Filippo Vanni – Gennaro Corrado – Daniela Gregorio –
Marialuisa Simona Gregalli – Ibrahim Nasser Al Mandoos

L'INTELLIGENZA ARTIFICIALE AL SERVIZIO DELLA SICUREZZA E LA GESTIONE DEL RISCHIO NEI SISTEMI A SUPPORTO DELLE DECISIONI. PROFILI PROBLEMATICI NELL'APPLICAZIONE DELLA NORMATIVA SUL TRATTAMENTO DEI DATI PERSONALI PER FINALITÀ DI POLIZIA E RAGIONI DI GIUSTIZIA

Quaderno della Rivista Trimestrale
della Scuola di Perfezionamento per le Forze di Polizia

I/2026

SOMMARIO

"Le macchine sono incredibilmente veloci, precise e stupide. Gli esseri umani sono incredibilmente lenti, inaccurati e intelligenti. L'insieme dei due costituisce una forza incalcolabile".

Albert Einstein

"There is a world before mass access to generative AI and one after, and they are not the same thing".

Sam Altman (the CEO of OpenAI)

Abstract

Il lavoro analizza l'applicazione dell'intelligenza artificiale (IA) nei sistemi di sicurezza e nell'ambito delle attività delle Forze di polizia, con particolare attenzione alle implicazioni giuridiche sul trattamento dei dati personali. La ricerca esplora l'evoluzione tecnica dell'IA dai primi esperimenti degli anni '50 fino alle attuali applicazioni basate su *machine learning* e *deep learning*, evidenziando come questa tecnologia sia passata da sistemi con capacità limitate a strumenti sofisticati in grado di analizzare grandi volumi di dati e supportare processi decisionali complessi. Il lavoro esamina le diverse implementazioni dell'IA nelle agenzie di *law enforcement* in Italia e in alcuni Paesi europei, confrontandole con il modello adottato negli Emirati Arabi Uniti, dove l'approccio alla sicurezza attraverso la tecnologia presenta caratteristiche davvero peculiari, non trascurando di evidenziare il delicato equilibrio tra esigenze di sicurezza pubblica e tutela dei diritti fondamentali. L'analisi mette in luce come i diversi contesti giuridici, culturali e sociali influenzino l'implementazione e la regolamentazione di queste tecnologie, suggerendo la necessità di un approccio che coniughi innovazione tecnologica e salvaguardia dei princìpi etici fondamentali.

Il documento identifica le diverse tipologie di sistemi di intelligenza artificiale utilizzati nel campo della sicurezza, distinguendo tra IA debole (ANI – *Artificial Narrow Intelligence*) – attualmente predominante nelle applicazioni di *law enforcement* – e IA forte (AGI – *Artificial General Intelligence*), ancora in fase di sviluppo teorico. I sistemi ANI vengono implementati per compiti specifici come il riconoscimento facciale, l'analisi predittiva della criminalità e l'elaborazione di grandi quantità di dati investigativi. La ricerca evidenzia come l'evoluzione dell'IA, caratterizzata da fasi alterne di rapido sviluppo e rallentamenti (i cosiddetti "inverni dell'IA"), abbia portato negli ultimi due decenni a progressi significativi grazie alla disponibilità di *big data*, all'aumento della potenza di calcolo e all'introduzione di *hardware* specializzati.

La tesi approfondisce quindi i quadri normativi che regolano l'utilizzo dell'IA nelle attività di polizia nei diversi contesti geografici analizzati. In ambito europeo, viene esaminata l'interazione tra il GDPR, la Direttiva 2016/680 (cd *Law Enforcement Directive*) e il recente *AI Act*, evidenziando le specifiche salvaguardie previste quando i sistemi di IA vengono impiegati per finalità di prevenzione e repressione dei reati. Per l'Italia, l'analisi si concentra sulle normative nazionali che disciplinano il trattamento dei dati personali da parte delle Forze

di polizia e sull'adeguamento delle prassi operative agli *standard* europei. Il caso degli Emirati Arabi Uniti viene presentato come modello alternativo, caratterizzato da un approccio decisamente più centrato sull'implementazione dell'IA nella sicurezza pubblica e da un diverso bilanciamento tra innovazione tecnologica e tutela della *privacy*.

Il documento analizza infine le implicazioni etiche dell'utilizzo dell'IA nei sistemi decisionali delle Forze di polizia, affrontando questioni come il rischio di *bias* algoritmici, la trasparenza dei processi decisionali automatizzati e l'*accountability*. La ricerca evidenzia la necessità di sviluppare *framework* etici condivisi che possano guidare l'implementazione di queste tecnologie nel rispetto dei diritti fondamentali, pur riconoscendo le diverse sensibilità culturali e giuridiche dei paesi esaminati. In chiave prospettica, vengono identificate le sfide future legate all'evoluzione tecnologica, come l'integrazione dei sistemi di sorveglianza con tecnologie emergenti quali il riconoscimento delle emozioni e l'analisi comportamentale predittiva, suggerendo l'importanza di un approccio proattivo alla regolamentazione che accompagni l'innovazione tecnologica alla tutela imprescindibile dei valori democratici e del rispetto dei diritti umani.

This dissertation deals with the application of artificial intelligence (AI) in security systems and within the context of police operations, with a particular focus on the legal implications for the processing of personal data. The research examines the technical evolution of AI from the early experiments of the 1950s to current applications based on machine learning and deep learning, thus highlighting how this technology has progressed from systems with limited capabilities to sophisticated tools capable of analyzing large volumes of data and supporting complex decision-making processes. This study takes into account the various implementations of AI within law enforcement agencies in Italy as well as in some European countries, comparing them with the model adopted in the United Arab Emirates, where the approach to security through technology has got truly distinctive characteristics, without neglecting to highlight the delicate balance between public security requirements and the protection of fundamental rights. This dissertation underlines that the different legal, cultural and social contexts influence the implementation and regulation of these technologies, thus pointing out the need for an approach that combines technological innovation with the safeguarding of fundamental ethical principles.

Moreover, this paper describes the different types of artificial intelligence systems used in the field of security, distinguishing between narrow AI

(ANI – Artificial Narrow Intelligence) – currently prevailing in law enforcement applications – and general AI (AGI – Artificial General Intelligence), which is still at the theoretical development stage. ANI systems are deployed for specific tasks such as facial recognition, predictive crime analysis and the processing of large volumes of investigative data. The research highlights how the evolution of AI, characterized by alternating phases of rapid development and slowdowns (the so-called 'AI winters'), has led to significant progress over the last two decades, thanks to the availability of big data, increased computing power and the introduction of specialized hardware.

Then, this dissertation examines in depth the regulatory frameworks governing the use of AI in police activities across the various geographical contexts analyzed. In relation to the European context, it examines the interaction between the GDPR, Directive 2016/680 (the so-called Law Enforcement Directive) and the recent AI Act, highlighting the specific safeguard measures provided for when AI systems are used for the purposes of crime prevention and law enforcement. As regards Italy, the analysis focuses on the national regulations governing the processing of personal data by the police forces and on the adjustment of operational practices to European standards. The case of the United Arab Emirates is presented as an alternative model, characterized by an approach that is decidedly more focused on the implementation of AI in public security and by a different balance between technological innovation and privacy protection.

Finally, the document analyzes the ethical implications of using AI in police forces' decision-making systems, addressing issues such as the risk of algorithmic bias, the transparency of automated decision-making processes, and accountability. This research stresses the need to develop shared ethical frameworks that can guide the implementation of these technologies whilst respecting fundamental rights and, at the same time, acknowledging the differing cultural and legal peculiarities of the countries examined. Looking ahead, it identifies future challenges linked to technological evolution, such as the integration of surveillance systems with emerging technologies, like emotion recognition and predictive behavioural analysis, thus suggesting the importance of a proactive regulatory approach that balances technological innovation with the essential protection of democratic values and respect for human rights.

Introduzione

In tempi come questi non ci è sembrato fuori luogo introdurre questo nostro lavoro con un richiamo all'approccio della Chiesa universale alla sfida dell'intelligenza artificiale. La *Rerum Novarum* di Leone XIII, pubblicata nel 1891, nacque come reazione alle drammatiche trasformazioni portate dalla rivoluzione industriale. L'enciclica affrontava dunque le conseguenze sociali ed economiche del capitalismo emergente proponendo una giustizia sociale che equilibrasse le forze del mercato con il bene comune. Oltre un secolo più tardi, il nuovo pontefice Leone XIV, nel suo primo intervento ai cardinali, ha definito l'intelligenza artificiale "una delle questioni più critiche che l'umanità deve affrontare", giacché pone sfide incommensurabili alla difesa della dignità umana, alla giustizia e al lavoro[1].

Questo suo intervento segue peraltro quello già reso da papa Francesco I nel suo inedito intervento al G7 a guida italiana nel giugno del 2024, che qui vale la pena riportare integralmente: "l'intelligenza artificiale è uno strumento *sui generis*. Mentre l'uso di un utensile semplice è sotto il controllo dell'essere umano, e solo da quest'ultimo dipende il suo buon uso, l'intelligenza artificiale può adattarsi autonomamente al compito che le viene assegnato e anche operare scelte indipendenti dall'essere umano per raggiungere l'obiettivo prefissato".

In queste parole si coglie con grande evidenza la fiducia nei progressi che questa nuova tecnologia può garantire, così come la preoccupazione per i suoi prevedibili effetti collaterali: dopo aver conquistato il "monopolio informativo" attraverso i motori di ricerca, i *social network* e le piattaforme *web*, i grandi attori della tecnologia stanno ora acquisendo anche una sorta di "monopolio interpretativo", grazie agli algoritmi posti alla base dei programmi di intelligenza artificiale generativa. Chi controlla gli algoritmi (e in particolare i ccdd *large language models* – LLM), in parte controlla la narrazione globale, avendo la potenzialità di produrre contenuti indistinguibili dal vero. In questo senso, la posta in gioco è la libertà e la verità nell'era degli algoritmi.

---

1 *Credo ut intelligam.* Questa massima, radicata nel pensiero di S.Agostino al cui ordine il nuovo Pontefice appartiene, insegna che la fiducia in una verità superiore ("credere") stimola l'intelletto ad allargare i propri orizzonti ("comprendere"), invece di chiuderli per paura. In questo risiede la scelta della Chiesa di incoraggiare i progressi della scienza e della tecnica, purché siano esercitati in modo responsabile e orientati a servire la persona umana e il bene comune.

Forse i due papi contemporanei avevano in mente un "algoritmo alternativo", in cui la persona umana non sia variabile trascurabile ma fine ultimo di una tecnologia digitale che non punti a "risucchiare" l'uomo, bensì a servirlo.

Di fronte a una simile sfida, mentre la Chiesa ha delineato un approccio incentrato sui princìpi morali e su una visione di lungo periodo[2], dall'altra parte l'Unione Europea ha scelto una via più burocratica e dirigistica: l'*AI Act* proposto dalla Commissione e approvato da Consiglio e Parlamento nel 2024 è il primo tentativo organico di regolamentare l'IA a livello mondiale tramite un complesso sistema di classificazione dei rischi e dei correlati obblighi normativi per sviluppatori e utenti. Questo approccio, pur certamente animato da intenti lodevoli (dalla tutela dei consumatori alla prevenzione degli abusi) ha tuttavia sollevato alcune critiche per il suo carattere potenzialmente punitivo e certamente precoce. In sostanza, nel nobile sforzo di fissare un elevato *standard* globale, l'UE ha in realtà regolamentato molto rapidamente e in modo minuzioso una tecnologia ancora in gestazione, col rischio di soffocare sul nascere l'innovazione digitale del continente. L'imposizione di rigidi vincoli tecnici su un settore in rapidissima evoluzione può portare infatti a congelarne lo sviluppo, col rischio di allontanare i talenti e gli investimenti verso ambienti normativi meno onerosi (quali USA e Asia), e di "ingessare" la creatività di chi resta tra burocrazia, certificazioni e timori di sanzioni.

Un domani molto prossimo ci disvelerà se la scelta eurounitaria sia stata o meno pagante sul piano delle tutele ma anche dello sviluppo e se, e in che termini, avrà invece bisogno di essere rapidamente emendata per renderla maggiormente "a prova di futuro".

---

2 Non è mancato chi ha persino "suggerito" al Papa la redazione di una nuova enciclica di settore, una moderna *Rerum "artificialium"* che interpreti le sfide etiche e sociali di un'epoca dominata dalla tecnologia digitale e dall'automazione.

Capitolo I
L'intelligenza artificiale: profili tecnici e princìpi etici

di Marialuisa Simona Gregalli*

1.1. Storia ed evoluzione

L'intelligenza artificiale (IA) è un ramo dell'informatica che si avvale di princìpi di matematica, statistica e neuroscienze per creare macchine intelligenti in grado di affrontare compiti complessi con un livello di autonomia e sofisticazione sempre maggiore. Il presente capitolo si propone di esplorare l'evoluzione storica e i profili tecnici che caratterizzano l'intelligenza artificiale, per poi soffermarsi sui princìpi etici fondamentali che ne dovrebbero guidare lo sviluppo e l'applicazione. L'obiettivo è fornire una panoramica generale dell'IA, evidenziando in particolare la sua natura duale di potente strumento tecnologico con significative implicazioni morali e sociali.

Preliminarmente, conoscere la storia dell'intelligenza artificiale è utile per apprezzare lo stato attuale di questo campo in rapida evoluzione e per discernere le sue potenzialità future. Vedremo che la traiettoria dell'IA, da antiche concezioni filosofiche[3] fino alle sofisticate applica-

* Dirigente della Polizia Penitenziaria, già frequentatrice del XL Corso di Alta Formazione presso la Scuola di Perfezionamento per le Forze di Polizia.

3 L'idea di creare esseri artificiali capaci di pensare e agire come gli umani o le divinità può essere fatta risalire a miti e leggende di varie civiltà antiche, ben prima che il termine "Intelligenza Artificiale" venisse coniato. Nella mitologia greca, ad esempio, il dio Efesto era il maestro artigiano che creò servitori meccanici, come il gigante di bronzo Talo, che sorvegliava l'isola di Creta, e le ancelle d'oro, che lo assistevano nella sua officina. Anche in altre culture si ritrovano narrazioni di automi e creature artificiali, come testimoniato da leggende dell'antico Egitto e della Cina.
Il termine *"automaton"* stesso deriva dal greco antico e significa "agire di propria volontà". Già nell'antichità, si ritrovano esempi di "automi", dispositivi meccanici in grado di muoversi autonomamente senza intervento umano. Esempi storici di automi meccanici, sebbene non rappresentassero IA nel senso moderno, illustrano una persistente fascinazione umana

zioni odierne, è caratterizzata da periodi di fervente ottimismo, seguiti da fasi di disillusione e rinnovato vigore. Esaminando le passate ondate di entusiasmo e le successive fasi di rallentamento, spesso definite *"inverni dell'IA"*, è possibile sviluppare una prospettiva più equilibrata che fornisce un contesto più chiaro per valutare le attuali capacità e i limiti dell'IA, nonché per le implicazioni etiche che ne derivano.

I primi esperimenti concreti nel campo dell'IA ebbero inizio a metà del XX secolo. Un contributo fondamentale nel passaggio dalle speculazioni filosofiche a una concettualizzazione più scientifica dell'IA fu fornito dal matematico e pioniere dell'informatica britannico Alan Turing. Nel suo influente articolo del 1950, *"Computing Machinery and Intelligence"*, Turing affrontò la domanda *"Le macchine possono pensare?"* proponendo una riformulazione più pratica: *"Esistono computer digitali immaginabili capaci di eccellere nel gioco dell'imitazione?"*. Questo *"gioco dell'imitazione"*, ora noto come *Test* di Turing, prevedeva che un valutatore umano tentasse di distinguere tra le risposte di un *computer* e quelle di un umano in una conversazione testuale. Se il valutatore non fosse stato in grado di effettuare una distinzione affidabile, si sarebbe potuto considerare che il *computer* avesse dimostrato un comportamento intelligente paragonabile a quello umano. L'idea di Turing di sostituire la domanda sull'essenza del "pensiero" con la sua manifestazione comportamentale esterna si rivelò un approccio pragmatico, che permise di avviare una linea di ricerca più concreta nell'IA.

