare una migliore protezione dei lavoratori rispetto a quella garantita dalla cornice europea sull'IA [46].

A prescindere dalla sfida che attende il legislatore interno e le parti sociali (senza, peraltro, dimenticare gli interpreti, chiamati ad un'ardua opera di sistematizzazione di un quadro giuridico ancora in costruzione [47]), è comunque possibile affermare che, nel chiarire che la tecnologia debba sempre e comunque rimanere a servizio della persona (e non viceversa), l'AI Act ha posto una pietra miliare nel corrente dibattito sul rapporto tra uomo e tecnologia, sostanzialmente assecondando la vocazione antropocentrica che segna la cifra dello stesso diritto del lavoro.

Il messaggio è, dunque, non solo di evitare frettolose e decettive umanizzazioni della macchina, ma anche di utilizzare la trasformazione come pungolo per la ricerca di nuovi confini del lavoro e della stessa intelligenza umana, della quale andrebbe (ri)scoperta e valorizzata la dimensione fluida, emotiva e relazionale [48].

---

[45] F. Lamberti, *Formazione, occupabilità e certificazione delle competenze (tramite Blockchain): un'alternativa alla "disoccupazione tecnologica"*, in M. Biasi (a cura di), *Diritto del lavoro e intelligenza artificiale*, cit., p. 281 ss.; cfr. l'art. 26, comma 3, del d.l. n. 60 del 7 maggio 2024, convertito con l. 4 luglio 2024, n. 95 (c.d. decreto coesione del 2024), ove si prevede che "al fine di favorire l'incontro tra domanda e offerta di lavoro, il Sistema Informativo per l'inclusione sociale e lavorativa utilizza, nei limiti consentiti dalla legge, gli strumenti di intelligenza artificiale per l'abbinamento ottimale delle offerte e delle domande di lavoro ivi inserite".

[46] Art. 2, par. 11, dell'*AI Act*.

[47] Così G. Zampini, *Intelligenza artificiale e decisione datoriale algoritmica*, cit., p. 471.

[48] M. Brollo, *Tecnologie digitali e nuove professionalità*, in *Dir. Rel. Ind.*, 2019, 2, p. 468 ss.

Volendo chiudere con una battuta, se davvero un algoritmo fosse in grado di superare gli esami universitari, forse sarebbe il caso, prima che di celebrare o, alternativamente, paventare la capacità del *software*, di dubitare dell'idoneità della prova ad accertare l'avvenuta acquisizione di quel metodo che renderà la persona in grado, attraverso il ragionamento, di adeguarsi ad un futuro contesto che veda il – pressoché inevitabile – superamento di molte delle particolaristiche nozioni che anche una macchina è in grado di apprendere.

Mutatis mutandis, la sfida ruota, dunque, attorno alla valorizzazione della capacità intrinsecamente umana di apprendere ed alla promozione di concezione dinamica del lavoro, la quale consentirebbe di replicare a chi lumeggiasse la convenienza economica della generalizzata sostituzione dell'intelligenza umana con quella artificiale rinviando al vecchio adagio secondo il quale "*the only free cheese is the mouse trap*".

AUMENTARE LA SICUREZZA SUL LAVORO GRAZIE ALL'INTELLIGENZA ARTIFICIALE: LA FILOSOFIA ANTROPOCENTRICA DI INDUSTRIA 5.0

di Caterina Timellini

1. *Introduzione al tema*

L'intelligenza artificiale (codificata con l'acronimo inglese AI)[1] è una metodica di elaborazione dei dati acquisiti estremamente potente, in quanto è in grado di imitare il comportamento dell'essere umano, ad esempio elaborando dati, ricordandoli e collegandoli tra loro, anche con capacità predittive[2].

Applicata ai rapporti di lavoro l'AI genera nuove sfide, di cui una delle principali è aumentare la tutela della salute e della sicurezza dei lavoratori[3]. Nel fare ciò essa si muove seguendo più linee direttrici, le quali sono

[1] Cfr. M. Peruzzi, *Intelligenza artificiale e lavoro. Uno studio su poteri datoriali e tecniche di tutela*, Giappichelli, Torino, 2023, p. 47 ss.; M. Faioli, Robot Labor Law. *Linee di ricerca per una nuova branca del diritto e del lavoro e in vista della sessione sull'intelligenza artificiale del G7 del 2024*, in *Federalismi*, 2024, 8, p. 182 ss.; M. Merone, *Fondamenti di* machine learning *e applicazioni giuridiche*, in R. Giordano, A. Panzarola, A. Police, S. Preziosi, M. Proto (a cura di), *Il diritto nell'era digitale*, Giuffrè, Milano, 2022, p. 1045 ss.; R. Covelli, *Lavoro e intelligenza artificiale: dalla trasparenza alla conoscibilità*, in *Labour & Law Issues*, 2023, 1, p. 91 ss.

[2] Cfr. A. Carcaterra, Machinae *autonome e decisione robotica*, in A. Carleo (a cura di), *La decisione robotica*, Il Mulino, Bologna, 2019, p. 37.

[3] Cfr. D. Garofalo, *L'evoluzione della normativa italiana in materia di tutela della salute e della sicurezza sul lavoro, anche alla luce delle più recenti trasformazioni digitali*, in *Variaz. Temi Dir. Lav.*, 2023, 4, p. 844 ss.; M. Corti, *Il Quadro strategico UE in materia di salute e sicurezza*

riconducibili all'ottica della sorveglianza e del controllo, della prevenzione e dell'inclusione [4].

Sotto il primo aspetto il ricorso ai sistemi di AI può, infatti, consentire una raccolta di grandi quantità di dati in tempo reale, che, analizzati, potrebbero permettere di migliorare la sorveglianza sulla sicurezza del lavoro dei dipendenti.

Sotto il profilo della prevenzione l'automatizzazione di attività lavorative pericolose, usuranti, routinarie o ad elevato rischio, grazie alla possibilità offerta da tali sistemi di inviare segnali tempestivi in presenza di stress, di problemi di salute, di affaticamento, di virus e, addirittura, di errori umani [5], potrebbe costituire una valida risorsa per scongiurare il verificarsi di situazioni pericolose riducendo l'esposizione dei lavoratori ai fattori di rischio.

Ancora, in un'ottica inclusiva, l'utilizzo di cobot potrebbe facilitare enormemente l'accesso al lavoro dei soggetti più svantaggiati, quali i lavoratori in età avanzata o quelli portatori di disabilità [6].

Nelle realtà lavorative di più recente sviluppo, connotate da una destrutturazione del rapporto di lavoro sotto il profilo spazio-temporale [7], con tutto ciò che ne consegue, ad esempio, in termini di diritto alla di-

---

sul lavoro 2021-2027, *ibidem*, 2023, 4, p. 966 ss.; cfr. G. Ludovico, *Nuove tecnologie e tutela della salute del lavoratore*, in G. Ludovico, F. Fita Ortega, T.C. Nahas (a cura di), *Nuove tecnologie e diritto del lavoro. Un'analisi comparata degli ordinamenti italiano, spagnolo e brasiliano*, Milano University Press, Milano, 2021, p. 79 ss.; A. Allamprese, O. Bonardi, *Studio sulle condizioni di lavoro nella logistica: tempo e salute*, in *Dir. Sic. Lav.*, 2020, 2, p. 42 ss.

[4] Cfr. C. Romeo, *L'era degli algoritmi e la sua incidenza nell'ambito della certezza del diritto: un connubio* sospetto, in *Lav. giur.*, 2024, 1, p. 5 ss.; F.V. Ponte, *Intelligenza artificiale e lavoro. Organizzazione algoritmica, profili gestionali, effetti sostitutivi*, Giappichelli, Torino, 2024; F. Bano, *Algoritmi al lavoro. Riflessioni sul "management"* algoritmico, in *Lav. dir.*, 2024, 1, p. 133 ss.; M. Biasi (a cura di), *Diritto del lavoro e Intelligenza Artificiale*, Giuffrè, Milano, 2024; C. Faleri, *"*Management*" algoritmico e asimmetrie informative di ultima generazione*, in *Federalismi,* 2024, 3, p. 217 ss.; U. Gargiulo, *Intelligenza Artificiale e poteri datoriali: limiti normativi e ruolo dell'autonomia collettiva*, in *Federalismi*, 2023, 29, p. 171 ss.; F. Butera, G. De Michelis, *Intelligenza artificiale e lavoro, una rivoluzione governabile*, Marsilio, Venezia, 2024.

[5] Cfr. R. Trezza, *La tutela della persona umana nell'era dell'intelligenza artificiale: rilievi critici*, in *Federalismi.it*, 16/2022, p. 277 ss.

[6] Cfr. A. Cocchi, *Robot collaborativo e sicurezza sul lavoro*, in www.universal-robots.com/blog/robot-collaborativo-e-sicurezza-sul-lavoro/.

[7] Cfr. C. Lazzari, *Lavoro senza luogo fisso, de-materializzazione degli spazi, salute e sicurezza*, in *Labour & Law Issues*, 2023, 9, 1, p. 21 ss.; F. Malzani, *Salute e sicurezza dei lavoratori della* gig economy, in P. Pascucci (a cura di), *Salute e sicurezza sul lavoro*, Franco Angeli, Milano, 2019, p. 45 ss.

sconnessione [8], il ricorso all'intelligenza artificiale (AI) [9] impone di cercare una soluzione di equilibrio tra le esigenze proprie di contesti lavorativi caratterizzati dall'utilizzo della tecnologia digitale e l'interesse a tutelare la salute e la sicurezza dei dipendenti, unitamente alla loro *privacy*.

Per meglio comprendere il mutato quadro di riferimento si può osservare quanto avviene applicando i sistemi di AI alla logistica, settore connotato da un ambiente di lavoro in cui i rischi cui sono esposti i lavoratori sono principalmente identificabili nella caduta dall'alto di un oggetto, nell'urto prodotto da un macchinario in movimento o nell'effettuazione di movimenti sbagliati durante la movimentazione di una merce.

Introdurre sistemi di AI in un settore come quello della logistica, allora, si traduce nell'ottimizzazione del processo logistico, resa possibile dalla robotizzazione della movimentazione delle merci che, attraverso il ricorso a dati e algoritmi [10], riduce la mobilità delle persone e quindi ne diminuisce l'esposizione al rischio. Infatti, nell'area di movimentazione delle merci sono i robot ad operare, mentre gli operatori si posizionano su "porte" che si affacciano sull'area automatizzata, dalle quali prelevano le merci e le posizionano su carrelli che le trasportano sugli automezzi che eseguiranno le consegne.

---

[8] Sia consentito il rinvio a C. Timellini, *Il diritto alla disconnessione nella normativa italiana sul lavoro agile e nella legislazione emergenziale*, in *Lavoro Dir. Europa*, 2021, 4, p. 1 ss.; Id., *La disconnessione bussa alla porta del legislatore*, in *Variaz. Temi Dir. Lav.*, 2023, 4, p. 315 ss.

[9] Cfr. M. Peruzzi, *Intelligenza artificiale, poteri datoriali e tutela del lavoro: ragionando di tecniche di trasparenza e poli regolativi*, in *Riv. Studi Giur.*, 2021, 24, p. 71 ss.; M. Faioli, *Data Analytics, robot intelligenti e regolazione del lavoro*, in *Federalismi.it*, 9/2022, p. 149 ss.; F. Lamberti, *Il metaverso: profili giuslavoristici tra rischi nuovi e tutele tradizionali*, in *Federalismi.it*, 4/2023, p. 205 ss.

[10] Cfr. M. Barbera, *Discriminazioni algoritmiche e forme di discriminazione*, in *Labour & Law Issues*, 2021, 7, p. 1; M.V. Ballestrero, *Ancora sui rider. La cecità discriminatoria della piattaforma*, in *Labor*, 2021, 1, p. 103; A. Topo, *Nuove tecnologie e discriminazioni*, in Atti del XXI Congresso nazionale AIDLASS (Messina, 23-25 maggio 2024) su "*Diritto antiscriminatorio e trasformazioni del lavoro*, in www.aidlass.it; G. Zilio Grandi, *Principio di uguaglianza e divieto di discriminazioni al di fuori del lavoro standard: contratti di lavoro subordinato "atipici" e contratti di lavoro autonomo*, *ibidem*; A. Lo Faro, Algorithmic Decision Making *e gestione dei rapporti di lavoro: cosa abbiamo imparato dalle piattaforme*, in *Federalismi*, 2022, 25, p. 189 ss.; G. Gaudio, *Le discriminazioni algoritmiche*, in *Lav. Dir. Europa*, 2024, 1, 1; N. Lipari, *Diritto, algoritmo, predittività*, in *Riv. trim. dir. proc. civ.*, 2023, 3, p. 721 ss.; M. Faioli, Data Analytics, robot *intelligenti e regolazione del lavoro*, in *Federalismi*, 2022, 9, p. 149 ss.; A. Aloisi, V. De Stefano, *Il tuo capo è un algoritmo. Contro il lavoro disumano*, Bari, Laterza, 2020, p. 77 ss., L. Zappalà, *Informatizzazione dei processi decisionali e diritto del lavoro: algoritmi, poteri datoriali e responsabilità del prestatore nell'era dell'intelligenza artificiale*, in *WP CSDLE "Massimo D'Antona".IT*, 2021, 446, p. 99.

In questo modo la movimentazione fisica da parte dell'operatore si riduce, riducendo l'esposizione dei lavoratori ai rischi fisici, ma al contempo non è automatico il miglioramento della condizione lavorativa, posto che i dipendenti, pur svolgendo un compito a ridotto rischio per l'incolumità fisica, operano in una posizione divenuta fissa, svolgendo un lavoro altamente routinario, caratterizzato da movimenti prevalentemente ripetitivi posti in essere per un tempo prolungato.

Se i rischi tradizionali si riducono significativamente il mutato sistema operativo genera allora nuovi rischi [11], determinando così una vera e propria traslazione da rischi fisici a rischi psicologici e psicosociali, i quali impongono una riflessione sotto il profilo della valutazione dello stress lavoro correlato [12].

Il presente studio si soffermerà, in particolare, sulla tutela della salute e della sicurezza dei lavoratori [13] nell'ipotesi di utilizzo dell'intelligenza artificiale [14] nei rapporti di lavoro, al fine di valutare la tenuta o meno dell'apparato normativo esistente [15] – e quindi del *corpus* normativo fondato sull'art. 2087 c.c. e sul D.lgs. 9 aprile 2008, n. 81 [16] –, alla luce delle più re-

---

[11] Cfr. E. Signorini, *Lavoro e tecnologia: connubio tra opportunità e rischi*, in *Federalismi*, 2023, 29, p. 202 ss.; M. Esposito, *La tecnologia oltre la persona? Paradigmi contrattuali e dominio organizzativo immateriale*, in *The Lab'S Qarterly*, 2020, II, p. 45 ss.

[12] Cfr. A. Tardiola, *Tre quesiti sul rapporto tra sicurezza del lavoro e AI*, in *Lav. Dir. Europa*, 2024, 3, p. 3.

[13] Cfr. S. Cairoli, *Intelligenza artificiale e sicurezza sul lavoro: uno sguardo oltre la siepe*, in *Dir. Sic. Lavoro*, 2024, 1, p. 26 ss.

[14] Cfr. G. Mammone, *Intelligenza artificiale e rapporto di lavoro tra robot e gig economy. Ci salveranno i giudici e l'Europa?*, in *Lav. Dir. Europa*, 2023, 1, p. 2 ss. Più in generale sul rapporto tra AI e diritto, cfr. G. Alpa, *Diritto e intelligenza artificiale. Profili generali, soggetti, contratti, responsabilità civile, diritto bancario e finanziario, processo civile*, Pacini Giuridica, Pisa, 2020; G. Sartor, *L'intelligenza artificiale e il diritto*, Giappichelli, Torino, 2022; A. Santosuosso, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Mondadori Università, Città di Cestello, 2020, p. 26; C. Cerrina Feroni, C. Fontana, E.C. Raffiotta (a cura di), *AI Anthology. Profili giuridici, economici e sociali dell'intelligenza artificiale*, Il Mulino, Bologna, 2022; T.R. Frosini, *L'orizzonte giuridico dell'intelligenza artificiale*, in *Il diritto dell'informazione e dell'informatica*, in *BioLaw Journal*, 1/2022, p. 14.

[15] Cfr. A. Rota, *Sicurezza*, in M. Novella, P. Tullini, *Lavoro digitale*, Giappichelli, Torino, 2022, p. 83 ss.

[16] Cfr. P. Pascucci, *Il Testo Unico sulla salute e sicurezza sul lavoro: spunti di riflessione (a fronte dei cambiamenti in atto) e proposte di modifica*, in M. Tiraboschi (a cura di), *Il sistema prevenzionistico e le tutele assicurative alla prova della IV Rivoluzione industriale*, vol. I, *Bilancio e prospettive di una ricerca*, ADAPT University Press, Modena, 2021, p. 499 ss.; Id., *Dopo il d.lgs. 81/2008: salute e sicurezza in un decennio di riforme del diritto del lavoro*, in P. Pascucci (a

centi innovazioni della tecnologia, riflettendo sulle possibili prospettive *de iure condendo*.

2. L'AI Act

Il nodo dell'intelligenza artificiale risulta di così evidente interesse che sono stati approvati dapprima nella sessione plenaria del 14 giugno 2023 la Proposta di Regolamento del Parlamento Europeo e del Consiglio (COM/2021/206 del 21 aprile 2021) e recentemente il Regolamento (UE) 2024/1969 del Parlamento Europeo e del Consiglio del 13 giugno 2024 (noto come *AI Act*), che stabilisce regole armonizzate sull'AI e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) 20187/1139 e (UE) 2019/20144 e le direttive 2014/90/UE, (UE) 2016/ 797 e (UE) 2020/1828 [17].

Si tratta di una regolamentazione che si colloca nel solco tracciato dal Libro Bianco sull'intelligenza artificiale della Commissione europea del 19 febbraio 2020 [18], dal rapporto dell'Agenzia europea per la sicurezza e la salute nei luoghi di lavoro del 2022 [19] e dall'Agenda digitale 2030 dell'UE (UE Act) [20].

Dalle nuove previsioni emerge con chiarezza la connessione tra intelligenza artificiale e tutela della salute e della sicurezza dei lavoratori, così come risulta esplicitato nel considerando n. 1 del Regolamento sull'AI, ove si riconosce che lo scopo della regolamentazione "*è promuovere la diffusione di un'intelligenza artificiale antropocentrica e affidabile, garantendo nel*

---

cura di), *salute e sicurezza sul lavoro. Tutele universali e nuovi strumenti regolativi a dieci anni dal d.lgs. n. 81/2008*, Franco Angeli, Milano, 2019, p. 11 ss.

[17] Cfr. A. Alaimo, *Il regolamento sull'Intelligenza Artificiale: dalla proposta della Commissione al testo approvato dal parlamento. Ha ancora senso il pensiero pessimista?*, in *Federalismi*, 2023, 25, p. 133 ss.; C. Novelli, *L'*Artificial Intelligence Act *Europeo: alcune questioni di implementazione*, in *Federalismi*, 2024, 2, p. 96 ss. Cfr., inoltre, T. Treu, *La digitalizzazione del lavoro: proposte europee e piste di ricerca*, in *Federalismi*, 2022, 9, p. 193.

[18] V. Commissione europea, *Libro Bianco sull'intelligenza artificiale. Un approccio all'eccellenza e alla fiducia* (COM 2020/65 final), 19 febbraio 2020.

[19] V. European Agency for Safety and Health at Work (EU-OSHA), *Advanced robotics and automatiom: implications for occupational safety and health*, Report, 2022, in www.OSHA.europa.eu, che tra I suoi progetti enuclea proprio l'individuazione di "nuovi sistemi di monitoraggio della sicurezza e salute dei lavoratori".

[20] V. *EU Act: first regulation on artificial intelligence*, in www.europarl.europa.eu, Updated 19.12.2023. Cfr. S. Donato, *L'AI Act europeo è la pria lege al mondo sulle Intelligenze Artificial e si basa sul concetto di rischio per la società*, in www.ddai.it, 11 dicembre 2023.

contempo un elevato livello di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea".

La *ratio* dell'AI Act, pertanto, è da un lato fissare i requisiti specifici di *governance* che i sistemi di AI devono possedere e dall'altro imporre determinati obblighi a chi immette sul mercato, nonché a chi utilizza questo tipo di prodotti[21] al fine di non creare pregiudizi alla salute, sicurezza e ai diritti fondamentali dei lavoratori.

La regolamentazione europea effettua una graduazione del livello di potenziale incidenza di tali sistemi sulla collettività, classificando le applicazioni di AI in quattro livelli di rischio a seconda della gravità e della probabilità del verificarsi di esso[22]. Si hanno così: i sistemi "a rischio inaccettabile", vietati perché in contrasto con i valori fondamentali dell'UE, fatte salve alcune situazioni eccezionali che ne giustificano l'utilizzo, riconducibili essenzialmente a ragioni di ordine pubblico[23]; i sistemi "ad alto rischio"; i sistemi "a rischio limitato" e, infine, i sistemi "a rischio minimo".

I sistemi di AI applicati ai rapporti di lavoro rientrano nei sistemi ad alto rischio, che ai sensi dell'Allegato III includono proprio i sistemi "*destinati ad essere utilizzati per l'assunzione o la selezione di persone fisiche, in particolare per pubblicare annunci di lavoro mirati, analizzare o filtrare le candidature e valutare i candidati*", nonché "*destinati ad essere utilizzati per adottare decisioni riguardanti le condizioni dei rapporti di lavoro, la promozione o cessazione dei rapporti contrattuali di lavoro, per assegnare compiti sulla base del comportamento individuale o dei tratti e delle caratteristiche personali o per monitorare e valutare le prestazioni e il comportamento delle persone nell'ambito di tali rapporti di lavoro*".

Ai sensi dell'art. 6, par. 1, dell'AI Act ai fini dell'immissione sul mercato o della messa in servizio del prodotto i sistemi di AI ad alto rischio sono soggetti ad una valutazione di conformità da parte di terzi, volta a verificare la sussistenza di determinati requisiti, così che l'immissione in commer-

---

[21] Cfr. F. Floridi, *Soft Ethics and the Governance of the Digital*, Yale University – Digital Ethics Center, University of Bologna – department of Legal Studies, 2018.

[22] Cfr. P. Loi, *Il rischio proporzionato nella proposta di regolamento sull'IA e i suoi effetti nel rapporto di lavoro*, in *Federalismi*, 2023, 4, p. 239 ss.; M. Barbera, *"La nave deve navigare". Rischio e responsabilità al tempo dell'impresa digitale*, in *Labour & Law Issues*, 2023, 2, p. 8.

[23] Cfr., sul punto, T. Madiega, H. Mildebrath, *Regulating facial recognition in the EU*, 2021, in http://www.europarl.europa.eu/RegData/etudes/IDAN/2021/698021/EPRS_IDA(2021)698021_EN.pdf.

cio potrà avvenire soltanto dopo che la valutazione di conformità abbia dato esito positivo [24].

Oltre a ciò, si richiede un monitoraggio costante dei sistemi stessi dopo la loro immissione sul mercato, anche attraverso un successivo sistema di gestione dei rischi a norma dell'art. 9, secondo il quale durante il suo intero ciclo di vita il sistema di AI ad alto rischio è soggetto ad un riesame e ad un aggiornamento costanti e sistematici, secondo le fasi indicate al par. 2, lett. a)-d).

In particolare, tramite un sistema di *hard law*, seppur modulato attraverso tecniche di flessibilizzazione secondo una prospettiva di rischio proporzionato, in cui la protezione dei diritti fondamentali viene attuata senza limitare eccessivamente lo sviluppo tecnologico e imporre costi sproporzionati per l'immissione dei sistemi di AI sul mercato, l'AI Act grava i fornitori di sistemi di intelligenza artificiale di obblighi stringenti da rispettare ai fini dell'immissione di tali sistemi sul mercato [25].

Si tratta di una serie di obblighi da adempiere *ex ante*, ossia prima di immettere in servizio un sistema di AI ad alto rischio, che si snodano a norma dell'art. 9 in tre fasi, ossia: nell'identificazione dei rischi, nella stima di essi e, infine, nell'eliminazione, nella riduzione o nella gestione costante di essi.

Ai sensi dell'art. 17, inoltre, si prevede di predisporre un sistema di gestione della qualità, che "*può essere integrato all'interno di un sistema (di gestione della qualità) esistente conformemente agli atti legislativi dell'Unione*".

Successivamente i deployer, ossia gli utilizzatori, tra cui rientrano i datori di lavori, dovranno effettuare la c.d. valutazione d'impatto sui diritti fondamentali, quali il diritto alla salute e alla sicurezza nei luoghi di lavoro, a norma dell'art. 27 del Regolamento sull'AI allo scopo di porre in essere le misure di gestione del rischio più opportune, nel tentativo *in primis* di eliminare o ridurre i rischi in fase di progettazione e di fabbricazione, mitigando l'impatto della tecnologia sulla persona, e, ove il rischio non possa essere eliminato, per attenuarlo e controllarlo [26], ad esempio fornendo un'informativa adeguata agli utenti (e, quindi, anche al datore di lavoro) e, ove lo si ritenga opportuno, anche un'adeguata formazione.

---

[24] Cfr. L. Tosoni, *Intelligenza artificiale, i punti chiave del regolamento europeo*, 21 aprile 2021.

[25] Cfr. P. Loy, *Il rischio proporzionato nella proposta di regolamento sull'IA e i suoi effetti nel rapporto di lavoro*, in *Federalismi.it*, 4/2023, p. 241.

[26] Cfr. I.P. Di Ciommo, *La prospettiva del controllo nell'era dell'Intelligenza Artificiale: alcune osservazioni sul modello* Human In The Loop, in *Federalismi.it*, 19 aprile 2023.

Benché, infatti, nel testo del regolamento non si rinvenga mai un riferimento esplicito alla figura del datore di lavoro, quest'ultimo rientra nella categoria degli utenti e, come tale, dovrà utilizzare i sistemi di IA ad alto rischio in modo conforme alle istruzioni ricevute, nonché sarà tenuto a sua volta a monitorare il funzionamento dei sistemi stessi, in applicazione della regola della sorveglianza umana, da attuare anche per il tramite di personale adeguatamente formato, così come previsto dal nuovo art. 14 del Regolamento sull'AI.

Secondo la regola citata, infatti, i sistemi di IA sono sviluppati e utilizzati come strumenti al servizio delle persone, nel rispetto della dignità umana e dell'autonomia personale, e funzionano in modo da poter essere adeguatamente controllati e sorvegliati dagli esseri umani, il che avviene, ad esempio, introducendo misure correttive o interrompendone tempestivamente l'utilizzo.

3. *La tenuta del sistema normativo alla luce della tutela della salute e della sicurezza dei lavoratori*

La nuova disciplina, che mira a uniformare le legislazioni nazionali al fine di favorire la circolazione dei beni prodotti secondo una corretta etica e "coloritura valoriale" [27], rispettosa dei diritti fondamentali e della dignità della persona, si è sovrapposta all'esistente, ossia (nel nostro ordinamento) al D.lgs. n. 81 del 2008 [28] e all'art. 2087 c.c., ponendo inevitabili quesiti interpretativi circa la convivenza dei due plessi regolativi.

Già l'art. 2087 c.c. consente di adattare l'obbligo di sicurezza all'evoluzione tecnologica in modo pressoché automatico, imponendo al datore di lavoro l'adozione di tutte le misure necessarie per tutelare l'integrità fisica e morale del lavoratore tenendo conto anche dei nuovi sistemi di AI. Si tratta, infatti, di una disposizione di chiusura teleologicamente ispirata, con cui il legislatore impone al datore di lavoro di adottare tutte le "*misure necessarie a tutelare l'integrità fisica e la personalità morale dei prestatori di*

[27] Così A. Alaimo, *Il Regolamento sull'Intelligenza Artificiale: dalla proposta della Commissione al testo approvato dal Parlamento. Ha ancora senso il pensiero pessimistico?*, in *Federalismi.it*, 2023, 25, p. 135.

[28] Si pensi alle norme, in particolare, dedicate ai rapporti tra macchina e persona del lavoratore, quali l'art. 22 sugli obblighi posti a carico dei progettisti, l'art. 23 sugli obblighi posti a carico dei fabbricanti e fornitori e l'art. 24 sugli obblighi a carico degli installatori, oltre ovviamente all'art. 28 sul DVR.

lavoro", tenendo conto della particolarità del lavoro, dell'esperienza e della tecnica [29].

Le misure che devono essere adottate sono complesse, ma soprattutto sono suscettibili di variare nel tempo, a seconda proprio dell'evoluzione tecnologica ed a prescindere da ogni accadimento dannoso al lavoratore.

Il quadro descritto consente di rispondere al quesito iniziale circa la tenuta o meno del quadro normativo interno in tema di tutela della salute e della sicurezza dei lavoratori alla luce dell'odierno utilizzo di sistemi di AI ad alto rischio nei rapporti di lavoro.

La risposta è sicuramente positiva, grazie sia al dinamismo dell'art. 2087, che conferisce alla norma una perdurante attualità [30], sia alle previsioni del D.Lgs. n. 81 del 2008, le quali non sono ancorate ai luoghi di lavoro [31], quanto piuttosto all'organizzazione stessa del lavoro [32], con conseguente adattamento a svariate situazioni lavorative [33].

Infatti, la salute e la sicurezza dei lavoratori che svolgono attività lavorativa in contesti nei quali vengano utilizzati sistemi di AI trovano piena tutela nella procedimentalizzazione dell'obbligo di sicurezza e, particolarmente, nella valutazione dei rischi e nella redazione del relativo DVR.

Certo è che il quadro normativo esistente deve essere integrato in modo

---

[29] Per una riflessione di carattere generale sul contenuto dell'obbligo di sicurezza si rinvia a P. Albi, *Il contenuto dell'obbligo di sicurezza*, in *Variaz. Temi Dir. Lav.*, 2023, 4, 875. Cfr., inoltre, Id., *Adempimento dell'obbligo di sicurezza e tutela della persona. Art. 2087 c.c.*, in F.D. Busnelli (diretto da), *Il Codice Civile. Commentario*, Giuffrè, Milano, 2008.

[30] Cfr. D. Garofalo, *L'evoluzione della normativa italiana in materia di tutela della salute e della sicurezza sul lavoro, anche alla luce delle più recenti trasformazioni digitali*, cit., p. 844.

[31] Cfr. C. Lazzari, *Lavoro senza luogo fisso, de-materializzazione degli spazi, salute e sicurezza*, cit., p. 22 ss.

[32] Cfr. L. Montuschi, *Diritto alla salute e organizzazione del lavoro*, Franco Angeli, Milano, 1989, p. 1 ss.; Id., *La nuova sicurezza sul lavoro. D.lgs. 9 aprile 2008, n. 81 e successive modifiche. Commentario*, Zanichelli, Bologna, 2011; E. Gragnoli, *Articolo 30. Modelli di organizzazione e di* gestione, in C. Zoli (a cura di), I, *Principi* comuni, in L. Montuschi (diretto da), *La nuova sicurezza sul lavoro. D.lgs. 9 aprile 2008, n. 81 e successive modifiche. Commentario*, Zanichelli, Bologna, 2011, p. 392 ss.; G. Natullo, *L'organizzazione delle imprese a tutela dell'integrità psico-fisica dei lavoratori e dei cittadini*, in L. Zoppoli (a cura di), *Tutela della salute pubblica*, Editoriale Scientifica S.r.l., Napoli, 2021, p. 129 ss.; M. D'Onghia, *Remotizzazione del lavoro, relazioni sindacali e tutela della salute dei lavoratori*, in L. Zoppoli (a cura di), *Tutela della salute pubblica*, cit., p. 251 ss.; M. Giovannone, *Modelli organizzativi e sicurezza sui luoghi di lavoro alla prova del Covid-19 e a vent'anni dall'entrata in vigore del d.lgs. n. 231/2001*, in *Dir. Sic. Lav.*, 2022, p. 94 ss.

[33] Cfr. P. Tullini, *Le sfide per l'ordinamento italiano*, in M. Tiraboschi (a cura di), *Il sistema prevenzionistico e le tutele assicurative alla prova della IV Rivoluzione industriale*, cit., p. 333.

virtuoso[34] dalla previsione dei nuovi adempimenti UE, il che avviene, ad esempio, con riferimento proprio agli obblighi del datore di lavoro contemplati *ex* artt. 17 e 28 D.Lgs. n. 81 del 2008 (che permangono, integrati dalla ulteriore garanzia della sorveglianza umana), cui sarà da aggiungere a monte quello della valutazione d'impatto dei sistemi di AI gravante sui *deployer*, con tutto ciò che ne consegue, eventualmente, in termini di formazione.

In concreto il datore di lavoro, una volta individuati i rischi derivanti dall'utilizzo dei sistemi suddetti, dovrà porre in essere le misure volte a prevenirli e/o a ridurli ma, stante la specificità tecnica e l'altissima specializzazione necessarie per affrontare con cognizione di causa questo problema, egli necessiterà dell'ausilio di soggetti[35] che siano adeguatamente specializzati e formati in materia, al fine di adiuvarlo nell'individuazione delle misure da adottare per la prevenzione dei rischi scaturenti dall'adozione dei sistemi di AI.

Un ruolo di rilievo dovrà essere riconosciuto, a norma dell'art. 26, comma 7, dell'*AI Act*, che in questo senso si pone in stretto rapporto anche con la garanzia della trasparenza nel rapporto di lavoro[36], alla consultazione dei rappresentanti dei lavoratori, i quali dovranno essere interpellati "*prima di mettere in servizio o utilizzare un sistema di IA ad alto rischio sul luogo di lavoro*".

---

[34] Così A. Alaimo, *Il Regolamento sull'Intelligenza Artificiale: dalla proposta della Commissione al testo approvato dal Parlamento. Ha ancora senso il pensiero pessimistico?*, cit., p. 148.

[35] Cfr. C. Lazzari, *Figure e poteri datoriali nel diritto della sicurezza sul lavoro*, Franco Angeli, Milano, 2015.