Tra gli anni '50 e '60, vennero sviluppati programmi che, per i tempi, apparivano sorprendenti: *computer* capaci di risolvere problemi di algebra, dimostrare teoremi di geometria e persino apprendere e parlare in inglese. Contemporaneamente, vennero realizzati i primi programmi capaci di giocare a dama. Nel 1951, Christopher Strachey scrisse un programma per il gioco della dama per il *computer* Ferranti Mark I,

---

per la creazione di artefatti autonomi. Tra questi si annoverano la colomba meccanica di Archita di Taranto (circa 400-350 a.C.), le creazioni idrauliche di Erone di Alessandria nel I secolo d.C., e, nel Rinascimento, i progetti di Leonardo da Vinci per un uomo robotico vestito da cavaliere tedesco. Il termine *"robot"*, come lo intendiamo oggi, fece la sua comparsa nel 1921 con l'opera teatrale di fantascienza *"Rossum's Universal Robots"* dello scrittore ceco Karel Čapek, che introdusse il concetto di "persone artificiali", chiamate appunto *robot*. Nel 1949, il libro *"Giant Brains, or Machines that Think"* di Edmund Callis Berkley paragonò i nuovi modelli di *computer* al cervello umano. La ricorrenza del tema degli automi in diverse culture e periodi storici suggerisce un desiderio umano fondamentale di superare i limiti della propria intelligenza e capacità fisica attraverso la creazione di macchine. Questa prospettiva storica aiuta a comprendere l'IA come una continuazione di questa antica aspirazione, desiderio che ha trovato una sua prima concretizzazione teorica e pratica a metà del XX secolo.

mentre Arthur Samuel ne sviluppò uno simile negli Stati Uniti nel 1952 per il prototipo IBM 701, integrando successivamente funzionalità che permettevano al programma di apprendere dall'esperienza. Nel 1954 il *Georgetown-IBM experiment* effettuò la traduzione completamente automatica di oltre sessanta frasi dal russo all'inglese; sebbene il sistema disponesse di un vocabolario limitato e di regole grammaticali semplificate, rappresentò un primo passo rilevante nel campo dell'elaborazione del linguaggio naturale. Nel 1955-56, Allen Newell e Herbert Simon svilupparono il *Logic Theorist*, un programma in grado di dimostrare teoremi matematici utilizzando la logica simbolica. Seppur rudimentali rispetto agli *standard* attuali, questi primi programmi alimentarono un forte ottimismo, dimostrando la capacità delle macchine di simulare forme di ragionamento logico, apprendimento e comprensione linguistica, compiti un tempo ritenuti esclusiva prerogativa umana.

L'estate del 1956 segnò inoltre un evento fondamentale nella storia dell'intelligenza artificiale: il *Dartmouth Summer Research Project on Artificial Intelligence*. Questo seminario, organizzato da John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon al *Dartmouth College*, è considerato il momento di nascita ufficiale dell'IA come disciplina accademica. Il *workshop* si basava sulla convinzione che ogni aspetto dell'apprendimento e dell'intelligenza potesse essere descritto in modo sufficientemente preciso da essere simulato da una macchina; l'obiettivo era ambizioso: insegnare alle macchine a usare il linguaggio, a formare astrazioni e concetti, a risolvere problemi complessi. La conferenza di Dartmouth si rivelò quindi cruciale, non solo per aver dato un nome al campo, ma anche per aver stabilito una visione decennale per la ricerca, promuovendo lo scambio di idee e alimentando grandi aspettative, che innescarono un'intensa fase di ricerca e sviluppo. Gli anni successivi alla conferenza di Dartmouth furono quindi caratterizzati da grande fiducia nelle potenzialità dell'IA, con molti ricercatori che, incoraggiati dai primi successi nel campo, prevedevano la creazione di macchine intelligenti come gli umani entro una generazione. Un esempio notevole di questi successi, oltre a quelli già menzionati, fu ELIZA (1966) di Joseph Weizenbaum, considerato uno dei primi *chatbot* capaci di simulare una conversazione umana attraverso l'elaborazione del linguaggio naturale.

Tuttavia, nonostante il fervore iniziale, gli anni '70 furono segnati da stagnazione e disillusione nella ricerca, un periodo noto come il primo *"inverno dell'IA"*. Le ambiziose promesse si scontrarono con le reali limitazioni dell'*hardware* e delle tecniche di programmazione disponibili all'epoca. I *computer* non disponevano della potenza di calco-

lo e della memoria sufficienti per affrontare compiti complessi. Il fallimento nel realizzare sistemi di traduzione automatica affidabili e le limitazioni intrinseche dei primi modelli di reti neurali contribuirono a creare un clima di scetticismo nei confronti del potenziale dell'IA[4]. La complessità di simulare l'intelligenza umana si rivelò quindi molto più grande del previsto. Una sfida significativa fu rappresentata dal paradosso di Moravec[5], che evidenziava come compiti apparentemente semplici per gli umani, come il riconoscimento di oggetti o la navigazione in ambienti non strutturati, si rivelassero estremamente difficili per le macchine, mentre compiti considerati "intelligenti", come la dimostrazione di teoremi, risultavano comparativamente più facili. Si riscontrarono inoltre notevoli difficoltà nel rappresentare la conoscenza del "senso comune"[6], essenziale per molte applicazioni dell'IA. La discrepanza tra le aspettative troppo alte create dall'ottimismo iniziale e le effettive capacità dei sistemi AI dell'epoca portò a una perdita di fiducia da parte del pubblico, degli investitori e dei governi. Un evento che ebbe un impatto significativo in questo periodo fu la pubblicazione del Rapporto Lighthill nel 1973[7]. Commissionato

---

4 Marvin Minsky e Seymour Papert nel libro *Percettroni: un'introduzione alla geometria computazionale*, del 1969, presentava un'analisi matematica rigorosa delle capacità e, soprattutto, delle limitazioni dei percettroni, un tipo di rete neurale artificiale sviluppata negli anni '50 e '60. Fu un'analisi matematica critica che dimostrò i limiti significativi dei percettroni a singolo strato, contribuendo a un temporaneo rallentamento della ricerca sulle reti neurali, ma fornendo anche spunti importanti per lo sviluppo futuro del campo.

5 Il paradosso di Moravec è un'osservazione fatta dai ricercatori di intelligenza artificiale Hans Moravec, Rodney Brooks, Marvin Minsky tra gli anni '70 e '80. Afferma che, contrariamente alle ipotesi tradizionali, le abilità di alto livello che richiedono ragionamento cosciente (come la logica, la matematica o la pianificazione) richiedono relativamente poca computazione, mentre le abilità senso_motorie di basso livello (come la percezione sensoriale, il riconoscimento facciale o la locomozione) richiedono enormi risorse computazionali. In altre parole, è più facile per le macchine eseguire compiti che gli umani trovano difficili, come risolvere equazioni complesse o giocare a scacchi, che eseguire compiti che gli umani trovano facili, come riconoscere un volto familiare o camminare su un terreno irregolare.

6 Il senso comune è quella vasta raccolta di conoscenze implicite che noi umani acquisiamo attraverso l'esperienza quotidiana. Comprende anche abilità come capire il contesto di una conversazione, interpretare espressioni facciali o prendere decisioni rapide in situazioni familiari. È una forma di conoscenza che diamo per scontata, ma che è estremamente complessa da codificare e insegnare a una macchina.

7 Michael James Lighthill (23 gennaio 1924 – 17 luglio 1998) è stato un matematico applicato britannico, noto per il suo lavoro pionieristico nel campo dell'aeroacustica e per aver scritto il rapporto Lighthill nel 1973, che affermava pessimisticamente che *"In nessuna parte del campo (dell'IA) le scoperte fatte finora hanno prodotto il grande impatto che era stato allora promesso"*, contribuendo al clima cupo dell'inverno dell'IA.

dal governo britannico, il rapporto criticava la mancanza di risultati pratici nella ricerca sull'IA e portò a significativi tagli di finanziamenti nel Regno Unito. La critica di Lighthill, basata sull'argomento che la ricerca sull'IA non aveva prodotto risultati tangibili che giustificassero gli ingenti investimenti, contribuì a minare la fiducia nel settore, portando i governi e le agenzie di finanziamento a reindirizzare le risorse verso aree ritenute più promettenti. Questo, combinato con le limitazioni tecnologiche, diede un duro colpo alla comunità dell'IA.

Negli anni '80 si assistette a un rifiorire dell'attenzione e dei capitali indirizzati all'intelligenza artificiale, dovuto in larga misura allo sviluppo dei "sistemi esperti", programmi progettati per emulare il processo decisionale umano in determinati ambiti determinati. Uno dei primi è stato XCON, usato per configurare *computer* su misura per i clienti[8]; altri sistemi, furono DENDRAL per l'analisi chimica e MYCIN per la diagnosi di infezioni del sangue. Questi programmi dimostrarono l'applicabilità efficace dell'IA nella risoluzione di problemi in domini specifici. L'abilità di tali sistemi nell'emulare il processo decisionale di esperti umani in campi ristretti fornì prove concrete del valore pratico dell'IA, riaccendendo l'entusiasmo, portando a un rinnovato ottimismo e a nuovi finanziamenti. Il governo giapponese lanciò l'ambizioso progetto *Fifth Generation Computer* nel 1981, investendo ingenti risorse con l'obiettivo di creare *computer* in grado di conversare, tradurre lingue e ragionare come gli esseri umani. Di risposta gli Stati Uniti lanciarono nel 1983 il programma decennale *Strategic Computing Initiative*, stanziando un miliardo di dollari. Questo periodo vide anche la nascita di aziende specializzate in *hardware*, come *Symbolics* e *Lisp Machines*, e in *software* per l'IA.

Questo rinnovato ottimismo fu però di breve durata. Tra la fine degli anni '80 e l'inizio degli anni '90, l'IA conobbe un secondo periodo di declino, il *"secondo inverno dell'IA"*. I sistemi esperti si rivelarono infatti "fragili", difficili da aggiornare e limitati a domini ristretti[9]. Il mer-

---

8 Ricevendo un ordine del cliente, XCON (*Digital Equipment Corporation*) selezionava automaticamente i componenti *hardware* e *software* necessari per creare un sistema funzionante e coerente in base alle specifiche del cliente. Assicurava che tutti i componenti necessari (CPU, memoria, periferiche, cavi, *software*, ecc.) fossero inclusi nell'ordine e che fossero compatibili tra loro.

9 Nonostante il successo iniziale, i sistemi esperti si rivelarono inadatti a risolvere problemi che richiedevano flessibilità e capacità di generalizzazione. La difficoltà di codificare tutta la conoscenza di un esperto in un insieme di regole e la "fragilità" di questi sistemi di fronte a *input* imprevisti ne limitarono l'applicazione su vasta scala.

cato delle macchine specializzate per l'esecuzione del linguaggio LIPS[10], preferito nella ricerca AI dell'epoca, crollò a causa della concorrenza di *workstation* più potenti ed economiche. Il rapido progresso dell'*hardware general-purpose*[11] rese obsolete le costose macchine specializzate per l'IA.

A partire dagli anni 2000, l'intelligenza artificiale ha vissuto una straordinaria rinascita, trainata dall'emergere dell'apprendimento automatico (*machine learning*) e, più recentemente, dell'apprendimento profondo (*deep learning*)[12]. Il *machine learning* ha permesso ai *computer* di apprendere dai dati senza una programmazione esplicita; in parallelo, il *deep learning* ha compiuto progressi eccezionali in numerosi ambiti, tra cui il riconoscimento delle immagini (anche grazie a *dataset* come ImageNet), i veicoli autonomi, l'elaborazione del linguaggio naturale (NLP), i *chatbot* (come Siri e Alexa), i sistemi di raccomandazione (come quelli utilizzati da Netflix e Amazon), il *trading* algoritmico, la diagnostica medica e molte altre applicazioni. Questa trasformazione è stata resa possibile grazie a una combinazione di fattori fondamentali: la disponibilità di enormi quantità di dati (*big data*), grazie all'espansione di *Internet* e la diffusione di dispositivi digitali che hanno reso disponibili enormi *dataset*; l'aumento esponenziale della potenza di calcolo; e l'introduzione di *hardware* specializzati, come le unità di elaborazione grafica (GPU) e tensoriale (TPU), che hanno reso l'elaborazione avanzata più accessibile ed efficiente, riducendo notevolmente i tempi di addestramento e rendendo possibile l'utilizzo su larga scala delle reti neurali profonde[13]. Le

---

10 Lisp (*List Processor*), ideato nel 1958 da John McCarthy, è una famiglia di linguaggi di programmazione che usa liste di azioni e simboli per dire al *computer* cosa fare, concentrandosi sulla manipolazione di concetti e informazioni, non solo numeri. Lisp è stato molto importante per l'intelligenza artificiale perché permetteva ai programmatori di esprimere idee complesse e di far "ragionare" i *computer* con simboli che rappresentavano la conoscenza.

11 Si definisce *general-purpose* (lett. "scopo generale" in inglese, traducibile "per uso generale") un dispositivo, strumento o meccanismo caratterizzato da una certa versatilità, adatto a molti impieghi e non specializzati per particolari esigenze. Al contrario vengono definiti *single purpose* o *special purpose* quei dispositivi che vengono realizzati direttamente per compiti specifici.

12 *Machine Learning* (ML) è un insieme di tecniche che permettono ai *computer* di imparare dai dati senza essere esplicitamente programmati. *Deep Learning* (DL) è un sottoinsieme del ML che utilizza reti neurali artificiali profonde (con molti strati) per apprendere in modo più sofisticato e gestire compiti complessi.

13 Le reti neurali profonde sono un tipo di modello di apprendimento automatico ispirato al funzionamento del cervello umano. La parola "profondo" si riferisce al fatto che queste reti sono composte da molti strati di nodi interconnessi (neuroni artificiali). Questa architettura a più livelli permette loro di apprendere rappresentazioni complesse e gerarchiche dei dati.

innovazioni algoritmiche sviluppate negli ultimi anni hanno così permesso di raggiungere risultati straordinari in compiti che un tempo sembravano irraggiungibili per le macchine. Tappe storiche come la vittoria di Deep Blue contro il campione di scacchi Garry Kasparov nel 1997, quella di Watson di IBM al *quiz* Jeopardy![14] nel 2011, e la sconfitta del campione di Go[15] Lee Sedol da parte di AlphaGo di DeepMind nel 2016, hanno segnato momenti emblematici dei progressi dell'intelligenza artificiale. Nel 2019 è stato poi rilasciato GPT-2, che rappresenta un passo significativo nella capacità delle macchine di generare testo coerente e di alta qualità. Il modello è stato addestrato su un vasto *corpus* di testo proveniente da *internet*, ed è in grado di generare frasi e paragrafi che spesso risultano difficili da distinguere da quelli scritti da esseri umani. Inoltre, l'emergere di modelli generativi, capaci di creare contenuti originali e realistici (testi, immagini, musica e video), come le reti avversarie generative (GAN)[16] introdotta nel 2014, DALL·E[17] rilasciato nel 2021 e ChatGPT[18] nel 2022, ha ulteriormente ampliato le potenzialità dell'IA, aprendo nuove opportunità nella produzione di contenuti artificiali, che spaziano dall'arte digitale alla scrittura automatica.

In sintesi, la combinazione di grandi volumi di dati, potenza computazionale e architetture neurali sempre più sofisticate ha consentito di superare molte delle barriere che in passato avevano limitato l'evoluzione del settore. Oggi, l'intelligenza artificiale è diventata una tecnologia perva-

---

14 Nel 2011, il *supercomputer* Watson di IBM ha partecipato al celebre *quiz* televisivo statunitense Jeopardy!, sfidando e sconfiggendo due dei campioni più forti del programma, Ken Jennings e Brad Rutter.