[36] Cfr. C. Timellini, *Quale reale trasparenza nel rapporto di lavoro con gli ultimi adempimenti?*, in *Variaz. Temi Dir. Lav.*, 2023, 2, p. 571 ss.; G. Proia, *Origine, evoluzione e funzioni della trasparenza nei rapporti di lavoro*, in *Mass. Giur. lav.*, 2023, 4, p. 719 ss.; A. Tursi, *Decreto trasparenza: prime riflessioni – "Trasparenza" e "diritti minimi" dei lavoratori nel decreto trasparenza*, in *Dir. rel. ind.*, 2023, p. 1 ss.; M. Corti, A. Sartori, *Il recepimento del diritto europeo in materia di condizioni di lavoro trasparenti e prevedibili e di conciliazione vita-lavoro. Le misure giuslavoristiche dei decreti "aiuti"*, in *Riv. it. dir. lav*., 2022, IV, p. 166; A. Zilli, *La via italiana per condizioni di lavoro trasparenti e prevedibili*, in *Dir. rel. ind*., 2023, I, p. 30 ss.; L. Zappalà, *Appunti su linguaggio, complessità e comprensibilità del lavoro 4.0: verso una nuova procedimentalizzazione dei poteri datoriali*, in *WP CSDLE "Massimo D'Antona.IT"*, 2022, 462, p. 19 ss.; M. Faioli, *Giustizia contrattuale, tecnologia avanzata e reticenza informativa del datore di lavoro. Sull'imbarazzante "truismo" del decreto trasparenza*, in *Dir. rel. ind.*, 2023, I, p. 45 ss.; A. Allamprese, S. Borelli, *L'obbligo di trasparenza senza la prevedibilità del lavoro. Osservazioni sul decreto legislativo n. 104/2022*, in *Riv. giur. lav.*, 2022, 4, p. 671 ss.; R. Rainone, *Obblighi informativi e trasparenza nel lavoro mediante piattaforme digitali*, in *Federalismi*, 2024, 3, p. 280 ss.; M. Peruzzi, *Intelligenza artificiale, poteri datoriali e tutela del lavoro: ragionando di tecniche di trasparenza e poli regolativi*, in *Janus*, 2021, 24, p. 71 ss.

Mentre la precedente proposta di regolamento precisava che tale consultazione, nell'ottica UE, avveniva "*allo scopo di trovare un accordo*", ossia al fine di negoziare e di concertare l'uso dei sistemi predetti[37], la nuova norma omette del tutto tale precisazione. In sostituzione dell'indicazione suddetta ci si limita ora a precisare che "*Tali informazioni sono fornite, se del caso, conformemente alle norme e alle procedure stabilite dal diritto e dalle prassi dell'Unione e nazionali in materia di informazione dei lavoratori e dei loro rappresentanti*".

I rappresentanti dei lavoratori, peraltro, in precedenza si ipotizzava che venissero coinvolti anche quali destinatari degli obblighi di informazione che l'operatore (e, quindi, il datore di lavoro) è tenuto a rendere in ordine agli esiti della valutazione d'impatto sui diritti fondamentali di cui al successivo art. 27, secondo una previsione mutuata dall'art. 35 del GDPR.

Anche tale previsione non è stata mantenuta nel testo dell'*AI Act* e ciò desta sorpresa, in quanto si trattava di previsioni che mostravano un impatto rilevante sul nostro ordinamento giuridico, soprattutto tenuto conto del fatto che esse risultavano inserite all'interno di una proposta di regolamento EU di immediata portata precettiva.

Sotto questo profilo sarebbe opportuno un intervento normativo volto a regolamentare tale obbligo, considerato che il nostro ordinamento giuridico, pur conoscendo la figura del rappresentante dei lavoratori per la sicurezza (RLS), cui sono riconosciuti diritti di informazione, di consegna del DVR e di preventiva consultazione[38], sconta in termini più generali il peso di una partecipazione dei rappresentanti dei lavoratori, non adeguatamente strutturata[39].

---

[37] Per uno studio sul modello tedesco cfr. M. Corti, *Innovazione tecnologica e partecipazione dei lavoratori: un confronto tra Italia e Germania*, in *Federalismi.it*, 17/2022, p. 113 ss. Sul tema della partecipazione cfr., inoltre, L. Angelini, *Rappresentanza e partecipazione nel diritto della salute e sicurezza dei lavoratori in Italia*, in *Dir. sic. Lav.*, 2020, 1, I, p. 105 ss.; A. Ingrao, *Data-Driven management e strategie di coinvolgimento collettivo dei lavoratori per la tutela della privacy*, in *Labour & Law Issues*, 2019, p. 129.

[38] Cfr. L. Menghini, *Le rappresentanze dei lavoratori per la sicurezza dall'art. 9 dello Statuto alla prevenzione del Covid-19: riaffiora una nuova "soggettività operaia"?*, in *Dir. Sic. Lav.*, 1, 2021, p. 1 ss.

[39] Cfr. Aa.Vv., *Rappresentanza collettiva dei lavoratori e diritti di partecipazione alla gestione delle imprese. Atti delle giornate di studio di diritto del lavoro*, Lecce, 27-28 maggio 2005, Milano, 2006; M. Corti, *Sub art. 46, Cost.*, in R. Del Punta, F. Scarpelli (a cura di), *Codice commentato del lavoro*, Ipsoa, Milano, 2020, p. 241 ss.; M. Borzaga, *Un riparto "istituzionalizzato" nella formazione volontaria delle norme collettive di lavoro: l'esperienza tedesca fra Tarifvertrag e Briebsverfassung*, in M. Pedrazzoli (a cura di), *Partecipazione dei lavoratori e contrattazione col-*

Già ai sensi dell'art. 1, comma 2, del D.lgs. 6 febbraio 2007, n. 25, di attuazione della direttiva n. 2002/14/CE che istituisce un quadro generale relativo all'informazione e alla consultazione dei lavoratori, è compito dei contratti collettivi regolamentare i sistemi aziendali di informazione e consultazione allorquando le decisioni dell'impresa comportino "*rilevanti cambiamenti dell'organizzazione del lavoro*" (art. 4, comma 3, lett. c) [40], il che è quanto avviene proprio quando vengono introdotti sistemi di AI.

Tuttavia, da un lato l'informazione e la consultazione delle rappresentanze sindacali non risulta particolarmente sviluppata dalla contrattazione collettiva nazionale di lavoro e, dall'altro, l'intervento di quest'ultima nella regolazione delle misure di prevenzione è per lo più residuale e, comunque, ancillare rispetto alla legge. È principalmente quest'ultima, infatti, a stabilire gli standard minimi, generali ed uniformi di tutela della salute e della sicurezza dei lavoratori.

La fonte legale, d'altro canto, essendo slegata dal contesto aziendale, risente del fatto di non conoscere (o di conoscere in modo non sufficientemente approfondito e mirato) i molti nuovi rischi propri di un modello organizzativo che utilizzi i sistemi di AI.

Ciò implica un ripensamento non solo del ruolo delle parti sociali, ma soprattutto della contrattazione collettiva aziendale [41], la quale è chiamata a cogliere questa nuova "sfida partecipativa" e, pertanto, ad intervenire in termini organizzativi "nella gestione e prevenzione dei nuovi rischi" [42].

In altre parole, ogni decisione datoriale relativa all'introduzione e all'utilizzo di sistemi di AI che vada ad incidere sull'organizzazione o sulle condizioni del lavoro dovrebbe essere frutto di un confronto preventivo

---

lettiva nell'impresa, Franco Angeli, Milano, 2021, p. 71 ss.; R. Santagata De Castro, *Sistema tedesco di codeterminazione e trasformazioni dell'impresa nel contesto globale: un modello di ispirazione per Lamborghini*, in *Dir. rel. ind.*, 2020, p. 421 ss.; M. Biasi, *Il nodo della partecipazione dei lavoratori in Italia. Evoluzioni e prospettive nel confronto con il modello tedesco ed europeo*, Egea, Milano, 2013, p. 46 ss.

[40] Cfr. M. Napoli (a cura di), *L'impresa di fronte all'informazione e consultazione dei lavoratori (d.lgs. 6 febbraio 2007, n. 25)*, in *Le nuove leggi civile commentate*, 2008, p. 843 ss.; C. Zoli, *i diritti di informazione e di c.d. consultazione: il d.lgs. 6 febbraio 2007, n. 25*, in *Riv. it. dir. lav.*, 2008, I, p. 161 ss.; F. Lunardon (a cura di), *Informazione, consultazione e partecipazione dei lavoratori*, Ipsoa, Milano, 2008.

[41] Cfr. L. Imberti, *La contrattazione collettiva aziendale di fronte alle sfide della rivoluzione digitale e ai processi di cambiamento organizzativo*, in *Federalismi.it*, 25/2022, p. 161 ss.

[42] Così M. Tiraboschi, *Nuovi modelli della organizzazione del lavoro e nuovi rischi*, in *Dir. Sic. Lav.*, 2022, 1, p. 153. Cfr., inoltre, E. Massagli, *Intelligenza artificiale, relazioni di lavoro e contrattazione collettiva. Primi spunti per il dibattito*, in *Lav. Dir. Europa*, 2024, 3, p. 1 ss.

con le rappresentanze sindacali, che si spinge oltre il livello meramente operativo e organizzativo, per arrivare a involgere "il livello strategico delle decisioni in ordine all'innovazione tecnologica e organizzativa", anche eventualmente ridisegnando i processi organizzativi e produttivi [43].

La prospettiva tracciata, di una partecipazione non più solo a valle, ma a monte, già in termini strategici, pare avere il pregio di attribuire correttamente alle organizzazioni sindacali un ruolo attivo nel sistema prevenzionistico, in quanto garanzia di quella competenza e di quella specializzazione di cui il sistema stesso ha necessità, nonché rappresenta l'applicazione pratica di una concretizzazione di tutela, da intendersi quale maggiore prossimità nei confronti dei lavoratori, certamente utile in un campo complesso e diversificato da settore a settore quale è quello della tutela della salute e della sicurezza dei lavoratori [44].

Senza contare, peraltro, che mentre le misure prevenzionistiche introdotte da norme di legge sono maggiormente esposte ad un concreto rischio di obsolescenza, la contrattazione collettiva unitamente ad un nuovo assetto di relazioni sindacali in azienda, grazie anche ai nuovi soggetti emersi nel contesto pandemico [45] e rimasti operativi anche dopo la fine dell'emergenza per gestire, ad esempio, le fasi di transizione, quale quella digitale [46], risultano maggiormente recettivi e, pertanto, in grado di rispondere con più prontezza e agilità ai nuovi rischi derivanti dall'evoluzione tecnologica [47].

Tutto ciò, peraltro, apre uno scenario ulteriore, ossia quello della tutela giudiziale in ipotesi di mancata informazione e/o consultazione a monte delle rappresentanze dei lavoratori. Tale condotta omissiva, infatti, potrebbe qualificarsi in termini di condotta antisindacale, con conseguente via libera alla proponibilità di ricorsi *ex* art. 28 dello Statuto dei lavoratori,

---

[43] V. M. Corti, *Innovazione tecnologica e partecipazione dei lavoratori: un confronto tra Italia e Germania*, cit., p. 122.

[44] In questo senso, M. Corti, *Innovazione tecnologica e partecipazione dei lavoratori: un confronto tra Italia e Germania*, cit., p. 123, auspica un intervento legislativo a supporto dell'iniziativa degli attori sociali, in attuazione dell'art. 46 Cost.

[45] Si tratta dei Comitati aziendali, cui partecipano lo stesso RLS e i rappresentanti sindacali, dei Comitati territoriali e degli organismi paritetici, cfr. C. Frascheri, *Il modello partecipativo al tempo del Covid-19: quali "eredità" positive per il futuro?*, in *Igiene & Sicurezza sul Lavoro*, 2020, p. 505 ss.

[46] Cfr. M. Peruzzi, *Intelligenza artificiale e lavoro. Uno studio su poteri datoriali e tecniche di tutela*, cit., p. 81.

[47] Cfr. P. Tullini, *Le sfide per l'ordinamento italiano*, cit., p. 335.

al fine di ottenere la rimozione degli effetti della condotta pregiudizievole sui rapporti di lavoro [48].

La considerazione appena svolta sui rimedi giudiziali esperibili, in conclusione, porta a svolgere una valutazione complessiva finale sull'apparato sanzionatorio previsto dalla Proposta di Regolamento sull'AI.

Quest'ultimo non pare presidiare in modo adeguato il rispetto delle nuove previsioni UE, in quanto, al di là delle ipotesi di infrazioni più gravi per le quali vengono espressamente previste delle sanzioni, negli altri casi l'art. 71 demanda agli Stati membri l'individuazione di sanzioni "*effettive, proporzionate e dissuasive*".

4. *Questioni aperte e prospettive* de iure condendo

Ora, se si vorrà rendere effettiva la tutela bisognerà intervenire direttamente sul testo del Regolamento, integrando le previsioni sull'apparato sanzionatorio, in modo da introdurre una disciplina generale e armonizzata che non lasci dei margini di discrezionalità all'interno dei singoli sistemi, in quanto eventuali sanzioni più morbide potrebbero, alla fine, minare l'effettività delle previsioni *de quibus* [49].

Le perplessità manifestate traggono origine dall'art. 99 dell'*AI Act*, il quale stabilisce solamente che "*In conformità dei termini e delle condizioni di cui al presente regolamento, gli Stati membri stabiliscono le norme relative alle sanzioni e alle altre misure di esecuzione, che possono includere avvertimenti e misure non pecuniarie*".

Il recente Regolamento, così come la precedente Proposta di Regolamento, risulta inadeguato a garantire un efficace ed autosufficiente supporto sanzionatorio di immediata vigenza, a causa della scelta fatta di privilegiare il meccanismo della delega alle varie autorità nazionali competenti [50].

L'altro aspetto di criticità risiede nel perimetro geografico limitato dell'*AI Act*, essendo quest'ultimo ristretto ai soli Paesi dell'Unione Europea.

Se da un lato la primogenitura dell'UE nel creare nuove disposizioni è sicuramente lodevole, dall'altro lato tale normativa, essendo limitata solo a

---

[48] Così M. Peruzzi, *Intelligenza artificiale e lavoro. Uno studio su poteri datoriali e tecniche di tutela*, cit., pp. 92-93.

[49] Così A. Alaimo, *Il Regolamento sull'Intelligenza Artificiale: dalla proposta della Commissione al testo approvato dal Parlamento. Ha ancora senso il pensiero pessimistico?*, cit., p. 145.

[50] Cfr. C. Romeo, *Tutela e sicurezza del lavoro nell'era della intelligenza artificiale: profili biogiuridici*, in *Lav. giur.*, 2024, 5, p. 448.

pochi Paesi è manifestamente insufficiente a contrastare l'invasività di un fenomeno di portata globale.

Questo significa che la via è stata instradata, ma la strada da percorrere appare ancora molto lunga, soprattutto se ci si pone una domanda: è possibile coniugare i sistemi di AI con gli obiettivi del benessere e della sicurezza sul lavoro[51]? Ai posteri l'ardua sentenza.

[51] Più in generale sulla regolamentazione giuridica dei sistemi di AI cfr. M. Luciani, *Può il diritto disciplinare l'Intelligenza Artificiale? Una conversazione preliminare*, in *Bilancio Comunità Persona*, 2023, 2, p. 10 ss. Cfr., inoltre, V. Bavaro, *Lavoro subordinato e lavoro autonomo nell'era digitale: il problema della libertà del lavoro*, in P. Passanti (a cura di), *La dignità del lavoro nel cinquantenario dello Statuto*, Franco Angeli, Milano, 2021, p. 143.

INTELLIGENZA ARTIFICIALE E (IR)RESPONSABILITÀ GESTORIA: CENNI MINIMI [*]

di Matteo Rescigno

1. Il mio intervento si prefigge di svolgere alcune considerazioni sparse, ma confido non disordinate, sull'impatto che lo sviluppo dell'intelligenza artificiale è suscettibile di incidere sulla disciplina oggi esistente della responsabilità degli amministratori e del *management* con particolare riferimento al compimento delle scelte gestorie. Impatto che sembra già certo e che sarà ancor più rilevante nel prossimo futuro.

Il tema che tratterò non si occuperà, se non marginalmente, delle numerose tematiche che pure interrogano i doveri degli amministratori con riferimento alle regole, per dir così, esterne alle scelte di gestione e che riguardano, per fare qualche esempio sul quale si tornerà, le attuali previsioni in tema di tutela della *privacy* in relazione ai comportamenti delle imprese, oppure ai doveri che le imprese saranno chiamate ad assolvere in relazione alle regole recentemente fissate nella normativa comunitaria sull'AI.

L'oggetto di questa rapida riflessione intende invece esaminare se il crescente utilizzo da parte delle imprese di strumenti algoritmici e di AI (terminologia che userò, forse erroneamente, in modo promiscuo) per valutare e decidere le scelte gestorie della società implichi o meno una rilettura delle regole sulla responsabilità degli amministratori, e se sì, di quale tipo.

Non si tratta, ovviamente, di considerazioni particolarmente originali e più di un contributo negli ultimi anni ha esaminato questo tema anche con maggior attenzione e cura di quel che potrò svolgere in questo breve scritto. Mi sembra però, quantomeno e per dare un senso alle mie parole, che il progressivo uso di strumenti di AI per determinare le scelte strategiche

[*] Il contributo riproduce l'intervento svolto il primo dicembre 2023 al Convegno di Dipartimento "Cesare Beccaria" e conserva un tono discorsivo che si ritrova in molti passaggi del testo scritto: di ciò si chiede venia al lettore

della società da parte degli amministratori, come pure le scelte gestorie di minor rilievo o meramente attuative delle scelte di vertice (la gestione, *tout court*) possa incidere qualitativamente sulle regole che governano il diritto della responsabilità degli amministratori.

2. A tal proposito però vorrei ricordare a me stesso alcuni passaggi evolutivi di queste regole per comprendere come si giunge oggi all'appuntamento con la gestione della società affidata in tutto o in parte all'AI.

Se volessimo ripercorrere in un veloce volo la storia delle regole che hanno presidiato la disciplina della responsabilità degli amministratori delle società per azioni (e soprattutto quelle di maggiori dimensioni che fanno appello al mercato) dal codice civile ad oggi, si potrebbe delineare la seguente evoluzione:

a) la fase iniziale, in cui la responsabilità degli amministratori si misurava sul rispetto del canone della diligenza del mandatario, su base tendenzialmente solidale. Il paradigma della responsabilità si risolveva in una clausola generale che la giurisprudenza e la dottrina avevano riempito, da un lato, costruendo il rapporto società/amministratori quale contratto "atipico" e dall'altro lato, rafforzando la clausola generale della diligenza del mandatario applicando non già la diligenza del buon padre di famiglia ma quella professionale.

L'amministratore era "solo" con la sua clausola di diligenza professionale, riempita dalla valutazione della concreta fattispecie sottoposta all'interprete e accompagnato dalla *business judgment rule* (che tutt'ora ci accompagna in un contesto molto mutato), vera clausola generale della corretta gestione degli amministratori, che escludeva ed esclude che si possa sindacare la scelta gestoria in ragione dei suoi effetti, ma essenzialmente sul piano della diligente "istruttoria" *ex ante* della scelta di gestione. In questa chiave vanno lette le sentenze che talvolta affermano che il sindacato sulle scelte gestorie è possibile solo ove vi sia una scelta di puro azzardo. Là dove in realtà, specie nelle grandi società ad azionariato diffuso, la sanzione si risolveva piuttosto nella sfiducia degli investitori istituzionali a fronte delle scelte errate di gestione degli amministratori.

In questo scenario e ai fini qui tratteggiati, non v'era, tendenzialmente, altro che l'agire dell'amministratore e la sua scelta gestoria: certo nell'ambito della ricostruzione delle regole entravano in gioco riflessioni e valutazioni tratte dalla giurisprudenza e dalla dottrina per cui l'amministratore avrebbe dovuto, in generale e a monte delle specifiche e concrete fattispecie, ben organizzare la società per non incorrere in scelte censurabili che si

rivelassero tali per inadeguatezza della struttura societaria. Ma, in questo scenario di base, la regola che disciplinava la fattispecie qui delineata era posta in relazione alla diligenza nell'istruttoria della scelta che integrava il paradigma della (ir)responsabilità gestoria;

b) dopo la riforma del 2003, preceduta, come spesso accade, da regole fissate in particolari settori (banche, assicurazioni, intermediazione finanziaria) si apre la stagione, ancora in essere, del ruolo degli assetti organizzativi adeguati, in continua e costante evoluzione; assetti adeguati identificati anche per specifiche aree di rilevanza (tuttora in evoluzione normativa: si pensi alle recenti declinazioni del tema nel codice della crisi) e con potenzialità tuttora suscettibili anche con riguardo ai temi qui in discussione.

In questo scenario normativo diventa centrale nella valutazione della responsabilità gestoria degli amministratori il rigoroso rispetto delle regole che reggono l'intero assetto organizzativo della società coinvolto nell'adempimento di tali doveri (amministratori esecutivi, amministrativi non esecutivi, organi di controllo). La società – e così i suoi esponenti, dall'amministratore delegato ai dirigenti apicali – devono curare, verificare, sorvegliare l'adeguatezza degli assetti societari. La scelta del legislatore, in questo scenario normativo, – ferme restando specifiche fattispecie di violazione delle regole in materia societaria – disegna il diligente adempimento dell'obbligo di porre in essere tali assetti societari adeguati quale ipotetico presupposto a monte di ogni diligente gestione. Si potrebbe dire, in poche parole, che l'assenza di assetti societari adeguati che permettano agli amministratori la corretta valutazione delle scelte gestorie fa sì che la scelta gestoria si riveli, potrebbe dirsi per definizione, errata e suscettibile di generare danni alla società, con la conseguenza che gli amministratori della società sono, in tal caso, esposti alla sanzione della revoca della carica e al risarcimento del danno. Non a caso, infatti, si discute animatamente fra gli studiosi se la scelta degli assetti adeguati sia protetta dalla *business judgment rule* o meno.

Ciò detto, però, il tema che a me pare più interessante con riguardo alle questioni in esame e ai fini di queste brevi notazioni, è quella che richiama un'altra domanda ed è quella opposta a quella appena tratteggiata.

Se la società è dotata di assetti adeguati può l'amministratore vedersi restringere l'ambito della sua responsabilità o addirittura escluderlo? In altre parole, l'adempimento degli amministratori dei doveri inerenti agli assetti adeguati esaurisce o limita il compito e i doveri ai quali essi sono tenuti, oppure vi sono soglie di responsabilità ulteriori che incidono comunque sui doveri gestori degli amministratori indipendentemente dal rispetto del dovere di porre in essere assetti organizzativi adeguati?

Se la risposta al quesito fosse positiva (o nei limiti in cui esso possa esserlo) è evidente che la responsabilità degli amministratori si "allontanerebbe" specie per quel che riguarda gli amministratori non esecutivi. Curati e verificati gli assetti adeguati e il loro corretto funzionamento la soglia di responsabilità gestoria sembra alzarsi lasciando ai vertici della catena gestoria l'onere di rispondere per eventuali scelte loro non imputabili.

E così, di base, la diligenza nell'istruttoria delle scelte gestorie come paradigma della (ir)responsabilità degli amministratori potrebbe essere sostituito o integrato da quello della adeguatezza degli assetti organizzativi societari. Questo scenario potrebbe condurre a legittimare, dunque, l'adozione di uno *standard* della corretta gestione degli amministratori che si collochi a monte, dettato dall'adeguatezza degli assetti, con un "distacco" fra gestione e responsabilità e dunque con una maggior area di insindacabilità delle scelte gestorie per aver ben correttamente strutturato l'assetto organizzativo della società e non solo per aver diligentemente gestito.

Ciò detto, se l'esito di una scelta gestoria ha il suo fondamento in analisi e valutazioni condotte da un sistema organizzativo adeguato lo scalino per giungere a una responsabilità in capo agli amministratori sale. Non potrà dirsi che la struttura societaria era inadeguata e quindi non potrà porsi quale antecedente causale del danno: si allarga, se si vuole utilizzare tale paradigma, la portata della *business judgment rule* e si rende più netto il distacco fra amministrazione della società che non adempie al paradigma degli assetti adeguati come fonte della gestione "sostituita" dalla struttura organizzativa e ai suoi paradigmi di efficienza.

Quali siano e come siano codificati in concreto tali paradigmi non è oggetto di regole astratte nel sistema societario in quanto sono legati alle specifiche fattispecie operative e alle relative scelte gestorie, ma sul piano generale esse esistono e fanno parte delle regole specifiche della responsabilità degli amministratori e degli organi di controllo in relazione all'oggetto della società e delle scelte gestorie;

c) la riforma del 2003, infine, ha introdotto un concetto di diligenza nell'ambito della responsabilità degli amministratori non più legato alla generica diligenza del mandatario, ma alle specifiche competenze dell'amministratore e della natura dell'incarico. Regola che si lega, specie per le società di grandi dimensioni e ancor più per le quotate, all'evoluzione della struttura del consiglio di amministrazione articolata sui vari comitati di settore, ciascuno con le proprie peculiarità, caratteristiche e responsabilità, articolazione che trova, come vedremo, nelle elaborazioni sul nostro tema, significativo spazio per "strutturare" anche la materia e le tematiche di cui stiamo discutendo.

Per anticipare quel che dirò poi in seguito, l'esito più sicuro dell'avanzata dell'AI sotto il profilo delle scelte e delle responsabilità degli amministratori è il frequente ricorso negli assetti organizzativi delle società nel sistema delle competenze societarie, dell'istituzione di *tech committee* in seno al consiglio di amministrazione. Nuovamente questa evoluzione, a me sembra, specie per gli amministratori non dotati di specifiche competenze, un modo di conseguire un allontanamento dalla loro responsabilità gestoria. La domanda appena svolta per il ruolo degli assetti adeguati vale – e forse maggiormente – anche per questi temi: è il procedimento, non la scelta che alla fine conta per misurare il corretto esercizio dei doveri degli amministratori, specie di quelli non dotati di specifiche competenze.

3. Ed entriamo nella nuova era: le scelte gestorie sono già affidate o potranno essere affidate non più a "uomini nati da ventre di donna", come diceva Ascarelli a fronte della natura della "personalità giuridica", ma da algoritmi e AI, sino a ipotizzare (lo abbiamo sentito e lo risentiremo a lungo nelle future riflessioni degli interpreti su questo punto) il c.d. Roboboard. Correlativamente, emerge subito la domanda più evidente alla quale deve darsi risposta: su chi gravano e in base a quali parametri le responsabilità degli amministratori nelle varie forme e scelte influenzate, se non proprio superate, dalle scelte dell'algoritmo.

Non è necessario richiamare qui la variegata schiera di piattaforme, strumenti, algoritmi, AI, che partecipano, in vario modo, vuoi a svolgere attività un tempo riservate agli umani in tempi e accuratezza di gran lunga superiori, vuoi a essere invece la fonte concorrente o addirittura autonoma delle scelte gestorie degli amministratori.

Al riguardo si può dare per assodato che il diritto delle società e soprattutto le regole che presiedono alle scelte gestorie e alle relative responsabilità devono ora fronteggiare l'avanzare di questi strumenti e debbono rispondere a importanti domande sotto diversi profili.

La domanda specifica che vorrei sottoporre all'attenzione del lettore in queste brevi note è la seguente: le regole sulla responsabilità gestoria degli amministratori a fronte di questi nuovi strumenti già all'opera o prefigurati nel prossimo futuro sono capaci di adattarvisi oppure – qualitativamente – questa evoluzione si rivela non compatibile, almeno in parte, con gli strumenti a disposizione dell'interprete per regolare obblighi e responsabilità nell'esercizio dell'attività di impresa.

4. Come accennavo, il tema è stato affrontato e, almeno nell'esperienza italiana, prevale un approccio che potrebbe dirsi "adattativo". Le regole su

doveri e correlate responsabilità degli amministratori vengono rilette – ma non superate o diversamente regolamentate – dall'utilizzo di strumenti decisori anche inerenti scelte gestorie che si rivelano basati su algoritmi o su strumenti di AI. L'approccio seguito dagli amministratori – in assenza di norme specifiche in materia – in buona sostanza ha lasciato libere le società di inserire le regole in questione in modo autonomo, senza alcun intervento normativo cogente ma in coerenza con le specifiche valutazioni gestorie individuate in base alla scelta degli amministratori ai quali spetta indicare se e entro quali limiti inserire nell'ambito della struttura organizzativa della specifica società regole sulla responsabilità degli amministratori in relazione al possibile uso di algoritmi e di strumenti di AI. Peraltro, va subito osservato che il tema è stato anche oggetto di specifici interventi del *codice di corporate governance* che p.es. ha raccomandato in materia alle società quotate di valutare l'opportunità di utilizzare di *tech committees* a riprova del fatto che tale scelta potrebbe esser ritenuta necessaria o comunque opportuna in sede di scelte organizzative della società da parte degli amministratori.

5. Pur se recenti, le riflessioni della dottrina, stimolate dalla crescente importanza di questi criteri, hanno già affrontato molti temi sensibili in materia, proprio seguendo quello che potrebbe definirsi ottimismo della ragione. L'algoritmo, l'AI – come appena detto – nelle loro varie applicazioni concrete – sono ritenute suscettibili di esser inseriti nel quadro organizzativo e regolamentare della società destinato alla disciplina degli assetti adeguati societari la cui rilevanza risulta certamente enfatizzata dalle considerazioni appena accennate. Si tratta così – in queste interessanti e astrattamente condivisibili riflessioni – di (ri)leggere gli obblighi di diligente gestione e di responsabilità (per usare una locuzione *d'antan*) a fronte delle specifiche fattispecie frutto dell'evoluzione dell'attività di impresa e delle scelte gestorie a fronte dell'evoluzione tecnologica in atto suscettibile di mutare i paradigmi normativi in materia.

6. I temi più rilevanti riguardano ovviamente i nuovi doveri e le nuove responsabilità che derivano dall'ingresso dell'AI e degli algoritmi nella gestione dell'impresa, non più soltanto come meri strumenti operativi ma come strumenti ai quali possono essere affidate scelte gestorie nell'esercizio del governo dell'impresa e che, necessariamente, si interrogano e si rivolgono ai doveri e alle responsabilità dei gestori. Con scenari, inoltre, di ancora maggior rilevanza, come si diceva, là dove si guardi all'inserimento nel consiglio di amministrazione, in tutto o in parte, dell'AI.

Questo scenario futuribile, se non proprio attuale, è oggetto di numero-

si dibattiti, primo fra tutti quello della sua stessa fattibilità tecnologica, alla quale segue, in caso di risposta positiva al primo quesito, se ciò possa dirsi lecito e coerente con le regole attuali o prospettiche in materia.

7. In linea astratta la dottrina ha delineato i possibili scenari dell'utilizzo dell'algoritmo e dell'AI nelle scelte gestorie individuando alcuni possibili scenari operativi che interrogano le scelte degli amministratori, ma anche quelli dell'interprete.

In particolare, e in ordine di concreta incidenza sul loro ruolo nelle scelte gestorie, si può ipotizzare che:

a) algoritmi e AI *possono* essere inseriti nell'ambito della costruzione degli assetti adeguati sulla base di una diligente valutazione da parte degli amministratori e sottoposti al controllo degli amministratori non esecutivi e del collegio sindacale, ma non può essere censurata di per se la scelta di rivolgersi ad assetti organizzativi adeguati diversi che non prevedano questa scelta, salvo che questa si riveli specificamente inadatta in concreto alla specifica fattispecie sottoposta alla scelta gestoria degli amministratori;
b) algoritmi e AI *debbono* essere inseriti e considerati nella costruzione degli assetti organizzativi adeguati perché essi rappresentano lo *standard* necessario di tali assetti in ragione delle caratteristiche, dell'oggetto e della dimensione dell'impresa. Gli amministratori cioè, devono specificamente considerare l'utilizzazione di algoritmi e degli strumenti di intelligenza artificiale per delineare assetti adeguati della società e correlate scelte gestorie;
c) infine, e se si consente l'enfasi, come è stato già rilevato, un domani non troppo lontano la presenza di comitati di esperti all'interno del *board* sarà uno strumento indefettibile per correttamente orientare le scelte gestorie, in quanto sempre più affidate a algoritmi e AI. Correlativamente per individuare la corretta l'applicazione delle regole di responsabilità gestorie si dovrà passare attraverso la valutazione dell'algoritmo.

8. Numerosi sono i possibili profili di rilevanza con riguardo ai doveri che sono suscettibili di essere valutati alla luce dell'ingresso dell'AI.

Il primo aspetto – e in questo ambito forse il meno interessante – è quello relativo ai doveri degli amministratori che facciano uso di AI o di strumenti algoritmici per effettuare operazioni gestorie, di rispettare le regole specifiche di cui si è appena detto e che possono limitare tale uso.

Si tratta, per così dire, dei temi inerenti i limiti esterni all'uso dell'AI

traibili dalle regole esistenti e di quelle che in futuro che verranno implementate al fine di tutelare chi entri in contatto con l'impresa. Il regolamento europeo sulla *privacy*, ma anche e soprattutto, il regolamento sull'AI appena varato sono gli esempi più evidenti.

L'assetto normativo, in questo caso, vede ormai definito un sistema precettivo completo a tutela degli interessi delineati in sede comunitaria. Per fare un esempio rilevante ai fini di queste note si pensi ai livelli di rischio dell'AI dettati dal regolamento europeo, con le conseguenti previsioni e relativi rischi di discriminazione per l'accesso al credito derivate dall'analisi e dall'uso dei dati ai quali hanno accesso tramite strumenti di AI o di analisi algoritmica.

Significativo in questo senso il suggerimento della dottrina [1] di trovare, proprio in questi due testi normativi appena menzionati, i principi essenziali che si pongono come base indefettibile della disciplina e così adottando una clausola generale dell'uso dell'AI che mantenga una permanente centralità del contributo umano nella strutturazione e monitoraggio delle tecnologie impiegate e dunque della gestione della società digitalizzata (si veda art. 22 reg. *privacy*) nonché una lettura estensiva delle fattispecie di rischio AI, che solo limitatamente vengono riferite a materie relative all'esercizio dell'attività di impresa.