15 Il Go è un gioco da tavolo di origine cinese, noto per la sua complessità strategica e il vasto numero di possibili mosse.

16 Le reti avversarie generative, meglio conosciute con l'acronimo GAN (*Generative Adversarial Networks*), sono un tipo di architettura di intelligenza artificiale basata su reti neurali profonde, introdotta nel 2014 da Ian Goodfellow e altri ricercatori. Le GAN sono particolarmente famose per la loro capacità di generare dati realistici, come immagini, video, musica o persino testo, a partire da *input* casuali. Sono tra le tecnologie alla base dei *deepfake*, della generazione di volti realistici mai esistiti, e di molte innovazioni creative nel campo dell'IA.

17 DALL·E è un modello di intelligenza artificiale sviluppato da OpenAI, progettato per generare immagini a partire da descrizioni testuali (*text-to-image*). Il nome è un gioco di parole che unisce Salvador Dalí, celebre artista surrealista, e WALL·E, il robot protagonista dell'omonimo film Pixar – un chiaro riferimento alla creatività e all'automazione.

18 ChatGPT è un'evoluzione che sfrutta modelli più avanzati (GPT-3 e GPT-4), ottimizzati per interazioni conversazionali più naturali e contestuali, con maggiore sicurezza e capacità di adattamento, rispetto GPT-2 che è un modello potente per generare testo, ma non è specificamente progettato per la conversazione o il mantenimento del contesto a lungo termine. In sintesi, ChatGPT è un modello evoluto, specializzato e migliorato per le conversazioni rispetto a GPT-2.

siva, con impatti rilevanti in quasi ogni ambito produttivo, scientifico e sociale, con applicazioni che si estendono dalla ricerca accademica alla vita quotidiana, trasformando il modo in cui interagiamo con la tecnologia.

Cronologia delle Pietre Miliari dell'IA

Anno	Evento Principale
1950	Proposta del *Test* di Turing
1956	Conferenza di Dartmouth: nascita ufficiale dell'IA
1966	Sviluppo del *chatbot* ELIZA
1980	Primo sistema esperto commerciale (XCON)
1997	Deep Blue batte Garry Kasparov
2011	Watson di IBM vincono a Jeopardy!
2012	AlexNet rivoluziona il riconoscimento delle immagini
2019	Rilascio di GPT-2
2022	Rilascio di ChatGPT

1.2. Tipologie e profili tecnici

L'intelligenza artificiale può essere classificata in diversi modi, a seconda dei criteri utilizzati. Alcune delle classificazioni più comuni riguardano: il livello di capacità; le funzionalità; l'apprendimento; l'architettura.

1.2.1. Classificazione per livello di capacità

In particolare, sulla base del grado di autonomia e complessità cognitiva, i sistemi di IA si distinguono in tre categorie principali: Intelligenza Artificiale Debole (o Ristretta), Intelligenza Artificiale Forte (o Generale) e Superintelligenza Artificiale.

- IA Debole (ANI – *Artificial Narrow Intelligence*)
  L'Intelligenza Artificiale Debole, nota anche come ANI o IA ristretta, si riferisce a sistemi di IA orientati a specifici obiettivi e progettati per eseguire compiti singoli. Sebbene questi sistemi possano apparire intelligenti, in realtà simulano il comportamento umano all'interno di un insieme limitato di parametri e contesti. Essi eccellono in compiti specifici grazie alla loro ca-

pacità di analizzare grandi quantità di dati e identificare modelli ricorrenti, ma non possiedono la capacità di apprendere in modo indipendente o di adattarsi a situazioni nuove e impreviste. La loro intelligenza è "ristretta" in quanto confinata al problema specifico per cui sono stati progettati e addestrati. La maggior parte delle applicazioni di IA attualmente in uso rientra nella categoria dell'ANI. Esempi comuni includono gli assistenti virtuali come Siri, Alexa e Cortana, i sistemi di raccomandazione utilizzati da piattaforme come Netflix e Amazon, le auto a guida autonoma nel loro stadio attuale di sviluppo, il *software* di riconoscimento facciale e di immagini, i filtri *anti-spam* e i sistemi di traduzione linguistica. Questi esempi dimostrano le applicazioni pratiche dell'ANI nella vita quotidiana e in vari settori industriali, evidenziando lo stato attuale della tecnologia IA e la sua crescente integrazione.

- IA Forte (AGI – *Artificial General Intelligence*)
  L'Intelligenza Artificiale Forte, anche denominata AGI o IA generale, si riferisce al concetto di una macchina dotata di un'intelligenza generale simile a quella umana, con la capacità di apprendere e applicare la propria intelligenza per risolvere una vasta gamma di problemi. Il raggiungimento dell'AGI richiede che i sistemi di IA possiedano capacità cognitive a livello umano in un'ampia varietà di compiti, tra cui la comprensione, l'apprendimento e l'applicazione di conoscenze in diverse situazioni. Ciò implica non solo l'esecuzione di compiti, ma anche la comprensione dei concetti sottostanti e la capacità di trasferire l'apprendimento da un dominio all'altro, rappresentando una sfida significativa per l'IA attuale. La capacità di generalizzare l'apprendimento è un elemento distintivo chiave rispetto all'ANI.
  Le caratteristiche teoriche dell'AGI includono la consapevolezza di sé, la capacità di ragionamento astratto, il trasferimento di conoscenze tra domini diversi, la comprensione contestuale e l'adattabilità a situazioni nuove. Allo stato attuale, l'AGI non è ancora stata raggiunta e rappresenta un obiettivo primario della ricerca sull'IA. Le sfide per raggiungere l'AGI sono molteplici e includono la replicazione delle emozioni umane, della creatività e del buon senso.
- Super IA (ASI – *Artificial Superintelligence*)
  La Superintelligenza Artificiale, o ASI, rappresenta un livello teorico di IA che supera le capacità cognitive umane in ogni campo.

L'ASI è un concetto puramente teorico che rappresenta un livello di intelligenza ben oltre la comprensione e le capacità umane. Le caratteristiche teoriche dell'ASI includono la consapevolezza di sé e la capacità di superare l'intelletto umano in termini di velocità, memoria, analisi dei dati, creatività e risoluzione dei problemi. È un concetto futuristico e potenzialmente rischioso, ampiamente esplorato nella fantascienza e nella filosofia.
La distinzione tra AGI e ASI risiede nel grado di intelligenza, con l'ASI che rappresenta un salto significativo oltre l'intelligenza generale a livello umano. Mentre l'AGI mira a replicare l'intelligenza umana, l'ASI mira a superarla praticamente in ogni dominio cognitivo, con implicazioni profonde per il futuro ruolo dell'IA nella società.

Principali differenze tra le tipologie di intelligenza artificiale

Tipo di IA	Definizione	Caratteristiche Chiave	Stato Attuale	Esempi/ Implicazioni Teoriche
ANI (Debole o Ristretta)	IA orientata agli obiettivi, progettata per compiti singoli	Specializzazione, mancanza di vera comprensione o coscienza, dipendenza da dati di *training*, limitazione al proprio dominio	Maggioranza delle applicazioni di IA attuali	Assistenti virtuali, sistemi di raccomandazione, auto a guida autonoma (livelli attuali), riconoscimento facciale
AGI (Forte o Generale)	Conferenza di Macchina con intelligenza generale che imita l'intelligenza e il comportamento umano, con capacità di apprendere e risolvere qualsiasi problemanascita ufficiale dell'IA	Consapevolezza di sé, ragionamento astratto, trasferimento di conoscenze tra domini, comprensione contestuale, adattabilità	Non ancora raggiunta, obiettivo primario della ricerca	*Robot* con capacità cognitive umane, sistemi in grado di apprendere e applicare conoscenze in modo flessibile
ASI (Super intelligenza)	IA che supera l'intelligenza e le capacità umane in ogni campo	Consapevolezza di sé, capacità di superare l'intelletto umano in velocità, memoria, analisi dati, creatività, *problem solving*	Puramente teorica	Sistemi con intelligenza e capacità che superano di gran lunga quelle umane, potenziali benefici e rischi esistenziali

1.2.2. Classificazione per funzionalità operative

Un'altra classificazione dell'IA si basa sulle sue funzionalità operative specifiche[19]. In base a questa si distinguono:

- Macchine reattive: sono la forma più semplice di IA, prive di memoria e capaci solo di reagire a stimoli specifici. Un esempio emblematico è Deep Blue, il sistema di scacchi di IBM che sconfisse Garry Kasparov, il quale basava le sue mosse sull'analisi della situazione attuale sulla scacchiera senza ricordare le partite precedenti.
- IA con memoria limitata: sono in grado conservare informazioni per un breve periodo, consentendo loro di prendere decisioni più informate basate su esperienze recenti. Molte auto a guida autonoma utilizzano questa forma di IA per tenere traccia della velocità e della direzione degli altri veicoli. Altro esempio sono i *chatbot*.
- IA con Teoria della Mente (*Theory of Mind AI*); rappresenta una fase futura e teorica dell'IA in cui i sistemi saranno in grado di comprendere i pensieri, le emozioni, le credenze e le intenzioni degli altri, consentendo interazioni sociali più naturali e complesse. La Teoria della Mente è spesso considerata un aspetto cruciale dell'IA forte, in quanto consentirebbe alle macchine di comprendere gli stati mentali degli altri e di utilizzare questa comprensione per prevedere il loro comportamento. Questa capacità permetterebbe un'interazione più naturale ed efficace tra umani e IA, nonché tra sistemi di IA stessi.
- IA con Autoconsapevolezza (*Self-Aware AI*): l'autoconsapevolezza è lo stadio evolutivo finale teorico dell'IA, in cui le macchine svilupperebbero una coscienza di sé e la capacità di comprendere i propri stati interni, oltre a quelli degli altri.

1.2.3. Classificazione per modalità di apprendimento

Gran parte delle applicazioni pratiche dell'IA si basano oggi sull'Apprendimento Automatico – *Machine Learning* (ML) –, un campo

19 Diverse architetture sono alla base dei sistemi di intelligenza artificiale, ognuna con le proprie caratteristiche e adatta a specifici tipi di problemi. Le architetture più comuni includono le reti neurali, gli alberi decisionali e gli algoritmi di apprendimento per rinforzo.

che si concentra sullo sviluppo di algoritmi[20] in grado di apprendere dai dati e migliorare le proprie prestazioni nel tempo, senza essere esplicitamente programmati per ogni compito.

All'interno di questa classificazione, il *Machine Learning* si suddivide ulteriormente in diverse tipologie di apprendimento automatico, tra cui:

- L'apprendimento supervisionato (*Supervised Learning*), in cui l'algoritmo apprende da dati etichettati con le risposte corrette. Gli algoritmi vengono addestrati su *set* di dati etichettati in cui ogni esempio è associato a un *input* e a un *output* corrispondente. Imparano da questi dati etichettati per fare previsioni su dati nuovi e mai visti prima. Esempi sono la classificazione di immagini (gatti *vs.* cani), previsione dei prezzi delle case, riconoscimento vocale.
- L'apprendimento non supervisionato (*Unsupervised Learning*), l'algoritmo impara a scoprire strutture o raggruppamenti nascosti nei dati, senza dati etichettati o *output* predefiniti. Esempi sono: *clustering* di clienti in base al comportamento d'acquisto, riduzione della dimensionalità dei dati, rilevamento di anomalie.
- L'apprendimento per rinforzo (*Reinforcement Learning*), l'algoritmo apprende il comportamento ottimale in ambienti complessi attraverso tentativi ed errori, guidati da ricompense e penalità. In questo paradigma di apprendimento, un agente interagisce con un ambiente e riceve un *feedback* sotto forma di ricompense o penalità in base alle azioni intraprese. L'agente apprende una strategia, o politica, che gli permette di massi-

20 Numerosi algoritmi sono utilizzati nel *machine learning*, ognuno con i propri punti di forza e applicazioni, tra cui la regressione lineare e logistica, gli alberi decisionali e le foreste casuali, le macchine a vettori di supporto (SVM) e il *K-Nearest Neighbors* (KNN). Gli alberi decisionali sono algoritmi intuitivi e versatili utilizzati sia per compiti di classificazione che di regressione Un albero decisionale ha una struttura gerarchica composta da un nodo radice, rami, nodi interni e nodi foglia. L'algoritmo divide i dati in modo ricorsivo in base alle caratteristiche, creando un percorso decisionale che porta a una previsione o a una classificazione. Esistono diversi algoritmi per la costruzione di alberi decisionali, adatti a compiti di classificazione e regressione. I metodi di *ensemble* come le *Random Forest* combinano più alberi decisionali per migliorare l'accuratezza e la robustezza delle previsioni. Le *Gradient Boosted Decision Trees* (GBDT) rappresentano un'evoluzione degli alberi decisionali tradizionali, che utilizzano una tecnica chiamata *boosting* per migliorare le prestazioni del modello. La struttura ad albero rende questi modelli facilmente interpretabili, mostrando il processo decisionale basato su diverse caratteristiche. I metodi di *ensemble* come le *Random Forest* e il *Gradient Boosting* combinano le previsioni di più alberi per ridurre l'*overfitting* e migliorare le prestazioni di generalizzazione, portando a modelli più affidabili.

mizzare la ricompensa cumulativa nel tempo, attraverso un ciclo continuo di esplorazione di nuove azioni e sfruttamento di quelle che in passato hanno portato a risultati positivi. L'apprendimento per rinforzo ha trovato numerose applicazioni in diversi settori, tra cui la robotica, dove gli agenti possono imparare a eseguire compiti complessi attraverso l'interazione fisica con l'ambiente, le auto a guida autonoma, che devono apprendere le regole della strada e navigare in scenari complessi, i giochi, dove gli agenti possono raggiungere prestazioni sovrumane come nel caso di AlphaGo, e la finanza, dove gli algoritmi possono imparare strategie di *trading* ottimali.

- L'apprendimento Semi-Supervisionato (*Semi-Supervised Learning*), l'algoritmo utilizza un *mix* di dati etichettati e non etichettati per apprendere. Questo approccio è utile quando l'etichettatura dei dati è costosa o richiede molto tempo. Trova applicazione nella classificazione di documenti di testo, riconoscimento di immagini, analisi del *sentiment*.

Una delle evoluzioni più significative del *Machine Learning* è rappresentata dall'apprendimento profondo (*Deep Machine*), un sottoinsieme del *machine learning* che impiega reti neurali artificiali con molteplici strati (*layer*) per analizzare e apprendere da grandi volumi di dati. A differenza degli algoritmi tradizionali, che richiedono la definizione manuale delle *feature* (caratteristiche) rilevanti tramite un processo noto come *feature engineering*, il *Deep Learning* consente al sistema di apprendere automaticamente rappresentazioni gerarchiche dei dati, grazie alla struttura stratificata delle reti neurali profonde. Questa architettura multilivello permette di identificare *pattern* complessi senza intervento umano diretto, rendendo i modelli di *Deep Learning* particolarmente efficaci nell'elaborazione di immagini, testi, audio e altri dati non strutturati. Tuttavia, tali modelli richiedono ingenti quantità di dati e risorse computazionali per essere addestrati in modo efficace. Analogamente agli approcci classici del *machine learning*, il *deep learning* può essere applicato attraverso le diverse modalità di apprendimento sopra illustrate (supervisionate, non supervisionate, semi-supervisionate e per rinforzo).

In sintesi, mentre il *machine learning* tradizionale è generalmente più interpretabile e adatto a *dataset* di dimensioni ridotte, il deep *learning* si distingue per le sue prestazioni superiori in compiti complessi, a fronte di una maggiore opacità nei processi decisionali e di una più elevata complessità computazionale.

Confronto tra *Machine Learning* e *Deep Learning*

Caratteristica	***Machine Learning***	***Deep Learning***
Definizione	Apprendimento dai dati senza programmazione esplicita	Sottoinsieme del ML con reti neurali profonde
Tipo di Dati	Strutturati o non strutturati	Principalmente non strutturati (grandi *dataset*)
Complessità dei Problemi	Adatto a problemi di media complessità	Adatto a problemi complessi (immagini, testo, audio)
Estrazione di *Feature*	Richiede intervento umano	Automatica attraverso le reti neurali
Numero di Strati	Generalmente pochi strati	Molti strati (reti neurali profonde)
Prestazioni con grandi *dataset*	Può saturarsi	Migliorano con l'aumento dei dati

1.2.4. Classificazione per architettura

Le reti neurali (ANN) sono modelli computazionali ispirati alla struttura del cervello umano e sono fondamentali per molte applicazioni del *Machine Learning* e, in particolare, di *Deep Learning*. Una rete neurale è composta da strati di neuroni (o nodi) interconnessi, tra cui uno strato di *input*, uno o più strati nascosti e uno strato di *output*. Le informazioni fluiscono attraverso la rete, con ogni neurone che elabora l'*input* utilizzando funzioni di attivazione e trasmette il risultato allo strato successivo. L'addestramento di queste reti neurali avviene tipicamente attraverso l'algoritmo di *backpropagation*, che propaga l'errore all'indietro attraverso la rete per aggiornare i pesi e i *bias* dei neuroni, e attraverso tecniche di ottimizzazione per migliorare le prestazioni del modello[21]. La "profondità" di una rete neurale, ovvero il numero di

21 Sinteticamente le reti neurali funzionano:
- *Input*: I dati vengono forniti allo strato di *input*.
- Pesi e *Bias*: Ogni connessione tra i neuroni ha un peso associato, che ne determina l'importanza. Ogni neurone ha anche un *bias* (o soglia di attivazione).
- Funzione di Attivazione: Ogni neurone riceve gli *input* dai neuroni del livello precedente, li moltiplica per i rispettivi pesi, somma i risultati e aggiunge il *bias*. Questa somma viene poi passata attraverso una funzione di attivazione, che introduce non-linearità e determina l'*output* del neurone.

strati nascosti, le consente di apprendere *pattern* sempre più complessi dai dati. Il *deep learning*, che utilizza reti neurali profonde, è stato un motore trainante dei recenti progressi nell'IA.

Esistono diversi tipi di architetture di reti neurali, ognuna progettata per specifici tipi di dati e compiti.

- Le reti neurali *feedforward* (FNN) sono costituite da strati di neuroni interconnessi in cui le informazioni fluiscono in un'unica direzione, dall'*input* all'*output*.
- Le reti neurali convoluzionali (CNN) sono particolarmente efficaci per l'elaborazione di dati simili a griglie, come le immagini e video, utilizzando strati convoluzionali per rilevare *pattern* e strati di *pooling* per ridurre la dimensionalità.
- Le reti neurali ricorrenti (RNN) sono progettate per elaborare dati sequenziali, come il linguaggio naturale e le serie temporali, grazie alla loro capacità di mantenere una "memoria" degli *input* precedenti che consentono alle informazioni di persistere nel tempo.
- Le reti di memoria a lungo termine (LSTM) sono un'evoluzione delle RNN in grado di gestire dipendenze a lungo termine nelle sequenze di dati.
- Le reti *Transformer* sono un'architettura introdotta nel 2017, particolarmente efficace per la gestione di dati sequenziali come il testo, grazie a meccanismi di attenzione che permettono alla rete di concentrarsi sulle parti più rilevanti dell'*input* e all'architettura *encoder-decoder*, che consente di elaborare sequenze in parallelo, superando i limiti delle RNN.
- Infine, le reti generative avversarie (GAN) comprendono due reti neurali, un generatore e un discriminatore, che competono tra loro per generare dati sintetici realistici.

È importante notare che queste classificazioni non sono sempre mutuamente esclusive e un sistema di IA può rientrare in più categorie contemporaneamente. Ad esempio, un sistema di guida autonoma avanzato potrebbe utilizzare reti neurali profonde (basate sull'ar-

---

- Propagazione in Avanti (*Forward Propagation*): L'informazione fluisce attraverso la rete, dallo strato di *input* agli strati nascosti fino allo strato di *output*.
- Apprendimento (*Training*): La rete impara regolando i pesi e i *bias* delle connessioni. Questo processo avviene fornendo alla rete grandi quantità di dati etichettati (apprendimento supervisionato) e confrontando l'*output* della rete con l'*output* desiderato. L'errore viene quindi propagato all'indietro attraverso la rete (retropropagazione) per aggiornare i pesi e i *bias* in modo da ridurre l'errore nelle previsioni future.

chitettura) con apprendimento supervisionato e per rinforzo (basato sull'apprendimento) e rientrare nella categoria di IA con memoria limitata (basata sulle funzionalità).

1.3. Princìpi etici

L'evoluzione storica dell'intelligenza artificiale, dai suoi albori concettuali fino ai progressi esponenziali registrati negli ultimi decenni, ha evidenziato la straordinaria capacità di questa tecnologia di trasformare profondamente le modalità di interazione tra l'essere umano e il mondo digitale. Attualmente, le infrastrutture tecniche dell'IA costituiscono strumenti altamente sofisticati in grado di offrire soluzioni innovative e perseguire sfide sempre più complesse e trasversali. In questo contesto in continuo e veloce progresso, sarà fondamentale proseguire nello sviluppo di modelli di IA sempre più potenti e versatili, orientati al raggiungimento dell'ambizioso obiettivo dell'intelligenza artificiale generale. Parallelamente, si rende imprescindibile avviare una riflessione critica e sistematica sulle implicazioni etiche, sociali e giuridiche connesse all'utilizzo di tali tecnologie. Lo sviluppo e l'applicazione dell'intelligenza artificiale sollevano infatti una serie di importanti questioni etiche che devono essere affrontate per garantire un utilizzo responsabile e benefico di questa tecnologia. Pertanto, un approccio multidisciplinare, che integri competenze tecniche, etiche, sociali e giuridiche, sarà essenziale per guidare il suo sviluppo futuro in una direzione che sia sostenibile, responsabile e al servizio dell'umanità. Analizzeremo pertanto le questioni più rilevanti tra quelle sollevate dalle principali linee guida internazionali e istituzionali, con particolare riferimento a quelli elaborati nei documenti dall'UNESCO, dall'OCSE e dall'Unione Europea[22].

---

22 Raccomandazione dell'UNESCO sull'Etica dell'Intelligenza Artificiale: L'UNESCO ha prodotto il primo *standard* globale sull'etica dell'IA nel novembre 2021, applicabile a tutti i 194 Stati membri. La raccomandazione pone al centro la protezione dei diritti umani e della dignità, basandosi su princìpi fondamentali come la trasparenza e l'equità e sottolineando l'importanza della supervisione umana dei sistemi di IA. I dieci princìpi chiave includono proporzionalità, sicurezza, *privacy*, *governance*, responsabilità, trasparenza, supervisione umana, sostenibilità, consapevolezza ed equità.
Princìpi sull'IA dell'OCSE: L'OCSE ha stabilito i primi *standard* intergovernativi sull'IA, promuovendo un'IA innovativa, affidabile e rispettosa dei diritti umani e dei valori democratici. Adottati nel 2019 e aggiornati nel 2024, comprendono cinque princìpi basati sui

Uno dei problemi centrali è l'equità. I sistemi di IA devono essere progettati per trattare tutti gli individui e i gruppi in modo equo, minimizzando i pregiudizi non intenzionali. È quindi essenziale affrontare e mitigare i pregiudizi algoritmici derivanti da dati di addestramento distorti, promuovendo la giustizia sociale e garantendo che i benefici dell'IA siano accessibili a tutti. Ciò richiede l'utilizzo di *set* di dati di addestramento diversificati e inclusivi. Il *bias* algoritmico, che si verifica quando gli algoritmi incorporano pregiudizi presenti nei dati di *training*, può portare a decisioni discriminatorie e l'impatto può essere significativo, influenzando decisioni in settori critici (come l'assunzione, la concessione di prestiti, la giustizia penale e l'assistenza sanitaria). Le fonti di questi *bias* possono essere molteplici, inclusi dati storici che riflettono disuguaglianze sociali, rappresentazione insufficiente di alcuni gruppi o errate ipotesi di progettazione. Per mitigare questi rischi, è fondamentale adottare strategie come la raccolta di dati di *training* bilanciati e rappresentativi, l'utilizzo di tecniche di *debiasing* per identificare e correggere i pregiudizi nei dati e negli algoritmi, e un monitoraggio continuo delle prestazioni dei sistemi di IA per rilevare eventuali comportamenti discriminatori.

Un altro principio etico fondamentale riguarda la necessità che gli utenti e le parti interessate comprendano il funzionamento e i processi decisionali dei sistemi di intelligenza artificiale, secondo i concetti noti come trasparenza ed esplicabilità. La trasparenza implica la capacità di rendere conto di come un sistema di IA arriva a una determinata conclusione, mentre la spiegabilità si riferisce alla possibilità di fornire motivazioni comprensibili per tali decisioni. Comprendere il processo decisionale di un'IA, soprattutto in contesti ad alto rischio come la giustizia o la medicina, è essenziale per consentire a esperti e utenti di valutare l'affidabilità delle sue raccomandazioni e di intervenire in caso di errori o decisioni inique. L'intelligenza artificiale esplicabile (XAI) emerge pertanto come un'area cruciale di ricerca e sviluppo di tecniche per superare la "scatola nera" rappresentata da alcuni

---

valori – crescita inclusiva, diritti umani, trasparenza, robustezza, responsabilità – e cinque raccomandazioni per i governi – investimento in R&S, promozione di ecosistemi, definizione di politiche, sviluppo di capacità, cooperazione internazionale.
Regolamenti e Quadri Normativi Governativi: Diverse nazioni e regioni stanno sviluppando regolamenti e quadri normativi sull'etica dell'IA. L'*AI Act* dell'UE adotta un approccio basato sul rischio per la regolamentazione dei sistemi di IA.
Paesi come gli Stati Uniti hanno pubblicato il "*Blueprint for an AI Bill of Rights*". Il Dipartimento della Difesa degli Stati Uniti ha adottato cinque princìpi etici per l'IA, mentre la Comunità di *Intelligence* ha stabilito i propri princìpi etici.

modelli di IA complessi, come le reti neurali profonde, e per rendere i sistemi di IA più comprensibili e responsabili. Il problema della "scatola nera", dove il ragionamento interno dell'IA rimane oscuro, solleva infatti preoccupazioni in termini di responsabilità e fiducia. È pertanto essenziale fornire informazioni significative sulle fonti dei dati, sui processi e sulla logica alla base dei risultati dell'IA, garantendo livelli di trasparenza appropriati al contesto specifico.

Ed ancora, la responsabilità e imputabilità sono concetti etici interconnessi che riguardano la capacità di attribuire colpa e di rispondere delle azioni compiute da un sistema di intelligenza artificiale. In un mondo in cui le IA prendono decisioni sempre più autonome, diventa cruciale definire chi è responsabile in caso di errori, danni o conseguenze indesiderate. Questa questione è particolarmente complessa a causa della natura opaca e dell'apprendimento dinamico di alcuni sistemi di IA. Stabilire meccanismi di imputabilità, che consentano di tracciare le decisioni, di identificarne le cause e di attribuire responsabilità, è fondamentale per garantire che l'IA sia utilizzata in modo etico e per prevenire abusi o usi impropri.

Anche la *privacy* e la protezione dei dati sono princìpi etici di primaria importanza nell'era dell'IA. I sistemi, soprattutto quelli basati sul *machine learning* e sul *deep learning*, richiedono enormi quantità di dati, spesso di natura personale e sensibile, per il loro addestramento e funzionamento. La protezione di queste informazioni, il rispetto della *privacy* degli individui e la garanzia della sicurezza dei dati da accessi non autorizzati o usi impropri sono quindi imperativi etici fondamentali. È necessario implementare politiche e tecnologie che assicurino la riservatezza, l'integrità e la disponibilità dei dati, in conformità con le normative vigenti e con i princìpi etici di rispetto per l'autonomia e la dignità delle persone, assicurando pratiche responsabili di raccolta, archiviazione e utilizzo dei dati per prevenire violazioni.

L'impatto dell'IA sul mercato del lavoro è un'ulteriore questione etica di grande rilevanza. Se da un lato l'IA ha il potenziale per automatizzare compiti ripetitivi e pericolosi, aumentare la produttività e creare nuove opportunità di lavoro, dall'altro lato suscita preoccupazioni riguardo alla possibile perdita di posti di lavoro, soprattutto in settori che si basano su mansioni routinarie. È quindi fondamentale considerare le implicazioni etiche e sociali dell'automazione guidata dall'IA e adottare misure per gestire la transizione del mercato del lavoro, come programmi di riqualificazione e di sostegno per i lavoratori che potrebbero essere colpiti.

Princìpi Etici Fondamentali dell'IA

Principio Etico	Definizione	Sfide Potenziali	Strategie di Mitigazione
Equità	Assenza di pregiudizi o discriminazioni nelle decisioni prese dall'IA	*Bias* nei dati di *training*, sottorappresentazione di gruppi, errate ipotesi di *design*	Raccolta dati bilanciata, tecniche di *debiasing*, monitoraggio continuo delle prestazioni
Trasparenza	Comprensione del processo decisionale dell'IA	Opacità di alcuni modelli (es. reti neurali profonde)	Sviluppo di tecniche di *Explainable AI* (XAI), documentazione chiara dei modelli
Responsabilità	Capacità di attribuire colpa e rispondere delle azioni dell'IA	Difficoltà nell'attribuire responsabilità a sistemi autonomi e complessi	Meccanismi di *accountability*, tracciamento delle decisioni, auditabilità dei sistemi
Privacy	Protezione delle informazioni personali utilizzate e generate dall'IA	Rischio di violazione dei dati, usi impropri, sorveglianza non consensuale	Implementazione di politiche e tecnologie per la protezione dei dati, anonimizzazione, crittografia
Impatto Lavoro	Considerazione delle implicazioni dell'IA sull'occupazione	Perdita di posti di lavoro, necessità di nuove competenze	Programmi di riqualificazione, sostegno per i lavoratori, creazione di nuove opportunità di lavoro in settori emergenti

Tra i pilastri chiave di una intelligenza artificiale etica sono inoltre da annoverare:

- la necessità della Supervisione e Autonomia Umana: è fondamentale mantenere il controllo umano sui sistemi di IA ed evitare un'eccessiva dipendenza da essi, assicurando che l'IA aumenti l'intelligenza umana piuttosto che sostituirla completamente. Il giudizio umano rimane essenziale, soprattutto quando l'IA influisce sui diritti fondamentali, ed è importante consentire agli individui di interagire e comprendere i sistemi di IA;
- Sicurezza e Robustezza: i sistemi di IA devono essere sviluppati per essere sicuri, affidabili e resilienti agli attacchi, proteggendo i sistemi e i dati dell'IA dalle minacce informatiche e dall'influenza ostile. È fondamentale garantire che i sistemi funzionino in modo appropriato in condizioni normali e av-

verse, implementando misure di *test* e garanzia durante l'intero ciclo di vita dell'IA;
- la Sostenibilità: è necessario valutare gli impatti ambientali e sociali delle tecnologie di IA, minimizzando il consumo energetico e promuovendo pratiche di sviluppo ecocompatibili. È importante considerare il ciclo di vita dei sistemi di IA dalla produzione allo smaltimento e allineare lo sviluppo dell'IA agli Obiettivi di Sviluppo Sostenibile;
- la Beneficenza e non maleficenza: l'IA deve essere sviluppata e utilizzata per promuovere il benessere umano e il bene comune, evitando usi che possano causare danni a persone o all'ambiente. È essenziale garantire che i benefici dei progressi siano realizzati e massimizzati adottando un approccio allo sviluppo e all'uso dell'IA incentrato sull'essere umano.