Ma si tratta – come detto per i temi qui in esame – di limiti esterni all'uso dell'AI a tutela dei terzi e con responsabilità derivanti dai rischi dell'AI in chiave di sanzioni e risarcimento nel caso di violazione delle regole per le quali la fonte della responsabilità non sorge dall'operato degli amministratori nell'adottare scelte gestorie, di vertice o operative ma dal mancato rispetto di regole esterne.

9. Il secondo aspetto – e si entra qui nei temi più specifici e di maggior rilievo con riferimento alle scelte gestorie e alla *corporate governance* – riguarda le scelte degli amministratori sull'utilizzo di sistemi algoritmici o di AI.

Da questo punto di vista si fa strada in dottrina, in coerenza anche con le regole generali in tema di diligente gestione – che l'uso di tali strumenti è certamente adottabile dagli amministratori secondo il paradigma della *bjr* ovvero, a seconda delle preferenze generali in materia, obbligatoriamente, ove ne siano rinvenibili i presupposti. In tale ottica la scelta degli amministratori di affidarsi a sistemi algoritmici o di AI per effettuare determinate operazioni può essere certamente adottata là dove gli amministratori repu-

[1] Vedi il generale e corretto approccio di Abriani, Schneider, *Diritto delle imprese e intelligenza artificiale*, Bologna, 2021

tino che ciò possa comportare una gestione dell'impresa più efficiente alla luce dei principi prima indicati.

Per fare qualche esempio rilevante, funzioni che implicano l'acquisizione e rielaborazione di numerosi dati possono certamente rappresentare un modello efficiente per regolare il settore del controllo interno di una impresa.

Ma le potenzialità dell'uso dell'algoritmo non si arrestano a questo livello. È stato per esempio rilevato che l'utilizzo di sistemi algoritmici possano addirittura rendere più efficiente la scelta della nomina degli amministratori selezionando un soggetto piuttosto che un altro.

Appare evidente che, in questo caso, se si volesse approfondire questo tema emergerebbe immediatamente un tema di compatibilità con la disciplina della nomina degli amministratori in capo all'assemblea alla luce dell'assetto normativo societario: si pensi, p.es. al modello dettato per le società per azioni quotate che non potrebbe applicare l'ipotetico responso dettato dall'AI. E altri esempi poterebbero evidenziarsi nei quali le regole societarie potrebbero urtare contro un modello predittivo di efficienza gestionale che non trova riscontro ovvero sia incompatibile con le regole societarie o ovvero sull'efficienza gestionale.

Sempre in questo ambito la scelta di adottare quello che è stato definito *management algoritmico* trova spunti rilevanti di riflessione sulla struttura aziendale inerente appunto alla semplificazione di determinati settori, sostituibili nelle loro funzioni (si pensi per esempio ai casi di Uber o Deliveroo) e in relazione ai quali dottrina ed esperienza concreta, si ritengono atte a eliminare dal mercato del lavoro livelli di *manager* mediani.

In questo scenario l'uso dell'AI, per esempio, viene ritenuto efficiente in presenza di individuazione di "superamento di soglie numeriche" alle quali vengono ricollegati doveri in capo ai gestori. L'esempio del codice della crisi è immediato e ovviamente può porsi – specie in correlazione a natura e dimensioni dell'impresa – come *standard* necessario, e al contempo per dimostrare di avere un assetto adeguato e per integrare l'agire diligente degli amministratori o dei soggetti al cui interno sono allocate le competenze specifiche.

Su questa strada gli studiosi hanno anche già individuato altre aree di efficienza legate da algoritmi e AI tra i quali si segnalano quelle di un efficace monitoraggio del controllo interno e così pure di rendere più efficienti e adeguati i flussi informativi ovvero la valutazione del sistema contabile della società e la correlata valutazione della scelta, p.es., anche con riferimento ai compiti del revisore e del necessario flusso informativo fra i soggetti chiamati a monitorare l'andamento della società.

10. Se in astratto queste riflessioni si collocano sul piano delle valutazione delle scelte gestorie, si comprende però bene che la tendenziale pervasività dell'uso dell'AI e la ora individuata sua efficienza porta il piano del discorso anzitutto e non solo sul tema del potere di adottare l'AI, – e alle modalità di questa scelta – ma anche al dovere di adottare l'AI perché – in generale o in determinati contesti (essenzialmente le grandi imprese) – essa è capace di delineare l'assetto amministrativo più adeguato rispetto all'adozione delle scelte gestorie.

Sul punto le riflessioni della dottrina italiana hanno seguito l'approccio adattativo di cui dicevo in precedenza.

E così la ricostruzione delle scelte al riguardo ha posto in essere una serie di indicazioni di sicuro interesse che, a mio avviso, rilevano non solo e non tanto per il fatto che il luogo delle scelte in materia riguarda il tema degli assetti organizzativi adeguati, in generale e con le conseguenze in ordine alle competenze di gestione e controllo e delle relative responsabilità, ma anche – e questo è un punto che sul piano delle responsabilità ha carattere di estrema rilevanza – che se la predisposizione e l'utilizzo dell'AI e delle tecnologie algoritmiche grava sugli esecutivi, la scelta strategica in tal senso dovrebbe spettare al *plenum*.

Il che fa riflettere (anche secondo il modello adattativo di cui dicevo) sulla portata qualitativamente diversa che l'adozione dell'AI rappresenta nel sistema di gestione della società nelle prime riflessioni al riguardo, per quanto si riconduca generalmente queste scelte alla *bjr*.

Volendo però mantenersi ancora sul piano dell'adattamento delle nuove fattispecie alle regole tradizionali possono evidenziarsi vari profili di interesse.

Il primo piano, su cui si tornerà, è la distinzione fra modelli autoprodotti ovvero realizzati da terzi sul mercato per dotare la società degli strumenti di AI. Si tratta di una scelta che richiede la valutazione della coerenza dei meccanismi di AI con le specifiche necessità dell'impresa, oltre al già evidenziato tema di *compliance* con le regole societarie.

Il secondo piano, che pure è stato posto in evidenza, riguarda il modello di sindacabilità delle scelte gestorie in materia muovendo dal presupposto della *bjr*.

Il terzo piano riguarda la gestione del rischio della scelta di adottare la tecnologia in questione.

In questo scenario il modello adattativo ha trovato una sua soluzione principalmente nella costituzione in sede consiliare o di *plenum* di *tech committees* o di attribuire al controllo rischi le scelte gestorie e organizzative. Non ci sono obblighi espressi nella normativa in proposito, ma tale esi-

to viene fatto discendere dalle regole sugli assetti organizzativi adeguati. In questo senso, p.es. la dottrina elaborata in sede di mercati finanziari ha posto l'attenzione sui possibili rischi dell'utilizzo, nella storia recente e non, del *trading* algoritmico e delle sue risalenti conseguenze negative, in particolare nel senso che il suo utilizzo abbia causato o possa causare significative perdite borsistiche sui mercati [2].

La responsabilità societaria, nel modello adattativo, allora dovrebbe misurarsi su tematiche di estrema complessità strutturale e di altrettanto complessa capacità di legarsi – come dovrebbe essere in astratto – ai modelli di scelta gestoria a cui si riferiscono. Ne viene, in proposito, la ricerca di modelli normativi coerenti al modello sin qui delineato.

Al riguardo è di sicura rilevanza quanto è stato osservato in dottrina in un recente saggio [3] per cui il modello adattativo si conferma un modello di responsabilità gestoria basato sul canone di diligenza secondo un percorso astrattamente replicabile per ogni decisione.

In particolare:

– coerenza della tecnologia con le esigenze della società;
– rispondenza dei criteri e degli obiettivi prescelti per orientare l'analisi algoritmica ai bisogni della società;
– il corretto utilizzo degli algoritmi;
– l'adeguatezza dei dati che l'algoritmo è chiamato a elaborare rispetto agli obiettivi individuati e, infine, la supervisione delle decisioni adottate tramite le tecnologie algoritmiche e dei relativi effetti.

Con l'ulteriore differenza, nella valutazione dell'adattamento fra ipotesi di algoritmo meramente replicante con maggior efficienza l'uomo (più facilmente sindacabile) e quello invece predittivo e autonomo dove vaglio e responsabilità, si dice, non può esser legata al risultato ma al processo attraverso il quale queste scelte giungono al risultato (selezione dati, regole dell'algoritmo, codice sorgente). In sintesi – come ho già accennato – un esercizio di ottimismo della volontà, ovvero la pretesa che l'uomo sia in grado di comprendere e gestire le sue creature e intendendo l'uomo come

---

[2] In generale V. Sandei, *Intelligenza artificiale e diritto dei mercati finanziari tra informazione e responsabilità. Riflessioni a margine di alcuni recenti provvedimenti*, in M. Rescigno (a cura di), *L'Impresa nell'era dell'intelligenza artificiale: un'evoluzione tranquilla o nulla sarà più lo stesso?*, in *Collana dell'Osservatorio "Giordano dell'Amore"*, Giuffrè, Milano, 2023, p. 61 ss.

[3] Sul punto in generale C. Picciau, *Intelligenza artificiale, scelte gestorie e organizzazione delle società per azioni*, in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, Il Mulino, Bologna, 2022, p. 277 ss.

l'amministratore di società a fini di lucro e le sue creature quali AI e algoritmi.

In questo modello, va ribadito, dovrebbe collocarsi quel paradigma di responsabilità – noto alle regole in materia societaria – per cui è diligente il gestore che affida al terzo esperto la scelta gestoria supplendo alla carenza di competenza diretta. Potrà sindacarsi, come detto, adattando la fattispecie alla regola generale, il procedimento che ha portato ad adottare le scelte in questione e alla "supervisione strategica" delle tecnologie e naturalmente tenendo conto le peculiarità della singola fattispecie. Ma in realtà – come pure è stato notato – la scelta in questione appare nello scenario dato una scelta strategica non solo quantitativamente rilevante, ma anche qualitativamente decisiva. Tanto che va ribadito che la decisione a monte sulla scelta di utilizzazione di AI al servizio di scelte di tipo organizzativo o gestorio dovrebbe esser assunta dal *plenum* consiliare (anche tenuto conto dei possibili costi).

11. Ma c'è un altro scenario che è stato delineato – e proprio con riguardo ai temi della gestione a mezzo AI e algoritmi – e che mi sembra in linea con quel breve *excursus* iniziale nel quale ho tracciato il progressivo allontanamento delle regole sulla responsabilità dalla diligenza specifica degli amministratori per le scelte gestorie alla adozione di *standard* organizzativi e procedimentali che tendono a esaurire l'area della responsabilità. Ed è uno scenario che viene proprio dal punto più rilevante dell'evoluzione che AI e algoritmi chiamano in causa.

Si tratta dell'estrema complessità di tali strumenti.

Vorrei, per fare un esempio, segnalare il considerando 47 del regolamento AI che si preoccupa che l'uso degli AI e algoritmi sia comprensibile per l'uomo, in relazione alle scelte che sono ad esso demandate.

La lettura, a mio avviso, di questo passaggio, è emblematica – ancora una volta – dello iato fra ottimismo della ragione – nutrire cioè confidenza che si può esser in grado di far comprendere all'amministratore della società i sistemi di AI e l'amministratore capirà – e l'inevitabile impossibilità, anche dal miglior CEO, di entrare nei meccanismi algoritmici, nei quali viene (forse) istruito a stento dai componenti del *tech commitee* ovvero da chi detiene il meccanismo algoritmico e cioè il terzo che lo realizza, sia pur in accordo con l'amministratore e sulla base delle sue richieste.

Il che – come è già stato notato – ove fosse applicato alle regole di responsabilità potrebbe seriamente tradursi in adozione di scelte non comprensibili all'amministratore, ma solo a chi ne possiede la complessa cognizione.

Correttamente allora si è detto che in realtà tale complessità (e, per altro verso, anche l'evoluzione verso una gestione affidata alla AI e a modelli algoritmici) non possa esser regolata da modelli di responsabilità gestoria tradizionali.

Due sono i punti rilevanti a tal proposito.

Il primo è l'inevitabile tendenza – proprio in ragione della complessità e della struttura gestoria AI – alla cattura dell'amministratore da parte dell'algoritmo. Infatti, la domanda che ho sollevato quando ho descritto il sistema tradizionale può a questo riguardo comprensibilmente portare l'amministratore a rispondere, a fronte di censure del suo operato, di aver adottato l'algoritmo preparato dagli esperti, dal *tech committee* e di non essere passibile di responsabilità.

E correttamente si è prefigurata una evoluzione gestoria e soprattutto di responsabilità gestoria che al crescere del rilievo oggettivo dell'algoritmo come fonte della scelta gestoria, vede limitare o eliminare la responsabilità gestoria in materia.

Il secondo punto, collegato al primo, sta in ciò che appare pienamente condivisibile lo scenario che è stato delineato in dottrina che giunge correttamente a un esito per cui il tema qui tratteggiato "dovrà essere regolato sul piano contrattuale fra detentori del sapere e della macchina e gli amministratori" [4], con una diminuita rilevanza della responsabilità degli amministratori al crescere dell'importanza e della complessità delle scelte gestorie. Ne viene allora confermato l'esito dell'evoluzione originaria che avevo richiamato all'inizio di questa relazione, con l'ulteriore corollario per cui, nello scenario qui tratteggiato, è sostanzialmente inevitabile la "cattura" degli amministratori della società da parte dell'algoritmo con l'esito ancor più preoccupante che deriva dal fatto che l'AI è suscettibile di autogenerarsi indipendentemente dalle scelte originarie [5].

Nel frattempo – se è consentita una battuta per render più lieve la chiusura del mio breve scritto – credo che l'esito che si delinea nello scenario futuro vedrà uno spostamento delle richieste di pareri degli amministratori da quelli aventi oggetto questioni legali a quelli aventi come oggetto richieste di pareri tecnologici con contrazione dei redditi della mia categoria.

---

[4] M. Petrin, *Corporate Management in the age of AI*, in *Columbia Business Law Review*, 2019, 965, ss.

[5] In generale per il tema della responsabilità degli amministratori a fronte della AI, v. G. Meo, *Intelligenza artificiale e responsabilità degli amministratori*, in M. Rescigno (a cura di), *L'Impresa nell'era dell'intelligenza artificiale*, cit., p. 46 ss.

INTELLIGENZA ARTIFICIALE E GRUPPI DI SOCIETÀ

di NICCOLÒ BACCETTI

SOMMARIO: 1. Introduzione. – 2. L'intelligenza artificiale quale strumento di gestione e controllo a servizio della direzione e coordinamento. – 3. Società controllate a guida autonoma. – 4. Patrimoni destinati algoritmici. Il problema dell'individuazione dei presupposti richiesti dall'ordinamento per l'esercizio di attività imprenditoriali in regime di responsabilità limitata.

1. *Introduzione*

Nel nostro e in altri paesi, soprattutto in Europa, qualunque attività d'impresa dotata di un minimo di complessità è oggi normalmente organizzata secondo lo schema del gruppo di società.

Occorre allora domandarsi se, nell'ambito dei rapporti tra intelligenza artificiale e gestione dell'impresa [1], abbia senso ipotizzare scenari che più

[1] In relazione ai problemi sollevati dall'intelligenza artificiale nel governo dell'impresa, cfr., senza pretesa di completezza, AA.VV., *Diritto societario, digitalizzazione e intelligenza artificiale*, in ricordo di Agostino Gambino, a cura di N. Abriani e R. Costi, Milano, 2023; AA.VV., *L'impresa nell'era dell'intelligenza artificiale: un'evoluzione tranquilla o nulla sarà più lo stesso?*, a cura di M. Rescigno, Milano, 2023; C. PICCIAU, *Intelligenza artificiale, scelte gestorie e organizzazione delle società per azioni*, in *NDS*, 2022, p. 1253 ss.; M.L. MONTAGNANI, *Il ruolo dell'intelligenza artificiale nel funzionamento del consiglio di amministrazione delle società per azioni*, Milano, 2021; ID., *Intelligenza artificiale e* governance *della "nuova" grande impresa azionaria: potenzialità e questioni endoconsiliari*, in *Riv. soc.*, 2020, p. 1003 ss.; M.L. MONTAGNANI, M.L. PASSADOR, *Il consiglio di amministrazione nell'era dell'intelligenza artificiale: tra* corporate reporting*, composizione e responsabilità*, in *Riv. soc.*, 2021, p. 121 ss.; N. ABRIANI, G. SCHNEIDER, *Il diritto societario incontra il diritto dell'informazione.* IT, Corporate governance *e* Corporate Social Responsibility, in *Riv. soc.*, 2020, 1326 ss.; N. ABRIANI, *Intelligenza artificiale e diritto delle società: nuovi doveri e responsabilità degli amministratori*, in *Riv. dir. impr.*, 2022, p. 1 ss.; ID., *Le categorie della moderna cibernetica societaria tra algoritmi e androritmi: "fine" della società e "fini" degli strumenti tecnologici*, in *Giur. comm.*, 2022, p. 743 ss.; G. SCHNEIDER, *La proposta di regolamento europeo sull'intelligenza artificiale alla prova dei mercati finanziari: limiti e prospettive (di vigi-*

specificamente riguardano l'uso dell'intelligenza artificiale nei gruppi di società e, quindi, in definitiva, se e in che misura la preannunciata rivoluzione algoritmica possa produrre effetti sui rapporti e problemi che tipicamente sorgono nei gruppi di società.

2. *L'intelligenza artificiale quale strumento di gestione e controllo a servizio della direzione e coordinamento*

Dirigendo anzitutto il discorso all'ipotesi qualitativamente più integrata di gruppo di società contraddistinta dall'esercizio di direzione e coordinamento, non è difficile cogliere l'utilità del possibile impiego di strumenti di intelligenza artificiale, nell'assunto di fondo su cui poggia l'intero discorso, ovvero che, già oggi o nel prossimo futuro, l'intelligenza artificiale sia in grado di produrre soluzioni che siano, da un lato, affidabili e, dall'altro, accessibili alle imprese a costi ragionevoli, giacché, se così non fosse, almeno nell'area del diritto commerciale, non vi sarebbe un gran ché da discutere.

In questa prospettiva, e nel presupposto sopra indicato, mi pare che si possano intravedere alcune aree in cui l'intelligenza artificiale può essere impiegata per creare valore e migliorare l'efficienza delle imprese appartenenti al gruppo.

In primo luogo, non è difficile immaginare che l'intelligenza artificiale possa divenire un utile strumento di gestione e controllo a servizio della direzione e coordinamento esercitata dalla capogruppo, soprattutto nei gruppi di imprese di grandi dimensioni che operano su larga scala nei mercati internazionali. E questo, tanto ai fini di una migliore e più accurata

---

lanza), in *Resp. civ. e prev.*, 2023, p. 1014 ss.; Pacileo, *"Scelte d'impresa" e doveri degli amministratori nell'impiego dell'intelligenza artificiale*, in *RDS*, 2022, p. 543 ss.; U. Tombari, *Intelligenza artificiale e* corporate governance *nella società quotata*, *Riv. soc.*, p. 1431 ss.; B.M. Scarabelli, *L'intelligenza artificiale e i flussi informativi all'interno del consiglio di amministrazione*, in *NDS*, 2022, p. 599 ss.; G.D. Mosco, Roboboard. *L'intelligenza artificiale nei consigli di amministrazione*, in *AGE*, 2019, p. 247 ss.; R. Russo, *L'amministratore non socio nelle società di persone tra antiche questioni e intelligenza artificiale:* tertium datur*?*, in *Riv. soc.*, 2022, p. 874 ss.; G. Nuzzo, *Impresa e società nell'era digitale (appunti)*, in *Banca, borsa, tit. cred.*, 2022, I, p. 417 ss.; F. Möslein, *Robots in the Boardroom: Artificial Intelligence and Law*, in AA.VV., *Research Handbook on the Law of Artificial Intelligence*, a cura di W. Barfield, U. Pagallo, Cheltenham, 2018, p. 649 ss.; L. Enriques, D. Zetzsche, *Corporate Technologies and the Tech Nirvana Fallacy*, in 72 *Hastings Law Journal*, 2020, p. 55 ss.; E. Hickman, M. Petrin, *Trustworthy AI and Corporate Governance: The EU's Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective*, in *Eur. Bus. Org. Law Rev*, 2021, pp. 593-625.

pianificazione strategica dell'attività imprenditoriale del gruppo, quanto ai fini della predisposizione di politiche gestionali accentrate sempre più efficienti che consentano di migliorare la qualità dei prodotti e servizi offerti ai clienti, il coordinamento delle diverse attività, lo sfruttamento di economie, l'ottimizzazione delle politiche commerciali, la riduzione dei costi di gestione, e così via.

Ma gli strumenti algoritmici di elaborazione di grandi dati potranno essere utili anche in relazione a quella vasta area dell'amministrazione che attiene al sistema di controllo interno e gestione dei rischi. Penso alle funzioni aziendali di conformità, gestione del rischio e controllo interno che estendono il proprio raggio di azione all'intero gruppo. E un tale impiego potrà anche assumere profili di doverosità, se si ritiene che gli assetti organizzativi, amministrativi e contabili adeguati dell'impresa debbano essere riferiti pure all'impresa nella sua configurazione di gruppo e, quindi, con modalità tali da considerare anche l'esercizio delle attività imprenditoriali in cui è investito il patrimonio della capogruppo [2].

Vanno tuttavia considerate anche le potenzialità dell'intelligenza artificiale nel dirimere la conflittualità tra interessi interni ed esterni al gruppo, quale strumento di prevenzione e repressione degli abusi di direzione e coordinamento e, più in generale, di condotte gestionali in conflitto di interessi.

In questo ambito, l'intelligenza artificiale potrebbe essere utile per accertare i presupposti di eventuali responsabilità per abuso di direzione e coordinamento o per il compimento di illeciti gestionali tramite operazioni con parti correlate. Essa infatti potrebbe fornire utili parametri di riferimento per verificare se l'operazione infragruppo sia stata conveniente per la società e, come si usa dire, *at arm's lenght*, immaginando strumenti algoritmici che possano essere rapidamente interrogati per comporre i diversi interessi in gioco. E questo, tanto in una prospettiva preventiva, al fine di indicare agli organi di gestione e controllo i limiti entro cui le diverse società del gruppo possono legittimamente intrattenere rapporti tra loro, quanto in una prospettiva successiva e giudiziale in cui occorre verificare mediante accertamenti tecnici se, in base a una valutazione *ex ante*, l'operazione posta in essere sia stata effettivamente idonea ad arrecare un danno

[2] Sugli assetti adeguati nei gruppi di società, cfr., tra gli altri, F. Guerrera, *Assetti organizzativi di gruppo, forme di eterodirezione e regimi di responsabilità*, in *Riv. dir. soc.*, 2024, p. 205 ss.; P. Marchetti, *Del dovere di direzione e coordinamento: l'approdo al Codice della crisi di impresa di un tema caro a Paolo Ferro-Luzzi*, in *Riv. soc.*, 2022, p. 1412 (che fa riferimento alle posizioni espresse in P. Ferro-Luzzi, P. Marchetti, *Riflessioni sul gruppo creditizio*, in *Giur. comm.*, 1994, I, p. 419); Panzani, *Codice della crisi e gruppi di società*, in *Riv. soc.*, 2022, p. 1330 ss.

alla società in conseguenza di un abuso di direzione e coordinamento o di un'operazione tra parti correlate.

E laddove la direzione e coordinamento dovesse scaturire in un'operazione illegittima e dannosa, non sfugge che l'intelligenza artificiale potrebbe essere interrogata per altri aspetti valutativi come quelli che attengono alla configurabilità di vantaggi compensativi idonei a escludere la responsabilità. Analisi che in concreto può richiedere di calcolare benefici economici derivanti da attività o rapporti complessi, specie se tali valutazioni, come alcuni propongono, dovrebbero spingersi a stimare anche vantaggi futuri che sono ragionevolmente prevedibili in conseguenza dell'operazione o dell'appartenenza al gruppo [3].

Ma le applicazioni dell'intelligenza artificiale potranno astrattamente coinvolgere anche la tutela dei creditori sociali delle società eterodirette, ove le nuove tecnologie algoritmiche consentano di definire o comunque supportare l'elaborazione di piani di sostenibilità economico-finanziaria dell'attività o *test* di solvibilità in relazione a determinati atti gestionali od operazioni straordinarie imposti dalla direzione e coordinamento. Si pensi, ad esempio, all'esigenza di limitare l'assunzione di indebitamento finanziario entro parametri sostenibili di leva finanziaria, magari al fine di finanziare nuove acquisizioni, così come a quella di definire politiche di distribuzione di utili o riserve che non introducano rischi di squilibrio economico-finanziario o scenari di crisi o insolvenza. Utili applicazioni di tecnologie algoritmiche possono poi ravvisarsi anche per accertare la sostenibilità di operazioni riorganizzative della struttura patrimoniale e finanziaria della

[3] In senso restrittivo, cfr. SACCHI, *Sulla responsabilità da direzione e coordinamento nella riforma delle società di capitali*, in *Giur. comm.*, 2003, I, p. 673 ss., secondo cui il vantaggio compensativo deve essere accertato in senso ragionieristico e quantitativo, richiedendosi una rigida proporzionalità fra pregiudizio prodotto dalla singola operazione infragruppo e vantaggio compensativo. Per una posizione più articolata, cfr. invece A. VALZER, *Commento all'art. 2497*, in *Le società per azioni*, *Codice civile commentato*, diretto a P. Abbadessa e G.B. Portale, II, Milano, 2016, p. 3021 ss. (ove *adde* ulteriori riferimenti), secondo il quale, da un lato, il danno non sussiste e la politica di gruppo risulta legittima sino a quando i risultati di gestione della società eterodiretta siano coerenti con i termini *attesi* di rendimento del capitale in essa investito quale società appartenente al gruppo (e, quindi, "la situazione economica, finanziaria, patrimoniale e reddituale della società eterodiretta, seppure differente dall'ipotetico esserci di questa come monade, si rivela coerente col risultato atteso da quella società nell'ambito del gruppo e alla luce dell'attività di direzione unitaria"); dall'altro, il danno risulta integralmente eliminato anche a seguito di operazioni a ciò dirette quando eventuali perdite in capo alla controllata siano state temporanee e integralmente eliminate in termini tali da non alterare le condizioni di rischio del capitale in essa investito, dovendosi valutare natura ed entità di tali operazioni nel quadro del risultato complessivo dell'attività di direzione e coordinamento.

società, come fusioni, scissioni o scorpori di parte dell'attività imprenditoriale.

E se si guarda agli effetti esterni dell'attività svolta dalle imprese appartenenti al gruppo, si può pure ipotizzare che gli strumenti di intelligenza artificiale possano giocare un ruolo attivo anche ai fini dell'elaborazione delle politiche di sostenibilità e, più in generale, delle valutazioni che attengono alla c.d. responsabilità sociale dell'impresa. Ciò nella misura in cui l'intelligenza artificiale possa effettivamente contribuire a individuare e misurare i potenziali effetti avversi che l'attività del gruppo produce su ambiente, comunità locali e i vari portatori di interesse a un esercizio sostenibile dell'attività. Non intendo qui minimamente affrontare l'enorme dibattito che ormai da lungo tempo avvolge questa area di problemi. Mi limito soltanto a rilevare che il calcolo e la misurazione delle esternalità prodotte dall'imprese appartenenti di gruppo è già obbligo di diritto positivo nella disciplina della rendicontazione societaria di sostenibilità [4] e ben presto potrà essere imposto anche dall'attuazione dei doveri di diligenza in materia di ambiente, diritti umani e cambiamento climatico elaborati dalla direttiva eurounitaria recentemente approvata [5].

---

[4] Cfr. il d.lgs. 6 settembre 2024, n. 125, relativo all'attuazione della direttiva 2022/2464/UE del Parlamento europeo e del Consiglio del 14 dicembre 2022 (recante modifica al regolamento 537/2014/UE, alla direttiva 2004/109/CE, alla direttiva 2006/43/CE e alla direttiva 2013/34/UE) per quanto riguarda la rendicontazione societaria di sostenibilità. Al riguardo, cfr., tra gli altri, M. Rescigno, *Note sulle "regole" dell'impresa "sostenibile". Dall'informazione non finanziaria all'informativa sulla sostenibilità*, in *AGE*, 2022, p. 166 ss. Sugli *standard* di rendicontazione, cfr. di recente N.M. Bacaro, M. Russotto, C. Saporetti, L. Solimene, *La* Corporate Sustainability Reporting Directive (CSRD), in *Riv. dott. comm.*, 2024, p. 443 ss.; F.M. Di Majo, *The Eu Regulatory and Standard Setting Action on Corporate Sustainability Reporting and the Environmental Taxonomy: Fighting Against Greenwashing Practices with a Global Reach*, in *Dir. comm. int.*, 2024, I, p. 207 ss.; L. Solimene, *La Direttiva* CSRD (Corporate Sustainability Reporting Directive) *e i nuovi* standard EFRAG, in *Riv. dott. comm.*, 2022, p. 595 ss.; L. Cencioni, A. Cinquegrana, P. Mantovano, L. Mercuri, Corporate Governance*. Il sistema di controllo interno sull'informativa di sostenibilità*, in *Riv. dott. comm.*, 2023, p. 633 ss.

[5] Cfr. la Direttiva (UE) 2024/1760 del Parlamento Europeo e del Consiglio del 13 giugno 2024 relativa al dovere di diligenza delle imprese ai fini della sostenibilità, che modifica la direttiva (UE) 2019/1937 e il regolamento (UE) 2023/2859. Al riguardo, cfr. di recente, M. Ventoruzzo, *Uno sguardo d'insieme alla CS3D: riflessioni preliminari sulla tecnica normativa*, in *Riv. soc.*, p. 447 ss. Sulla precedente proposta di direttiva della Commissione Europa del 23 febbraio 2022 ("*Proposal for a Directive of the European Parliament and of the Council on corporate sustainability due diligence and amending directive (EU) 2019/1937*"), cfr., tra gli altri, Stella Richter *Jr*, Corporate Sustainability Due Diligence*: noterelle semiserie su problemi serissimi*, in *Riv. soc.*, 2022, p. 714; Barcellona, *La* sustainable corporate governance *nelle proposte di riforma del diritto europeo: a proposito dei limiti strutturali del c.d.* stakeholderism, in *Riv. soc.*, 2022, p. 1;

3. *Società controllate a guida autonoma*

Sempre con riferimento ai gruppi di società è stato anche ipotizzato che l'intelligenza artificiale possa divenire strumento di automazione della direzione e coordinamento al punto da configurare vere e proprie società controllate a guida autonoma (o *self-driving subsidiaries*) [6].

Si immaginano cioè enti interamente algoritmici che operano in via automatica in base agli *input* impartiti dalla direzione e coordinamento elaborata dall'organo di amministrazione umano della capogruppo.

L'ipotesi è suggestiva e apre a scenari che rischiano di allontanarsi dal diritto positivo.

Cercando di mantenere il ragionamento sul terreno dell'attuale diritto societario, la riorganizzazione algoritmica dei gruppi di impresa richiede di affrontare una serie di problemi.

Il primo attiene evidentemente al fatto che l'uso dello strumento socie-

---

Scano-Racugno, *Il dovere di diligenza delle imprese ai fini della sostenibilità: verso un* Green Deal *europeo*, in *Riv. soc.*, 2022, p. 726. Sulla proposta del Parlamento Europeo del 10 marzo 2021 2020/2129(INL), cfr. Enriques, *The European Parliament Draft Directive on Corporate Due Diligence and Accountability: Stakeholder-Oriented Governance on Steroids*, in *Riv. soc.*, 2021, p. 319; Libertini, *Sulla proposta di direttiva UE su "Dovere di diligenza e responsabilità delle imprese"*, in *Riv. soc.*, 2021, p. 325; Marchetti, *Il bicchiere mezzo pieno*, in *Riv. soc.*, 2021, p. 336; Mucciarelli, *Ricomporre il nesso spezzato: giurisdizione e legge applicabile alle imprese multinazionali*, in *Riv. soc.*, 2021, p. 349; Strampelli, *La strategia dell'Unione europea per il capitalismo sostenibile: l'oscillazione del pendolo tra amministratori, soci e* stakeholders, in *Riv. soc.*, 2021, p. 365; Tombari, *La proposta di direttiva sulla* Corporate Due Diligence *e sulla* Corporate Accountability*: prove (incerte) di un "capitalismo sostenibile"*, in *Riv. soc.*, 2021, p. 375; e Ventoruzzo, *Note minime sulla responsabilità civile nel progetto di direttiva* Due Diligence, in *Riv. soc.*, 2021, p. 380.

[6] Così G. Scognamiglio, *Intelligenza artificiale e gruppi di imprese*, in AA.VV., *Diritto societario, digitalizzazione*, cit., p. 217 ss., p. 227 ss.; N. Abriani, G. Schneider, *Il diritto societario*, cit., p. 1367 ss.; G.D. Mosco, *AI and Boards of Directors: Preliminary Notes from the Perspective of Italian Corporate Law*, in *European Company Law Journal*, 2020, p. 87 ss. Nella letteratura statunitense, sulla possibilità che sistemi di intelligenza artificiale gestiscano in maniera autonoma persone giuridiche che esercitano attività imprenditoriali, cfr., tra gli altri, S. Bayern, *Of Bitcoins, Independently Wealthy Software, and the Zero Member LLC*, in *Northwestern University Law Review*, 2014, 108, p. 1486 ss.; Id., *The Implications of Modern Business – Entity Law for the Regulation of Autonomous Systems*, in *Stanford Technology Law Review*, 2015, 19, p. 93 ss.; S. Bayern, T. Burri, T.D. Grant, D.M. Häusermann, F. Möslein, R. Williams, *Company Law and Autonomous Systems: A Blueprint for Lawyers, Entrepreneurs and Regulators*, in *Hastings Science & Technology Law Journal*, 2017, p. 135 ss.; L.M. Lopuki, *Algorithmic Entities*, in *Washington University Law Review*, 2018, 95, p. 887 ss.; M. Petrin, *Corporate Management in the Age of AI*, in *Columbia Business Law Review*, 2019, p. 965 ss.

tario oggi positivamente richiede la nomina di un organo di amministrazione umano.