Capitolo II
La regolamentazione dell'intelligenza artificiale in ambito italiano ed europeo

di Filippo Vanni*

2.1. *Artificial Intelligence Act*: il regolamento (UE) 2024/1689

L'*Artificial Intelligence Act* nasce dalla necessità di armonizzare la regolamentazione dell'intelligenza artificiale (IA) all'interno dell'Unione europea, bilanciando innovazione, sicurezza e diritti fondamentali. La Commissione Europea ha presentato la proposta normativa nell'aprile 2021, ispirandosi al modello *"risk-based"*, che classifica i sistemi di IA in base al potenziale impatto sui diritti e sulla sicurezza, con l'obiettivo di prevenirne usi dannosi (es. sorveglianza di massa) e promuovere un cd "ecosistema digitale etico".

L'adozione di sistemi di IA ha un forte potenziale per apportare benefici alla società nel suo complesso, favorire la crescita economica e migliorare l'innovazione e la competitività globale. Tuttavia, in alcuni casi, le caratteristiche specifiche di alcuni sistemi possono creare nuovi rischi per la sicurezza degli utenti, inclusa la sicurezza fisica e per i diritti fondamentali. In tale quadro, alcuni potenti modelli di IA ampiamente utilizzati potrebbero persino comportare dei rischi sistemici.

Secondo al Commissione, ciò può favorire incertezza giuridica e un'adozione potenzialmente più lenta delle tecnologie di intelligenza artificiale da parte di autorità pubbliche, imprese e cittadini, a causa della mancanza di fiducia rispetto all'esistenza dei richiamati rischi. Inoltre, risposte normative disomogenee da parte delle autorità nazionali rischierebbero di frammentare il mercato interno.

* Colonnello dell'Arma dei Carabinieri, già frequentatore del XL Corso di Alta Formazione presso la Scuola di Perfezionamento per le Forze di Polizia.

Sempre secondo l'opinione di palazzo Berlaymont, per rispondere a queste sfide è stato necessario un intervento legislativo volto a garantire un mercato interno pronto ad accogliere le potenzialità dei sistemi di intelligenza artificiale, in cui sia i vantaggi che i rischi siano adeguatamente previsti e affrontati[23]. Il nuovo quadro giuridico si applica sia agli attori pubblici che a quelli privati all'interno e all'esterno dell'UE, purché il sistema di intelligenza artificiale sia immesso sul mercato dell'Unione o il suo utilizzo abbia un impatto sulle persone che si trovano all'interno dei confini unionali.

Gli obblighi possono riguardare sia i fornitori (ad esempio, lo sviluppatore di uno strumento di *screening* dei *curricula*) sia gli utilizzatori di sistemi di IA (ad esempio, una banca che acquista tale strumento di *screening*). Al contrario, le attività di ricerca, sviluppo e prototipazione che si svolgono prima dell'immissione sul mercato di un sistema di IA non sono soggette alle disposizioni del nuovo Regolamento, così come sono esenti anche i sistemi progettati esclusivamente per scopi militari, di difesa o di sicurezza nazionale.

Come anticipato, la legge sull'intelligenza artificiale contiene un approccio basato sul rischio livellato su più indicatori:

- rischio inaccettabile: un insieme molto limitato di usi particolarmente dannosi dell'IA che violano i valori dell'UE perché incidono sui diritti fondamentali e che vengono pertanto vietati. Tra questi, lo sfruttamento delle vulnerabilità delle persone, la manipolazione e l'utilizzo di tecniche subliminali, l'attribuzione di un punteggio sociale per scopi pubblici e privati, la polizia predittiva basata esclusivamente sulla profilazione delle persone, l'estrazione non mirata di immagini facciali da *internet* o dalle telecamere di sorveglianza per creare o espandere *database*, il riconoscimento delle emozioni nei luoghi di lavoro e nelle istituzioni educative, a meno che non siano sottese ragioni mediche o di sicurezza (ad esempio, monitoraggio dei livelli di stanchezza di un pilota), la categorizzazione biometrica delle persone fisiche per dedurne la razza, le opinioni politiche, l'appartenenza sindacale, le convinzioni religiose o filosofiche o l'orientamento sessuale. Sono infine incluse in tale categoria l'identificazione biometrica remota in tempo reale in spazi accessibili al pubblico da parte delle for-

23 Così la presentazione della proposta da parte della Commissione europea, sul sito https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_1683.

ze dell'ordine, fatta salve alcune limitate eccezioni procedurali che dopo avremo modo di esporre;
- ad alto rischio: un numero limitato di sistemi di IA definiti nella proposta, che potrebbero avere un impatto negativo sulla sicurezza delle persone o sui loro diritti fondamentali (come tutelati dalla Carta dei diritti fondamentali dell'UE). In allegato alla legge sono riportati gli elenchi dei sistemi di IA ad alto rischio, in modo tale da poter essere aggiornati e rivisti per allinearli all'evoluzione del settore. Tali sistemi di IA ad alto rischio includono, ad esempio, quelli che valutano se un individuo è in grado di ricevere un determinato trattamento medico, di ottenere un determinato lavoro o un prestito bancario. Altri sistemi di IA ad alto rischio sono quelli utilizzati dalle forze dell'ordine per profilare le persone o valutarne il rischio di commettere un reato;
- rischio specifico per la trasparenza: per promuovere la fiducia di persone e imprese, è importante garantire la trasparenza nell'uso dell'intelligenza artificiale. Pertanto, l'*AI Act* introduce requisiti specifici di trasparenza per alcune applicazioni, ad esempio laddove vi sia un chiaro rischio di manipolazione (l'uso di *chatbot* ne è un caso) o *deepfake*. Gli utenti, in sostanza, devono essere consapevoli di interagire con una macchina;
- rischio minimo: la maggior parte dei sistemi di IA può essere sviluppata e utilizzata nel rispetto della legislazione vigente, senza ulteriori obblighi di legge. I fornitori di tali sistemi possono scegliere volontariamente di aderire a codici di condotta volontari.

La proposta legislativa della Commissione contenente il nuovo regolamento sull'intelligenza artificiale viene presentata nell'aprile 2021, con grande enfasi anche mediatica. Da allora ha preso avvio il complesso procedimento legislativo unionale che, come da disposizioni dei Trattati, ha dapprincipio coinvolto Consiglio UE, dove i rappresentanti dei Governi degli Stati membri (attraverso esperti inviati dai Ministeri interessati) hanno negoziato le rispettive posizioni, spesso divergenti. In tale quadro, si deve segnalare il caso di paesi come Italia, Francia e Germania, preoccupati per l'impatto del nuovo *corpus* normativo sull'innovazione del settore, e di Belgio e Paesi Bassi che invece hanno spinto per tutele più rigorose dei diritti delle persone[24].

---

24 https://europa.today.it/attualita/italia-no-regole-intelligenza-artificiale.html. "Niente limiti all'intelligenza artificiale": Italia, Germania e Francia bloccano la legge Ue. Il Parlamento

Con particolare riferimento all'Italia, in tutti i consessi tecnici, strategici o politici, formali o informali, il Paese ha costantemente sostenuto come tale tecnologia sia destinata a ricoprire un ruolo sempre maggiore nel supporto alle attività preventive e investigative delle Forze di polizia e delle analisi forensi, pur nella consapevolezza di essere a favore di Forze di polizia "tecnologicamente supportate" piuttosto che "tecnologicamente guidate".

Partendo da tali presupposti, il nostro Paese (con l'appoggio della Francia) ha fortemente sostenuto un testo che:

- lasciasse impregiudicata la facoltà per le Forze di polizia, già prevista nella prima versione approvata dal Consiglio, di utilizzare sistemi di identificazione biometrica in spazi pubblici *"in real time"* previa delega dell'autorità giudiziaria, per l'accertamento dei fatti e la repressione dei reati tassativamente indicati dalla norma;
- consentisse l'uso dell'IA per l'identificazione biometrica da remoto *"ex post"*, senza alcuna preliminare autorizzazione dell'AG;
- escludesse dall'elenco dei sistemi ad alto rischio gli algoritmi per l'individuazione di *deepfake*, per il *crime analytics* e per la verifica dell'autenticità dei documenti di viaggio da parte delle autorità di polizia, quali essenziali strumenti di supporto tecnologico alle attività di *law enforcement*.

A seguito di un lungo dibattito, gli Stati membri hanno infine trovato un compromesso accettabile, votando e approvando quindi il Regolamento all'unanimità: il Consiglio ha quindi adottato la propria posizione ufficiale nel dicembre 2022, precisando – quale *red line* – la necessità che dalla regolazione fosse esclusa la sicurezza nazionale[25].

Allo stesso modo ha proceduto il co-legislatore unionale, il Parlamento Europeo, che ha approvato la propria posizione formale nel giugno 2023, con 99 voti a favore, 28 contrari e 93 astensioni.

Quel consesso, tuttavia, ha voluto incidere in maniera decisamente significativa sull'uso degli strumenti *"AI-driven"* da parte delle Forze di polizia. In particolare, i deputati hanno modificato

---

europeo chiede norme rigide di trasparenza per i *big* come ChapGpt. Ma i governi dei tre principali Paesi del blocco temono che queste blocchino lo sviluppo delle imprese di casa.

25 Nei negoziati unionali si parla di *red line* quando l'eccezione posta da uno Stato o da un organo comunitario (come in questo caso il Consiglio) costituisce una condizione imprescindibile per l'approvazione del testo.

in modo sostanziale l'elenco dei divieti sugli usi dei sistemi di IA, introducendovi:

- sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico, proponendo di vietarli sempre anche alle Forze di polizia;
- sistemi di identificazione biometrica a distanza *"ex post"*, consentiti alle sole Forze di polizia solo per il perseguimento di reati gravi e solo previa autorizzazione giudiziaria;
- sistemi di categorizzazione biometrica che utilizzano caratteristiche sensibili (ad esempio, sesso, razza, etnia, cittadinanza, religione, orientamento politico);
- sistemi di polizia predittiva (basati su profili, ubicazione o comportamenti criminali passati);
- sistemi di riconoscimento delle emozioni ad opera delle Forze di polizia, per la gestione delle frontiere, nei luoghi di lavoro e nelle istituzioni educative;
- scorporo indiscriminato di dati biometrici dai *social media* o dalle telecamere a circuito chiuso per creare *database* di riconoscimento facciale (in violazione del diritto alla *privacy*).

I deputati hanno infine ampliato la classificazione delle *aree ad alto rischio* per includervi i danni alla salute, alla sicurezza, ai diritti fondamentali o all'ambiente. Hanno inoltre aggiunto alla *lista ad alto rischio* i sistemi di IA per influenzare gli elettori nelle campagne elettorali e proposto la creazione di un *Ufficio per l'Intelligenza artificiale* dell'Unione europea, al posto del *Comitato per l'IA* presente nel testo originario dalla Commissione, approvato dal Consiglio.

A seguito dell'approvazione delle due posizioni da parte dei co-legislatori sono state avviate le fasi ccdd di "trilogo tecnico"[26]: dopo negoziati intensi, nel dicembre 2023 si è finalmente raggiunta una posizione di compromesso. Il regolamento è stato quindi definitivamente adottato nel marzo 2024 e pubblicato nella Gazzetta ufficiale UE del 24/04/2024[27].

Con specifico riferimento alle facoltà attribuite alle Forze di polizia, il testo vigente contiene queste principali previsioni:

---

26 Riunioni tecniche congiunte tra i *rapporteurs* del Parlamento europeo (On. Brando Benifei – italiano, gruppo *Socialisti e Democratici*– e Ioan-Dragos Tudorache – romeno, gruppo *Renew Europe*), componenti della presidenza di turno del Consiglio e membri esperti della Commissione.

27 Regolamento (UE) n. 2024/1689.

- principio del doppio controllo (*4eyes rule*): nessuna decisione può essere presa sulla base dei soli risultati provenienti da un sistema di IA "ad alto rischio", a meno che non sia confermata da un doppio controllo di due persone fisiche, esperte nella materia. Come richiesto, tra gli altri, dall'Italia, tale regola è esclusa per i sistemi in uso alle Forze di polizia, nonché per esigenze di controllo della migrazione, dei confini o di asilo;
- registrazione dei sistemi di IA in uso alle Forze di polizia: gli organismi di *law enforcement* degli Stati membri devono registrare i sistemi di IA in uso presso una sezione non pubblica e sicura di un *database* dedicato, istituto presso la Commissione europea;
- gli Stati membri devono individuare una "autorità di sorveglianza di mercato" per la supervisione dei sistemi di IA ad alto rischio in uso alle Forze di polizia, cui notificare, assieme al Garante per la *privacy*, l'uso di tali sistemi. Come richiesto, tra gli altri, dall'Italia, in nessun caso tali autorità possono interferire con l'indipendente esercizio della funzione giudiziaria o con l'esercizio delle funzioni di polizia giudiziaria;
- uso dell'IA per l'identificazione biometrica remota in *real-time* (RBI): per i suoi riflessi operativi, è la parte più importante del testo e quella dove i confronti col Parlamento sono stati più serrati. Il Parlamento, decisamente rigido sul punto, ha chiesto da subito un divieto totale di uso di questi sistemi anche da parte delle Forze di polizia. Come fortemente richiesto, tra gli altri, dall'Italia, tale pratica è stata invece ammessa, pur alle seguenti condizioni: preventiva autorizzazione giudiziaria[28]; ricerca di vittime dei reati di rapimento, traffico di esseri umani, sfruttamento sessuale, ricerca di minori scomparsi[29]; prevenzione di una specifica, sostanziale e imminente minaccia alla vita o alla salute di una persona; prevenzione di una minaccia *autentica e presente* o *autentica e prevedibile* di un attacco terroristico. Su richiesta dell'Italia, è stato cancellato il termine *"imminente"* che, nelle intenzioni del Parlamento, avrebbe dovuto caratterizzare la minaccia

28 In caso di urgenza le attività investigative possono essere avviate, salva la necessità di ottenere autorizzazione giudiziaria entro le successive 24 ore.

29 L'Italia e altri grandi Stati membri avevano richiesto un catalogo decisamente più ampio, ma la Presidenza non è stata in grado di convincere il Parlamento sul punto.

terroristica[30]; la ricerca o l'identificazione di una persona sospettata di aver commesso uno dei reati indicati specificamente nel provvedimento[31]; preventiva predisposizione di un *impact assessment* sui diritti fondamentali, da redigere a cura della forza di polizia che utilizza il sistema[32];

- uso dell'IA per l'identificazione biometrica *ex-post* (*ex post* RBI): anche questa parte del provvedimento ha importanti riflessi operativi. Tuttavia, nonostante gli sforzi negoziali di Italia, Francia, Germania e Grecia, il testo richiede la *preventiva autorizzazione giudiziaria o di altra autorità amministrativa* anche per l'utilizzo di tale sistema, come preteso fin da subito dal Parlamento europeo[33].

Il provvedimento prevede altresì che gli Stati membri possano decidere di approvare misure legislative più restrittive per l'uso dell'IA sia in attività di RBI che di *ex-post* RBI.

Sulla base delle previsioni del testo di compromesso rimane vietato l'uso dell'IA per attività di "polizia predittiva" (ovvero per valutare il rischio che una persona possa divenire vittima o autore di un reato), per influire sulla libertà di autodeterminazione (con strumenti quali poligrafi, narcoanalisi, ecc.), per valutare l'attendibilità delle prove nell'ambito delle indagini o di un processo penale.