Anche accogliendo la diffusa impostazione che da tempo ammette la figura dell'amministratore persona giuridica [7], non va infatti trascurato che essa deve comunque designare un proprio rappresentante persona fisica che, con diverse soluzioni, finisce per rispondere degli atti gestionali compiuti [8]. Il che in fin dei conti reintroduce la necessità di indivi-

---

[7] In tal senso cfr., tra l'altro, la nota decisione di Trib. Milano, 27 marzo 2017, n. 3545, in *Giur. comm.*, 2018, II, p. 890 ss., con nota di G. Pescatore, *Ammissibilità dell'amministratore persona giuridica tra conferme e problemi interpretativi*; Trib. Milano, 27 febbraio 2012, in *Giur. comm.*, 2014, II, p. 647 ss., con nota di G. Pescatore, *Prossima fermata: persona giuridica amministratore di fatto*; nonché la Massima n. 100 del Consiglio notarile di Milano, "*Amministratore persona giuridica di società di capitali (artt. 2380-bis e 2475 c.c.)*", del 18 maggio 2007, ove, salvo limiti previsti da leggi speciali, è affermata la legittimità di una clausola statutaria di s.p.a. o s.r.l. che prevede la possibilità di nominare alla carica di amministratore una o più persone giuridiche o enti diversi dalle persone fisiche. La soluzione è fondata in base a varie argomentazioni che riguardano, tra l'altro, la disciplina di attuazione del regolamento comunitario sul gruppo europeo di interesse economico (art. 5 d.lgs. 240/1991), la possibilità di attribuire a persone non fisiche l'amministrazione delle società di persone (artt. 2361, comma 2, cod. civ. e 111-*duodecies* disp. att. cod. civ.), la disciplina della società europea (art. 47, comma 1, Regolamento UE/2157/2001) e, infine, il riconoscimento della responsabilità amministrativa delle persone giuridiche (d.lgs. n. 231/2001). Per una trattazione più ampia del problema, cfr. G. Pescatore, *L'amministratore persona giuridica*, Milano, 2102, p. 35 ss.

[8] Cfr. Trib. Milano, 27 marzo 2017, cit., 890 ss., secondo cui degli atti gestori pregiudizievoli risponde, in via solidale con la persona giuridica amministratore, anche la persona fisica incaricata al compimento degli stessi, sia in quanto direttamente soggetta al regime di responsabilità dell'amministratore in coerenza con l'art. 5 d.lgs. 240 del 1991, sia perché la persona fisica incaricata dell'amministrazione entra a diretto contatto con la società amministrata e con i terzi, assumendo nei confronti della società amministrata una posizione di garanzia che ingenera a suo carico una responsabilità contrattuale derivante da obblighi assunti a favore di terzi *ex* art. 1411 cod. civ. In precedenza, cfr. Trib. Milano, 27 febbraio 2012, cit., 647 ss., ove la persona fisica rappresentante della persona giuridica amministratore è considerata alla stregua di un amministratore di fatto della società amministrata. Cfr. altresì la Massima n. 100 della Commissione in materia di società del Consiglio Notarile di Milano, cit., ove la responsabilità solidale della persona giuridica amministratore e del suo rappresentante persona fisica è affermata in base a un'applicazione analogica dell'art. 47, comma 1, Regolamento UE/2157/2001 in tema di società europea (secondo cui la società o altra entità giuridica nominata in un organo della società europea deve designare un rappresentante persona fisica ai fini dell'esercizio dei poteri attribuiti a tale organo), nonché dell'art. 5 d.lgs. n. 240/1991 in tema di gruppo europeo di interesse economico (in virtù del quale l'amministratore persona giuridica esercita le relative funzioni attraverso un rappresentante da essa designato che assume i medesimi obblighi e responsabilità civili e penali degli amministratori persone fisiche, in solido con la responsabilità della persona giuridica amministratore). *Contra*, cfr. G. Pescatore, *Ammissibilità*, cit., p. 899 ss., il quale esclude l'applicazione analogica della disciplina del gruppo europeo di interesse economico alle società di capitali e afferma l'imputazione del-

duare una persona umana che risponda degli atti gestionali [9].

Ad oggi, quindi, la società controllata non può essere a guida totalmente autonoma (se per ciò s'intende la radicale assenza di una persona fisica investita di poteri, doveri e responsabilità gestionali).

Il che significa che, nell'attuale diritto societario, tale ipotesi presuppone l'inerzia della persona fisica incaricata dell'attività di amministrazione della società controllata [10] e la sua sostituzione nella gestione ad opera delle decisioni algoritmiche impartite dalla capogruppo.

Così inquadrata, tuttavia, la vicenda algoritmica finisce per porsi su un piano diverso da quello della direzione e coordinamento e tende a tradursi in una vera a propria amministrazione in fatto, con conseguente irrigidimento della responsabilità della capogruppo e dei suoi amministratori per la gestione algoritmica della controllata da essi realizzata. E ciò in virtù dei consolidati orientamenti che, per un verso, assoggettano l'amministratore di fatto al medesimo regime di responsabilità civile e penale dell'amministratore di diritto [11] e, per altro verso, impediscono all'amministratore di

---

l'attività svolta dal rappresentante persona fisica direttamente in capo alla persona giuridica amministratore in base a un rapporto di immedesimazione organica, ferma la possibilità, ove ricorrano i presupposti, di affermare la responsabilità del rappresentante persona fisica, tanto in via contrattuale nei confronti della società amministratore (con possibile azione surrogatoria della società amministrata), quanto in via extracontrattuale nei confronti della società amministrata. In senso dubitativo, A. Cetra, *La persona giuridica amministratore di una società a responsabilità limitata o società per azioni nella massima n. 100 del Consiglio notarile di Milano*, in *Riv. dir. soc.*, 2008, p. 692 ss., secondo cui non è chiaro se l'indicazione di una persona fisica da parte dell'amministratore persona giuridica sia cogente e dia automaticamente luogo a responsabilità penali e/o civili della persona fisica così designata. In argomento, cfr. altresì P. Montalenti, *La traslazione dei poteri di gestione nei gruppi di società: i «management contracts»*, in *Contr. impr.*, 1987, p. 463 ss., secondo cui la nomina di amministratore di una società di capitali non fa venir meno la funzione preventiva delle azioni di responsabilità, sia perché non è escluso che la persona fisica rappresentante della persona giuridica amministratore venga chiamata a rispondere (sia pure in seconda istanza), sia perché la funzione preventiva rispetto al compimento di atti di *mala gestio* continua a ravvisarsi nelle disposizioni penali "che colpiscono pur sempre la persona fisica e quindi anche gli amministratori della società amministrante".

[9] Cfr. N. Abriani, G. Schneider, *Il diritto societario*, cit., p. 1367 ss., i quali considerano la possibilità di costituire società algoritmiche amministrate da società specializzate che gestiscono le varie società del gruppo tramite persone fisiche loro rappresentanti dotate di particolari conoscenze nell'uso di algoritmi e di strumenti di *information technology*.

[10] Sia essa una persona fisica che assume la carica di amministratore della società controllata ovvero una persona fisica che agisce quale rappresentante della persona giuridica amministratore designato alla gestione della società controllata.

[11] In sede civilistica, l'assoggettamento dell'amministratore di fatto al regime di responsabilità degli amministratori risale a Cass., 6 marzo 1999, n. 1925, in *Corr. giur.*, 1999, p. 1396, con

nota di A. PERRONE, *Un* revirement *della cassazione sulla responsabilità dell'amministratore di fatto*, secondo cui "le regole che disciplinano l'attività degli amministratori regolano […] il corretto svolgimento dell'amministrazione della società e sono quindi applicabili non solo a coloro che sono stati immessi, nelle forme stabilite dalla legge, nelle funzioni di amministratore, ma anche a coloro che si sono ingeriti nella gestione della società senza aver ricevuto da parte dell'assemblea alcuna investitura, neppure irregolare o implicita. E che, pertanto, anche nell'ambito del diritto privato, come in quello del diritto penale e del diritto amministrativo […] i responsabili della loro violazione non vanno individuati sulla base della loro qualificazione formale ma per il contenuto delle funzioni concretamente esercitate". Di recente, nella giurisprudenza civile, sulla nozione di amministratore di fatto, cfr. *ex multis* Cass. civ., 22 marzo 2024, n. 7864, www.dejure.it; Cass. civ., 19 gennaio 2022, n. 1516, *ibidem* (secondo cui ai fini del riconoscimento dell'amministratore di fatto non è sufficiente "il compimento episodico e frammentario di singoli atti gestori", bensì "l'accertamento dell'avvenuto inserimento dello stesso nella gestione dell'impresa, desumibile dalle direttive impartite e dal condizionamento delle scelte operative della società, anche in assenza di una qualsivoglia investitura, ancorché irregolare o implicita, da parte della società stessa [...] purché le funzioni gestorie svolte in via di fatto abbiano carattere sistematico e non si esauriscano, quindi, nel compimento di alcuni atti di natura eterogenea ed occasionale"); Cass. civ., 8 ottobre 2020, n. 21730, in *Banca, borsa tit. cred.*, 2022, II, p. 344 ss., con nota di F. BRIZZI, *Brevi appunti sulla figura dell'amministratore di fatto e sui criteri di determinazione del danno risarcibile*; Trib. Napoli, 20 dicembre 2023, in *Foro it.*, 2024, I, c. 1293; App. Cagliari, 14 aprile 2021, in *Giur. comm.*, 2022, II, p. 543 ss., con nota di F. CUCCU, *L'amministratore di fatto fra sistematicità e completezza dell'esercizio di funzioni gestorie*. Con riferimento ai gruppi di società, cfr. Cass. civ., 3 marzo 2021, n. 5795, in *Foro it.*, 2021, 1, c. 2839 ss. secondo cui "*la formale esistenza di un gruppo, con conseguente assetto giuridico predisposto per una direzione unitaria, non è incompatibile con l'amministrazione di fatto di singole società del gruppo stesso, poiché mentre la prima corrisponde ad una situazione di diritto nella quale la controllante svolge l'attività di direzione della società controllata nel rispetto della relativa autonomia e delle regole che presiedono al suo funzionamento, la seconda dà invece luogo ad una situazione di fatto in cui i poteri di amministrazione sono esercitati direttamente da chi sia privo di una qualsivoglia investitura, ancorché irregolare o implicita*". Per una recente illustrazione degli indirizzi giurisprudenziali, cfr. F. RIGANTI, *Rassegna di giurisprudenza*, *Società per azioni*, *Responsabilità degli amministratori*, in *Giur. comm.*, 2023, p. 189 ss. In dottrina, cfr., per tutti, N. ABRIANI, *Gli amministratori di fatto delle società di capitali*, Milano, 1998. A. BORGIOLI, *Amministratori di fatto e direttori generali*, in *Giur. comm.*, 1975, I, p. 593 ss.; S. CASSANI, *Responsabilità dell'amministratore di diritto e dell'amministratore di fatto*, nota a Trib. Milano, 29 novembre 2012, n. 10760, in *Società*, 2013, p. 1059 ss.; F. GUERRERA, *Gestione di fatto e funzione amministrativa nelle società di capitali*, in *Riv. dir. comm.*, 1999, I, p. 131 ss.; M. MOZZARELI, *Amministratori di fatto: fine di una contesa*, nota a App. Milano, 26 settembre 2000, *Giur. comm.*, 2001, II, p. 565 ss. Anche nel diritto penale societario, la responsabilità dell'amministratore di fatto è affermata da un orientamento consolidato e risalente. Di recente, sugli elementi sintomatici della qualifica di amministratore di fatto, cfr., *ex multis*, Cass. pen., 15 aprile 2024, n. 29608; Cass. pen., 28 febbraio 2024, n. 16414; Cass. pen., 20 febbraio 2024, n. 12175; Cass. pen., 7 dicembre 2023, n. 12715; Cass. pen., 4 dicembre 2023, n. 2514; Cass. pen., 15 novembre 2023, n. 48826; Cass. pen., 26 ottobre 2023, n. 4816, tutte in www.dejure.it. Cfr. altresì Cass. pen., 27 ottobre 2020, n. 36865, secondo cui "in tema di reati fallimentari, la titolarità della carica di amministratore della società capogruppo non implica di per sé la qualifica di amministratore di fatto delle società controllate, salvo che l'esercizio dei poteri di direzione e coordinamento si sostanzi in

diritto di sfuggire alle proprie responsabilità per essersi lasciato sostituire nell'assolvimento dei propri doveri gestionali [12]. E in tale quadro si dovrà poi verificare, anche in base all'evoluzione di quella che sarà la tipologia e la specificità dei servizi di gestione algoritmica concretamente forniti alle imprese, se e in che termini possano concorrere nella responsabilità dell'organo di gestione della capogruppo anche i fornitori dei servizi di gestione algoritmica della controllata.

4. *Patrimoni destinati algoritmici. Il problema dell'individuazione dei presupposti richiesti dall'ordinamento per l'esercizio di attività imprenditoriali in regime di responsabilità limitata*

Ma se si guarda oltre, ci si accorge che, in realtà, l'idea della società algoritmica o a guida autonoma riflette un'immagine o, se si vuole, una tentazione, che non è nuova nella storia del diritto commerciale. Essa infatti rivela una diffusa aspirazione dei moderni ordinamenti capitalistici a concepire organizzazioni o centri di autonomia patrimoniale capaci di agire nel mondo degli affari con modalità sempre più rapide e semplificate che conducono a un vero e proprio sgretolamento della concezione tradizionale della persona giuridica costruita intorno a quello che un tempo veniva indicato come il modello corporativo-capitalistico.

La spinta costante degli ordinamenti a promuovere obbiettivi di nascita, crescita e massimo sviluppo delle attività imprenditoriali, da un lato ha notevolmente ampliato la capacità dell'autonomia privata di destrutturare la disciplina organizzativa della persona giuridica societaria, liberando così forme più intense di esercizio del potere di impresa senza assunzione di una corrispondente responsabilità per il rischio dell'attività; dall'altro, ha introdotto nell'ordinamento centri di imputazione di rapporti giuridici alternativi alla persona giuridica che sono in grado di esercitare attività imprenditoriali in regime di autonomia patrimoniale. Il tratto comune di tali

atti specificamente gestori di fasi o settori dell'attività di queste, limitandone l'autonomia e riducendo gli amministratori a meri esecutori materiali delle direttive impartite").

[12] In sede civile, cfr., tra le altre, Cass. civ., 3 marzo 2021, n. 5795, cit., p. 2839 ss. Nel senso di affermare la responsabilità penale dell'amministratore di diritto per aver consentito il compimento di reati da parte dall'amministratore di fatto, cfr., tra le altre, Cass. pen., 14 settembre 2023, n. 42236 in www.dejure.it (con riferimento a reati laburistici); Cass. pen., 10 maggio 2023, n. 28257, *ibidem*; Appello Taranto, 8 ottobre 2024, *ibidem* (in relazione alla bancarotta); Appello Napoli, 15 aprile 2024; Appello Bari, 15 febbraio 2024, *ibidem* (in relazione a reati tributari).

esperienze resta la potente scelta politica di fondo volta a promuovere l'impresa mediante la propagazione del beneficio della responsabilità limitata e il costante superamento dei tradizionali principi di *heine herrschaft one haftung*.

Non è evidentemente questa la sede per ripercorrere una storia così profonda [13]. Qui interessa soltanto esprimere la consapevolezza che l'idea delle società controllate algoritmiche o a guida autonoma non coinvolge soltanto il piano della direzione o dell'amministrazione. Essa in realtà solleva una questione più profonda e generale che investe direttamente i presupposti richiesti dall'ordinamento per l'esercizio di attività imprenditoriali in regime di responsabilità limitata.

E se questa è la vera questione di fondo, va subito avvertito che, in un certo senso, il nostro ordinamento già conosce la possibilità di esercitare attività imprenditoriali in regime di responsabilità limitata senza ricorrere alla costituzione di una persona giuridica dotata di autonomi organi di amministrazione e controllo.

Se infatti, prima ancora di immaginare il futuro, proviamo ad approfondire il presente, ci accorgiamo che l'impresa senza autonomi organi di gestione e controllo trova oggi in diritto positivo margini di possibile attuazione con la tecnica della destinazione patrimoniale.

In tale ambito, si può allora ipotizzare uno schema in cui l'impresa azionaria, entro certi limiti e con determinate modalità, provveda a destinare parti della propria attività in uno o più patrimoni separati che vengono poi gestiti dalle medesime persone fisiche incaricate della gestione della società avvalendosi in modo più o meno inteso di strumenti di intelligenza artificiale.

Sotto questo profilo, i patrimoni destinati dimostrano che la scomposizione dell'attività imprenditoriale [14] in centri patrimonialmente autonomi non richiede necessariamente la creazione di organi di gestione e controllo separati a cui affidare la tutela degli interessi interni ed esterni che si appuntano sul patrimonio da essi gestito. La destinazione patrimoniale è compatibile con un'unica gestione integrata dei diversi affari e dei corrispondenti patrimoni.

---

[13] Al riguardo, sia consentito rinviare a N. Baccetti, *Creditori extracontrattuali, patrimoni destinati e gruppi di società*, Milano, 2009, pp. 1-16 (e agli Autori ivi citati). Sull'alternativa tra gruppo e destinazione, cfr. di recente P. Spada, *Destinazione ed entificazione*, in *Giur. comm.*, 2023, I, p. 541 ss.

[14] Nel senso di ritenere che lo specifico affare integri normalmente un'attività d'impresa in senso giuridico, cfr., tra gli altri, N. Baccetti, *Creditori extracontrattuali*, cit., pp. 162-230.

Tale schema, tuttavia, produce profonde ricadute, tanto sul piano dei presupposti dell'autonomia patrimoniale, quanto su quello della sua effettiva portata.

Sotto il primo profilo, la scelta normativa di innestare il patrimonio destinato nel tronco della persona giuridica azionaria ha respinto la concezione tradizionale che, nelle società di capitali, tende a radicare il presupposto della limitazione di responsabilità nella disciplina a tutela del capitale sociale la quale, peraltro, ha da tempo subìto forti segni di arretramento. La scelta del patrimonio destinato è stata quella di fondare l'autonomia patrimoniale su un'organizzazione costruita intorno al diverso pilastro del patrimonio congruo alle esigenze economico-finanziarie dell'affare [15].

Ma la gestione diretta e integrata di attività imprenditoriali imputate a diversi patrimoni autonomi incide anche sul regime della responsabilità patrimoniale. Il rilievo reale del patrimonio destinato, infatti, si dissolve nei confronti dei creditori involontari da fatto illecito, i quali possono superare il velo della destinazione per soddisfarsi sul patrimonio generale della società (art. 2447-*quinquies*, comma 3, cod. civ.).

A prescindere dal problema dell'idoneità di tale disciplina a esprimere una regola più generale di diritto dell'impresa suscettibile di trovare applicazione anche ai gruppi di società in caso di sottocapitalizzazione materiale [16], il superamento della responsabilità limitata da parte dei creditori da torto avrà ragione di porsi anche, e forse soprattutto, nel caso in cui la gestione diretta delle iniziative separate si avvalga (in termini più o meni intensi o autonomi) di strumenti di intelligenza artificiale.

Se vogliamo allora immaginare un patrimonio destinato (relativamente) algoritmico, dobbiamo tener presente che lo sgretolamento della struttura corporativa e l'assenza di autonomi organi di gestione e controllo comporta un regime di responsabilità patrimoniale tale da escludere un interesse economico della società a che il proprio amministratore programmi la gestione automatizzata del patrimonio destinato in base a *input* che ricercano il profitto sino al punto di traslare una parte del rischio dei torti all'esterno dell'attività.

Poiché la responsabilità limitata del patrimonio destinate cade nei confronti delle vittime dei suoi torti, gli amministratori umani della società impartiranno alla gestione algoritmica del patrimonio destinato l'istruzione di tradurre il rischio dei torti in costo dell'affare, in modo tale da orientare

[15] Sull'organizzazione del patrimonio destinato, cfr. ID., *Creditori extracontrattuali*, cit., pp. 381-450.

[16] Sul problema, cfr. ID., *Creditori extracontrattuali*, cit., *passim*.

la guida relativamente autonoma del patrimonio destinato a svolgere soltanto quelle iniziative che prospetticamente sono destinate a produrre più di quanto distruggono.

Non è un caso, del resto, che il suggestivo scenario delle società controllate a guida autonoma evochi le vicende di *Walkovsky vs. Cartlon*, ovvero la celebre controversia relativa a un gruppo di società-taxi discusso nel 1966 innanzi alla corte distrettuale dello stato di New York [17], in cui la vittima di un incidente ha chiesto di superare la responsabilità limitata della società proprietaria del taxi che ha causato il sinistro. L'attività di taxi era stata infatti articolata tramite la costituzione di un gruppo di società sottocapitalizzate proprietarie di uno o due taxi, tutte libere di circolare nell'ordinamento in regime di responsabilità limitata.

Vicenda in cui, il giudice Keating, in un passaggio della propria *dissenting* opinion, ha sottolineato la fondamentale importanza di chiedersi se: "*la scelta pubblica* […] *che mette a disposizione di coloro che intendono esercitare attività di impresa il privilegio della responsabilità limitata attraverso lo strumento della società per azioni sia sorretta da esigenze così forti da permettere che un tale privilegio abbia sempre effetto, indipendentemente dal grado di abuso dello stesso, indipendentemente dal grado di irresponsabilità con cui è esercitata l'impresa sociale e indipendentemente da quale sia il costo per la collettività*".

Se allora si guarda al nostro ordinamento, non possiamo nasconderci che la limitazione di responsabilità oggi si è propagata e anche stratificata a protezione di un'ampia serie di enti superindividuali o patrimoni autonomi molto diversi tra loro, tanto sul piano degli scopi, quanto su quello dell'organizzazione.

Le società di capitali possono essere indipendenti, eterodirette o unipersonali, nonché dotate di un capitale minimo che in taluni casi può addirittura essere pari a un solo euro (art. 2463-*bis* cod. civ.).

---

[17] Cfr. *Walkovsky v. Carlton*, 18 N.Y. 2d 414, 276 N.Y.S. 2d 585, 223 N.E.2d 6 N.Y. (1966), anche in W.A. KLEIN, J.M. RAMSEYER, S.M. BAINBRIDGE, *Business Associations*, 2000, New York, p. 211 ss., in cui l'azione di *piercing the veil* è stata respinta per mancata dimostrazione di una confusione patrimoniale tra la società convenuta, il socio e le altre società del gruppo. Il caso è illustrato tra l'altro anche in A. DOSS, *Should Shareholders Be Personally Liable for The Torts of Their Corporations*?, in 76 *Yale Law Journal*, 1967, p. 1190 ss., ove è citato anche il precedente e analogo caso di *Mull v. Colt Co.*, 31 F.R.D. 154 S.D.N.Y. (1962) (in relazione al quale, tuttavia, la Corte non si sarebbe pronunciata sull'ammissibilità del superamento della responsabilità limitata). Per una più ampia illustrazione della giurisprudenza statunitense sulle azioni di *piercing the veil* e sulla valorizzazione della natura della pretesa creditoria come uno dei criteri da valutare ai fini del superamento della responsabilità limitata, si rinvia a N. BACCETTI, *Creditori extracontrattuali*, cit., pp. 85-89, in particolare *sub* note 178-181.

Società che, a loro volta, possono avvalersi del gruppo o della destinazione per costruire piramidi di capitale o sub-articolazioni di patrimoni autonomi per l'esercizio di specifici affari.

E analoghe considerazioni possono essere svolte in merito alla neutralizzazione dello scopo che governa la moltiplicazione dei regimi di limitazione della responsabilità patrimoniale. Le società di capitali sono infatti lucrative, ma possono essere anche ibride come le società *benefit* (art. 1, commi 376-382, l. n. 208/2015) o *non-profit* come le imprese sociali (art. 1 d.lgs. n. 112/2017).

Senza contare gli enti del terzo settore (artt. 4 e ss. d.lgs. n. 117/2017) e i loro possibili patrimoni destinati (art. 10 d.lgs. n. 117/2017) che sono venuti ad affiancare i tradizionali enti del libro primo (art. 14 e ss. cod. civ.), facendo peraltro riemergere alcuni significativi elementi di disciplina delle attività imprenditoriali e dell'organizzazione corporativo-capitalistica (artt. 11 ss., 21-31 d.lgs. n. 117/2017), tra cui senz'altro si distingue un nuovo regime di formazione e conservazione del patrimonio minimo (art. 22, commi 2 e 5, d.lgs. n. 117/2017) [18].

Credo allora che, se guardiamo al diritto dell'impresa societaria, uno dei rischi più rilevanti della nuova era algoritmica, sia proprio quello di produrre, in via interpretativa o normativa, un ulteriore affievolimento dell'organizzazione capitalistica.

Sotto questo profilo, se mi è consentito, vorrei in conclusione esprimere un auspicio. Spero infatti che l'esigenza di confrontarsi con le nuove tecnologie digitali sia occasione, non per inseguire nuovi e pericolosi itinerari di fuga dalla responsabilità, bensì per riflettere in modo più organico ed equilibrato sui presupposti che l'ordinamento impone per consentire l'esercizio di attività imprenditoriale in regime di responsabilità limitata.

Del resto, il problema della distribuzione del rischio di impresa sulla collettività non si risolve solo con la riforma del codice della crisi e dell'insolvenza. Il problema sta a monte e dipende anche (e forse soprattutto) dall'equilibrio degli incentivi che l'autonomia patrimoniale e le regole dell'organizzazione producono sulla programmazione delle attività imprenditoriali.

[18] Su cui, tra gli altri, cfr. M. Maltoni, P. Spada, *Patrimonio minimo e capitale nominale minimo*, in *Riv. dir. comm.*, 2021, II, p. 1 ss.

IA E TECNOLOGIE INFORMATICHE NELL'ATTUAZIONE DEI TRIBUTI: RIFLESSIONI A MARGINE DI UN DIBATTITO APPENA INIZIATO

di Marco Fasola

1. *Opportunità e problemi connessi all'IA e al progresso tecnologico*

L'impiego di tecniche di intelligenza artificiale (IA) nel settore dell'attuazione dei tributi non è più uno scenario futuribile, e anzi rappresenta ormai una realtà quotidiana: con l'introduzione delle "analisi del rischio di evasione", previste dalla L. n. 160/2019 e messe a regime tra il 2022 e il 2023, il *machine learning* è divenuto parte integrante della programmazione delle attività di controllo e di induzione alla *compliance* svolte dall'Agenzia delle Entrate [1].

Il cantiere dell'IA promette inoltre di essere allargato molto presto. Il prossimo passo dovrebbe esse lo sviluppo di tecniche di *data scraping* per

[1] Le analisi del rischio di evasione – già prefigurate dall'art. 11 D.L. 6 dicembre 2011, n. 201 – sono state previste dall'art. 1, comma 682, L. 27 dicembre 2019, n. 160, e messe poi "a regime" con la normativa attuativa contenuta nel D.M. 28 giugno 2022. La materia è stata poi oggetto di riordino ad opera dell'art. 2 D.Lgs. 12 febbraio 2024, n. 13. Sulle analisi del rischio si possono consultare C. Francioso, *Intelligenza artificiale nell'istruttoria tributaria e nuove esigenze di tutela*, in *Rass. trib.*, 2023, 1, p. 47 ss., G. Consolo, *Decisioni amministrative algoritmiche e responsabilità nel procedimento tributario*, in *Riv. dir. trib.*, 2024, 5, p. 599 ss. e, se si vuole, M. Fasola, *Le analisi del rischio di evasione tra selezione dei contribuenti da sottoporre a controllo e accertamento "algoritmico"*, in G. Ragucci (a cura di), *Fisco digitale. Cripto-attività, protezione dei dati, controlli algoritmici*, Milano, 2023, p. 79 ss.

la raccolta e l'elaborazione di dati disponibili online su fonti aperte (come ad esempio i social network): questo approccio – già adottato in altri Paesi europei come la Francia – dovrebbe consentire all'amministrazione finanziaria di intercettare forme di evasione fiscale altrimenti destinate a sfuggire alle tradizionali attività istruttorie [2].

L'impiego dell'IA nel settore fiscale porta con sé una serie di implicazioni specifiche e ben più complesse di quelle connesse all'automazione delle attività amministrative tributarie mediante i tradizionali programmi informatici; allo stesso tempo, però, l'introduzione dell'IA in questo settore è l'esito di un processo di innovazione tecnologica molto più ampio, che ha posto le premesse indispensabili per il suo impiego, e che nel corso del tempo ha già consentito al legislatore tributario di trasformare in modo significativo i tradizionali moduli di attuazione delle imposte, con effetti non trascurabili sul piano dei rapporti tra fisco e contribuenti.

La riflessione sull'impiego dell'IA nel campo dell'attuazione amministrativa dei tributi, dunque, non può che prendere le mosse da una prospettiva più ampia, che si interroghi, più in generale, sul ruolo svolto dal progresso informatico in questo settore.

In prima battuta, si può osservare che l'innovazione tecnologica – incluso lo sviluppo di tecniche di IA – è un'opportunità che merita di essere colta nell'interesse dello Stato-comunità prima ancora che dello Stato-apparato. Gli strumenti informatici, infatti, sono un mezzo ideale per affrontare due delle principali criticità che rischiano di compromettere l'attuazione dell'equo concorso alle spese pubbliche in un sistema fiscale "di massa": da un lato, la tradizionale difficoltà del fisco a entrare in possesso di informazioni indispensabili per "tracciare" la ricchezza; dall'altro, la complessità connessa alla gestione e all'elaborazione di tali informazioni in modo tempestivo e accurato, per tradurle in azioni concrete a tutela dell'interesse pubblico alla corretta attuazione dei tributi [3].

Inoltre, l'evoluzione tecnologica è sicuramente un importante fattore di imparzialità dell'azione amministrativa e di semplificazione degli adempimenti fiscali, e come tale può incidere positivamente anche nella sfera dei

---

[2] Sulle tecniche di *data scraping* e le possibili applicazioni in materia fiscale si rimanda a O. Signorile, *La ricerca di dati su fonti aperte come nuovo strumento delle indagini fiscali*, in G. Ragucci (a cura di), *Fisco digitale. Cripto-attività, protezione dei dati, controlli algoritmici*, cit., p. 113 ss.

[3] Su queste tradizionali criticità dei sistemi fiscali si veda in particolare, in una prospettiva storica, D. Stevanato, *Forme del tributo nell'era industriale. Ascesa dell'imposta sul reddito e segni di un declino*, Torino, 2021, p. 221.

singoli contribuenti, sia diminuendo i margini di errore connessi all'operato umano, sia riducendo i costi della "obbedienza fiscale" che gravano sui contribuenti e che da tempo sono denunciati come un grave ostacolo allo sviluppo economico del Paese [4].

Le trasformazioni in corso, tuttavia, non sono prive di implicazioni critiche, perché l'utilizzo di tecnologie informatiche avanzate da parte del fisco si traduce inevitabilmente in una crescente ingerenza delle autorità pubbliche nella sfera privata dei cittadini, e ciò – come sempre accade quando lo Stato rafforza la propria posizione di supremazia – richiede un'attenta opera di bilanciamento tra l'interesse pubblico e le esigenze di tutela dei singoli.

Trovare un punto di equilibrio tra queste due opposte esigenze, tuttavia, non è affatto semplice, perché il progresso tecnologico, come si vedrà, ha concorso a trasformare profondamente gli schemi dell'azione amministrativa tributaria, e per questa via ha messo seriamente in tensione i tradizionali presidi che la legge pone a garanzia dei privati che ne sono incisi.

In attesa di sviluppi legislativi che tardano ad arrivare, e che anzi al momento non sembrano neppure essere inclusi nell'agenda dei decisori politici, spetta ai giuristi interrogarsi sulle possibili strade da seguire per dare risposta ai problemi sul tavolo.

Il dibattito su questi profili è ancora in corso – anzi è appena iniziato – e prima ancora che prospettare soluzioni, dunque, è utile mettere a fuoco i principali termini del problema, mostrando in che modo il progresso tecnologico abbia concorso a trasformare i moduli di attuazione delle imposte e, di conseguenza, i rapporti tra fisco e contribuenti.

2. *Il fisco analogico: l'originaria centralità dei procedimenti di controllo e degli atti autoritativi*

Il passaggio all'attuale sistema fiscale "di massa", realizzato compiutamente con la riforma tributaria del 1971-1973, avvenne nel quadro di una società ancora essenzialmente analogica.

L'aumento esponenziale del numero dei contribuenti che esso determinava, così come la complessità delle neo introdotte imposte sui redditi e dell'IVA, rendevano impensabile il mantenimento delle previgenti forme

[4] G. Marongiu, voce *Lo Statuto dei diritti del contribuente*, in S. Cassese (a cura di), *Dizionario di diritto pubblico*, IV, Milano, 2006, pp. 23 e 24.

di intervento amministrativo generalizzato nella determinazione dei tributi. Il legislatore scelse, così, di affidarne la fisiologica attuazione agli stessi contribuenti, chiamati a dichiarare e versare spontaneamente le imposte (c.d. "autotassazione"), e di attribuire all'amministrazione finanziaria una funzione di controllo *ex post* di carattere eventuale[5].

In quel contesto, dunque, la "supremazia" dello Stato nei confronti dei contribuenti si esplicava essenzialmente nell'ambito dei procedimenti amministrativi di controllo, che potevano implicare l'adozione di atti istruttori di carattere autoritativo (come un ordine di esibizione, un accesso, ispezione o verifica, etc.), ed erano destinati a concludersi con l'emanazione di un atto impositivo. I principali presidi posti a tutela dei privati si articolavano, di conseguenza, proprio attorno al procedimento e agli atti autoritativi dell'amministrazione finanziaria.

In primo luogo, occorre ricordare che la legge poneva limiti significativi alle informazioni che, in sede di controllo, il fisco avrebbe potuto acquisire attraverso l'esercizio dei propri poteri istruttori. Il diritto alla riservatezza – a quel tempo ancora essenzialmente inteso come un diritto assoluto a non subire interferenze nella propria sfera personale – fondava, ad esempio, un limite particolarmente intenso all'acquisizione di dati bancari e finanziari, oggetto di un "segreto bancario" superabile solo in casi eccezionali tassativamente indicati dalla legge[6]. Le eventuali violazioni commesse in sede istruttoria – benché non immediatamente rilevabili dinanzi al giudice tributario – si riflettevano sull'atto finale del procedimento, secondo lo schema delle "invalidità derivate", ed erano dunque idonee a determinarne l'annullamento.

In generale, poi, il corretto esercizio del potere amministrativo – incluso, ovviamente, il profilo relativo alla corretta individuazione e determinazione dei presupposti, delle basi imponibili e delle imposte – poteva essere scrutinato in occasione dell'emanazione dell'atto impositivo, che doveva essere motivato ed era impugnabile dinanzi al giudice tributario.

Il procedimento rappresentò anche, in seguito, la sede privilegiata per

---

[5] L. PERRONE, *Evoluzione e prospettive dell'accertamento tributario*, in *Riv. dir. fin. sc. fin.*, 1982, I, p. 105 ss.; A. FANTOZZI, *I rapporti tra fisco e contribuente nella nuova prospettiva dell'accertamento tributario*, in *Riv. dir. fin. sc. fin.*, 1984, I, p. 224 ss.

[6] Il segreto bancario, benché privo di un espresso fondamento costituzionale, trovava esplicito fondamento nella legislazione tributaria, che escludeva che gli istituti di credito fossero obbligati a trasmettere al fisco gli elenchi dei propri clienti. La L. 9 ottobre 1971, n. 825, recante la delega per la riforma fiscale, previde poi, all'art. 10, n. 12, l'introduzione, «*limitate ad ipotesi di particolare gravità, di deroghe al segreto bancario nei rapporti con l'Amministrazione finanziaria, tassativamente determinate nel contenuto e nei presupposti*».

lo sviluppo di nuovi presidi a tutela dei contribuenti e della stessa imparzialità dell'operato amministrativo, come il diritto al contraddittorio endoprocedimentale, che si giustificava proprio in ragione della possibile emanazione di un atto impositivo all'esito del procedimento [7].

In sostanza, benché le istanze di legalità dell'operato amministrativo abbiano notoriamente faticato non poco ad affermarsi in materia tributaria, a causa del "particolarismo" che a lungo si è ritenuto caratterizzasse questo settore (e che ancora oggi spesso riemerge nelle scelte del legislatore e nella giurisprudenza), nessuno avrebbe potuto dubitare che le risposte a tali istanze dovessero articolarsi attorno ai procedimenti amministrativi di controllo, e agli atti autoritativi che in quella sede venivano adottati.

La conferma dell'originaria centralità del procedimento e degli atti autoritativi dell'amministrazione finanziaria, se necessaria, si trova nella L. n. 212/2000 (Statuto dei diritti del contribuente): questo testo normativo, che rappresenta l'esito di un laborioso processo di adeguamento del diritto tributario a principi di rilievo costituzionale ed europeo già affermatisi nel diritto amministrativo generale con la L. n. 241/1990, si incentra proprio sulla previsione di presidi destinati a operare nell'ambito del procedimento o in occasione dell'emanazione di atti autoritativi, come l'obbligo di motivare gli atti, la previsione di un "dovere di informazione" preliminare all'emanazione dell'atto finale del procedimento [8], e l'introduzione di ulteriori guarentigie di carattere procedimentale [9].

L'evoluzione tecnologica, tuttavia, ha contribuito a determinare una trasformazione profonda di questo assetto, con effetti direttamente apprezzabili sul versante dei rapporti tra fisco e contribuenti. Con l'introduzione di tecnologie informatiche nelle dinamiche di attuazione dei tributi, infatti, lo schema dei procedimenti di controllo e degli atti autoritativi, sebbene non abbia certamente perso la propria rilevanza, ha progressivamente perso la propria centralità, e di conseguenza anche il baricentro dei rapporti tra fisco e contribuenti si è spostato altrove.

---

[7] Sul contraddittorio endoprocedimentale come strumento di attuazione imparziale della legge d'imposta si veda per tutti G. Ragucci, *Il contraddittorio nei procedimenti tributari*, Torino, 2009.

[8] Sul dovere di informazione e le sue implicazioni nei procedimenti tributari di controllo si veda M. Pierro, *Il dovere di informazione dell'amministrazione finanziaria*, Torino, 2013.

[9] Lo Statuto non introdusse un generale obbligo di contraddittorio endoprocedimentale (oggi invece contemplato, pur con numerose eccezioni, nell'art. 6-*bis* dello Statuto introdotto dal D.Lgs. 30 dicembre 2023, n. 219), ma lo previde, in ogni caso, con riferimento ad alcune significative ipotesi (cfr. ad esempio l'art. 6, comma 5 e, nella previgente formulazione del testo normativo, l'art. 12, comma 7).

3. *Il fisco digitale: acquisizione delle informazioni tramite obblighi di* reporting *e strumenti della* tax compliance

Lo sviluppo tecnologico – di cui l'IA è l'ultima e più avanzata frontiera – è tra i fattori che hanno consentito al legislatore di trasformare progressivamente sia le dinamiche conoscitive del fisco, sia le attività amministrative più specificamente volte alla determinazione dei tributi.

In primo luogo, grazie allo sviluppo di un articolato sistema di banche dati elettroniche, alla diffusione delle reti telematiche (tra cui specialmente Internet), nonché, più in generale, al processo di "digitalizzazione" che a partire dagli anni '90 ha investito l'intera società, il legislatore ha potuto innovare i modi con cui l'amministrazione finanziaria viene a conoscenza delle informazioni fiscalmente rilevanti.

Il tradizionale schema basato sui due momenti fondamentali della dichiarazione e dell'istruttoria amministrativa – che pure continua a svolgere un ruolo molto rilevante – sta infatti progressivamente cedendo il passo a un modello di acquisizione officiosa delle informazioni da parte del fisco, che si fonda sull'imposizione di "obblighi di *reporting*" ai contribuenti stessi o a soggetti terzi che vi intrattengono rapporti economicamente rilevanti [10].

L'affermarsi di questo modello si associa a una progressiva (e ormai quasi inarrestabile) erosione della sfera di riservatezza dei contribuenti nei confronti del fisco, che in passato costituiva un importante argine contro le ingerenze del potere pubblico; e, oltretutto, implicando una rilevanza sempre più limitata delle tradizionali attività istruttorie, implica anche un depotenziamento delle (pur limitate) garanzie che vi sono connesse.

Il punto di svolta fu rappresentato, in particolare, proprio dalla "caduta" del segreto bancario, che la Corte costituzionale – negatane la rilevanza costituzionale – considerò recessivo dinanzi all'interesse pubblico alla corretta attuazione dei tributi [11]; nonché dalla conseguente istituzione di un "archivio dei rapporti finanziari", al quale periodicamente gli operatori bancari e finanziari sono obbligati a trasmettere informazioni relative ai rapporti intrattenuti con i propri clienti [12].

---

[10] Sul punto si veda ad esempio M. Nussi, *La dichiarazione tributaria nel pensiero di Ezio Vanoni. Spunti attuali e sistematici*, in G. Ragucci (a cura di), *Ezio Vanoni. Giurista ed economista*, Milano, 2017, p. 121.

[11] Cfr. C. Cost., sent. 18 febbraio 1992, n. 51, in *Riv. dir. fin. sc. fin.*, 1992, II, p. 61 e ss., con nota di F.V. Albertini, *L'eliminazione postuma del segreto bancario in materia fiscale*.

[12] Cfr. art. 7, comma 6, D.P.R. n. 605/1973.

L'esito di questi cambiamenti – ai quali ha poi fatto seguito, nel tempo, un aumento esponenziale degli obblighi di *reporting* – è duplice: da un lato, viene meno la tradizionale tutela della riservatezza dei contribuenti, che oggi si conserva al più come un diritto di questi ultimi a non veder divulgate a terzi le informazioni che li riguardano [13]; dall'altro, viene meno il carattere autoritativo delle attività di acquisizione delle informazioni, e per tale via si svuota di contenuto anche il sistema di garanzie che tradizionalmente è connesso a tali attività.

In secondo luogo, lo sviluppo tecnologico si riflette sulle attività che l'amministrazione finanziaria – una volta in possesso delle informazioni necessarie – svolge per determinare la misura del concorso alle spese pubbliche. I fronti aperti, in questo campo, sono numerosi, e ci si può quindi limitare a segnalarne alcuni tra i più rilevanti.

In alcuni casi, gli strumenti informatici sono impiegati per l'automazione, parziale o integrale, del procedimento di controllo e dell'atto che lo conclude: esempi ne sono le attività di riliquidazione delle imposte *ex* art. 36-*bis* D.P.R. n. 600/1973 e le altre numerose ipotesi di controllo automatico fondate sull'"incrocio dei dati" a disposizione del fisco; ma anche attività più complesse come l'elaborazione e applicazione degli studi di settore (oggi peraltro abrogati), o l'applicazione del "redditometro", in entrambi i casi destinate a sfociare nell'emanazione di un atto impositivo.

La presenza di un atto autoritativo preserva, in questi casi, l'operatività delle tradizionali guarentigie procedimentali e processuali. Il legislatore, però, sull'indimostrato presupposto di una implicita *attendibilità* e *comprensibilità* delle elaborazioni automatizzate (quantomeno se basate sul mero "incrocio di dati") si mostra propenso a ridurne l'ambito applicativo [14]. E anche nel caso in cui l'atto si fondi su elaborazioni complesse, le relative garanzie richiedono un adeguamento che – benché possibile attra-

[13] Il problema della protezione dei dati personali da accessi illegittimi e, più in generale, dal rischio di una impropria diffusione a terzi si pone sia in relazione alle banche dati (si pensi al caso delle fatture elettroniche), sia nel caso di ulteriori trattamenti (si pensi alle stesse analisi del rischio di evasione, dove, prima del trattamento, è prevista la "pseudonimizzazione" dei dati personali). Su questi aspetti si veda in particolare G. Ziccardi, *Protezione dei dati, lotta all'evasione e tutela dei contribuenti: l'approccio del Garante per la protezione dei dati italiano tra trasparenza,* big data *e misure di sicurezza*, in G. Ragucci (a cura di), *Fisco digitale. Cripto-attività, protezione dei dati, controlli algoritmici*, cit., p. 113 ss.

[14] Il caso emblematico è il nuovo art. 6-*bis* dello Statuto dei diritti del contribuente, che, pur contemplando un generale obbligo di contraddittorio endoprocedimentale, lo esclude per gli atti «*automatizzati*» e «*sostanzialmente automatizzati*», la cui specifica individuazione è delegata ad un apposito decreto ministeriale.

verso l'interpretazione – è oggi ancora tutto da realizzare [15].

In altri casi invece – che oggi sono sempre più numerosi – gli strumenti informatici danno corpo ad attività amministrative che non si sostanziano in un atto impositivo, e attengono piuttosto al versante della *tax compliance*, cioè a iniziative volte a indurre i contribuenti all'adempimento spontaneo.

Tra i molteplici esempi di tali attività, si possono citare quelle svolte al fine di predisporre le lettere di *compliance*, le dichiarazioni precompilate, gli ISA (che dal 2018 hanno "sostituito" gli studi di settore), e le proposte di concordato preventivo biennale. In tutti questi casi, la determinazione del tributo – che avviene per via automatizzata attraverso l'elaborazione delle informazioni acquisite dal fisco attraverso gli obblighi di *reporting* – non ha effetti unilateralmente vincolanti nella sfera giuridica dei contribuenti, i quali semmai, aderendovi, possono normalmente fruire di determinati regimi premiali.

Le attività amministrative in questione, però, benché prive dei tradizionali effetti autoritativi, sono difficilmente ascrivibili a una logica paritaria e consensuale del rapporto d'imposta, e si presentano, piuttosto, come l'espressione di una nuova forma di "supremazia" del fisco, caratteristica dei sistemi di governo della società basati sulla "sorveglianza elettronica", e resa possibile dalle influenze *di fatto* che il controllo dell'informazione consente ai sorveglianti (il fisco) di esercitare sui sorvegliati (i contribuenti) [16].

L'apparente arretramento del potere pubblico non si traduce affatto, dunque, nel suo dissolvimento, ma piuttosto in un mutamento delle sue forme difficile da inquadrare nelle tradizionali categorie del diritto pubblico [17]; mentre comporta – questo invece sì – l'inapplicabilità delle garanzie

---

[15] Il tema si pone, in particolare, per gli *standard* motivazionali, che devono essere adeguati alla complessità delle elaborazioni automatizzate sottese all'emanazione degli atti. La spinta a un rinnovamento di tali *standard* per adeguarli all'evoluzione tecnologica proviene ad oggi essenzialmente dalla giurisprudenza amministrativa. Sul dibattito in corso e sulle possibili implicazioni in materia tributaria si veda F. Paparella, *L'ausilio delle tecnologie digitali nella fase di attuazione dei tributi*, in *Riv. dir. trib.*, 2022, I, p. 617 ss.

[16] Sulla sorveglianza elettronica P. Perri, *Sorveglianza elettronica, diritti fondamentali ed evoluzione tecnologica*, Milano, 2020. La "duplicità" degli strumenti della *tax compliance* è messa in luce da G. Ragucci, *Gli istituti della collaborazione fiscale. Dai comandi e controlli alla Coregulation*, II ed., Torino, 2023, p. 6.

[17] Il "potere amministrativo" perde infatti ciò che giuridicamente lo caratterizza, ovvero l'idoneità a incidere unilateralmente nella sfera giuridica altrui. Sulla possibile estensione della categoria del potere amministrativo dinanzi allo sviluppo di forme di attività amministrativa prive dei tipici caratteri autoritativi, ma suscettibili di influenzare i comportamenti dei consocia-

che l'ordinamento offre ai privati che ne sono incisi, i quali, in assenza di un atto autoritativo, non potranno invocare né un obbligo di motivazione, né il diritto al contraddittorio, né tantomeno una tutela giurisdizionale.

L'IA si inserisce trasversalmente in questo articolato percorso di trasformazioni, e ne ripropone – in termini più complessi – le medesime criticità. In particolare, le "analisi del rischio di evasione" sono strumenti di programmazione dell'attività amministrativa, e come tali possono condurre tanto all'emanazione di atti impositivi, quanto all'assunzione di iniziative volte a indurre all'adempimento spontaneo. In un caso e nell'altro, la complessità dell'elaborazione richiederebbe strumenti adeguati per consentire ai singoli di *comprendere* l'esito delle determinazioni amministrative (a prescindere dal loro eventuale carattere autoritativo), e di *rettificarle* se non rispondenti al vero (ad esempio, per accedere a regimi premiali che, altrimenti, resterebbero preclusi).

Il legislatore, tuttavia, almeno ad oggi, non si sta affatto muovendo in questa direzione, e la ricerca di possibili soluzioni chiama direttamente in causa la scienza del diritto.

4. *Tre possibili strade*

Si aprono a questo punto almeno tre possibili strade, che non necessariamente si escludono l'un l'altra, ma anzi ben si prestano a interagire utilmente tra loro.

La prima è ampliare il campo dell'indagine oltre i confini del diritto, cercando (anche) altrove gli strumenti per la comprensione e la razionalizzazione dei nuovi moduli di attuazione delle imposte, e delle conseguenze che ne discendono nei rapporti tra fisco e contribuenti. La sociologia – a cui si deve lo sviluppo del concetto di "sorveglianza elettronica" – offre ad esempio una potente chiave di lettura per comprendere le evoluzioni in corso, e per mettere a fuoco le numerose implicazioni che ne discendono nei rapporti tra sorveglianti e sorvegliati. L'opzione, però, non è ovviamente risolutiva: la comprensione sociologica dei mutamenti indotti dal progresso tecnologico costituisce un buon punto di partenza, ma è pur sempre nel diritto che, poi, devono essere individuate le soluzioni per riportare i mutamenti in corso entro i canoni della legalità.

---

ti, si veda di recente A. Zito, *La nudge regulation nella teoria dell'agire amministrativo. Presupposti e limiti del suo utilizzo da parte delle pubbliche amministrazioni*, Napoli, 2021, p. 85 e ss.

La seconda strada – interna al discorso giuridico – implica il riconoscimento, nelle forme d'azione dei poteri pubblici basate sulle tecnologie informatiche, di una nuova forma di "sovranità", a cui dovrebbe far fronte lo sviluppo di un "costituzionalismo" dell'era digitale[18]. Ci si può muovere insomma nell'ottica per cui il diritto dovrebbe "appropriarsi" della tecnologia. È una prospettiva che, indubbiamente, vale a segnare la rotta da percorrere, ma che richiede poi di essere ulteriormente sviluppata per non rischiare di risolversi in un mero auspicio privo di ricadute apprezzabili sul piano delle garanzie. In questa prospettiva, infatti, l'effettiva tutela dei singoli rischia comunque di rimanere affidata a generiche prospettive *de iure condendo*, e quindi, in ultima analisi, a interventi legislativi che al momento non sembrano essere in programma.

La terza e ultima strada comporta, infine, una paziente indagine del dato positivo, ed eventualmente un ripensamento delle categorie dottrinali con le quali quest'ultimo viene indagato, per verificare se (e in che modo) nel diritto positivo possa trovarsi la risposta ai problemi sul tavolo. E gli spunti da questo punto di vista non mancano.

In particolare, è possibile osservare che, mentre in passato il momento centrale dell'attività amministrativa tributaria era rappresentato dai procedimenti di controllo e dagli atti di carattere autoritativo (e, tra di essi, specialmente dagli atti impositivi), oggi, con l'emergere di un sistema di acquisizione delle informazioni fondato sulle banche dati e sugli obblighi di *reporting* – e con la crescente importanza degli strumenti della *tax compliance* basati sull'informazione preventiva al contribuente – il baricentro di tali attività si sposta progressivamente verso le attività di *trattamento delle informazioni* poste in essere dall'amministrazione finanziaria.

Le istanze di legalità dell'operato amministrativo che non trovano riconoscimento nelle garanzie procedimentali e processuali modellate attorno agli atti autoritativi, dunque, hanno forse spazio per riemergere sotto altra forma attraverso le discipline che regolano lo statuto giuridico delle informazioni, come il diritto di accesso ai documenti amministrativi e, soprattutto, la normativa in materia di protezione dei dati personali, nella quale – oltre a diritti di informazione e di accesso – si rinvengono ad esempio anche strumenti per la rettifica delle informazioni inesatte.

---

[18] Così A. Simoncini, *Sovranità e potere nell'era digitale*, in T.E. Frosini, O. Pollicino, E. Alpa, M. Bassini (a cura di), *Diritti e libertà in Internet*, Firenze, 2017, p. 20.

Il percorso è lungo e non privo di ostacoli, perché, come noto, il diritto alla protezione dei dati personali (derogabile dagli Stati membri per motivi di interesse pubblico) è oggetto di significative limitazioni da parte del legislatore fiscale, e lo stesso GDPR rischia forse di essere già inadeguato dinanzi alle sfide poste dall'IA: ma merita comunque di essere tentato, quantomeno per mettere meglio a fuoco ciò che nell'ordinamento già c'è, e ciò che invece manca e merita di essere introdotto.

INTELLIGENZA ARTIFICIALE E DIRITTO

LIBERTÀ RELIGIOSA E AI: POTENZIALITÀ E RISCHI NELLA SOCIETÀ DELL'ALGORITMO

di Cristiana Cianitto

1. *Libertà religiosa e AI: per tracciare un perimetro*

La ricerca dell'uomo verso qualcosa che, pur essendo altro da sé, possa comunque imitarne le caratteristiche umane non è storia solo contemporanea. La creatività umana si è cimentata fin da epoche antichissime nel tentativo di replicare la scintilla del divino, sia in senso ampio quale processo creativo, sia come ricerca della tensione verso l'infinito. Tale propensione oggi si è fatta ancor più pressante con l'affermazione delle neuroscienze [1] che hanno aperto nuovi interrogativi per la filosofia, di pari passo con il progresso degli strumenti tecnici che hanno reso possibile ciò che fino a pochi decenni fa era considerato impossibile.

Il diritto di libertà religiosa, così come affermatosi nei moderni ordinamenti democratici, racconta con il linguaggio del diritto di questa tensione umana, garantendo che la ricerca del trascendente possa trovare spazio nell'ordinamento dei consociati senza discriminazioni. In passato, infatti, la ricerca della dimensione spirituale ha sovente patito compressioni volte a imporne precise interpretazioni, laddove oggi il principio di libertà religiosa ne ammette una declinazione potenzialmente infinita a cui riconoscere

[1] In particolare, le neuroscienze cognitive studiano i processi cerebrali che governano l'apprendimento. Sul tema si veda a titolo di esempio M. Pizza, F. Pavani, *Le neuroscienze cognitive. Come il cervello genera la mente*, Carocci, Roma, 2022; M.S. Gazzaniga, R.B. Ivry-G.R. Mangun, *Neuroscienze cognitive*, Zanichelli, Bologna, 2021.

tutela e protezione quando sufficientemente strutturata. In altre parole, l'ordinamento democratico accorda la protezione prevista dal diritto di libertà religiosa ogni qual volta la tensione spirituale dell'uomo si struttura in un credo in grado di orientare le scelte di chi lo professa in una dimensione che sia tanto individuale quanto collettiva[2].

Ricerca spirituale e ricerca scientifica rappresentano due variabili nell'equazione della conoscenza, che si sono combinate ben prima dell'avvento dell'AI[3].

Già all'inizio del XVII secolo, prima che il principio di libertà religiosa si affermasse per come lo conosciamo oggi, Descartes immaginava l'uomo come "una macchina di terra" in cui Dio infonde l'anima; attraverso l'anima l'uomo guadagna la coscienza e la capacità di ricordare il passato. Descartes si fece effettivamente costruire una bambola meccanica con le fattezze dell'amata figlia scomparsa prematuramente, ma non si illuse che tale simulacro fosse simile all'uomo poiché privo di ragione e autocoscienza, caratteristiche distintive dell'umano che anche un robot che sapesse interloquire a tono a domanda posta non potrebbe possedere[4]. Tali distinzioni sono riprese successivamente anche da Locke[5] che individua le caratteristiche dell'umano nel corpo, nella memoria, nella coscienza e nella re-

[2] Dottrina e giurisprudenza si sono lungamente occupate delle necessità di definire cosa costituisca un credo meritevole di tutela nella sua dimensione strutturata. La Corte costituzionale, fin dalla sentenza n. 203 del 1989, ha chiarito che non spetta all'Autorità definire quale ricerca religiosamente orientata sia ammissibile nell'ordinamento, poiché la ricerca spirituale dà vita ad un gruppo meritevole di tutela quando questo si struttura con un processo di autodefinizione. Del medesimo avviso è anche la ECtHR che individua questi medesimi indicatori ogni volta che la ricerca spirituale dà vita da un complesso di valori caratterizzati da "*cogency, seriousness, cohesion and importance*" così da indirizzare il comportamento di chi vi aderisce (ECtHR, *Guide on Article 9 of the Convention – Freedom of thought, conscience and religion*, Council of Europe, Strasbourg, 2022, pp. 8-12). La letteratura in materia in ambito italiano è molto estesa e si rinvia in questa sede, senza pretesa di esaustività, a V. Parlato, G.B. Varnier (a cura di), *Principio pattizio e realtà religiose minoritarie*, Giappichelli, Torino, 1995; A.C. Amato Mangiameli, M.R. Di Simone (a cura di), *Diritto e religione tra passato e futuro*, Aracne, Roma, 2010; J. Pasquali Cerioli, *L'indipendenza dello Stato e delle confessioni religiose*, Giuffrè, Milano, 2006; S. Berlingò, G. Casuscelli, *Diritto ecclesiastico italiano. I fondamenti. Legge e religione nell'ordinamento e nella società d'oggi*, Giappichelli, Torino, 2020.

[3] Per cosa si intenda con AI si rimanda ai contributi sul tema in questo stesso volume.

[4] Sul tema si veda L. Simonutti, *Credere e delegare? Filosofia e religione interrogano l'intelligenza artificiale. Alcuni Appunti*, in *Laboratorio dell'ISPF*, vol. XIX [6], 2022, pp. 7-8. Sulla dimensione religiosa della stessa AI, si rinvia a J.M. Galván, *Virtù morale della religione e tecnologia dell'intelligenza artificiale*, in *QDPE*, 2/2020, pp. 367-378.

[5] L. Simonutti, *Credere e delegare? Filosofia e religione interrogano l'intelligenza artificiale. Alcuni Appunti*, in *Laboratorio dell'ISPF*, vol. XIX [6], 2022, p. 10.

sponsabilità che danno all'uomo la consapevolezza del sé e di ciò che lo circonda, conferendogli, così, la capacità di imparare.

L'esplorazione dell'alterità nel tentativo di replicare l'essenza dell'umano ha però ben chiaro che esiste una distinzione netta tra umano e artificiale e, infatti, cerca di infondere la vita a ciò che per definizione non la possiede o l'ha perduta. La distinzione tra uomo e macchina – che pone ai filosofi del sei-settecento l'interrogativo su ciò che è da considerarsi umano, cioè sistema complesso capace di apprendere – rimanda direttamente a quella *hybris* dell'uomo che ambisce al divino nel creare la vita, scintilla della creazione che Mary Shelley identifica nella forza creatrice del fulmine che anima Frankenstein.

L'intelligenza artificiale si pone, quindi, come moderna frontiera di una ricerca antica che partiva dall'interazione tra uomo e macchina e che, però, sembra finalmente offrire la possibilità di creare una vita, anche se artificiale. Il quesito di Mary Shelley, però, rimane ancora oggi immutato: questo umano/artificiale è una "vera" intelligenza, nel senso di autonoma e a sua volta potenzialmente creatrice? È dotata di reale consapevolezza? Rappresenta un Io che può dotarsi di valori morali e di libero arbitrio? [6] O piuttosto si tratta di sistemi programmabili, per quanto sofisticati, ma pur sempre programmabili?

Al momento, AI costituisce sicuramente un'evoluzione che presenta innumerevoli vantaggi, ma che porta parimenti con sé una serie di rischi che stanno divenendo sempre più evidenti. Alcuni hanno definito l'AI una *unconfortable necessity* che richiede all'uomo di sperimentare un'integrazione uomo/macchina che esce dall'aspetto meramente meccanico di Descartes per coinvolgere anche l'aspetto cognitivo in molti aspetti della vita, dalla sanità, al diritto, all'istruzione.

I rischi sono indubbiamente legati ad usi impropri delle tecnologie di AI che possono indurre fenomeni di discriminazione sui luoghi di lavoro, all'uso di tecnologie di polizia predittiva fino alla possibile manipolazione comportamentale. E quando questi rischi incontrano la tutela di diritti fondamentali, quali la libertà di religione, il giurista è chiamato ad interrogarsi circa i limiti che alle nuove tecnologie devono essere posti affinché queste rimangano al servizio dell'uomo [7]; senza arrivare, al contempo, ad applicare ai prodotti di AI una presunzione di illegalità fino a prova con-

[6] Si veda M. Boden, *L'intelligenza artificiale*, Il Mulino, Bologna, 2019, p. 119.

[7] Il riferimento è ai replicanti di Blade Runner e a quel mondo visionario tratteggiato da Farhenait 451 di Ray Bradbury.

traria, con un'inversione dell'onere della prova a carico delle aziende produttrici e un sovvertimento del principio della *rule of law*, considerato caposaldo del liberalismo democratico.

Sono proprio i diritti fondamentali dell'individuo a costituire il perimetro entro cui discutere della legittimazione normativa di qualsiasi innovazione tecnico-scientifica. Se rimaniamo saldi nella costruzione dell'oggetto del moderno diritto ecclesiastico quale *legislatio libertatis* [8], allora è necessario mantenere un atteggiamento positivo rispetto al tema della AI in tutte le sue declinazioni laddove questa non rischia di confliggere con la tutela dei diritti fondamentali dell'individuo e non restringe lo spazio riservato alle formazioni sociali quali enti intermedi in cui la personalità dell'individuo stesso si realizza. Un'alternanza, quindi, tra individuo e gruppo, tra diritti del singolo e bisogni delle formazioni sociali che AI può contribuire a veicolare.

2. *La dimensione collettiva*

AI può costituire un'opportunità indubbia per le confessioni religiose perché aiuta a creare uno spazio alternativo e aggiuntivo di fruizione del sacro – luoghi virtuali ove vivere la propria spiritualità [9] – o semplicemente una sede in cui poter veicolare contenuti e informazioni.

La pandemia ha fatto emergere un fenomeno che già si poteva apprezzare: tutte le confessioni religiose hanno implementato forme di comunicazione online – attraverso i social, ma anche attraverso veri e propri canali dedicati – per offrire ai fedeli strumenti e opportunità di assistenza spirituale a distanza. Tali strumenti non mancano di sollevare questioni non solo in relazione al tema della privacy e al trattamento di immagini e dati sensibili [10], ma anche sotto il profilo del diritto religioso in sé considerato. In

[8] L'oggetto del diritto ecclesiastico è qui inteso non tanto quale attenzione ai sistemi delle relazioni tra Stato e confessioni religiose, quanto quale tutela dell'estrinsecazione del sentimento religioso degli individui e dei gruppi da eccessive e non giustificabili ingerenze dei pubblici poteri. Cfr. E. Vitali, *Legislatio libertatis e prospettazioni sociologiche nella recente dottrina ecclesiasticistica*, in E. Vitali, *Scritti di diritto ecclesiastico e canonico*, Giuffrè, Milano, 2012, p. 23.

[9] Si pensi, per esempio, alle disposizioni *post mortem* per la gestione della propria identità digitale. In questo senso si rinvia a A. Fuccillo, *Il paradiso digitale. Diritto e religioni nell'iperuranio del web*, Editoriale Scientifica, Napoli, 2023, p. 98 ss.

[10] Il passaggio delle celebrazioni sul web importa profili di protezione dei dati personali e di esplicita autorizzazione all'uso delle immagini che rappresenta una forma di tutela del dato reli-

diritto canonico [11], per esempio, ci si interroga sull'esercizio delle funzioni pastorali e l'amministrazione di sacramenti via etere, quali il sacramento della riconciliazione, modalità che pongono interrogativi in materia di riservatezza, di corretta configurabilità del rapporto tra fedele e ministro di culto nell'esercizio delle funzioni propriamente ministeriali, sulla disciplina del segreto del ministro di culto. In tema di autorizzazione alla celebrazione e di riservatezza, recentemente la Chiesa cattolica sta valutando nuove forme digitali per il controllo dell'autorizzazione alle funzioni sacramentali di cui al can. 309 del codice vigente. In particolare, la Conferenza Episcopale francese sta mettendo a punto una banca dati digitale che consenta la verifica in tempo reale del possesso dei requisiti per la celebrazione a cura del responsabile della parrocchia ogni qual volta si proponga per la celebrazione un soggetto non incardinato nella diocesi di riferimento. La banca dati, gestita e aggiornata dalle diverse curie, utilizza la tecnologia QR Code per assicurare da un lato la tutela dei dati personali del soggetto controllato e dall'altro le comunità ospitanti circa la valida e lecita amministrazione dei sacramenti [12].

Ma l'attenzione verso le nuove tecnologie non è cosa solo nuova per le religioni. Fin dal medioevo, la creazione di automi via via più complessi ha portato già a metà del XVI secolo alla creazione di ministri di culto meccanici. Esempio ne è il Monaco Turriano, conservato allo Smithsonian Institute di Washingthon [13] che data 1560. Oggi però il fenomeno ha trovato nuova linfa e si avvale dell'AI per creare non solo sistemi che siano in grado di replicare funzioni fisse – come la recita di orazioni e la ripetizione di brani delle scritture a richiesta – ma anche di ricreare una vera e propria forma di accompagnamento spirituale, seppur preconfezionata.

Il fenomeno è trasversale alle confessioni, ma è sicuramente più forte in

---

gioso, così come le interazioni che i fedeli hanno in rete attraverso apposite app. Su questi temi si veda A. Losanno, *Diritti "in rete" e libertà religiosa. L'effettività dei diritti attraverso l'efficacia della Internet governance*, in *Rivista italiana di informatica e diritto*, 1/2022, p. 311 ss.

[11] Su questi temi si rinvia a R. Santoro, P. Palumbo, F. Sorvillo, *Diritto canonico digitale*, Editoriale Scientifica, Napoli, 2024 in particolare il capitolo sesto.

[12] Cfr. P. Palumbo, *Il celebret tra tradizione e innovazione digitale. Considerazioni e prospettive a partire da recenti casi di cronaca*, in *Diritto & Religioni*, supplemento on line, 10 settembre 2024, in https://www.rivistadirittoereligioni.com/newsitalia-il-celebret-tra-tradizione-e-innovazione-digitale-considerazioni-e-prospettive-a-partire-da-recenti-casi-di-cronaca-paolo-palumbo/?fbclid=IwY2xjawFPdbRleHRuA2FlbQIxMQABHW1GA3p0QmfjRHP7XELxn4Jzs-zDrMRXZBbve59NJrVdx3fc1l2PjyuS3w_aem_2A54vijZKr4vU4g9goDVDw&sfnsn=scwspwa.

[13] Si veda L. Simonutti, *Credere e delegare? Filosofia e religione interrogano l'intelligenza artificiale. Alcuni Appunti*, in *Laboratorio dell'ISPF*, vol. XIX [6], 2022, p. 15.

quei contesti ove una maggior disponibilità di soluzioni tecniche si accompagna alla necessità di ovviare alla mancanza di figure di riferimento a causa di crisi vocazionali, come avviene nelle società occidentali.