Al contrario, il regolamento IA non si applica:

- ai sistemi prodotti all'interno o all'esterno dell'UE che siano usati *esclusivamente* per scopi afferenti alla sicurezza nazionale, alla difesa e alle esigenze militari degli Stati membri[34];

---

30 Il testo originariamente voluto dal Parlamento recitava: *"specific and imminent threat of a terrorist attack"*.

31 Terrorismo, traffico di esseri umani, sfruttamento sessuale di minori e pedopornografia, traffico illecito di stupefacenti e sostanze psicotrope, traffico illecito di armi, munizioni o esplosivi, omicidio o lesioni personali gravi, traffico illecito di organi umani o tessuti, traffico illecito di materiale nucleare o radioattivo, rapimento o sequestro di persona/ostaggi, crimini che ricadono sotto la giurisdizione della CPI, sequestro di nave o aeromobile, violenza sessuale, reati ambientali, rapina con l'uso di armi, sabotaggio, partecipazione a organizzazione criminale coinvolta in uno o più dei crimini indicati sopra.

32 Come richiesto dall'Italia, in caso di urgenza la Forza di polizia può avviare le attività investigative, rinviando – senza ritardo – la predisposizione della citata analisi di impatto.

33 È parere di chi scrive che, per l'uso di intelligenza artificiale per l'identificazione biometrica *ex-post*, sarebbe stato preferibile lasciare libere le Forze di polizia di agire di iniziativa, al fine di velocizzare le attività investigative.

34 Questa formulazione, sostenuta anche dall'Italia attraverso gli attaché delle Agenzie per le informazioni e la sicurezza distaccati in RappUE, potrebbe incontrare la forte opposizione della Francia in ragione della presenza dell'avverbio *esclusivamente*. Quello Stato avrebbe

- ai sistemi in uso alle Forze di polizia per l'individuazione dei ccdd *"deepfake"*, per l'analisi criminale (cd *"crime analytics"*) e per l'individuazione di documenti falsi.

Al fine di esaminare e studiare le esperienze e le buone prassi in materia di uso dell'IA, la Commissione viene autorizzata a istituire un Ufficio europeo per l'Intelligenza artificiale (*European AI Office*), composto da un rappresentante per Stato membro[35] che avrà il compito, tra gli altri, di indicare all'esecutivo dell'Unione eventuali esigenze di modifica del Regolamento sulla base delle concrete esperienze maturate all'interno degli Stati.

In conclusione, l'*AI Act* rappresenta senza dubbio un modello pionieristico nella *governance* dell'intelligenza artificiale, con un innovativo approccio proporzionale al rischio. Tuttavia, la reale efficacia dello strumento potrà essere valutata solo nei prossimi mesi e dipenderà dall'implementazione negli Stati membri del relativo Regolamento e dalla capacità delle autorità di vigilanza (ad esempio l'*European AI Board*) di mediare tra innovazione, sicurezza e diritti fondamentali. Le eccezioni che sono state introdotte per il lavoro delle Forze di polizia, in ogni caso, rimangono i nodi più critici e quelli destinati a influenzare il dibattito giuridico e operativo dei prossimi anni.

2.2. Il disegno di legge di recepimento del regolamento europeo sull'intelligenza artificiale (Atto Senato 1146 – Atto Camera 2316)

Il 20 maggio 2024 la Presidenza del Consiglio invia alle Camere (precisamente al Senato della Repubblica) un disegno di legge di iniziativa governativa contenente il recepimento normativo interno del nuovo Regolamento UE in materia di intelligenza artificiale. Nel volgere di alcune settimane il Senato ne avvia l'esame, effettua un ampio ciclo di audizioni di esperti in materia[36] e infine ne completa l'approvazione il 20 marzo 2025, apportando alcune modifiche al testo originariamente predisposto dal Governo.

---

preferito una formulazione più ampia, in considerazione del fatto che la sicurezza nazionale è una competenza esclusiva degli Stati membri.

35 Ai lavori può prendere parte anche la Commissione e il Garante europeo per la *privacy*, senza diritto di voto. I rappresentanti degli Stati membri permangono in carica per 3 anni, rinnovabili una sola volta.

36 Tra questi, il Garante per la protezione dei dati personali, personale delle Agenzie per la cybersicurezza nazionale e per l'Italia digitale, il Presidente della Commissione sull'IA istituita presso la Presidenza del Consiglio.

Secondo il procedimento legislativo previsto dal nostro ordinamento, il disegno di legge è stato ora trasmesso alla Camera dei deputati e – se interverranno modifiche – dovrà poi essere ritrasmesso nuovamente al Senato per la conforme approvazione definitiva.

Nel dettaglio, il disegno di legge governativo approvato dal Senato – secondo le direttrici tracciate dal Regolamento UE in materia – si propone di regolamentare la ricerca, la sperimentazione, lo sviluppo, l'adozione e l'applicazione di sistemi e modelli di intelligenza artificiale attraverso un approccio antropocentrico. La normativa mira a garantire che l'utilizzo dell'IA avvenga in modo corretto, trasparente e responsabile, bilanciando le opportunità offerte da queste tecnologie con la necessaria vigilanza sui rischi economici, sociali e sui diritti fondamentali.

In tale quadro, il testo legislativo stabilisce che le attività legate all'intelligenza artificiale devono rispettare i diritti fondamentali garantiti dalla Costituzione italiana e dal diritto dell'Unione europea, conformandosi ai princìpi di trasparenza e proporzionalità, sicurezza e protezione dei dati personali, riservatezza e accuratezza, non discriminazione e parità di genere, sostenibilità.

Come accennato, al fine di raggiungere questi obiettivi un aspetto centrale del provvedimento è il rispetto dell'autonomia e del potere decisionale dell'essere umano, enfatizzando così il concetto di *"human in the loop"*.

Al fine di un perfetto allineamento col nuovo quadro normativo unionale, il disegno di legge chiarisce che non sono prodotti nuovi obblighi rispetto a quelli già previsti dal regolamento europeo, al primario fine di evitare duplicazioni normative e favorire una più uniforme e certa applicazione del diritto nel delicato settore.

Il provvedimento pone particolare enfasi sulla cybersicurezza lungo tutto il ciclo di vita dei sistemi e modelli di IA, prevedendola come "precondizione essenziale" per garantire il rispetto dei diritti fondamentali. In quest'ottica, una delle disposizioni più significative riguarda la cd "sovranità digitale": i sistemi di intelligenza artificiale destinati all'uso in ambito pubblico devono essere installati su *server* ubicati nel territorio nazionale, con l'eccezione di quelli impiegati all'estero nell'ambito di operazioni militari. Questa misura mira a garantire la sovranità e la sicurezza dei dati sensibili dei cittadini.

In materia di tutela dei minori e di consenso informato, il testo introduce specifiche tutele nell'accesso alle tecnologie di IA: per i minori di 14 anni, l'accesso a tali tecnologie richiede il consenso di chi esercita la responsabilità genitoriale, mentre le persone di minore età

tra i 14 e i 18 anni potranno esprimere autonomamente il proprio consenso per il trattamento dei dati personali, purché le informazioni siano facilmente accessibili e comprensibili.

Il disegno di legge contiene poi numerose puntuali disposizioni in materia di sanità e disabilità, ricerca scientifica, sviluppo economico e competitività, lavoro, pubblica amministrazione[37]. Per quanto riguarda l'impiego dei sistemi di intelligenza artificiale nell'attività giudiziaria, sono sempre riservate al magistrato le decisioni che riguardano l'interpretazione e l'applicazione della legge, la valutazione dei fatti e delle prove nonché l'adozione dei provvedimenti. Per questo tipo di decisioni, che costituiscono il nucleo fondamentale e più sensibile dell'esercizio della giurisdizione, viene esclusa pertanto qualsiasi possibilità di fare ricorso all'intelligenza artificiale. L'autorizzazione alla sperimentazione e l'impiego dei sistemi di intelligenza artificiale negli uffici giudiziari ordinari è concessa dal Ministero della giustizia, che provvede sentite le autorità nazionali cui sono demandate le attività di controllo e sorveglianza (l'Agenzia per l'Italia digitale e l'Agenzia per la cybersicurezza).

Con specifico riguardo al settore della sicurezza nazionale e della difesa, il provvedimento esclude dall'ambito applicativo delle nuove regole le attività connesse ai sistemi e modelli di intelligenza artificiale predisposti per finalità di sicurezza nazionale, di cybersicurezza e di difesa. Si tratta delle attività promosse dal Dipartimento delle informazioni per la sicurezza (DIS), dall'Agenzia informazioni e sicurezza esterna (AISE), dall'Agenzia informazioni e sicurezza interna (AISI), ma anche dalle Forze di polizia laddove dirette alla tutela della sicurezza nazionale (l'art. 6 del disegno di legge richiama espressamente le operazioni ccdd sotto copertura condotte ai sensi della legge n. 146/2006). A fronte dell'esclusione dall'ambito applicativo della disciplina, rimane comunque fermo l'obbligo del rispetto dei diritti fondamentali e delle libertà previste dalla Costituzione nonché dello svolgimento con metodo democratico della vita istituzionale e politica.

---

37 Viene definita la *governance* italiana sull'intelligenza artificiale, con disposizioni in materia di Strategia nazionale per l'intelligenza artificiale. La Strategia deve essere approvata con cadenza almeno biennale dal Comitato interministeriale per la transizione digitale (CITD). La Strategia nazionale deve mirare a favorire la collaborazione tra le amministrazioni pubbliche e i soggetti privati relativamente allo sviluppo e all'adozione di sistemi di intelligenza artificiale, coordinare l'attività della pubblica amministrazione in materia, promuovere la ricerca e la diffusione della conoscenza in materia di intelligenza artificiale, indirizzare le misure e gli incentivi finalizzati allo sviluppo imprenditoriale e industriale dell'intelligenza artificiale.

Nonostante le previsioni abbastanza puntuali presenti nel Regolamento UE che questo disegno di legge si propone di recepire, al fine di dettare una disciplina ancora più dettagliata l'art. 24 lett. *h)*, contiene una delega al Governo per adottare, entro un anno dall'entrata in vigore del provvedimento di cui si discute, un decreto legislativo volto a prevedere "un'apposita disciplina per l'utilizzo di sistemi di intelligenza artificiale per l'attività di polizia". Si dovrà dunque attendere l'adozione di quest'ultimo provvedimento (che sarà certamente redatto col contributo di tutti i Ministeri da cui le Forze di polizia nazionali dipendono) per avere un quadro chiaro degli esatti spazi di manovra che saranno concessi alle agenzie nazionali di *law enforcement* per le attività di indagine criminale condotte con l'aiuto dell'IA[38].

2.3. Linee guida per l'adozione, l'acquisto, lo sviluppo di sistemi di intelligenza artificiale nella pubblica amministrazione

Dal 18 febbraio al 30 marzo 2025 sono state in consultazione pubblica le linee guida per l'adozione dell'IA nella pubblica amministrazione italiana. Queste ultime, che saranno emanate dall'Agenzia per l'Italia digitale (AgID) hanno come principale obiettivo quello di fornire un modello per l'adozione di sistemi *IA-driven* all'interno delle amministrazioni pubbliche, nel rispetto dei princìpi di conformità, etica, qualità, affidabilità, innovazione, sostenibilità, formazione e organizzazione delle risorse.

Il documento, in particolare, si inserisce nel quadro del Codice dell'Amministrazione Digitale (CAD), è stato emanato ai sensi del d.p.c.m. 12 gennaio 2024, e contiene alcuni princìpi chiave per l'utilizzo dell'IA:

- *governance* etica: l'IA deve rispettare valori fondamentali come trasparenza, equità e diritti umani;
- classificazione dei rischi: i sistemi di IA vengono valutati in base al livello di rischio per garantire sicurezza e affidabilità;
- protezione dei dati personali: il rispetto del GDPR è centrale nell'implementazione di soluzioni IA;
- sicurezza cibernetica: viene data grande importanza alla protezione contro minacce come attacchi informatici.

---

38 Tali "spazi di manovra" non potranno certamente travalicare i limiti dettati sul punto dal Regolamento UE, richiamati al precedente paragrafo 2.1.

Per conformarsi a questi princìpi, le amministrazioni pubbliche sono chiamate a definire chiaramente le aree prioritarie di applicazione dell'IA, individuando o creando strutture organizzative dedicate alla gestione dei relativi progetti. In tale quadro, la formazione del personale sarà cruciale e determinante per garantire un uso efficace delle tecnologie IA; per questo le pubbliche amministrazioni sono incoraggiate a investire risorse nello sviluppo delle competenze necessarie per utilizzare e gestire tali innovativi sistemi.

Capitolo III
L'intelligenza artificiale e la protezione dei dati personali. Profili problematici in materia di trattamento per finalità di polizia e ragioni di giustizia

di Daniela Gregorio*

3.1. Inquadramento valoriale: la dignità della persona come fondamento della coeva regolazione normativa in materia digitale

«Ho pensato di prendere il nome di Leone XIV. Diverse sono le ragioni, però principalmente perché il Papa Leone XIII, con la storica Enciclica *Rerum novarum*, affrontò la questione sociale nel contesto della prima grande rivoluzione industriale; e oggi la Chiesa offre a tutti il suo patrimonio di dottrina sociale per rispondere a un'altra rivoluzione industriale e agli sviluppi dell'intelligenza artificiale, che comportano nuove sfide per la difesa della dignità umana, della giustizia e del lavoro[39]». Il 10 maggio 2025, a due giorni dalla sua elezione, il Pontefice Leone XIV, nel suo discorso al collegio cardinalizio, ha utilizzato esattamente queste parole per spiegare le ragioni che lo hanno indotto a scegliere di instaurare una strada di continuità con un Papa eletto nel lontano 1878.

La crescente pervasività dell'intelligenza artificiale, che ormai ha assunto caratteri di infiltrazione progressiva, ineludibile ed esponenziale nella vita quotidiana di ogni essere umano, è stata, quindi, l'inaspettata motivazione principale, che ha determinato persino la scelta del nome di un Pontefice della Chiesa cattolica.

* Vice Questore della Polizia di Stato, già frequentatrice del XL Corso di Alta Formazione presso la Scuola di Perfezionamento per le Forze di Polizia.

39 https://www.vatican.va/content/leo-xiv/it/speeches/2025/may/documents/20250510-collegio-cardinalizio.html.

Gli sviluppi tecnologici ed informatici coevi, infatti, sono diventati uno dei temi principali della riflessione filosofica e giuridica, non soltanto per i prevedibili e significativi problemi di occupazione derivanti dal loro impiego, ma anche per via dei potenziali impatti negativi sulla società contemporanea e sulle singole persone, in special modo in presenza di decisioni automatizzate o laddove i dati raccolti siano relativi all'orientamento sessuale, alle convinzioni religiose o ad altre categorie di dati particolari, che possono alimentare *bias* negativi. Per *bias* si intende ogni ragione che porta un algoritmo a un risultato diverso da quello cui dovrebbe condurre il suo normale funzionamento, vale a dire un ragionamento basato sull'utilizzo di un'intelligenza 'analogica' e dati oggettivi, correttamente letti e impiegati; in sostanza si concretizza in un trasferimento di pregiudizi all'interno di un *software*, mediante dati di addestramento che, in un secondo momento, creano modelli di riferimento dannosi, che, per giunta, tendono ad amplificarsi nel tempo, essendo potenzialmente capaci di creare veri e propri focolai discriminatori.

Sicuramente l'utilizzo di un sistema di calcolo così potenziato è un'opportunità che non può essere persa per le molteplici possibilità di uso, che possono migliorare le condizioni di vita e che, d'altronde, oramai sono già ampiamente sfruttate, anche per le finalità di polizia e per ragioni di giustizia; tuttavia, è necessario «garantire che i diritti umani siano rafforzati e non minati dall'intelligenza artificiale [e ciò] è uno dei fattori chiave che definiranno il mondo in cui viviamo[40]».

Si può senz'altro affermare, dunque, che, alla luce del principio personalista del costituzionalismo contemporaneo, la dignità umana deve essere posta al centro di qualsivoglia valutazione etica e sociale e deve essere il faro e il fondamento della disciplina giuridica, di tutto l'impianto normativo statale e sovranazionale, per l'azione e la tutela giuridica, che va garantita ai singoli[41].

---

40 Consiglio d'Europa, *Unboxing artificial intelligence: 10 steps to protect human rights, 14 maggio 2019*, in cui si afferma l'esigenza di prevedere una valutazione dell'impatto sui diritti umani (*Human rights impact assessments*, HRIA) da parte delle Autorità pubbliche che acquisiscono e sviluppano sistemi di IA, con particolare attenzione per i gruppi di persone che registrano un maggiore rischio di impatto dal ricorso all'IA.

41 A livello internazionale, il preambolo della Carta delle Nazioni Unite riafferma la fede dei popoli «nei diritti fondamentali dell'uomo, nella dignità e nel valore della persona umana»; il principio della dignità della persona umana riecheggia anche nel Preambolo e nell'articolo 1 della Dichiarazione universale dei diritti umani e nella Convenzione europea dei diritti dell'uomo (Cedu). Nell'ambito dell'Unione europea, la dignità personale

Si impone, pertanto, l'esigenza di individuare opportuni princìpi di *governance* dell'IA, come, ad esempio, la necessità che sia assicurata sempre la supervisione umana, ma è anche indispensabile porsi nell'ottica di un utilizzo che tuteli la sicurezza, la trasparenza, il benessere sociale e ambientale, la diversità, la non discriminazione, l'equità e la *riservatezza dei dati personali*.

Il legislatore europeo, in effetti, ha dedicato una particolare attenzione all'intangibilità della dignità personale, ponendo tale valore al centro di due strumenti normativi, che sono cruciali nella contemporanea regolazione del fenomeno digitale, vale a dire il Regolamento (UE) 2016/679 del Parlamento Europeo e del Consiglio del 27 aprile 2016 (cd *General Data Protection Regulation* – GDPR), relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, e il Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale (cd *AI Act*).

In un mondo in cui le nostre identità sono "fatte di dati", la protezione dei dati personali diventa, infatti, strumentale alla difesa irrinunciabile della dignità personale: l'obiettivo perseguito dal legislatore europeo (sia con il GDPR, sia con l'*AI Act*) è stato proprio quello di trovare un punto di equilibrio tra protezione della dignità personale e sostegno all'innovazione tecnologica, vale a dire «mantenere l'uomo al centro dello sviluppo tecnologico, promuovere l'umanesimo digitale e incoraggiare uno sviluppo antropocentrico dell'IA[42]».

3.2. La disciplina generale della protezione dei dati personali. Il *General Data Protection Regulation*

Il GDPR costituisce l'asse portante del nuovo sistema di regole sulla protezione dei dati personali, applicabile a qualsiasi trattamento[43], sia

---

permea il dettato della Carta dei Diritti Fondamentali dell'Unione europea: nel preambolo viene affermato che l'Unione si fonda sui valori indivisibili di dignità umana, di libertà, di uguaglianza e di solidarietà.

42 Cfr. Ferrina Ceroni G., *La dignità della persona: tra GDPR E AI ACT*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024, "Intelligenza artificiale: impiego e implicazioni sotto il profilo tecnologico, etico e giuridico", 2024.

43 A tal proposito, l'art. 4, paragrafo 1, punto 2), GDPR stabilisce che il trattamento debba intendersi come «qualsiasi operazione o insieme di operazioni, compiute con o senza l'ausilio di processi automatizzati e applicate a dati personali o insiemi di dati personali,

nel settore pubblico, sia nel settore privato. La sua entrata in vigore[44] ha innalzato i livelli di protezione delle persone fisiche rispetto al trattamento dei dati personali e responsabilizzato i titolari dei trattamenti stessi. Il legislatore europeo è stato ben consapevole del fatto che la disciplina della protezione dei dati personali involve la tutela di tutti i diritti e le libertà fondamentali riconosciuti alle persone fisiche, come, ad esempio, il diritto al rispetto della vita privata e familiare, la libertà di pensiero, la libertà d'informazione, ecc., che dalla protezione dei dati ottengono tutela mediata.

All'interno del GDPR sono enucleati i princìpi, che devono essere rispettati in ogni fase dello sviluppo del trattamento mediante l'attuazione da parte dei titolari di concrete misure tecniche ed organizzative e l'adempimento di specifici obblighi. In particolare, il titolare del trattamento deve garantire che lo stesso sia effettuato in modo «lecito[45], corretto[46]

---

come la raccolta, la registrazione, l'organizzazione, la strutturazione, la conservazione, l'adattamento o la modifica, l'estrazione, la consultazione, l'uso, la comunicazione mediante trasmissione, diffusione o qualsiasi altra forma di messa a disposizione, il raffronto o l'interconnessione, la limitazione, la cancellazione o la distruzione».

44 Tale atto normativo è divenuto obbligatorio in tutti i suoi elementi e direttamente applicabile negli Stati membri dell'Unione europea dal 25 maggio 2018.

45 È lecito un trattamento che non violi norme generali e norme specifiche dell'ordinamento giuridico, pertanto è necessario o il *consenso dell'interessato* per una o più specifiche finalità o un'altra base legittima prevista dalla legislazione, vale a dire le condizioni tassative, collocate sul medesimo livello di importanza rispetto al consenso, disciplinate *ex* art. 6 GDPR; queste ultime si configurano quando il trattamento deve essere strumentale *all'esecuzione di un contratto o di misure precontrattuali* adottate su richiesta dell'interessato o anche all'*adempimento di un obbligo legale al quale è soggetto il titolare del trattamento*; vi può essere anche l'ipotesi in cui il trattamento sia necessario alla *salvaguardia degli interessi vitali dell'interessato* o di un'altra persona fisica oppure sia connesso *all'esecuzione di un compito di interesse pubblico o all'esercizio di pubblici poteri di cui è investito il titolare del trattamento*; vi è, infine, l'ultimo caso in cui il trattamento sia dovuto ai fini del *perseguimento del legittimo interesse del titolare del trattamento* o di terzi, a condizione che non prevalgano gli interessi o i diritti e le libertà fondamentali dell'interessato, in particolare se quest'ultimo è un minore.
In tema di IA, tuttavia, si pone una significativa problematica in merito al consenso dell'interessato, come causa primaria di liceità al trattamento, poiché, se si considera la natura e il funzionamento dei sistemi di intelligenza artificiale nella fase di istruzione, come nelle ipotesi di *web scraping*, in cui viene effettuata una vera e propria pesca a strascico di informazioni sul *web*, compresi i dati personali, risulta davvero impossibile immaginare una preliminare acquisizione di una manifestazione di volontà esplicita, attiva, basata su una completa informazione circa le finalità, le modalità e i tempi di trattamento.

46 Il principio di *correttezza* implica un concetto di rispetto di norme etiche, deontologiche non "codificate", ovvero osservanza di accorgimenti per rendere equilibrate le singole posizioni, adeguando il trattamento alle esigenze reciproche, oltre lo stretto dato normativo.

e trasparente[47]», che sia garantita la «limitazione della finalità[48]», la «minimizzazione dei dati[49]», l'«esattezza[50]», la «limitazione della conservazione[51]», l'«integrità e riservatezza[52]»; inoltre, del rispetto delle predette modalità il titolare del trattamento ne risponde direttamente sulla base del principio innovativo di responsabilizzazione (*accountability*), concetto che esprime, in una parola, l'attribuzione dell'obbligo giuridico di osservare il Regolamento e le altre fonti giuridiche e dell'onere processuale della prova di averlo fatto.

I predetti princìpi non sono, pertanto, solo delle mere enunciazioni teoriche, ma pongono degli obblighi concreti a carico dei soggetti coinvolti dal trattamento dei dati personali. Si pensi, a titolo esemplificativo, al principio di trasparenza, che si traduce nell'obbligo di rendere informativa agli interessati per renderli consapevoli della circostanza che sia in corso un trattamento di dati che li riguarda e di

---

47 Il concetto di *trasparenza* si concretizza nel dovere di tutelare la consapevolezza dell'interessato; può riferirsi alla tracciabilità del dato a favore dell'interessato, alle informazioni fornite alla persona prima dell'inizio del trattamento, alla facile accessibilità dei dati durante il trattamento, ma anche alle informazioni fornite agli interessati a seguito di una richiesta di accesso ai dati che li riguardano.

48 Il principio di *limitazione della finalità* si esprime nel senso che i dati personali devono essere «raccolti per finalità determinate, esplicite e legittime, e successivamente trattati in modo che non sia incompatibile con tali finalità [...]» (cfr. art. 5, paragrafo 1, lettera b), GDPR). Da questo deriva che un nuovo scopo che sia emerso *medio tempore* e che non sia compatibile con le finalità iniziali deve avere la propria base giuridica specifica e non può fondarsi sul fatto che i dati fossero stati inizialmente acquisiti o trattati per un'altra finalità legittima.

49 Il principio di *minimizzazione dei dati* implica una riduzione al minimo del numero dei dati, del tipo dei trattamenti, dei soggetti che a vario titolo possono avere contezza dei dati, del periodo di conservazione.

50 Il principio di *esattezza* comporta che i dati personali debbano essere «esatti e, se necessario, aggiornati; devono essere adottate tutte le misure ragionevoli per cancellare o rettificare tempestivamente i dati inesatti rispetto alle finalità per le quali sono trattati» (cfr. art. 5, paragrafo 1, lettera d), GDPR).

51 Il principio di *limitazione della conservazione* determina che i dati personali siano «conservati in una forma che consenta l'identificazione degli interessati per un arco di tempo non superiore al conseguimento delle finalità per le quali sono trattati» (cfr. art. 5, paragrafo 1, lettera e), GDPR). Una volta, quindi, venute meno le finalità per le quali i dati sono stati raccolti e trattati, gli stessi devono essere cancellati.

52 Il principio di *integrità e riservatezza* implica che i dati personali siano «trattati in maniera da garantire un'adeguata sicurezza […], compresa la protezione, mediante misure tecniche e organizzative adeguate, da trattamenti non autorizzati o illeciti e dalla perdita, dalla distruzione o dal danno accidentali» (cfr. art. 5, paragrafo 1, lettera f), GDPR). La sicurezza del trattamento si concretizza proprio in specifici obblighi gravanti sul trattamento del titolare, in modo da minimizzare i rischi di distruzione e perdita del dato (violazione di disponibilità), di modifica indebita (violazione di integrità), di divulgazione o accesso non autorizzati (violazione della riservatezza).

quali diritti possono esercitare. O ancora, la minimizzazione impone ai titolari di effettuare delle scelte operative volte a ridurre, per quanto possibile, la raccolta e l'uso di dati personali, prevenendo prassi di trattamenti a strascico o massivi di dati. Infine, si considerino i princìpi di integrità e riservatezza, i quali da astratti valori si traducono in misure concrete, quali installare un anti-*malware*, mantenere copie di *back-up* dei dati, chiudere un armadio a chiave e così via.

L'efficacia delle norme giuridiche in generale e dei diritti degli interessati in particolare, tuttavia, dipende in larga misura dall'esistenza di meccanismi adeguati alla loro attuazione. Nell'era digitale, il trattamento dei dati si è diffuso capillarmente e diventa sempre più difficile da comprendere per gli individui, soprattutto laddove si faccia ricorso a sistemi implementati con tecniche di intelligenza artificiale. Al fine di ridurre gli squilibri di potere tra gli interessati e i titolari del trattamento, ai singoli sono stati riconosciuti determinati diritti, affinché possano esercitare un maggiore controllo sul trattamento delle loro informazioni personali[53].

Il punto di partenza deve essere individuato nel diritto all'informazione che deve essere garantita dal titolare del trattamento a monte dello stesso e deve essere resa in forma concisa, trasparente, intelligibile e facilmente accessibile, con un linguaggio semplice e chiaro, in particolare nel caso di informazioni destinate ai minori. Il diritto all'autodeterminazione informativa si declina, poi, concretamente nell'attribuzione al soggetto interessato dal trattamento di un novero di altri diritti azionabili in relazione ai propri dati personali, come l'accesso, la rettifica, la cancellazione, l'opposizione, la limitazione di trattamento e la portabilità dei dati (artt. 15 a 22, GDPR) e sono funzionali all'esercizio di altri diritti e libertà fondamentali di rango costituzionale[54].

Le persone fisiche sono anche tutelate rispetto ai processi decisionali automatizzati: a mente dell'art. 22 GDPR, infatti, 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

---

53 Così *Manuale del diritto europeo in materia di protezione dei dati*, *Publications Office of the European Union*, 2018, p. 225.

54 Solo a titolo di esempio: accesso alla propria cartella clinica per la tutela del proprio diritto alla salute, la richiesta di cancellazione e rettifica dei dati personali allo scopo di proteggere l'identità, la riservatezza o la reputazione della persona fisica lesa da una notizia pubblicata *online*.

significativamente sulla sua persona. Il titolare del trattamento deve garantire all'interessato il rispetto del diritto di ottenere l'intervento umano correttivo, di esprimere la propria opinione e di contestare la decisione e ciò assume particolare importanza in relazione alle prassi sempre più invasive dei trattamenti automatizzati, della profilazione e degli usi dell'IA. Sul punto, il Garante per la protezione dei dati personali si è espresso più volte, garantendo sempre l'applicazione di tre fondamentali princìpi: 1) la conoscibilità dei processi decisionali automatizzati, 2) la non esclusività della decisione algoritmica e 3) la non discriminazione algoritmica.

Come già precisato, a carico del titolare del trattamento e degli altri soggetti eventualmente coinvolti[55] gravano alcuni adempimenti, tipizzati dalla disciplina unionale, quali l'obbligo di tenuta di un registro delle attività di trattamento, l'adozione di misure tecniche e organizzative adatte a garantire un livello di sicurezza adeguato al rischio e il dovere di notificare l'avvenuta violazione dei dati personali all'Autorità di controllo e di darne comunicazione all'interessato.

Particolarmente importante è, inoltre, l'obbligo di svolgimento della valutazione d'impatto sulla protezione dei dati, che deve intervenire quando un tipo di trattamento, allorché preveda, in particolare, l'uso di nuove tecnologie, considerati la natura, l'oggetto, il contesto e le finalità del trattamento, può presentare un rischio elevato per i diritti e le libertà delle persone fisiche; il titolare del trattamento deve effettuare, in questo caso, prima di procedere allo stesso, una valutazione dell'impatto dei trattamenti previsti sulla protezione dei dati personali. Quest'adempimento si inserisce nel nuovo principio della responsabilizzazione del titolare. La *ratio* di quest'adempimento risiede nella necessità che sia garantito un *surplus* di ponderazione e di giustificazione a fronte di trattamenti ritenuti più pericolosi di altri[56].

---

55 Si fa riferimento alle figure del responsabile del trattamento (art. 28, GDPR) e del responsabile della protezione dei dati personali (artt. da 37 a 39, GDPR), sui quali gravano profili di responsabilità anche autonoma.

56 L'art. 35, paragrafo 3, GDPR richiede, in particolare, una valutazione nelle ipotesi di:
a) trattamenti automatizzati, compresa la profilazione, di aspetti personali relativi a persone fisiche condotto in modo sistematico e globale;
b) il trattamento, su larga scala, di categorie particolari di dati personali o di dati relativi a condanne penali, ai reati e alle connesse misure di sicurezza;
c) la sorveglianza sistematica su larga scala di una zona accessibile al pubblico (ad es. la videosorveglianza cittadina).

Oltre alle ipotesi, meramente esemplificative, già previste dalla normativa, sulla base di quanto stabilito dall'art. 35, paragrafo 4, GDPR[57], il Garante ha adottato il provvedimento n. 467 dell'11 ottobre 2018, con il quale ha diffuso un elenco delle tipologie di trattamenti soggetti al requisito di una valutazione d'impatto sulla protezione dei dati. Si tratta di un elenco che non ha pretese di esaustività, adottato sulla base di quanto contenuto nelle «Linee guida in materia di valutazione d'impatto sulla protezione dei dati e determinazione della possibilità che il trattamento possa presentare un rischio elevato ai fini del regolamento (UE) 2016/679», approvate il 4 aprile 2017 dal "Gruppo di lavoro articolo 29 per la protezione dei dati".