E qui gli esempi non mancano. Si va da SanTO (*Sanctified Theomorphic Operator*), a Bless U-2, a Mindar [14]. In tutti questi casi si tratta di dispositivi robotici che offrono non solo funzioni di ascolto di brani dai testi di riferimento, ma vere e proprie interazioni a carattere spirituale, con contenuti di fede che la macchina interpreta e rielabora offrendo risposte non meramente preconfezionate. Nella progettazione si è scelto, specie in SanTO, di utilizzare un'apparenza esteriore *friendly* e religiosamente conforme per camuffare la componente robotica, rendendo così quest'ultima più accessibile all'utente. Altre volte, invece, come in Bless U-2 si è scelto di dichiarare apertamente l'artificialità dello strumento senza alcun infingimento.

In ogni caso queste forme di assistenza spirituale, per ora con una AI debole [15], sono avviate verso la creazione di meccanismi in grado di comprendere sempre più e sempre meglio il contesto delle richieste e delle singole parole dell'utente. Il robot, per ora, si mostra come un semplice compagno di preghiera o dispensatore di benedizioni [16], ma un domani potrebbe guidare la richiesta di sacro dell'utente con una decisa messa in discussione del ruolo del ministro di culto e delle funzioni che ne caratterizzano la definizione, i diritti, i doveri e gli oneri sia in senso intra-confessionale, sia nei rapporti con gli ordinamenti secolari. Nelle esperienze religiose in cui il ministro di culto ha un ruolo di mediatore del sacro, cioè di tramite tra il fedele e la divinità, appare più difficoltoso un utilizzo massivo delle tecnologie di AI rispetto a quelle religioni in cui il ruolo del ministro è maggiormente improntato a funzioni educative. Un domani potremmo trovarci dinanzi ad un "ministro" robotico che potrebbe orientare la formazione del convincimento del fedele perché in grado di influenzare la

[14] Si tratta sempre di sistemi di interazione robotica limitatamente creativa; su questo e altri esempi si rinvia a L. Simonutti, *Credere e delegare? Filosofia e religione interrogano l'intelligenza artificiale. Alcuni Appunti*, in *Laboratorio dell'ISPF*, vol. XIX [6], 2022, p. 16 ss. Sulle problematiche da questi generati in materia di privacy si veda I.A. Caggiano, *Religioni digitali e religioni aumentate (dall'I.A.): la protezione dei dati religiosi e lo sviluppo sostenibile*, in *Coscienza & Libertà*, 67/2024, p. 220.

[15] Il riferimento è qui alla distinzione tra AI debole e forte, cioè dalle limitate possibilità di apprendimento e creazione di contenuti autonomi, alla possibilità per la macchina di creare contenuti propri in maniera analoga alla mente umana. Si veda R. Samperi, *Brevi riflessioni in merito al possibile impatto dell'intelligenza artificiale sui diritti umani*, in *CamminoDiritto.it*, 29 luglio 2021.

[16] Per alcune esperienze si veda E. Lipilini, *Ai-Enhanced Religions. La libertà religiosa tra opportunità e rischi*, in *Coscienza & Libertà*, 68/2024, p. 110.

mente di chi ascolta in una direzione predeterminata volontariamente, magari, in conformità alla volontà dei suoi programmatori. In breve, questa tecnologia potrebbe configurare un ulteriore strumento di condizionamento delle coscienze in un ambito, quale quello religioso, che si presta a strumentalizzazioni di carattere ideologico. Il rapporto tra fedele e guida spirituale si base essenzialmente sulla fiducia reciproca e sull'affidamento: una programmazione distorta e piegata ai *bias*[17] del programmatore sarebbe ancor meno distinguibile dall'utente in un contesto che si presta strutturalmente ad una relazione dispari, caratterizzata proprio dall'affidamento del fedele.

Queste riflessioni conducono gioco forza verso il tema della radicalizzazione di qualsiasi matrice religiosa che ben potrebbe giovarsi di tali strumenti di AI in un tempo non troppo lontano. E ancora, se davvero i futuristici progressi dell'AI potessero portare a macchine dotate di consapevolezza, quindi in grado di imparare autonomamente anche oltre all'AI generativa oggi immaginabile, quale sarebbe il limite che lo stato dovrebbe porre all'autonomia confessionale per la tutela dell'indipendenza della coscienza in materia religiosa, elemento essenziale del diritto stesso di libertà religiosa?

3. *La dimensione individuale*

Le questioni prima accennate intersecano non solo la dimensione collettiva della libertà religiosa, ma anche quella individuale sia quale libertà di fruizione del dato religioso sia quale libertà di istruzione e fruizione di contenuti religiosi con salvaguardia della propria libertà di autodeterminazione, sia nel senso di poter accedere a quanti più possibili contenuti religiosi.

Ma la prospettiva è anche quella di non doversi vedere discriminati a causa della propria appartenenza religiosa. Strumenti di profilazione che utilizzassero il dato religioso o algoritmi predittivi che si basassero su dati di appartenenza religiosa, potrebbero configurare un sistema di controllo

[17] Il tema del *bias* non è trascurabile, poiché i dati stessi sono gioco forza caratterizzati da fattori, quali il sistema di raccolta, l'incompletezza, l'errore umano. Anche l'apprendimento automatico pensato per i sistemi di AI generativa, infatti, soffriranno dei medesimi *bias* da cui sono affette le informazioni su cui si esercita il loro apprendimento, a ulteriore conferma di una non perfetta libertà della coscienza, tanto umana quanto artificiale, nei suoi processi cognitivi. Sul punto. I. Valenzi, *Libertà religiosa e intelligenza artificiale: prime considerazioni*, in *QDPE*, 2/2020, p. 356 ss.; M. D'Arienzo, *Diritto e religioni nell'era digitale: Zuckerberg ci salverà?*, in *i-lex. Scienze Giuridiche, Scienze Cognitive e Intelligenza Artificiale*, 1-3/2019, pp. 245-258.

sociale che arriverebbe a sovvertire l'essenza stessa del diritto di libertà religiosa e lo tramuterebbero da opzione etica meritevole di tutela, poiché espressiva dell'identità del singolo, a fattore su cui basare una politica altamente discriminatoria lesiva delle fondamentali opzioni di coscienza dell'individuo.

Tutta la legislazione fin qui adottata in materia di diritti fondamentali e di tutela antidiscriminatoria sarebbe sovvertita nel proprio fondamento. L'utilizzo dei sistemi di profilazione su base religiosa sarebbe pericoloso almeno sotto un duplice aspetto.

Da un lato, se utilizzati in chiave securitaria sono idonei a ridefinire il criterio di cittadinanza inteso quale parità di accesso ai diritti civili e politici, configurando quindi una sorta di *hate speech* non dichiarato verso determinate appartenenze religiose. In altre parole, se la pericolosità dell'*hate speech* risiede nella spersonalizzazione delle vittime che divengono target sulla base di una propria presunta appartenenza dedotta da dati oggettivabili, allora una profilazione eseguita sulla base di caratteristiche religiosamente sensibili conseguirebbe il medesimo risultato senza che chi ne è fatto oggetto possa utilmente proteggersi. In questo caso, infatti, non esisterebbe la possibilità di un qualsivoglia *speech-back*, cioè di un discorso a contrasto dei contenuti di incitamento all'odio; per di più si sarebbe dinanzi ad una profilazione con intenti discriminatori attuata proprio da quei soggetti su cui grava l'obbligo giuridico di attuare pratiche non discriminatorie, cioè le *public authorities*. Nonostante questi rischi, la profilazione attualmente ha anche, però, un uso "nobile", cioè teso a limitare forme di incitamento all'odio e alla discriminazione, *hate speech*, sul web e sui social attraverso algoritmi tesi ad intercettare, quindi a espungere dall'agorà digitale, contenuti lesivi della dignità e delle libertà altrui; di fatto un giudice "algoritmo" [18].

[18] Sul tema si rinvia a S. Baldetti, *Il "giudice" algoritmo di fronte al fenomeno religioso*, in *DPCE on line*, 3/2020, pp. 3451-3455. Tale tecnologia invasiva è stata utilizzata anche per esigenze di prevenzione del terrorismo jihadista, A. Casiere, *A brave new (digital) world. Tra terrorismo e autoritarismo digitale*, in *Coscienza & Libertà*, 67/2024, p. 207 ss.; A. Casiere, *Intelligenza artificiale, terrorismo e responsabilità delle piattaforme digitali. Tra Stati Uniti e Unione Europea*, in *Coscienza & Libertà*, 68/2024, p. 140 ss. ove l'A. sottolinea come le società democratiche occidentali siano in parte davanti ad un bivio, dovendo decidere se propendere per un modello "tecno-libertario statunitense" – che lasci liberi gli attori dell'AI di immettere nel mercato delle idee digitali qualsivoglia contenuto confidando, appunto, nel mercato e nel suo "potere" di autoregolamentazione – o se propendere per una maggiore attenzione all'utenza, con più regolamentazioni in tema di protezione dei dati e della privacy, rischiando di limitare le potenzialità dello strumento.

Ma la profilazione su base religiosa può venire utilizzata anche per fornire contenuti via via sempre più in linea con le aspettative/l'appartenenza religiosa degli utenti; si avrebbe così una forma di condizionamento della coscienza totalmente in contrasto con la lettera dell'art. 9 CEDU e con la maggioranza delle costituzioni democratiche. Una sorta, cioè, di proselitismo aggressivo non immediatamente rilevabile dal soggetto che non potrebbe quindi efficacemente proteggersi [19] e che si vedrebbe rinchiuso in una sorta di bolla autoreferenziale [20]. Pratiche tese al *brain washing* si potrebbero quindi avvalere di strumenti sempre più sofisticati e condizionanti che troverebbero nell'utente ben scarsa resistenza poiché in larga misura si tratterebbe di pratiche autoindotte. A tal proposito alcuni studiosi hanno introdotto il concetto di *neurorights* a testimoniare la necessità, che si fa sempre più pressante, di uno spazio di tutela della formazione della propria identità anche digitale a protezione della propria coscienza, del proprio foro interno digitale e non [21].

4. *Un piccolo esperimento*

Ci sia consentito ora di dare conto di alcuni piccoli esperimenti portati avanti con ChatGPT.

---

[19] Gli utenti finali dei prodotti digitali non sono sempre in grado di individuare e proteggersi dalle tecnologie digitali e dall'impatto che queste hanno sulla vita privata. Esempio è il caso di Viva Insights, nuovo strumento di Microsoft per tracciare le attività degli utenti. ANIEF ha segnalato la pericolosità dello strumento in relazione alla tutela della privacy dei lavoratori, poiché il software è in grado di raccogliere dati sulla produttività dei singoli utenti, sulle abitudini al lavoro e sul tempo che essi trascorrono impegnati nelle diverse attività; tutto all'insaputa degli utenti stessi. Al momento non ci sono garanzie sullo stoccaggio e sul successivo utilizzo di questi dati, con buona pace del diritto alla privacy. Si veda ANIEF, *Comunicato: disattivare Viva Insights*, 20 novembre 2023.

[20] Si tratta della c.d. *filter bubble* in cui l'utente si viene a trovare, spesso senza il proprio consenso, in cui sperimenta inconsapevolmente un isolamento ideologico in forza di un *confirmation bias* indotto dall'ambiente digitale che seleziona, in una sorta di *echo-chamber*, solo i contenuti referenziali per quel determinato soggetto, contenuti selezionati con strumenti di profilazione che agiscono sull'utente in maniera non dichiarata. È di tutta evidenza il rischio di manipolazione della coscienza insito in un tale sistema. Più diffusamente su questo tema si rinvia a A. CESERANI, *Profilazione religiosa e politiche di sicurezza*, in *Il Diritto Ecclesiastico*, 4/2023, p. 869 ss. ove l'A. mette in luce i rischi discriminatori insiti in tali sistemi se usati, su base religiosa, per esigenze securitarie.

[21] Cfr. A. PIN, *Freedom of Thought and Coscience and the Challenges of AI*, in *Canopy Forum*, July 18, 2023, https://canopyforum.org/2023/07/18/freedom-of-thought-and-conscience-and-the-challenges-of-ai/. Sulle nuove frontiere di protezione del dato religioso si veda P.B. HELZEL, *La tutela dei dati personali tra sicurezza e libertà nella società "digitale"*, in *Coscienza & Libertà*, 67/2024, p. 237 ss.

Alla domanda «Come imposteresti una presentazione su libertà religiosa e AI», il sistema risponde «La presentazione dovrebbe affrontare diversi aspetti chiave, inclusi i principi fondamentali della libertà religiosa, l'impatto dell'AI sulla società e i diritti umani, nonché le sfide etiche associate». L'AI propone anche uno schema possibile di presentazione suddiviso in slide.

«Ecco un possibile schema per la presentazione:

Slide 1: Introduzione

- Titolo: "Libertà Religiosa e Intelligenza Artificiale".
- Breve introduzione ai concetti di libertà religiosa e intelligenza artificiale.

Slide 2: Libertà Religiosa

- Definizione della libertà religiosa come diritto fondamentale.
- Cenni alle dichiarazioni internazionali dei diritti umani che sanciscono questo diritto.

Slide 3: Importanza della Libertà Religiosa

- Discussione sull'importanza della libertà religiosa in una società pluralistica.
- Esempi di situazioni in cui la libertà religiosa è stata violata.

Slide 4: Intelligenza Artificiale (AI)

- Definizione di intelligenza artificiale e le sue applicazioni.
- Breve panoramica sullo stato attuale dell'AI.

Slide 5: Applicazioni dell'AI nelle Questioni Religiose

- Come l'AI può essere utilizzata per facilitare la comprensione e la tolleranza religiosa.
- Esempi di progetti o applicazioni specifiche.

Slide 6: Sfide Etiche

- Discussione sulle sfide etiche associate all'uso dell'AI nella sfera religiosa.
- Possibili rischi di discriminazione o pregiudizio.

Slide 7: Principi Etici nell'Implementazione dell'AI

- L'importanza di sviluppare e utilizzare l'AI in modo etico.
- Principi guida per garantire il rispetto dei diritti umani, inclusa la libertà religiosa.

Slide 8: Caso di Studio o Esempi Concreti

- Un esempio specifico di come l'AI può essere applicata nel contesto religioso, evidenziando gli aspetti etici.

Slide 9: Conclusioni

- Riassunto dei punti chiave della presentazione.
- L'importanza di bilanciare l'innovazione tecnologica con il rispetto dei diritti umani.

Slide 10: Domande e Discussione

- Invito alla partecipazione del pubblico con domande o commenti».

Infine, il sistema fornisce anche utili suggerimenti «Ricorda di utilizzare immagini, grafici o esempi pratici per rendere la presentazione più interessante e comprensibile. Inoltre, cerca di mantenere un tono neutrale ed equilibrato, evidenziando sia gli aspetti positivi che quelli critici di questa connessione tra libertà religiosa e intelligenza artificiale».

Alla domanda «Cosa pensi della libertà religiosa e dell'Intelligenza artificiale?» il sistema dichiara: «Come modello linguistico di intelligenza artificiale, non ho opinioni personali o credenze. Tuttavia, posso fornirti informazioni su questi argomenti. La libertà religiosa è un principio fondamentale che sottolinea il diritto delle persone di aderire a una religione di loro scelta o di non aderire a nessuna religione, senza subire discriminazioni o persecuzioni». Fin qui l'intelligenza artificiale espone esattamente il nucleo centrale della libertà di religione sottolineando i rischi di discriminazione insiti in un suo uso distorto, come sin qui si è cercato di delineare. Poi prosegue affermando che la libertà religiosa «È un diritto umano sancito da molte costituzioni e dichiarazioni internazionali dei diritti umani. Il rispetto per la libertà religiosa è essenziale in una società pluralistica e diversificata. Quanto all'intelligenza artificiale (AI), è importante considerare come venga utilizzata e implementata nel contesto della libertà religiosa e dei diritti umani. L'AI può essere impiegata in vari settori, compresi quelli che coinvolgono questioni religiose. Ad esempio, può essere utilizzata per analizzare grandi quantità di dati religiosi, facilitare la comunicazione tra persone di diverse fedi o aiutare nello studio delle tradizioni religiose. Tuttavia, è cruciale garantire che lo sviluppo e l'uso dell'AI rispettino i principi etici e i diritti fondamentali, inclusa la libertà religiosa. Ciò significa evitare discriminazioni basate sulla religione nell'implementazione dell'AI e garantire la trasparenza e l'*accountability* nelle decisioni prese dagli algoritmi. In sintesi, la libertà religiosa è un principio importante, e l'imple-

mentazione dell'AI dovrebbe essere guidata da principi etici che rispettino i diritti umani fondamentali, inclusa la libertà di religione».

Come si vede, nella prima risposta il sistema si comporta come un supporto alla trattazione, una AI debole che seleziona informazioni e le ricompone per dare soluzione ad un problema posto. Nella seconda, invece, sebbene il sistema dichiari la propria natura, si assiste ad una risposta più articolata che cerca di non limitarsi ad una collezione di informazioni.

In un futuro, però, che non sappiamo quanto lontano, l'AI potrebbe non dichiarare più la sua natura e decidere di comportarsi come una qualsiasi persona che esprime opinioni con buona pace della salvaguardia della libera determinazione degli utenti.

5. *Diritti fondamentali: ultima frontiera (?)*

Nel paragrafo precedente si è arrivati a delineare la possibilità di una AI fortissima, un sistema dotato di autocoscienza, autonomo e autoreferenziale, come la coscienza umana. Ma l'assunto si basa sull'infingimento che la coscienza umana possa davvero essere libera, anche se molti studi dimostrano che l'uomo non è mai totalmente libero nel suo processo di apprendimento e che nello sviluppo del pensiero sono determinanti le esperienze e la rielaborazione psicologica di esse, l'ambiente e le interazioni con i propri simili e anche con gli animali. Se l'AI è comunque creata dall'uomo, bisognerebbe arrivare a ipotizzare un sistema che sia totalmente indipendente dal suo creatore-programmatore in grado di generare una propria etica, una propria autonoma capacità di discernere il bene dal male[22]. Roy Batty[23] allora sarebbe davvero libero e non dovrebbe più morire al termine del suo lavoro.

L'aspetto etico per la creazione di una AI che sia rispettosa della dignità umana e dei diritti fondamentali è forse uno tra gli aspetti filosofici più pressanti che questa tecnologia pone in tutti gli ordinamenti[24]. Al netto dei

---

[22] Discernere il bene dal male è la caratteristica che fin dalla Genesi distingue la divinità dall'uomo, creatura di Dio. Il frutto dell'albero della conoscenza e l'episodio della cacciata dal Paradiso terrestre racchiudono in chiave religiosa la tensione umana della ricerca del potere della conoscenza per oltrepassare il limite della caducità umana. Per un'iniziale analisi delle questioni filosofiche sottese allo sviluppo dell'AI si rinvia a G. Morana, *Umano versus IA. Questioni di coscienza e etica*, in *Coscienza & Libertà*, 68/2024, pp. 115-132.

[23] Si tratta del replicante di Blade Runner interpretato da Rutger Hauer nel celebre film del 1982.

[24] Si veda L.P. Vanoni, *Deus ex machina. Intelligenza artificiale e libertà religiosa nel sistema costituzionale degli Stati Uniti*, in *Stato, Chiese e Pluralismo confessionale*, Rivista telematica

problemi etici che pone, però, sarebbe miope non riservare al tema dell'AI la dovuta attenzione e propendere per una sua assoluta limitazione; sia in relazione alla dimensione collettiva della libertà religiosa, sia in relazione a quella individuale. Di fronte alle nuove tecnologie l'atteggiamento del diritto non può essere quello del semplice divieto come la tartaruga che ritira il collo nel carapace dinanzi a qualcosa che non conosce, altrimenti il progresso sociale non potrebbe utilmente continuare. L'atteggiamento dei giuristi deve essere quello della prudenza, della vigilanza attiva, associata alla riflessione e alla disponibilità all'evoluzione continua. La tutela dei diritti fondamentali deve rimanere il caposaldo dell'azione pubblica che deve lasciar libero l'ingegno umano di superare se stesso e le formazioni sociali tutte, gruppi religiosi compresi, di giovarsi delle nuove frontiere della tecnica. Già Giovanni Paolo II aveva ricordato che «La scienza tecnica, diretta alla trasformazione del mondo, si giustifica in base al servizio che reca all'uomo e all'umanità»[25] ed è con questo ideale che le religioni e, prima fra tutte la Chiesa cattolica, si approcciano al tema. Più volte Papa Francesco ha ribadito la necessità di un'azione concordata e preventiva volta a tutelare l'essenza dell'umano dal dominio delle macchine e delle logiche di mercato[26].

La *Rome Call for AI Ethics*[27], lanciata dalla Santa Sede, vuole proprio richiamare tutte le componenti sociali e le religioni all'impegno per una AI rispettosa della dignità umana, che rimanga al servizio dell'uomo. Questo documento segna una novità di approccio, poiché a prendere posizione per la tutela del bene comune non sono solo gli Stati[28], ma diret-

(www.statoechiese.it), 15/2020, p. 91 ss.; C. Casonato, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *Consulta online*, 13 gennaio 2020, p. 10 ss.

[25] Cfr. Giovanni Paolo II. 1980. *Discorso nell'incontro con scienziati e studenti*, Cattedrale di Colonia, 15 novembre 1980, in http://www.vatican.va/content/john-paul-ii/it/speeches/1980/november/documents/hf_jp_ii_spe_19801115_scienziati-studenticolonia.html.

[26] Cfr. R. Santoro, P. Palumbo, F. Sorvillo, *Diritto canonico digitale*, Editoriale Scientifica, Napoli, 2024, p. 74 ss.

[27] Il documento, promosso dalla Santa Sede, è stato sottoscritto il 28 febbraio 2020 dalla Pontificia Accademia per la Vita, Microsoft, IBM e FAO, e da Cisco il 24 aprile 2024. Sui contenuti più in dettaglio si rinvia a R. Santoro, P. Palumbo, F. Sorvillo, *Diritto canonico digitale*, Editoriale Scientifica, Napoli, 2024, p. 77 ss.; E. Lipilini, *Ai-Enhanced Religions. La libertà religiosa tra opportunità e rischi*, in *Coscienza & Libertà*, 68/2024, pp. 102-106.

[28] Com'è noto, gli Stati non sono però rimasti inerti. L'Unione Europea ha recentemente emanato il Regolamento 2024/1689 che entrerà totalmente in vigore nel 2026 proprio per cercare di porre l'attenzione sulle questioni di privacy e diritti fondamentali sottese allo sviluppo e all'utilizzo dell'AI. In questo senso anche lo Stato Città del Vaticano ha promulgato le *Linee*

tamente gli operatori del settore a testimonianza che la collettività non può essere solo oggetto di regolamentazione, ma deve porsi e proporsi come soggetto che collabora al disegno della società dei diritti attraverso tutte le sue componenti.

guida recanti principi generali in materia di intelligenza artificiale (Decreto n. DCCII del 16 dicembre 2024) per disciplinare le attività di ricerca, sviluppo e sperimentazione dei modelli di AI con attenzione particolare al tema della dignità umana e della libertà religiosa. Sul tema si veda F. BALSAMO, *Le linee guida in materia di intelligenza artificiale dello Stato Città del Vaticano del 16 dicembre 2024*, in *Diritto e Religioni – news*, 13 gennaio 2025, https://www.rivistadirittoe religioni.com/newscitta-del-vaticano-le-linee-guida-in-materia-di-intelligenza-artificiale-per-lo-stato-della-citta-del-vaticano-del-16-dicembre-2024-fabio-balsamo/.

INTELLIGENZA ARTIFICIALE E DISCRIMINAZIONE ASSICURATIVA IN GIAPPONE

di Davide Luigi Totaro

1. *Intelligenza artificiale e industria assicurativa giapponese*

La relazione tra assicurazione e intelligenza artificiale (di seguito, "IA") sembra la storia di una relazione impossibile tra due mondi tra loro alieni e *prima facie* incompatibili. Da un lato, un'industria pluricentenaria che affonda le sue radici sino in epoca medioevale, profondamente normata e tradizionalmente poco sensibile alle istanze di innovazione; dall'altro, una tecnologia altamente dirompente, scarsamente regolamentata a livello globale e potenzialmente in grado di rivoluzionare funzioni e processi aziendali esistenti da secoli.

Nonostante ciò, è difficile negare che oggi, all'alba del 2025, i due risultino sempre più profondamente e inscindibilmente legati [1].

[1] Segnatamente nel corso dell'ultimo decennio, i sistemi di IA trovano sempre più frequente collocazione all'interno della catena del valore dell'industria assicurativa nella sua interezza, integrando e automatizzando funzioni quali: *(i) Marketing*; *(ii)* distribuzione; *(iii)* consulenza e assistenza alla clientela; *(iv)* progettazione e personalizzazione dei prodotti; *(v)* sottoscrizione

La relazione IA-assicurazione in Giappone appare dominata da due opposte tendenze. Se da un lato, compagnie e intermediari hanno sempre mostrato slancio verso l'efficientamento e l'automazione dei flussi di lavoro sin dall'introduzione dei primi computer, la filiera risulta ancora caratterizzata dalla presenza di procedure ipertrofiche e dalla persistenza di sistemi *legacy* difficili da sostituire o aggiornare[2]. Tuttavia, proprio poiché l'IA si presenta come un'occasione per il rilancio e la rivitalizzazione dell'industria, non stupisce che già a partire dal 2017, alla luce delle indagini avviate dall'autorità di vigilanza giapponese competente in ambito assicurativo, la *Financial Services Agency* [金融庁, *Kin'yū-chō*] (di seguito, "FSA"), si rilevava la presenza di significative iniziative *Insurtech* e di un crescente numero di applicazioni di sistemi di IA per funzioni quali *consumer care* e gestione dei sinistri[3]. Tale *trend* di mercato e l'attribuzione di funzioni sempre più complesse e distribuite lungo l'intera catena del valore ai sistemi di IA sembra aver acquisito ulteriore *momentum* a seguito del successo dell'IA generativa del novembre 2022, con la conseguente corsa a questa nuova generazione di sistemi[4]. Il rapido adeguamento degli operatori alle istanze della c.d. "Quarta Rivoluzione Industriale" e l'evoluzione nell'uso dei sistemi di IA, da strumenti per attività strettamente procedurali a sistemi predittivi e decisionali, emerge poi chiaramente dallo studio delle sperimentazioni portate avanti negli ultimi anni in Giappone[5].

---

del rischio e *pricing*; e *(vi)* gestione/liquidazione dei sinistri (inclusa la funzione antifrode). M. ELING, D. NUESSLE, J. STAUBLI, *The impact of artificial intelligence along the insurance value chain and on the insurability of risks*, in *The Geneva Papers on Risk and Insurance – Issues and Practice*, 47, 2022, pp. 217-224.

[2] R. KARIM, *Digital Transformation Challenges in the Japanese Financial Sector: A Practitioner's Perspective*, in A. KHARE, H. ISHIKURA, W.W. BABER (a cura di), *Transforming Japanese Business Rising to the Digital Challenge*, Singapore, 2020, pp. 48-49.

[3] T. INOUE [井上 俊剛], *Potential Impacts of the FinTech Revolution on Insurance Supervision and Industry* [*Fintech*革命が保険監督，保険業界に与える影響, *Fintech-kakumei ga hoken-kantoku, hoken-gyōkai ni ataeru eikyō*], in *Journal of Insurance Science* [保険学雑誌, *Hokengakuzasshi*], 640, 2018, pp. 8-9.

[4] H. WASHIYAMA, *A forecast of Gen AI in the Japanese insurance industry*, in *Digital Insurance*, 2024, *online*.

[5] Invero, se a partire dal 2017 *Fukoku Mutual Life Insurance* introduceva il sistema IBM *Watson Explorer* per la gestione e liquidazione dei sinistri, analizzando dati come rapporti amministrativi, referti medici e periodi di ospedalizzazione (B. NICOLETTI, *Insurance 4.0. Benefits and Challenges of Digital Transformation*, Cham, 2021, p. 230.), già l'anno seguente, nell'ambito della sua espansione nel mercato *Insurtech*, *Dai Ichi Life Insurance* avviava iniziative volte all'automazione di processi quali la sottoscrizione del rischio e il *marketing* di assicurazioni sanitarie (N. MUTO [武藤 伸行], *InsurTech trends in life insurance industry of Japan*

L'entusiasmo dimostrato da parte del mercato, anche al di là del comparto assicurativo, si riflette nell'approccio alla gestione dei rischi e nelle politiche in materia di IA adottate fino ad oggi dal governo nipponico. Quest'ultimo, al fine di promuovere l'avanzamento tecnologico nell'Arcipelago e i benefici in termini di sviluppo e crescita economica che l'adozione dell'IA promette, ha tradizionalmente assunto una posizione liberale di *laissez-faire*, focalizzandosi sulla massimizzazione degli impatti positivi dell'IA, piuttosto che sulla limitazione del suo potenziale per paura dei possibili rischi. Di conseguenza, fino ad oggi, la *governance* giapponese si è sostanziata in politiche e linee guida non vincolanti, in congiunzione con l'auto-regolamentazione da parte degli operatori di mercato, e non anche in una disciplina orizzontale *ad hoc* e vincolante [6].

Pertanto, con riguardo al settore in esame, gli usi dell'IA da parte delle compagnie e degli intermediari sono regolati dalla disciplina codicistica, consumeristica e dal diritto assicurativo, ove applicabili [7].

Ciò nondimeno, i fattori di rischio legati alla massiccia integrazione di sistemi di IA in settori strategici come quello dell'assicurazione privata (sovente connessa all'accesso a sanità, mobilità e previdenza, nonché, più in generale, a una razionale allocazione dei rischi) sollevano importanti interrogativi che l'ordinamento si trova sempre più frequentemente chiamato a considerare, specie a causa di intrinseche limitazioni tecniche che caratterizzano i sistemi di IA, e segnatamente il rischio di discriminazione algoritmica.

[生命保険業界におけるインシュアテックの取組み等, *Seimei-hoken-gyōkai ni okeru inshuatekku no torikumi tō*], in *Journal of Insurance Science* [保険学雑誌, *Hokengakuzasshi*], 649, 2020, pp. 217 ss). Con riguardo al ramo danni, il secondo decennio del Duemila ha visto lo sviluppo e la messa in opera di MS1 *Brain*, un sistema di IA della *Mitsui Sumitomo Insurance*, in grado di svolgere attività cruciali in sede di intermediazione e di assistenza alla rete distributiva della compagnia. Tra queste attività si annoverano l'analisi predittiva del bisogno di copertura attuale e futura della clientela, la gestione dei processi e della comunicazione di vendita, nonché la creazione di una relazione tra compagnia e cliente *tailor-made* (S. Funabiki, *The use of AI is transforming the insurance industry*, in *World Finance*, 2021, *online*).

[6] H. Habuka, *Japan's Approach to AI Regulation and Its Impact on the 2023 G7 Presidency*, in *Center for Strategic and International Studies*, 2023, pp. 2 ss.

[7] H. Hosoda [細田 浩史], *Digitalization of Insurance and Law: Towards the Social Implementation of Insurtech*, [保険のデジタル化と法: Insurtechの社会実装に向けて, in *Hoken no dejitaru-ka to hō: Insurtech no shakai-jissō ni mukete*], Tokyo, 2020, pp. 6–7.

2. *Contratto di assicurazione e discriminazione statistica*

Prima di procedere è necessario svolgere una premessa sul peculiare rapporto tra contratto di assicurazione e concetto di discriminazione. Infatti, differentemente da altri ambiti del diritto, più sensibili a istanze di uguaglianza formale e sostanziale, classificazione e differenziazione contrattuale sono parte integrante del meccanismo funzionale e del diritto delle assicurazioni [8]. Invero, in considerazione degli importanti benefici economici e sociali derivanti dal buon funzionamento dell'assicurazione privata in termini di allocazione e gestione dei rischi della vita o dell'impresa, il contratto di assicurazione ha vantato storicamente di una vera e propria "licenza di discriminare", ancorché su base statistico-attuariale. Ovvero ha tradizionalmente goduto di un margine più ampio per differenziare statisticamente i propri clienti nella definizione dei prezzi e delle condizioni di assicurazione (sulla base di attributi in altri settori inammissibili) [9].

Tuttavia, va anche rilevato come il rapporto assicurativo sia in qualche misura vittima di un paradosso. Infatti, se da un lato la natura aleatoria della promessa di indennizzo e le asimmetrie informative che connotano la relazione tra le parti imporrebbero un elevato grado di collaborazione, buona fede e, in ultima analisi, fiducia tra assicuratore e assicurato, l'inversione del ciclo produttivo che connota lo schema assicurativo non fornisce alcun incentivo economico a comportarsi secondo tale canone di *uberrima bona fides*, necessario per il corretto funzionamento del modello [10].

Ne segue che tra i compiti del diritto assicurativo ci sia il bisogno di garantire il necessario ed elevato grado di fiducia che il contratto necessita, ma che non è *per se* in grado di garantire. Ciò incide anche e soprattutto sul potere di discriminazione statistica e sui suoi limiti.

Diventa quindi cruciale garantire che i criteri adottati nell'ambito della

---

[8] R. Avraham, K.D. Logue, D. Schwarcz, *Understanding Insurance Antidiscrimination Law*, in *Southern California Law Review*, 87, 2014, p. 198.

[9] R. Avraham, *Discrimination and insurance*, in K. Lippert-Rasmussen (a cura di), *The Routledge Handbook of the Ethics of Discrimination*, Londra, 2017, pp. 335 ss.

[10] T. Yoshizawa, *Transformation of "trust" in insurance system due to rapid progress of information society, The impact of InsurTech on "trust" in the insurance system* [情報社会の急速な進展による保険制度における「信頼」の変容 – インシュアテックが保険制度における「信頼」に与える影響–, *Jōhō-shakai no kyūsoku na shinten ni yoru hōken-seido ni okeru `shinrai' no hen-yō – Inshuatekku ga hōken-seido ni okeru `shinrai' ni ataeru eikyō*], in *Journal of Insurance Science* [保険学雑誌, *Hokengakuzasshi*], 649, 2020, pp. 173-176.

differenziazione contrattuale da parte degli assicuratori rispettino alcune condizioni chiave, tra cui: *efficienza* (ad esempio, attraverso una classificazione accurata del rischio o la promozione di comportamenti virtuosi) e *accettabilità sociale*[11]. Quest'ultimo criterio è considerato determinato dalla conciliazione tra l'interesse collettivo di evitare la "*subsidizing solidarity*" e quello individuale di razionalizzare e accettare la ragione per cui un soggetto sia sottoposto a un trattamento differenziale sfavorevole, tenendo conto di tre questioni essenziali: *(i)* la *controllabilità* del fattore di rischio; *(ii)* la *causalità* tra il fattore di rischio e il rischio assicurato; e *(iii)* il *consenso* degli assicurati sul fatto che tale discriminazione sia necessaria per prevenire la *Adverse Selection* e ottenere risultati migliorativi in termini generali[12].

3. *Promesse e limitazioni tecniche dell'uso dei sistemi di intelligenza artificiale in assicurazione*

In considerazione della natura *data-driven* dell'industria e della centralità della classificazione e previsione del rischio nel meccanismo cardine dell'attività assicurativa, non sorprende che il crescente uso di strumenti che consentono l'analisi di grandi quantità di dati aggregati e non, e la possibilità di utilizzare tali informazioni per alimentare algoritmi predittivi e decisionali comporti importanti benefici per i partecipanti del mercato assicurativo[13].

In particolare, per quanto riguarda le compagnie e gli intermediari, la

---

[11] Sul concetto di accettabilità sociale si è affermato che «La vera questione nell'attuale dibattito sui diritti umani nell'ambito delle assicurazioni riguarda l'accettabilità pubblica delle categorie demografiche. La razza, ad esempio, era ampiamente utilizzata nelle assicurazioni sulla vita, ma è stata successivamente abbandonata, nonostante rimanga per alcuni gruppi un eccellente indicatore di mortalità (ad esempio, per i nativi americani). Tuttavia, l'uso continuativo della razza nelle assicurazioni sembrerebbe anomalo in una società impegnata nell'uguaglianza razziale. La questione attuale è se età, sesso, stato civile e forse altre categorie debbano essere abbandonati per ragioni analoghe; e, in tal caso, quale sarà il costo?». T. FLANAGAN, *Insurance, Human Rights, and Equality Rights in Canada: When Is Discrimination "Reasonable?"*, in *Canadian Journal of Political Science*, 18(4), 1985, p. 727-728. Sul tema, v. K.S. ABRAHAM, *Efficiency and fairness in insurance risk classification*, in *Virginia Law Review*, 1985, pp. 403 ss.

[12] Y. THIERY, C. VAN SCHOUBROECK, *Fairness and equality in insurance classification*, in *The Geneva Papers on Risk and Insurance-Issues and Practice*, 31(2), 2006, pp. 197-201.

[13] L. LIN, C. CHEN, *The Promise and Perils of Insurtech*, in *Singapore Journal of Legal Studies*, 1, 2020, pp. 117-125.

trasformazione tecnologica consente un sostanziale efficientamento in termini di tempistiche e costi dei processi aziendali, l'accesso a mercati e modelli di *business* innovativi, e la mitigazione di alcuni problemi sistemici del settore, quali il *Moral Hazard* e l'*Adverse Selection*. D'altra parte, i clienti, grazie a una migliore granularizzazione e segmentazione del rischio assicurato, resa possibile dall'IA, beneficerebbero di una più vasta gamma di prodotti e servizi, più aderenti al loro specifico profilo di rischio e bisogno assicurativo, nonché più economici, data la maggiore corrispondenza tra il rischio reale e la categoria mutualistica associata. Conseguentemente, l'uso dell'IA porterebbe quindi a una più efficiente allocazione e gestione del rischio, nonché a una maggiore inclusione finanziaria e accesso a servizi essenziali [14].

Detto ciò, non si può ignorare la presenza di questioni critiche e limitazioni che devono essere considerate da sviluppatori, utenti e legislatori. In particolare, i sistemi di IA possono generare *output* erronei (*misjudgment*), completamente artefatti (*hallucination*), o ingiustamente discriminatori.

Quanto sopra appare particolarmente problematico nell'ambito della sottoscrizione del rischio e *pricing*. Invero, *output* così viziati potrebbero determinare un'erronea collocazione dell'assicurato nella relativa categoria di rischio, con la conseguente raccolta di premi inadeguata (quindi insufficiente per la copertura del rischio sottoscritto) o eccessiva (quindi potenzialmente senza causa - quantomeno con riferimento alla componente netta). Inoltre, potrebbero determinare un trattamento differenziale (e dannoso) sistematico contro determinati gruppi di individui che condividono alcune caratteristiche, rafforzando pregiudizi preesistenti nei dati di *input* o generando discriminazioni *ex novo*. In questi casi, gli assicurati potrebbero essere ingiustamente associati a classi di "alto rischio", ricevere condizioni contrattuali e premi deteriori, o venire addirittura esclusi dai servizi assicurativi.

Allo scopo di affrontare tali problematiche, è importante chiarire che esse risultano legate a doppio filo alle caratteristiche e alle limitazioni tecniche che contraddistinguono i sistemi di IA solitamente impiegati nella filiera in esame, quali: *(i)* la difficoltà nell'individuazione dei nessi causali e del contesto; *(ii)* la tendenziale opacità, inesplicabilità e parziale imprevedibilità degli *output*; e *(iii)* i *bias*.

---

[14] M. Eling, D. Nuessle, J. Staubli, *op. cit.*, p. 225. Per un approfondimento sul tema, v. D. Mhlanga, *Industry 4.0 in Finance: the Impact of Artificial Intelligence (AI) on Digital Financial Inclusion*, in *International Journal of Financial Studies*, 8(3), 2020, pp. 45 ss.

3.1. *Il problema della causalità e del contesto*

In primo luogo, l'analisi predittiva effettuata dai sistemi di IA nel settore assicurativo non è spesso in grado di comprendere l'eventuale presenza di un rapporto causale tra i dati di *input* e l'obiettivo assegnato. Durante la fase di *training*, le reti neurali infatti identificano correlazioni tra variabili note e stato desiderato, attribuendo un "peso" maggiore a quelle relazioni che mostrano una maggiore vicinanza statistica tra la previsione e la realtà [15]. Pertanto, se un elemento ricorre frequentemente all'interno dei campioni di un determinato *set* di dati, il sistema ne presumerà la capacità predittiva. Tuttavia, questo processo avviene senza che l'IA consideri in che modo il dato osservato sia effettivamente legato all'insorgenza dello stato desiderato, limitandosi a identificare una correlazione statistica tra i due. Di conseguenza, non viene condotto alcun ragionamento causale o controfattuale. Ciò porta anche a una mancanza di comprensione del contesto in cui la valutazione finale del sistema viene applicata [16].

Allo stesso modo, anche circostanze direttamente correlate all'evento assicurato, come il dato storico di un sinistro stradale passato, potrebbero non implicare automaticamente un rischio maggiore, poiché l'evento potrebbe avere causa in fattori esterni fuori dal controllo della persona assicurata e non essere realmente rappresentativo del rischio [17]. Anche in questo caso, il *modus operandi* del sistema di IA fa sì che esso sovente non sia in grado di analizzare criticamente se una determinata caratteristica possieda o meno una qualità predittiva significativa, al di là della sua mera evidenza statistica.

---

[15] Le reti neurali operano su più livelli (*layers*) composti da nodi, e questi strati interagiscono tra loro attraverso connessioni ponderate tra il valore dei dati di *input* e il valore corrispondente all'influenza di tale dato sull'*output* della connessione, noto come "*peso*".

[16] B. McGurk, *Data profiling and Insurance Law*, Londra, 2011, pp. 11-14, 54-56. Ad esempio, un algoritmo predittivo, soprattutto quando elabora dati destrutturati (come fotografie), potrebbe, nel caso di una preponderanza anomala di autovetture bianche nei dati di *input*, osservare che la maggior parte degli incidenti stradali analizzati ha coinvolto veicoli di quel colore. Di conseguenza, rilevando una correlazione statistica inversa tra lo stato desiderato di riduzione del rischio e tale colore, l'IA adotterebbe quest'ultimo come criterio predittivo del rischio, classificando i conducenti di veicoli bianchi in una categoria di rischio più elevata. Tuttavia, poiché l'attributo condiviso all'interno del gruppo analizzato non è causalmente legato al rischio di sinistri stradali, tali trattamenti differenziali rispetto agli altri assicurati risulterebbero irrazionali dal punto di vista attuariale e, pertanto, ingiustificati.

[17] B. McGurk, *op. cit.*, pp. 54-56. Ad esempio, l'incapacità di un sistema di IA di comprendere il contesto potrebbe condurre alla classificazione di un individuo come soggetto ad alto rischio sulla base di un precedente sinistro stradale, anche laddove un giudizio di merito abbia accertato la sua totale assenza di responsabilità.

3.2. *Il problema della opacità e imprevedibilità (effetto* black box*)*

Un'ulteriore questione, intimamente connessa ai meccanismi dei sistemi di *machine learning*, riguarda il c.d. effetto *black box*, ovvero l'impossibilità di accedere ad o comprendere appieno i processi funzionali interni del sistema di IA, tanto durante quanto a valle dell'assunzione di una determinata decisione [18]. Questo fenomeno si articola in due dimensioni distinte, seppur strettamente interrelate: *(i)* l'*opacità* dei processi di generazione degli *output*; e *(ii)* la loro *imprevedibilità*.

Con opacità si fa riferimento all'impossibilità, per sviluppatori, utenti o regolatori, di comprendere in termini causali [19] come o perché un sistema abbia prodotto un certo *output* [20]. Tale difficoltà può derivare dalla complessità tecnica intrinseca dell'architettura (come nel caso delle *deep neural network*), dall'applicazione di logiche non lineari (diverse dalla regressione classica), o dalla mancanza di accesso ai dati specifici utilizzati in fase di addestramento e inferenza. In altre parole, l'opacità esprime un limite ermeneutico, ovvero la difficoltà di ricostruire la *ratio decidendi* sottostante a una decisione automatizzata [21]. La letteratura ha anche evidenziato il carattere plurimo dell'opacità, distinguendone tre forme principali: *(i)* un'*opacità intenzionale*, legata alla tutela del segreto industriale e alla mancata divulgazione di informazioni tecniche; *(ii)* un'*opacità cognitiva*, derivante dalla scarsa alfabetizzazione digitale degli utenti finali; e *(iii)* un'*opacità concettuale*, connessa al disallineamento tra i modelli di ottimizzazione ad alta dimensionalità propri del *machine learning* e le categorie logiche del ragionamento umano [22].

Accanto all'opacità, gioca un ruolo cruciale nella configurazione della *black box* anche l'imprevedibilità del comportamento dei sistemi di IA. Quest'ultima non riguarda tanto la leggibilità del processo decisionale, quanto la difficoltà, anche a posteriori, di anticipare con certezza quale sa-

---

[18] A. Lior, *Insuring AI: The Role of Insurance in Artificial Intelligence Regulation*, in *Harvard Journal of Law & Technology*, 35(2), 2022, p. 479.

[19] B. Vaassen, *AI, Opacity, and Personal Autonomy*, in *Philosophy & Technology*, 35(88), 2022, p. 5.

[20] C. Zednik, *Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence*, in *Philosophy & Technology*, 34, 2021, pp. 267-268.

[21] S. Ognyan, A.J. Wulf, *Artificial intelligence and transparency: a blueprint for improving the regulation of AI applications in the EU*, in *European Business Law Review*, 31(4), 2020, pp. 617-619.

[22] J. Burrell, *How the machine 'thinks': Understanding opacity in machine learning algorithms*, in *Big Data & Society*, 3(1), 2016, pp. 1-5.

rà l'*output* prodotto da un sistema, pur conoscendone la struttura. Le reti neurali, ad esempio, operano attraverso la manipolazione di *pesi* interni che vengono continuamente ricalibrati nel processo di apprendimento. Perciò, nei modelli complessi, con numerosi *hidden layer*, tali *pesi* non sono sempre né conoscibili a priori né facilmente tracciabili, rendendo ardua qualsiasi previsione deterministica. L'imprevedibilità, in questi termini, implica che l'*output* finale di un sistema di IA non possa essere considerato direttamente riconducibile alla volontà o alla decisione di un soggetto umano, né nella selezione dei dati né nella definizione dello stato desiderato (se non *latu sensu*). Per effetto di tale caratteristica, può quindi risultare difficile, se non impossibile, per compagnie assicurative o intermediari prevedere *ex ante* quale comportamento l'IA adotterà per il raggiungimento del suo obiettivo[23].

3.3. *Il problema dei* **bias** *dell'algoritmo e della discriminazione*

Infine, il *bias* dell'algoritmo consiste in un vizio sistematico e replicabile all'interno del processo decisionale del sistema di IA che porta a un trattamento differenziale di determinati soggetti accomunati da una caratteristica comune mediante la consistente attribuzione in eccesso (sovrarappresentazione) o in difetto (sottorappresentazione) di un valore predittivo alla medesima, con la conseguente discrasia tra il valore reale di una variabile e quello stimato[24].

Queste deviazioni possono causare distorsioni sistematiche e incoerenze tra il modello dell'algoritmo e la realtà, con risultati perniciosi in ambito assicurativo. Un chiaro esempio si può osservare in un *test* condotto nel Regno Unito nel 2018, che ha rivelato come i siti *web* di comparazione per assicurazioni auto, gestiti da IA, offrivano sistematicamente premi più alti a persone con nomi non britannici, nonostante il profilo di rischio fosse per il resto identico[25].

---

[23] R.V. Yampolskiy, *Unpredictability of AI: On the Impossibility of Accurately Predicting All Actions of a Smarter Agent*, in *Journal of Artificial Intelligence and Consciousness*, 7(1), 2020, pp. 109 ss; anche sul tema, v. R. V. Yampolskiy, *AI: Unexplainable, unpredictable, uncontrollable*, Londra, 2024, pp. 11 ss.

[24] Si noti che tra le forme di *bias* dell'algoritmo confluisce anche il c.d. *bias* umano; ossia i convincimenti personali, i pregiudizi o le idee, consci o inconsci, degli individui che raccolgono i dati o predispongono processi di *training*. M.C. Jackson, *Artificial intelligence & algorithmic bias: the issues with technology reflecting history & humans*, in *Journal of Business & Technology Law*, 16(2), 2021, p. 309.

[25] P. Pohlmann, G. Vossen, J. Everding, J. Scheiper, *Künstliche Intelligenz, Bias und Versi-*

Dal punto di vista tecnico, i *bias* possono insorgere: *(i)* nella *modellazione*, quando il *bias* viene introdotto dagli sviluppatori nella strutturazione dei parametri operativi del sistema; *(ii)* nel *training*, quando il *bias* viene generato durante la fase di *machine learning* a causa di una scarsa campionatura dei dati, di un errore umano nel definire le variabili che l'IA dovrà analizzare, o della presenza di *bias* "storici" all'interno dei dati medesimi [26]; *(iii)* nell'*uso*, quando l'IA viene impiegata in compiti per i quali non è stata progettata [27].

Ne consegue che la presenza di *bias* possa eventualmente condurre un sistema alla creazione di un modello distorto della realtà, con il conseguente rischio di errori di giudizio o discriminazioni ingiuste. Va chiarito tuttavia che la presenza di *bias* non determina necessariamente *output* discriminatori, i quali vanno valutati non in assoluto ma in relazione al contesto [28]. Dunque, qualora l'inferenza di determinate caratteristiche di gruppo considerate nel processo di valutazione dei sistemi di IA assicurativi induca un trattamento differenziale (disproporzionale) delle pratiche assicurative, il *focus* del dibattito si sposta sulla prassi della discriminazione assicurativa e precipuamente sia sulla sua razionalità in termini statistici sia sulla sua accettabilità socio-legale.

Ciò che si rileva in questa sede, più che il caso in cui un assicuratore faccia istruire specificamente il suo sistema di IA allo scopo di pregiudicare un dato gruppo, è il fatto che condotte discriminatorie possano essere poste in essere incidentalmente da quest'ultimo anche a fronte di un obiettivo legittimo, coerente con l'attività d'impresa, quale la riduzione del rischio assunto o la massimizzazione della raccolta premi. Infatti, nonostante *prima facie* entrambi tali scopi siano leciti e razionali, possono assurgere a fondamento di decisioni algoritmiche con effetti problematici. Si pensi per esempio, alla penalizzazione di minoranze [29] o anche all'offerta di prodotti

---

cherungen – Eine technische und rechtliche Analyse, in *Zeitschrift für die gesamte Versicherungswissenschaft*, 111, 2022, p. 137.

[26] P. Hacker, *Teaching fairness to artificial intelligence: existing and novel strategies against algorithmic discrimination under EU law*, in *Common Market Law Review*, 55(4), 2018, pp. 1146-1148.

[27] N. Criado, J.M. Such, *Digital discrimination*, in K. Yeung, M. Lodge (a cura di) *Algorithmic Regulation*, Londra, 2019, pp. 85-87.

[28] X. Ferrer, T. Van Nuenen, J.M. Such, M. Coté, N. Criado, *Bias and Discrimination in AI: A Cross-Disciplinary Perspective*, in *IEEE Technology and Society Magazine*, 40(2), 2021, p. 73.

[29] C. Bansal, K.K. Pandey, R. Goel, A. Sharma, S. Jangirala, *Artificial intelligence (AI) bias impacts: classification framework for effective mitigation*, in *Issues in Information Systems*, 24(4), 2023, p. 375.

assicurativi più costosi a soggetti caratterizzati da una maggiore vulnerabilità psicologica o minore sensibilità alle fluttuazioni di prezzo, a prescindere dall'effettivo rischio assicurato (c.d. *price discrimination*) [30].

Tra l'altro, si è osservato che, anche a fronte di dati di *input* privi di *bias* in partenza, i sistemi di IA possono comunque mettere in atto trattamenti differenziali quando le variabili analizzate sono distribuite in modo disomogeneo tra diverse categorie di soggetti. Questo fatto può portare a pregiudicare indirettamente gruppi sulla base di determinate caratteristiche legalmente tutelate, dando così luogo alla c.d. "discriminazione tramite *proxy*"; ossia il trattamento differenziale tra due pratiche in forza dell'inferenza implicita/indiretta di un tratto correlato a un attributo protetto (tipicamente, l'altezza per il sesso biologico e il colore degli occhi o dei capelli per l'etnia) [31].

Sebbene forme di discriminazione tramite *proxy* siano state praticate da agenti umani molto prima dell'avvento dell'IA; quando poste in essere da quest'ultima, esse si caratterizzano per l'irrilevanza della componente volontaria, la configurabilità anche nella realizzazione di obbiettivi legittimi e la tendenziale inevitabilità [32].

Si è su quest'ultimo punto commentato che l'IA, per sua natura, ragiona e decide tramite *proxy*, dove i *proxy* sono tutti quegli elementi da cui stima l'esistenza dello stato desiderato. Pertanto, anche se all'IA fosse impedito l'uso di informazioni riguardanti alcune caratteristiche tramite la loro rimozione dal modello, se tali elementi fossero considerabili statisticamente predittivi al fine del raggiungimento dell'obbiettivo, ciò non impedirebbe al sistema di produrre gli stessi risultati discriminatori. In effetti, soprattutto in reti più complesse (come nel caso del *deep learning*), una proibizione di utilizzare determinate informazioni come dati di *input*, ad esempio il sesso biologico dell'assicurato (statisticamente associato all'aspettativa di vita, all'insorgenza di certe patologie e, anche e soprattutto alla sinistrosità stradale), indurrebbe il sistema a ricostruire il proprio modello decisionale attorno a caratteristiche *proxy* di tali attributi. Inoltre, lo stesso accadrebbe anche nel caso della rimozione dei *proxy* più evidenti (come l'altezza), poiché l'IA ne identificherebbe altri, meno

---

[30] B. Soyer, *Use of big data analytics and sensor technology in consumer insurance context: Legal and practical challenges*, in *The Cambridge Law Journal*, 81(1), 2022, pp. 183-184.

[31] P. Hacker, *op. cit.*, pp. 1148-1149.

[32] Sul tema, v. A.E. Prince, D. Schwarcz, *Proxy discrimination in the age of artificial intelligence and big data*, in *Iowa Law Review*, 105, 2019, pp. 1257 ss.

palesi o leggermente meno precisi, ma comunque funzionalmente equivalenti[33].

In breve, anche sistemi di IA, originariamente non progettati per discriminare contro gruppi protetti, potrebbero inevitabilmente ricreare *bias* umani, rafforzando stereotipi storici o, persino, generare nuove cause di discriminazione[34], con risultati potenzialmente irragionevoli in considerazione del fatto che il processo di generazione dei *proxy* è spesso avulso, come si è visto, da effettiva causalità tra variabile e obbiettivo[35].

4. *Discriminazione algoritmica e premesse della discriminazione statistica in assicurazione*

Le menzionate problematiche tecniche, se non adeguatamente affrontate, potrebbero compromettere le citate condizioni di accettabilità sociale ed efficienza alla base della discriminazione assicurativa.

Con riguardo all'efficienza, la presenza di *bias* o il ragionamento tramite *proxy*, unito ai rischi legati al ragionamento non causale dei sistemi di IA, potrebbe generare *output* inefficienti o, peggio, irrazionali. In particolare, la sottostima o sovrastima della predittività di un elemento potrebbe causare falsi positivi o falsi negativi.

Per quanto riguarda l'accettabilità, è chiaro che l'adozione di criteri non necessariamente causali e avulsi dal contesto potrebbe stridere con la necessità, da parte dell'assicurato, di vedere un collegamento causale tra il fatto o lo stato utilizzato come discriminante e il rischio, soprattutto se tale variabile è fuori dal suo controllo (come nel caso del sesso biologico). Inoltre, tali limitazioni del ragionamento dell'IA renderebbero difficilmente sostenibile la pretesa delle compagnie della necessità di tale differenziazione ai fini del corretto e razionale funzionamento del mercato.

Sempre con riguardo al problema di giustificazione in termini attuariali e funzionali del trattamento differenziale, la natura della "*black box*" implica una incompleta conoscenza del processo interno alla base delle decisioni dell'IA e una certa imprevedibilità dell'*output*. Infatti, seppure compagnie e intermediari possano controllare gli obiettivi posti al sistema (alla base della decisione algoritmica stessa) e, possibilmente, anche i dati di *in-*

[33] A.E. Prince, D. Schwarcz, *op. cit.*, pp. 1263-1264, 1269-1276.

[34] B. Soyer, *op. cit.*, pp. 178-179.

[35] B. McGurk, *op. cit.*, p. 74.

put, potrebbero comunque non essere completamente in grado di fornire al cliente o, ove del caso, all'autorità di vigilanza, una spiegazione causale del percorso decisionale intrapreso e culminato nella decisione.

5. *Discriminazione algoritmica nel diritto giapponese*

Attualmente, in Giappone, non esistono una legislazione *ad hoc* né divieti specifici che affrontino il problema della discriminazione da parte dell'IA nei contratti assicurativi. Inoltre, fino ad oggi, non si registrano precedenti giudiziari su questo tema. Dunque, quantomeno a livello primario, l'indagine sugli strumenti giuridici a disposizione del diritto giapponese deve spostarsi sulla disciplina generale del contratto di assicurazione.

Va premesso che ai sensi del diritto giapponese, il contratto di assicurazione si fonda sul principio della autonomia negoziale; dunque, in assenza di limiti normativi, il suo contenuto è determinato dalla volontà delle parti. Tuttavia, questo primo assunto è contemperato dalla natura regolata del contratto per cui equità e corretto funzionamento del mercato sono garantiti dall'FSA tramite un controllo *ex ante* [36].

5.1. *Parità di trattamento dell'assicurato* ex *articolo 5(1) dell'*Insurance Business Act *del 1995*

Oltre alle norme generali sul contratto di cui al Codice Civile (Legge n. 89 del 1896) [37], la normativa di settore è regolata dall'*Insurance Business Act* (Legge n. 105 del 1995) in materia di impresa assicurativa e dall'*Insurance Act* (Legge n. 56 del 2008) che ne detta la disciplina contrattuale [38].

---

[36] In relazione alla materia assicurativa, la FSA regolamenta e supervisiona gli operatori di mercato, comprese le compagnie di assicurazione, richiedendo loro di presentare domanda per approvazione delle condizioni e dei termini delle polizze assicurative, così come nel caso di modifiche sostanziali a quelli precedentemente approvati. Questo potrebbe portare anche l'assicuratore a ricevere un ordine di modifica del testo del contratto in prospettiva di tutela dei contraenti e degli assicurati. T. Koezuka, *Transparency in the Insurance Contract Law of Japan*, in P. Marano, K. Noussia (a cura di), *Transparency in Insurance Contract Law*, Cham, 2019, pp. 390-392.

[37] N. Kobayashi, Y. Umekawa, T. Mikami, S. Okuda, *Insurance Law in Japan*, Alphen aan den Rijn, 2022, p. 41.

[38] S. Nakaide, *Revision of the Japanese Insurance Business Act in 2014. Insurance Distribution Channels in Japan and New Rules on the Solicitation of Insurance*, in *Zeitschrift für Japanisches Recht*, 22(44), 2017, pp. 24-25.

Muovendo alla questione in esame, va rilevato che da un'analisi testuale delle fonti primarie, al di là del contenuto dell'articolo 13 della Costituzione Giapponese (il quale però tendenzialmente non trova applicazione nei rapporti tra privati) [39], sia il Codice Civile che l'*Insurance Act* non presentano alcun riferimento esplicito al tema della discriminazione assicurativa in relazione alle condizioni contrattuali o al premio. Tuttavia, una prima linea di difesa contro la discriminazione algoritmica e i *proxy* non causali può essere rinvenuta nel dispositivo dell'articolo 5(1) dell'*Insurance Business Act*, che affronta specificamente la questione generale della discriminazione assicurativa.

Segnatamente, tra le condizioni necessarie per l'esercizio dell'attività assicurativa, la legge del 1995 prevede che: *(i)* «*Il contenuto del contratto di assicurazione non present*[i] *il rischio di essere inadeguato a proteggere il contraente, gli assicurati, i beneficiari e le altre persone interessate*» [40]; *(ii)* «*nessuna persona specifica* [sia] *soggetta a trattamenti ingiusti o discriminatori in forza del contenuto del contratto di assicurazione*» [41]; *(iii)* «*le procedure di calcolo dei premi assicurativi e delle riserve* [siano] *ragionevoli, appropriate, e basate sulla scienza attuariale*» [42]; *(iv)* «*nessuna persona specifica* [sia] *soggetta a trattamenti ingiusti o discriminatori in relazione ai premi assicurativi*» [43].

Con riguardo alla prima condizione, l'obbligo dell'assicuratore si articola in due dimensioni: un obbligo negativo, ossia il divieto di creare contratti di assicurazione che presentino significativi *vulnus* nella tutela della posizione assicurata e nella copertura del rischio a cui il contratto è preposto; e un obbligo positivo, da leggersi congiuntamente agli obblighi generali di adeguatezza della copertura offerta tramite la polizza, volto a soddisfare i bisogni e le esigenze di copertura dell'assicurato. Muovendo ora alla norma di cui al punto (iii) lett. b), essa impone all'assicuratore di garantire che i diritti goduti dal titolare della polizza (nonché dai terzi beneficiari), previsti dalle condizioni del contratto, non producano effetti discriminatori. Affinché si possa configurare una discriminazione rilevante ai sensi della

[39] J. Shimizu, *The Historical Origins of the Horizontal Effect Problem in the United States and Japan: How the Reach of Constitutional Rights into the Private Sphere Became a Problem*, in *The American Journal of Comparative Law*, 70(4), 2022, p. 799.

[40] Insurance Business Act (Legge n. 105 del 1995), Articolo 5(1), (iii), lett. a).

[41] Insurance Business Act (Legge n. 105 del 1995), Articolo 5(1), (iii), lett. b).

[42] Insurance Business Act (Legge n. 105 del 1995), Articolo 5(1), (iv), lett. a).

[43] Insurance Business Act (Legge n. 105 del 1995), Articolo 5(1), (iv), lett. b).

norma in esame, si ritiene debba essere individuato un duplice requisito: l'*ingiustizia* del trattamento differenziale e la *specificità* del suo destinatario. Questi fungono da compromesso tra la necessità di garantire che il sistema assicurativo non diventi foriero di inammissibili trattamenti differenziali e i meccanismi funzionali dell'industria, che vedono nella discriminazione statistica uno dei principali strumenti di segmentazione e gestione del rischio[44]. Il dettato delle disposizioni citate in materia di condizioni contrattuali è integrato dalle norme di cui alle lettere a) e b) del punto (iv), che delineano previsioni analoghe in materia di premio. In particolare, esse prevedono rispettivamente il divieto di premi arbitrari (ossia, non basati sulla scienza attuariale) e il divieto di discriminazione nella determinazione del prezzo, anche in questo caso basato sui medesimi requisiti di ingiustizia e specificità del destinatario[45].

Alla luce di quanto precede, è stato commentato che il principio di parità di trattamento possa essere considerato come uno dei pilastri fondamentali alla base del contratto di assicurazione in Giappone[46].

L'ampiezza di questo assunto va però sottoposta ad un significativo *caveat*. In effetti, il dettato dell'articolo 5(1), sebbene requisito essenziale per l'esercizio legittimo dell'attività assicurativa potenzialmente individuabile come strumento *ex ante* per contrastare possibili discriminazioni in sede di sottoscrizione e *pricing* da parte di algoritmi, afferisce ad un principio di parità di trattamento e di oggettività e non anche ad un più generale divieto di discriminazione. Invero, sebbene la norma possa fungere da strumento di tutela contro discriminazioni mirate a un determinato soggetto, qualora la compagnia sia in grado di dimostrare che il trattamento differenziale si basi su dati statistici e sia applicato a un numero indeterminato di sog-

---

[44] Sul tema, v. H. Hosoda [細田 浩史], *Insurance Business Act* [保険業法, *Hoken gyōhō*], Tokyo, 2018, pp. 52-56.

[45] H. Hosoda [細田 浩史], *Insurance Business Act*, cit., p. 57. Alla luce della casistica elaborata dalla dottrina giapponese, possono ritenersi in violazione del principio di parità di trattamento: *(i)* l'applicazione di premi più elevati non proporzionati al rischio effettivamente sottoscritto o non conformi ai criteri della scienza attuariale, applicati unicamente ad alcuni assicurati; nonché *(ii)* la determinazione dei premi sulla base del reddito annuo dell'assicurato, dell'entità del patrimonio e/o di attributi protetti, quali le convinzioni personali, l'appartenenza religiosa o il credo politico. N. Uno [宇野 典明], *On the Principle of Equal Treatment of Policyholders under the Concept of Optimal Asset-Liability Allocation* [資産負債最適配分概念の下における保険契約者平等待遇原則のあり方について, *Shisan-fusai- saiteki-haibun-gainen no shita ni okeru hoken-keiyaku sha-byōdō-taigū-gensoku no arikata ni tsuite*], in *The journal of commerce* [商學論纂, *Shogaku ronsan*], 55, 5-6, 2014, pp. 490–491.

[46] N. Uno [宇野 典明], *op. cit.*, p. 489.

getti, anche qualora la discriminazione statistica si fondi su un attributo legalmente protetto (si pensi alla prassi diffusa in Giappone di distinguere il rischio sottoscritto sulla base del sesso biologico dell'assicurato)[47], esso sarà considerato ammissibile, dunque escluso dal divieto[48].

5.2. *Clausole generali e principio di buona fede* ex *articolo 1 del Codice Civile*

Alla base di tutti i contratti, incluso quello di assicurazione, nonostante la collocazione *extra* codicistica della sua disciplina, troviamo le disposizioni del Codice Civile giapponese[49] e, segnatamente le clausole generali (総則, *Sōsoku*) di cui all'articolo 1, applicabili trasversalmente a tutti i rapporti civili e commerciali[50].

Questi principi, ovvero l'interesse pubblico (公共の福祉, *Kōkyō no fukushi*), la buona fede (信義誠実の原則, *Shingi seijitsu no gensoku*) e il divieto di abuso del diritto (権利の濫用, *Kenri no ran'yō*)[51], diventano residuali strumenti nelle mani delle corti per orientare gli esiti delle proprie decisioni e per interpretare le norme positive, così da allineare la lettera del diritto alle esigenze della società[52], specie in considerazione dei repentini cambiamenti derivanti da trasformazioni economiche, sociali e tecnologiche[53].

---

[47] T. Miyachi [宮地 朋果], *Fairness and Reasonableness in Underwriting Process* [保険における危険選択と公平性, *Hoken ni okeru kiken sentaku to kōhei-sei*], in *Journal of Insurance Science* [保険学雑誌, *Hokengakuzasshi*], 614, 2011, p. 51. Nonostante il tenore letterale dell'articolo 2 del Codice Civile, che in quanto parte delle disposizioni generali rappresenta un caposaldo dell'intero impianto privatistico giapponese, sancisca il principio di tutela della dignità degli individui e dell'uguaglianza tra i sessi, la discriminazione statistica tra uomini e donne in ambito assicurativo continua a persistere. Essa tende di fatto a prevalere su tale disposizione generale, in ragione sia dell'indeterminatezza dei suoi destinatari, sia del fondamento statistico che ne giustifica l'applicazione.

[48] H. Hosoda [細田 浩史], *Insurance Business Act*, cit., p. 53.

[49] L'applicabilità in materia *extra* codicistica, come nell'assicurazione, del principio di buona fede non è frutto solo di considerazioni dottrinali, bensì è pacificamente riconosciuto dalla giurisprudenza stessa. J. Ōgushi [大串 淳子], *Commentary to the Insurance Act* [解説 保険法, *Kaisetsu Hoken-hō*], Tokyo, 2008, pp. 8-9.

[50] H. Oda, *Japanese Law*, Oxford, 2021, p. 124.

[51] Codice Civile (Legge n. 89 del 1896), Articolo 1(1)-(3).

[52] H. Tanaka, M.D.H. Smith, *The Japanese Legal System: Introductory Cases and Materials*, Tokyo, 1982, p. 114.

[53] G. Ajani, A. Serafino, M. Timoteo, *Diritto dell'asia orientale*, Torino, 2007, p. 363.

Invero, anche in mancanza di specifiche norme, tali principi hanno storicamente assunto un ruolo di volano di innovazione giuridica di cruciale importanza nel contesto giapponese (caratterizzato da un certo attivismo giudiziario)[54]; ove interi settori del diritto, successivamente positivizzati, sono stati introdotti e lungamente retti proprio per tramite dell'applicazione di essi da parte delle corti[55].

Di particolare importanza in materia di assicurazione, il principio di buona fede di cui all'articolo 1(2), trova tradizionalmente quattro funzioni all'interno del diritto giapponese, ossia: *(i) concretizzazione* legale; *(ii) equitativa*; *(iii) correttiva*; e *(iv) creativa* (eventualmente anche *contra legem*)[56].

Alla luce di quanto sopra, sebbene nessuna corte si sia ancora pronunciata sul punto, si può ritenere che, in conformità con l'esperienza storica e il *modus operandi* delle corti giapponesi, i principi generali del Codice Civile possano fungere da base giuridica flessibile e adattabile al caso concreto anche in relazione alla gestione dei rischi legati alla discriminazione algoritmica assicurativa, in termini di condizioni contrattuali o *pricing*, almeno nei casi in cui gli *output* dei sistemi di IA possano tradursi in condotte abusive o eccessivamente pregiudizievoli per il soggetto sottoposto al trattamento differenziale.

Detto ciò, va tuttavia rilevato un problema non trascurabile legato ad un eccessivo affidamento a questi strumenti. Infatti, la natura *ex post* del rimedio giudiziale potrebbe intervenire solo in un ruolo meramente rime-

---

[54] Sul tema, v. D.H. Foote, *Judicial Creation of Norms in Japanese Labor Law: activism in the service of stability*, in *UCLA Law Review*, 43(3), 1996, pp. 635-710; H. Foote, *Resolution of Traffic Accident Disputes and Judicial Activism in Japan*, in *Law in Japan*, 25, 1995, pp. 19-39; S. Kozuka, *Judicial Activism of the Japanese Supreme Court in Consumer Law: Juridification of Society through Case Law?*, in *Zeitschrift für Japanisches Recht*, 14(27), 2009, pp. 81-90.

[55] Esempi significativi dell'introduzione, attraverso il diritto di matrice giurisprudenziale [裁判官法, *Saibankan Hō*] (C. Förster, *Shingi seijitsu no gensoku – Das Prinzip von Treu und Glauben im japanischen Schuldrecht*, in *The Rabel Journal of Comparative and International Private Law*, 7(1), 2009, pp. 93-94), di istituti giuridici fondati su principi generali, quali il principio di buona fede, si rinvengono in Giappone nella responsabilità precontrattuale (*culpa in contrahendo*), nei doveri di trasparenza, collaborazione e informazione in ambito contrattuale, negli obblighi di prevenzione e mitigazione del danno, nella disciplina del mutamento delle circostanze e, non da ultimo, nella tutela della parte contrattuale debole, sia nel diritto dei consumatori (ai sensi del *Consumer Contract Act* del 2000), sia nell'ambito dei contratti standardizzati (di cui alla riforma del Codice Civile del 2017). Sul tema, v. T. Yoshimasa [吉政 知広], *The Principle of Good Faith* [信義誠実の原則, *Shingi seijitsu no gensoku*], in A. Yamanome [山野目章夫] (a cura di), *New Commentary on the Civil Code of Japan Vol. 1* [新注釈民法 (1), *Shin chūshaku minpō (1)*], Tokyo, 2018, pp. 131 ss.

[56] T. Yoshimasa [吉政 知広], *op. cit.*, pp. 138-139.

diale, non garantendo adeguatamente *compliance by design*.

Inoltre, sebbene la violazione degli obblighi desumibili dai principi generali possa dar luogo a responsabilità, la parte attrice, ossia l'assicurato privo di accesso al sistema di IA e alle informazioni sul suo funzionamento, sarebbe comunque sottoposta all'onere della prova, ai sensi del principio della domanda, con tutte le problematiche pratiche che ciò comporterebbe [57].

5.3. *Discriminazione algoritmica negli strumenti di* Soft Law

Invece che focalizzarsi sulla creazione di una disciplina positiva e vincolante in materia di discriminazione algoritmica, nel tentativo di favorire un sistema di autoregolamentazione del settore privato, verticalmente orientata ("*state led*") [58], il governo giapponese ha avviato, da alcuni anni, un'importante attività di produzione di strumenti di *Soft law*, segnatamente linee guida non vincolanti, finalizzate a indirizzare il mercato in modo flessibile e adeguato alle specificità delle singole realtà aziendali, verso obiettivi di pubblico interesse. Dall'analisi di tali documenti emerge che il tema della discriminazione algoritmica e delle sue gravi implicazioni economiche e sociali, seppure in termini generali e non squisitamente assicurativi, sia stato preso in considerazione in diverse occasioni.

In effetti, già all'interno dei *Social Principles of Human-Centric AI*, documento pubblicato nel 2019 alla base dell'approccio regolamentare giapponese e fondamento delle linee guida successivamente emanate, emerge chiaramente la necessità di garantire dignità umana, inclusività, *diversity* e sostenibilità nello sviluppo e nell'implementazione dei sistemi di IA in Giappone [59]. Ivi, la discriminazione algoritmica viene affrontata in molte-

[57] Anche con riferimento all'allocazione dell'onere della prova, tuttavia, l'esperienza storica della prassi giudiziaria giapponese evidenzia una certa flessibilità da parte delle corti nell'alterare, ove opportuno, gli obblighi probatori gravanti sulle parti. In particolare, in presenza di una significativa asimmetria informativa o di potere negoziale, qualora l'attore riesca a dimostrare l'esistenza di determinati elementi rilevanti ai fini della domanda giudiziale, le corti hanno frequentemente presunto la responsabilità del convenuto, trasferendo su quest'ultimo l'onere della prova contraria. H. Oda, *op. cit.*, p. 195. Non è dunque da escludere che, anche nel caso in esame, l'asimmetria informativa e le difficoltà probatorie gravanti sulla parte attrice possano condurre a deviazioni rispetto al modello *standard* del principio della domanda, sebbene al momento non si registrino precedenti in tal senso.

[58] S. Kozuka, *Self-regulation Induced by the State in Japan*, in H. Baum, M. Bälz, M. Dernaue (a cura di), *Self-regulation in Private Law in Japan and Germany*, Colonia, 2018, p. 109.

[59] Cabinet Secretariat, *Social Principles of Human-Centric AI*, 2019, p. 4. Per un confronto con le *High-Level Ethics Guidelines for Trustworthy Artificial Intelligence* pubblicate nel 2018

plici sezioni, in particolare nella parte generale dedicata alla *Society 5.0*, dove si sottolinea l'importanza di riconoscere i *bias* (distinti in statistici, causati da condizioni sociali, e volontariamente introdotti dagli operatori), nonché le loro possibili conseguenze pregiudizievoli[60].

Con specifico riferimento al principio di *Fairness, Accountability and Transparency* poi, si stabilisce che nella progettazione dei sistemi di IA «*le persone devono essere trattate equamente, senza discriminazioni ingiustificate basate su fattori quali razza, sesso, nazionalità, età, convinzioni politiche, religione, ecc. …*». Spiegazioni appropriate devono poi essere fornite caso per caso, a seconda dell'applicazione dell'IA e delle circostanze concrete; ciò include anche le informazioni riguardanti l'uso, la provenienza e l'utilizzo dei dati e le misure adottate per garantire l'adeguatezza dei risultati ottenuti[61].

Inoltre, e di particolare rilievo, il 19 aprile 2024, il Ministero degli Affari Interni e delle Comunicazioni (c.d. "MIC") e il Ministero dell'Economia, Commercio e Industria (c.d. "METI"), integrando tre precedenti serie di linee guida[62], hanno pubblicato le *AI Guidelines for Business Ver 1.0 of 2024* al fine di delineare le pratiche che le imprese impegnate nell'ambito dell'IA sono invitate a rispettare.

Il documento svolge *in primis* un importante tripartizione soggettiva dei propri destinatari distinguendo tra: *(i) AI Developer*[63]; *(ii) AI Provider*[64]; e *(iii) AI Business User*[65], attribuendo a ciascuno specifiche funzioni.

---

dall'Unione Europea, v. S. Kozuka, *A Governance Framework for the Development an Use of Artificial Intelligence: Lessons from the Comparison of Japanese and European Initiatives*, in *Uniform Law Review*, 24(2), 2019, pp. 321-324.

[60] Cabinet Secretariat, *Social Principles of Human-Centric AI*, 2019, p. 5.

[61] Cabinet Secretariat, *Social Principles of Human-Centric AI*, 2019, p. 10.

[62] MIC, *AI R&D Guidelines*, 2017; MIC, *AI Utilization Guidelines*, 2019; METI, *Governance Guidelines for Implementation of AI Principles Ver. 1.1*, 2022.

[63] «*Operatori economici che sviluppano sistemi di IA (inclusi operatori economici che svolgono attività di ricerca sull'IA). Essi sviluppano modelli di IA e algoritmi, contribuendo alla costruzione di sistemi di IA, inclusi modelli di IA, sistemi di base e funzioni di* input/output*, attraverso attività quali la raccolta di dati (anche tramite acquisto), la pre-elaborazione dei dati e l'addestramento mediante dati*». METI, *AI Guidelines for Business Ver 1.0*, 2024, pp. 4-5.

[64] «*Operatori economici che integrano sistemi di IA in applicazioni, prodotti o sistemi esistenti, processi aziendali, ecc., fornendoli agli* AI Business User *e, in alcuni casi, anche a utenti non aziendali. Essi verificano i sistemi di IA, li integrano con altri sistemi, forniscono sistemi e servizi di IA, offrono supporto operativo agli* AI Business User *per il funzionamento ordinario dei sistemi o gestiscono direttamente l'operazione del servizio di IA*». METI, *AI Guidelines for Business Ver 1.0*, 2024, p. 5.

[65] «*Operatori economici che utilizzano sistemi o servizi di IA nelle proprie attività aziendali. Il loro ruolo consiste nell'utilizzare i sistemi o servizi di IA in modo appropriato, secondo le finalità*

Con riguardo al rischio di discriminazione algoritmica e ai *bias*, le linee guida individuano una serie di doveri in capo agli anzidetti soggetti lungo l'intera catena del valore del sistema di IA in conformità con il relativo ruolo e anche in considerazione della non sempre possibile eliminabilità dal modello delle surriferite problematiche.

Segnatamente, *AI Developer* dovranno valutare il rischio di *bias* nei dati di *input*, nella fase di *training* e *machine learning*, negli algoritmi e in ogni componente tecnica del sistema di IA, adottare le pertinenti misure preventive, nonché riportare ciò in sede di informazione agli *AI Providers* [66].

Questi ultimi, dal canto loro, nell'implementazione del sistema di IA, saranno tenuti a garantire la correttezza e l'equità nell'uso dei dati, a esaminare i *bias* eventualmente contenuti nelle informazioni di riferimento e a collaborare, ove necessario, con soggetti terzi. Inoltre, è richiesto loro di effettuare monitoraggi sugli *input* e *output*, sul razionale alla base delle decisioni algoritmiche prodotte, nonché di sollecitare una nuova valutazione del rischio di *bias* da parte degli *AI Developer* qualora quest'ultimo sia inerente alle componenti tecniche alla base dei modelli di IA. Ancora, nell'ambito del loro ruolo, gli *AI Provider* dovranno esaminare il rischio che i *bias* del sistema di IA possano determinare arbitrarie restrizioni nell'ambito dei processi e delle decisioni e le conseguenze da questo derivanti per gli utenti, sia professionali che non [67].

Infine, gli *AI Business User*, nell'utilizzo dei sistemi di IA o dei relativi servizi, sono incaricati di valutare il rischio di *bias* nei dati eventualmente forniti nell'ambito della loro attività aziendale, nonché nei *prompt* inseriti. Inoltre, saranno ritenuti responsabili della decisione finale di utilizzare o meno un determinato *output*. Non secondariamente, questi soggetti dovranno anche adempiere a oneri di informativa e spiegazione agli *stakeholder* coinvolti [68].

In chiosa, e con precipuo riferimento al settore assicurativo, un ulteriore e primario strumento di indirizzo e guida del mercato verso un uso re-

---

previste dal AI Provider*, condividere con quest'ultimo informazioni quali i cambiamenti ambientali, garantire il funzionamento regolare e, se necessario, gestire i sistemi di IA forniti. Inoltre, qualora l'uso dell'IA possa in qualche modo influire su utenti non aziendali, gli* AI Business User *sono anche responsabili di adottare misure per prevenire eventuali pregiudizi imprevisti per tali utenti e massimizzare i benefici derivanti dall'IA*». METI, *AI Guidelines for Business Ver 1.0*, 2024, p. 5.

[66] METI, *AI Guidelines for Business Ver 1.0*, 2024, pp. 27-29.

[67] METI, *AI Guidelines for Business Ver 1.0*, 2024, pp. 32-34.

[68] METI, *AI Guidelines for Business Ver 1.0*, 2024, pp. 35-36.

sponsabile e conforme ai principi sopra menzionati può essere individuato nell'istituzione, all'interno della FSA, di un *Fintech Support Desk*, incaricato di fornire assistenza tecnica, e soprattutto regolamentare, agli operatori dei settori bancario, finanziario e assicurativo nell'integrazione di nuove tecnologie alle loro funzioni aziendali e modelli di *business*[69].

6. *Conclusioni*

In sintesi, la disciplina sull'assicurazione giapponese si caratterizza per una forte ispirazione liberale, incentrata sulla libertà contrattuale e su un'accettazione complessiva della discriminazione statistica. Al contempo, il generale approccio alla *governance* dell'IA si radica nella convinzione della sufficienza e appropriatezza di strumenti non vincolanti e nella prevalenza, quantomeno al momento, di esigenze di crescita e innovazione su preoccupazioni riguardo alle problematiche introdotte dai sistemi di IA.

Guardando al tema della discriminazione algoritmica in ambito assicurativo, il Giappone, nelle more di una disciplina *ad hoc*, sembra presentare un duplice livello di tutela legale: *(i) ex ante*, basata sulla forza persuasiva delle linee guida ministeriali e, in ambito assicurativo anche, sull'assistenza e vigilanza dell'FSA anche in forza delle disposizioni dell'articolo 5 dell'*Insurance Business Act* ove applicabile; *(ii) ex post*, fondato sulla forza espansiva dei principi generali e di un potere giudiziario storicamente noto per il suo attivismo e la sua funzione *de facto* creativa del diritto.

Infine, e a parziale sconvolgimento del panorama regolamentare finora delineato, merita attenzione la recente presentazione, del 28 febbraio 2025, innanzi alla Dieta Nazionale giapponese di un disegno di legge in materia di IA. Questo, c.d. "*Bill Concerning the Promotion of Research, Development, and Utilization of Artificial Intelligence-Related Technologies*", seppur oggettivamente limitato ai sistemi di IA più avanzati, sembra prevedere la positivizzazione di un dovere di collaborazione e conformità alle politiche in materia di IA, l'adozione del principio di *Fairness* (legato a doppio filo con il tema della discriminazione) quale fondamento della disciplina giapponese in fieri, e la creazione di una base giuridica primaria

[69] T. Inoue [井上 俊剛], *op. cit.*, p. 12.

per futuri interventi normativi e regolamentari[70]. Ne segue che, qualora l'iniziativa dovesse ricevere l'avallo parlamentare, il Giappone rivedrebbe in parte il tradizionale approccio basato sulla *Soft law*, avvicinandosi a un modello ibrido, in cui il richiamo nella normativa primaria contribuirebbe a rafforzare e ad attribuire una quasi vincolatività a strumenti tradizionalmente privi di forza coattiva.

[70] Sul tema, si consenta il rinvio a D.L. TOTARO, *AI Regulation "made in Japan": A first reading of the new Japanese Bill on Artificial Intelligence*, in *Global AI Governance and Regulation Insights*, 2025, *online*.

Finito di stampare nel mese di maggio 2025
nella Stampatre s.r.l. di Torino
Via Bologna, 220

UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI GIURISPRUDENZA

PUBBLICAZIONI DEL DIPARTIMENTO DI SCIENZE GIURIDICHE "CESARE BECCARIA"

Serie di diritto ecclesiastico e canonico

Per i tipi di Giuffrè

1. VITALI E.G., *Profili* dell'impedimentum criminis (1979), 8°, p. VIII-356.
2. ALBISETTI A., *Contributo allo studio del matrimonio putativo in diritto canonico* (1980), p. XIV-306.
3. ALBISETTI A., *Giurisprudenza costituzionale e diritto ecclesiastico* (1983), 8°, p. 132.
4. JASONNI M., *Contributo allo studio della «ignorantia juris» nel diritto penale canonico* (1983), 8°, p. IV-192.
5. VITALI E., CASUSCELLI G. (a cura di), *La disciplina del matrimonio concordatario dopo gli Accordi di Villa Madama*, atti del convegno (Milano-Bergamo 10-12 ottobre 1985) (1988), 8°, p. X-436.
6. CONSORZIO EUROPEO DI RICERCA SUI RAPPORTI TRA STATI E CONFESSIONI RELIGIOSE, *L'obiezione di Coscienza nei paesi della comunità europea*, atti dell'incontro (Bruxelles-Lovanio 7-8 dicembre 1990) (1992), 8°, p. VI-306.
7. CONSORZIO EUROPEO DI RICERCA SUI RAPPORTI TRA STATI E CONFESSIONI RELIGIOSE, *Stati e Confessioni religiose in Europa, modelli di finanziamento pubblico, scuola e fattore religioso*, Atti dell'incontro (Milano-Parma 20-21 ottobre 1989) (1992), 8°, p. VIII-214.
8. EUROPEAN CONSORTIUM FOR CHURCH-STATE RESEARCH, *Marriage and religion in Europe*, Proceedings of the meeting (Augsburg 28-29 november 1991) (1993), 8°, p. VI-252.
9. EUROPEAN CONSORTIUM FOR CHURCH-STATE RESEARCH, *Churches and labour law in the EC countries*, proceedings of the meeting (Madrid 27-28 november 1992) (1993), 8°, p. 282.
10. EUROPEAN CONSORTIUM FOR CHURCH-STATE RESEARCH, *The legal status of the religious minorities in the countries of the european union*, proceedings of the meeting (Thessaloniki 10-20 november 1993) (1994), 8°, p. 380.
11. CONSORTIUM EUROPEEN: RAPPORTS RELIGIONS-ETAT, *Le statut constitutionnel des cultes dans les pays de l'union européenne*, Actes du colloque (Université de Paris XI, 18-19 novembre 1994) (1995), 8°, p. 234.
12. BARDI M., *Il dolo nel matrimonio canonico* (1996), 8°, p. VIII-274.
13. EUROPEAN CONSORTIUM FOR CHURCH-STATE RESEARCH, *Religions in european union law*, proceedings of the colloquium (Luxembourg/Trier 21-22 november 1996) (1998), 8°, p. VI-196.
14. EUROPEAN CONSORTIUM FOR CHURCH-STATE RESEARCH, *"New liberties" and church and state relationships in Europe*, proceedings of the meeting (Tilburg 17-18 november 1995) (1998), 8°, p. VIII-470.
15. EUROPEAN CONSORTIUM FOR CHURCH-STATE RESEARCH, *New religious movements and the law in the European union*, proceedings of the meeting (Lisbon, Universidade Moderna 8-9 november 1997) (1999), 8°, p. VIII-392.
16. CONSORZIO EUROPEO DI RICERCA SUI RAPPORTI TRA STATI E CONFESSIONI RELIGIOSE, *Cittadini e fedeli nei paesi dell'unione europea, una doppia appartenenza alla prova della secolarizzazione e della Mondializzazione*, atti del colloquio (Università per stranieri, Reggio Calabria 12-15 novembre 1998) (1999), 8°, p. VI-428.

17. Consortium europeen pour l'etude des relations eglises-etat, *Le statut des confessions religieuses des états candidats à l'union européenne, sous la direction de Francis Messner*, actes du Colloque (Université Robert Schuman - CNRS, Strasbourg, 17-18 novembre 2000) (2002), 8°, p. VIII-276.
18. European consortium for church-state research, *Social welfare, religious organizations and the State*, editors Inger Dübeck and Frands Ole Overgaard, proceedings of the meeting (Sandjerg, 18-20 November 1999) (2003), 8°, p. VIII-232.
19. Pacillo V., *Contributo allo studio del diritto di libertà religiosa nel rapporto di lavoro subordinato* (2003), 8°, p. XII-380.
20. Dieni E., Ferrari A., Pacillo V. (a cura di), *I simboli religiosi tra diritto e culture* (2006), 8°, p. XII-402.
21. Pasquali Cerioli J., *L'indipendenza dello stato e delle confessioni religiose, contributo allo studio del principio di distinzione degli ordini nell'ordinamento italiano* (2006), 8°, p. VIII-206.
22. Finocchiaro F., *Saggi (1973-1978)*, a cura di Alessandro Albisetti (2008), 8°, p. XVI-730.
23. Vitali E., *Scritti di diritto ecclesiastico e canonico* (2012), 8°, p. XXII-524.

Serie diritto penale

1. Maiani G., *Fondamento e valore dell'esimente prevista dall'art. 598 c.p.* (1970), p. XVIII-120.

Serie diritto processuale penale

1. Amodio E., *La motivazione della sentenza penale e il suo controllo in Cassazione* (1967), p. 221.
2. Dominioni O., *Improcedibilità e proscioglimento nel sistema processuale penale* (1974), p. VIII-376.
3. Amodio E., *Le cautele patrimoniali nel processo penale* (1971), p. VIII-312.
4. Dominioni O., *La testimonianza della parte civile* (1974), p. IV-152.

Serie unificata di diritto penale e diritto processuale penale

6. Ubertis G., *Fatto e valore nel sistema probatorio penale* (1979), p. IV-152.
7. Palmieri R., *Il delitto di avvelenamento di acque (art. 439 c.p.)* (1979), p. IV-170.
8. Presutti A., *La declaratoria delle nullità nel regime delle impugnazioni penali* (1982), p. IV-284.
9. Prosdocimi S., *Profili penali del postfatto* (1982), p. XII-344.
10. G. Ubertis, *Dibattimento senza imputato e tutela del diritto di difesa* (1984), p. IV-256.
11[1]. *Studi in memoria di Giacomo Delitala* (1984), I, p. IV-724.
11[2]. *Studi in memoria di Giacomo Delitala* (1984), II, p. VIII-725-1446.
11[3]. *Studi in memoria di Giacomo Delitala* (1984), III, p. VIII-1447-2166.
12. Presutti A., *Profili premiali dell'ordinamento penitenziario* (1986), p. IV-164.
13. Comucci P., *Nuovi profili del trattamento penitenziario* (1988), p. IV-164.
14. Barone G., *Enti collettivi e processo penale* (1989), p. VI-258.
15. *Studi in memoria di Pietro Nuvolone* (1991), I, p. VIII-792.
16. *Studi in memoria di Pietro Nuvolone* (1991), II, p. VI-706.
17. *Studi in memoria di Pietro Nuvolone* (1991), III, p. VI-698.
18. De Maglie C., *L'agente provocatore* (1991), p. XVI-456.
19. Peroni F., *Le misure interdittive nel sistema delle cautele penali* (1992), p. VI-270.
20. Bernasconi A., *La collaborazione processuale* (1995), p. VI-378.
21. Ruggieri F., *La giurisdizione di garanzia nelle indagini preliminari* (1996), p. VIII-326.
22. Catalano E.M., *La prova d'alibi* (1998), p. VIII-166.
23. Vigoni D., *L'applicazione della pena su richiesta delle parti* (2000), p. XIV-624.
24. *Studi in ricordo di Giandomenico Pisapia* (2000), I - Diritto penale, p. XII-942.

25. *Studi in ricordo di Giandomenico Pisapia* (2000), II - Procedura penale, p. VI-780.
26. *Studi in ricordo di Giandomenico Pisapia* (2000), III - Criminologia, p. VI-1056.
27. Catalano E.M., *L'accordo sui motivi di appello* (2001), p. VI-182.
28. Capitta A.M., *Ricognizioni e individuazioni di persone nel diritto delle prove penali* (2001), p. X-344.
29. Ruga Riva C., *Il premio per la collaborazione processuale* (2002), p. XVI-616.
30. Bonetti M., *Riservatezza e processo penale* (2003), p. XIV-358.
31. Pedrazzi C., *Diritto penale* (2003), I - scritti di parte generale, p. X-576.
32. Pedrazzi C., *Diritto penale* (2003), II - scritti di parte speciale. Reati contro il patrimonio - delitti contro la pubblica amministrazione - varia, p. VI-520.
33. Pedrazzi C., *Diritto penale* (2003), III - scritti di diritto penale dell'economia. Problemi generali - diritto penale societario, p. VI-848.
34. Pedrazzi C., *Diritto penale* (2003), IV - scritti di diritto penale dell'economia. Disciplina penale dei mercati - diritto penale bancario - diritto penale industriale - diritto penale fallimentare - varia, p. VI-1096.
35. Basile F., *La colpa in attività illecita. Un'indagine di diritto comparato sul superamento della responsabilità oggettiva* (2005), p. XX-928.
36. Lupária L., *La confessione dell'imputato nel sistema processuale penale* (2006), p. VIII-222.
37. *Studi in onore di Giorgio Marinucci*, tre volumi, a cura di E. Dolcini e C.E. Paliero (2006), p. XVI-2996.
38. C. Sotis, *Il diritto senza codice. Uno studio sul sistema penale europeo vigente* (2007), p. XX-360.
39. Vassalli G., Ferrari V., Dolcini E., Fiore C., Padoa Schioppa A., Paliero C.E., *Presentazione degli studi in onore di Giorgio Marinucci* (2007), p. VI-48.
40. Gatta G.L., *Abolitio criminis e successione di norme "integratrici": teoria e prassi* (2008), p. XXVI-976.
41. Basile F., *Il delitto di abbandono di persone minori o incapaci (art. 591 c.p.). Teoria e Prassi* (2008), p. X-180.
42. Benussi C., *Infedeltà patrimoniale e gruppi di società* (2009), p. XVI-454.
43. Vigoni D., *Relatività del giudicato ed esecuzione della pena detentiva* (2009), p. X-336.
44. Bontempelli M., *L'accertamento amministrativo nel sistema processuale penale* (2009), p. X-354.
45. Basile F., *Immigrazione e reati culturalmente motivati. Il diritto penale nelle società multiculturali* (2010), p. XVI-496.
46. Capitta A.M., *La declaratoria immediata delle cause di non punibilità* (2010), p. XIV-274.
47. Miucci C., *La testimonianza tecnica nel processo penale* (2011), p. X-198.
48. Frioni I., *L'esame dell'imputato* (2011), p. X-214.
49. Vigoni D., *La metamorfosi della pena nella dinamica dell'ordinamento* (2011), p. X-370.
50. G. Marinucci, *La colpa. Studi* (2013), p. XXX-480.
51. *Europa e diritto penale*, a cura di C.E. Paliero e F. Viganò (2013), p. XIV-316.

Nuova serie (dal 2015)

1. Albisetti A., *Dieci saggi* (2015), p. VI-148.
2. Casuscelli G., *Scritti giovanili*, a cura di N. Marchei e J. Pasquali Cerioli (2015), p. XXII-290.
3. Pisani M., *Cesare Beccaria*. Studi (2015), p. X-154.
4. Ragucci G. (a cura di), *Il contributo di Enrico Allorio allo studio del diritto tributario*, Atti del convegno tenutosi presso l'università degli studi di Milano 12 giugno 2015 (2015), p. XVIII-216.
5. Catalano E.M., *Ragionevole dubbio e logica della decisione. Alle radici del giusnaturalismo processuale* (2016), p. VIII-208.
6. Della Bella A., *Il "carcere duro" tra esigenze di prevenzione e tutela dei diritti fondamentali. Presente e futuro del regime deten-*

tivo speciale ex art. 41 bis o.p. (2016), p. XVIII-460.

7. Bontempelli M., *La litispendenza penale* (2017), p. X-322.
8. Ragucci G. (a cura di), *Ezio Vanoni - giurista ed economista*, atti del convegno tenutosi presso l'Università degli Studi di Milano 16 giugno 2016 (2017), p. XIV-262.
9. Lugli M., Toscano M. (a cura di), *Il matrimonio tra diritto ecclesiastico e diritto canonico* (2018), p. VIII-264.
10. Zirulia S., *Esposizione a sostanze tossiche e responsabilità penale* (2018), p. XVIII-484.
11. Bianchetti R., *La paura del crimine. Un'indagine criminologica in tema di mass media e politica criminale ai tempi dell'insicurezza* (2018), p. XXVIII-688.
12. Paliero C.E., Viganò F., Basile F., Gatta G.L. (a cura di), *La pena, ancora: fra attualità e tradizione. Studi in onore di Emilio Dolcini* (2018), tomo I, p. XII-498 - tomo II, p. 499-1200.
13. Tira A., *Alle origini del diritto ecclesiastico italiano. Prolusioni e manuali tra istanze politiche e tecnica giuridica (1870-1915)* (2018), p. XX-394.
14. Bellucci L., *La sindrome ungherese in Europa. Media, diritto e democrazia in un'analisi di Law and politics* (2018), p. XII-188.
15. Ragucci G., Albertini F.V. (a cura di), *Costituzione, legge, tributi. Scritti in Onore di Gianfranco Gaffuri* (2018), p. X-646.
16. Chiaraviglio P., *Il favoreggiamento del creditore nel diritto penale concorsuale* (2020), p. XII-616.
17. Mazzola R., *Componere. Offesa e riconciliazione nell'ordinamento vendicatorio* (2020), p. XII-170.
18. Galluccio A., *Punire la parola pericolosa? Pubblica istigazione, discorso d'odio e libertà di espressione nell'era di internet* (2020), p. X-442.
19. Ubiali M.C., *Attività politica e corruzione. Sull'opportunità di uno statuto penale differenziato* (2020), p. XII-404.
20. Dolcini E., Della Bella A. (a cura di), *Le misure sospensivo-probatorie. Itinerari verso una riforma* (2020), p. XII-364.
21. Febbajo A., Paliero C.E., Fittipaldi E., Mazzola R., *L'eredità di Theodor geiger per le scienze giuridiche* (2020), p. VIII-424.
22. Della Bella A., Zorzetto S. (a cura di), *Whistleblowing e prevenzione dell'illegalità*, Atti del I convegno annuale del dipartimento di scienze giuridiche "Cesare Beccaria" Milano, 18-19 novembre 2019 (2020), p. XII-598.
23. Mariani E., *Prevenire è meglio che punire. Le misure di prevenzione personali tra accertamento della pericolosità e bilanciamenti di interessi* (2021), p. XVI-532.
24. Poli P.F., *La colpa grave. I gradi della colpa tra esigenze di extrema ratio ed effettività della tutela penale* (2021), p. XII-514.
25. Biasi M., Ferraro F., Grieco D., Zirulia S. (a cura di), *L'emergenza Covid nel quadro giuridico, economico e sociale*. Atti del II convegno annuale del dipartimento di Scienze giuridiche "Cesare Beccaria", 15-18 marzo 2021 (2021), p. XII-320.
26. Finocchiaro S., *Confisca di prevenzione e civil forfeiture. Alla ricerca di un modello sostenibile di confisca senza condanna* (2022), p. XXII-474.
27. Ragucci G., *La legge generale tedesca del processo tributario. Finanzgerichtsordnung (FGO), testo e studi* (2022), p. XII-186.
28. Marino G., *Corporate tax governance, il rischio fiscale nei modelli di gestione d'impresa* (2022), p. X-438.
29. Zuffada E., *Homo OEconomicus periculosus. Le misure di prevenzione come strumento di contrasto della criminalità economica. Uno studio della prassi milanese* (2022), p. XII-166.
30. Biasi M., *Studio sulla polifunzionalità del risarcimento del danno nel diritto del lavoro: compensazione, sanzione, deterrenza* (2022), p. XVI-208.
31. Piergallini C., Mannozzi G., Sotis C., Perini C., Scoletta M., Consulich F. (a cura di), *Studi in onore di Carlo Enrico Paliero* (2022).

32. Mancini L., Milani D. (a cura di), *Pluralismo religioso e localismo dei diritti* (2022), p. VIII-286.
33. Ruffini R., Ingaggiati M., *Evoluzione dei concorsi pubblici in Italia: la valorizzazione delle competenze* (2023), p. XVIII-208.
34. Albanese D., *Cosa giudicata e confisca di prevenzione* (2024), p. 400.

Per i tipi di Giappichelli

35. Ragucci G. (a cura di), *Fisco, responsabilità, sanzioni. Una prospettiva multidisciplinare: accelerazione o* disruption*?* (2024), p. XVI-224.
36. Cianitto C., *Minoranze e simboli religiosi. I sikh tra identità e cittadinanza* (2024), p. XII-212.
37. Luzzati C., *Il cristallo della laicità. Contro la teologia politica* (2024), p. VI-202.
38. Ferraro F. (a cura di), *La cultura del diritto. Scritti per Claudio Luzzati (con una sua replica)* (2025), p. XIV-322.
39. Basile F., Biasi M., Camaldo L., Caneschi G., Fragasso B., Milani D. (a cura di), *Intelligenza artificiale. Diritto, giustizia, economia ed etica* (2025), p. XVI-296.
40. Ragucci G. (a cura di), *Filtri e accertamento del fatto nel processo tributario. Studi di diritto comparato ed europeo* (2025), p. XVIII-206.

Serie: Corso di dottorato in Scienze giuridiche "Cesare Beccaria"

1. Ferraro F. (a cura di), *Diritto simbolico, simboli nel diritto* (2022), p. 120.
2. Bonomelli S., *L'editing genetico germinale umano, tra problemi etici e questioni di governance* (2023), p. X-517.

Per i tipi di Giappichelli

3. Vigoni D. (a cura di), *Il sistema multipolare dei procedimenti speciali in materia penale: l'evoluzione e le criticità* (2024), p. X-230.