La valutazione deve essere eseguita prima di procedere al trattamento; nei casi in cui il titolare non sia in grado di individuare misure idonee ad attenuare il rischio, residuando un "rischio elevato", non è possibile avviare il trattamento ed è necessaria la consultazione preventiva del Garante, il quale, qualora ritenga che lo stesso non sia conforme alla normativa di riferimento fornisce un parere scritto.

Nello specifico, se il Garante giunga alla conclusione che la valutazione sia stata condotta correttamente ma, ciò nonostante, permanga un rischio residuo, vieterà il trattamento. Se, viceversa, si determini nel senso di una valutazione non condotta correttamente o che le misure di minimizzazione siano insufficienti, può indicare dei rimedi per attenuare i rischi residui e ricondurli a un livello accettabile. È previsto, peraltro, un termine di otto settimane (sei per i trattamenti per finalità di polizia) prorogabile di ulteriori sei settimane (quattro nell'ipotesi disciplinate dal D.Lgs. 51/2018), trascorsi i quali e in assenza di una comunicazione, si intende reso un parere positivo, sebbene siano ammessi pareri tardivi negativi.

3.2.1. Le limitazioni all'uso di dati biometrici e di categorie particolari

Particolari condizioni di liceità sono previste, poi, per le *"categorie particolari di dati personali"*[58] *per i quali vige un divieto di trattamento*,

---

57 Art. 35, paragrafo 4, GDPR: «L'autorità di controllo redige e rende pubblico un elenco delle tipologie di trattamenti soggetti al requisito di una valutazione d'impatto sulla protezione dei dati ai sensi del paragrafo 1 […]».

58 Si fa riferimento a dati personali che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche o l'appartenenza sindacale, nonché i dati genetici, dati

a meno che non si verifichino una serie tassativa di "eccezioni" (art. 9, GDPR), tra le quali si possono citare: l'interessato ha prestato il proprio consenso esplicito al trattamento; qualora i dati personali siano stati da lui resi manifestamente pubblici; per il perseguimento di finalità politiche, filosofiche, religiose o sindacali da parte di una fondazione, associazione o altro organismo senza scopo di lucro; per assolvere gli obblighi ed esercitare i diritti specifici del titolare del trattamento o dell'interessato in materia di diritto del lavoro e della sicurezza sociale e protezione sociale; per tutelare un interesse vitale dell'interessato o di un'altra persona fisica, qualora egli si trovi nell'incapacità fisica o giuridica di prestare il proprio consenso; per accertare, esercitare o difendere un diritto in sede giudiziaria o ogniqualvolta le autorità giurisdizionali esercitino le loro funzioni giurisdizionali; per motivi di interesse pubblico rilevante sulla base del diritto dell'Unione o degli Stati membri; per finalità di medicina preventiva o di medicina del lavoro, valutazione della capacità lavorativa del dipendente, diagnosi, assistenza o terapia sanitaria o sociale ovvero gestione dei sistemi e servizi sanitari o sociali[59]; per motivi di interesse pubblico nel settore della sanità pubblica; per il perseguimento di fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici.

3.3. Il trattamento dei dati personali per finalità di polizia e per ragioni di giustizia. La direttiva (UE) 2016/680

Per quanto concerne i trattamenti operati dalle pubbliche amministrazioni, è stato già messo in luce che tra i presupposti di liceità il GPDR include anche "l'esecuzione di un compito di interesse pubblico o connesso all'esercizio di pubblici poteri" di cui è investito il titolare del trattamento, entro i limiti in cui esso si configuri come necessario a tali fini; così facendo, il Regolamento – in quanto immediatamente efficace in tutti gli Stati membri dell'Unione europea – individua una condizione, un presupposto per il trattamento dei dati personali

---

biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona (art. 9, comma 1, GDPR).

59 Questa eccezione si applica soltanto se tali dati sono trattati da o sotto la responsabilità di un professionista soggetto al segreto professionale conformemente al diritto dell'Unione o degli Stati membri o alle norme stabilite dagli organismi nazionali competenti o da altra persona anch'essa soggetta all'obbligo di segretezza conformemente al diritto dell'Unione o degli Stati membri o alle norme stabilite dagli organismi nazionali competenti.

che risulta autosufficiente, nella misura in cui siano soddisfatte le prescrizioni poste dalla medesima fonte, quali il rispetto dei princìpi di cui all'art. 5 e delle altre disposizioni rilevanti. L'attività di trattamento compiuta dall'amministrazione è connotata, quindi, dalla presenza di tratti autoritativi, potendo il soggetto pubblico agire indipendentemente dal consenso del soggetto interessato e operare anche una significativa compressione dei diritti dei soggetti interessati, nei casi e con le modalità che si analizzeranno in seguito.

Il d.l. 8 ottobre 2021, n. 139, inoltre, ha provveduto ad una risistemazione della disciplina dei trattamenti operati nello svolgimento di compiti di interesse pubblico o connessi all'esercizio di pubblici poteri; sono state introdotte, infatti, diverse modifiche. In primo luogo, la formulazione dell'art. 2-*ter*, comma 1, Codice della *privacy* è stata aggiornata, prevedendo che la base giuridica rilevante per i trattamenti relativi all'esecuzione di compiti a rilevanza pubblicistica possa essere fondata non solo su una norma di legge o di regolamento, ma *anche su atti amministrativi generali*. In secondo luogo, è stato introdotto un nuovo comma 1-*bis* nel medesimo art. 2-*ter*, Codice della *privacy*, che stabilisce che il trattamento dei dati personali da parte delle pubbliche amministrazioni e soggetti ad esse parificati «è anche consentito se necessario per l'adempimento di un compito svolto nel pubblico interesse o per l'esercizio di pubblici poteri a essa attribuiti» nel rispetto dell'art. 6 GDPR. Si precisa, quindi, che, se la finalità del trattamento non è specificata nelle norme, i medesimi soggetti la individuano, dandone adeguata pubblicità.

Per quanto concerne la disciplina giuridica applicabile ai trattamenti dei dati personali effettuati per "finalità di giustizia penale" e per "finalità di polizia", occorre richiamare dapprima l'art. 23 GDPR, che prevede che l'applicazione di talune disposizioni del regolamento, relative ai diritti degli interessati e agli obblighi dei titolari del trattamento, possa essere limitata, sebbene tali limitazioni dovrebbero essere considerate eccezioni rispetto alla disciplina generale, quindi dovrebbero essere interpretate in modo restrittivo e applicate esclusivamente in circostanze specifiche e solo se sono soddisfatte determinate condizioni.

Le limitazioni all'attuazione della disciplina contenuta nel GDPR trovano giustificazione, in particolare, per salvaguardare:

a) la sicurezza nazionale;
b) la difesa;
c) la sicurezza pubblica;

d) la prevenzione, l'indagine, l'accertamento e il perseguimento di reati o l'esecuzione di sanzioni penali, incluse la salvaguardia contro e la prevenzione di minacce alla sicurezza pubblica;
e) altri importanti obiettivi di interesse pubblico generale dell'Unione o di uno Stato membro, in particolare un rilevante interesse economico o finanziario dell'Unione o di uno Stato membro, anche in materia monetaria, di bilancio e tributaria, di sanità pubblica e sicurezza sociale;
f) la salvaguardia dell'indipendenza della magistratura e dei procedimenti giudiziari;
g) le attività volte a prevenire, indagare, accertare e perseguire violazioni della deontologia delle professioni regolamentate;
h) una funzione di controllo, d'ispezione o di regolamentazione connessa, anche occasionalmente, all'esercizio di pubblici poteri;
i) la tutela dell'interessato o dei diritti e delle libertà altrui;
j) l'esecuzione delle azioni civili.

L'art. 23 GDPR prescrive, in ogni caso, che debba essere garantito il rispetto dell'essenza dei diritti e delle libertà fondamentali, che la misura sia necessaria e proporzionata, in una società democratica, per salvaguardare, tra l'altro, importanti obiettivi di interesse pubblico generale dell'Unione o di uno Stato membro e che venga assicurato il rispetto di alcune disposizioni di carattere generale, compreso il diritto degli interessati ad essere informati della limitazione, a meno che ciò possa compromettere la finalità stessa; anche in situazioni eccezionali, la protezione dei dati personali non può essere limitata nella sua interezza, ma deve essere mantenuta in tutte le misure di emergenza, di cui all'articolo 23 GDPR, contribuendo così al rispetto dei valori generali di democrazia e Stato di diritto, nonché dei diritti fondamentali sui quali si fonda l'Unione[60].

È stabilito, inoltre, il requisito dell'esistenza di una misura legislativa, che impone la vigenza di una norma unionale o nazionale, che legittimi il ricorso alla disciplina dell'art. 23 GDPR (a titolo temporaneo o permanente, a seconda dei presupposti di applicazione), non essendo sufficiente una valutazione d'iniziativa del titolare del trattamento della sussistenza dei requisiti fissati dal GDPR.

Oltre a questa disciplina derogatoria del regime comune, bisogna considerare che, qualora debba essere effettuato un trattamento a

60 Cfr. Linee guida 10/2020 sulle limitazioni a norma dell'articolo 23 del regolamento generale sulla protezione dei dati, adottate il 13 ottobre 2021, dall'*European data protection board* (EDPB - Comitato europeo per la protezione dei dati), p. 6.

fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, incluse la salvaguardia e la prevenzione di minacce alla pubblica sicurezza, la normativa di riferimento non è il GDPR, bensì la direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio del 27 aprile 2016[61], così come attuata dal D.Lgs. 18 maggio 2018, n. 51 nell'ordinamento giuridico nazionale.

La direttiva sulla protezione dei dati, destinata alla polizia e alle autorità giudiziarie penali, tutela i dati personali relativi a diverse categorie di interessati coinvolti in procedimenti penali, quali testimoni, persone informate dei fatti, vittime, indiziati e complici[62]. Le autorità di polizia e le autorità giudiziarie penali sono tenute a rispettare le disposizioni della direttiva quando trattano tali dati per finalità di applicazione della legge.

L'applicabilità della direttiva si estende sia al trattamento dei dati a livello nazionale e al trattamento in ambito transfrontaliero tra polizia e autorità giudiziarie degli Stati membri, sia ai trasferimenti internazionali da parte delle autorità competenti verso Paesi terzi e organizzazioni internazionali. Non si applica al trattamento dei dati personali da parte di istituzioni, organi, uffici e agenzie dell'U.E.

Nello specifico sistema tracciato per i trattamenti "di polizia" deve, tuttavia, essere considerato che durante il predetto trattamento il dato può cambiare la propria natura nel corso del proprio 'ciclo vitale', potendo essere acquisito in un procedimento penale e trattato dall'Autorità giudiziaria o dal Pubblico Ministero secondo le norme del codice di procedura penale e non più della direttiva 'LED' e del successivo decreto di attuazione.

Ciò premesso, occorre precisare che il principio generale è che il trattamento di dati genetici, dati personali concernenti reati, procedimenti e condanne penali e qualsiasi misura di sicurezza correlata, dati biometrici che identificano in modo univoco una persona, nonché

---

61 Nota come direttiva 'LED' (*Law Enforcement Directive*) o anche "Direttiva di polizia".

62 Vds. art. 6, direttiva 'LED', che statuisce: «Gli Stati membri dispongono che il titolare del trattamento […] operi una chiara distinzione tra i dati personali delle diverse categorie di interessati, quali: a) le persone per le quali vi sono fondati motivi di ritenere che abbiano commesso o stiano per commettere un reato; b) le persone condannate per un reato; c) le vittime di reato o le persone che alcuni fatti autorizzano a considerare potenziali vittime di reato, e d) altre parti rispetto a un reato, quali le persone che potrebbero essere chiamate a testimoniare nel corso di indagini su reati o di procedimenti penali conseguenti, le persone che possono fornire informazioni su reati o le persone in contatto o collegate alle persone di cui alle lettere a) e b)».

eventuali dati personali 'sensibili', è consentito solo qualora esistano garanzie adeguate alla prevenzione dei rischi che il trattamento di tali dati può comportare per gli interessi, i loro diritti e libertà fondamentali, segnatamente il rischio di discriminazione.

L'art. 1, paragrafo 2, chiarisce, infatti che «[…] gli Stati membri: a) tutelano i diritti e le libertà fondamentali delle persone fisiche, in particolare il diritto alla protezione dei dati personali; b) garantiscono che lo scambio dei dati personali da parte delle autorità competenti all'interno dell'Unione, qualora tale scambio sia richiesto dal diritto dell'Unione o da quello dello Stato membro, non sia limitato né vietato per motivi attinenti alla protezione delle persone fisiche con riguardo al trattamento dei dati personali». È evidente, quindi, che in questo settore è richiesto alle autorità pubbliche, competenti nelle materie sottoposte alla disciplina *de qua*, un contemperamento tra i diritti e libertà personali delle persone fisiche e il perseguimento delle finalità istituzionali di prevenzione, indagine e repressione dei reati e di prevenzione di minacce alla pubblica sicurezza, che possono richiedere lo scambio di dati personali all'interno dell'Unione, di per sé non limitabile o escludibile per ragioni attinenti alla protezione dei dati personali.

La direttiva si basa, in larga misura, sui princìpi e sulle definizioni contenuti nel regolamento generale sulla protezione dei dati, tenendo conto della specificità dei settori della polizia e della giustizia penale. I princìpi applicabili al trattamento per finalità di polizia, infatti, sono sovrapponibili a quelli previsti dal GDPR; pertanto, la raccolta di dati personali deve essere effettuata per quanto è strettamente necessario per la prevenzione di un pericolo concreto o per la repressione di un reato specifico. Ogni ulteriore raccolta di dati dovrebbe basarsi su una legislazione nazionale specifica. Il trattamento dei dati appartenenti alle cosiddette categorie particolari, inoltre, dovrebbe essere limitato allo stretto necessario nel contesto di una particolare indagine.

Rispetto al trattamento dei dati a fini commerciali, che è disciplinato dal regolamento, il trattamento di dati connesso alla sicurezza, può richiedere un certo livello di flessibilità. Per esempio, garantire agli interessati lo stesso livello di protezione in termini di diritto di informazione, accesso o cancellazione dei propri dati personali potrebbe significare che qualsiasi operazione di controllo effettuata a fini di contrasto diventerebbe inefficace nel contesto dell'applicazione della legge. La direttiva non contiene, pertanto, il principio di trasparenza. In modo analogo, è necessario che anche i princìpi della minimizzazione dei dati e della limitazione della finalità, che impongono di limitare

i dati personali solo a quanto necessario rispetto alle finalità per le quali sono trattati e che siano trattati per finalità legittime e specifiche, vengano applicati in modo flessibile nei trattamenti di dati relativi alla sicurezza. Le informazioni raccolte e conservate dalle autorità competenti per un caso specifico possono rivelarsi, infatti, estremamente utili per la risoluzione di casi futuri[63].

Per quanto concerne i diritti esercitabili dagli interessati, anche nel campo del trattamento per finalità di polizia è riconosciuta una serie di diritti analoga a quella contemplata dal regolamento, cui tuttavia la direttiva 'LED' (e il decreto legislativo di attuazione) riserva una regolamentazione peculiare e differenziata, determinata dalle evidenti esigenze di segretezza correlate allo svolgimento dell'attività di polizia giudiziaria e di sicurezza, nonché di giustizia penale.

Il diritto di informazione, quindi, si caratterizza per la sua generalità, nel senso che viene garantita a favore di una collettività imprecisata di soggetti, mediante il ricorso anche a pubblicazioni su siti *internet*, inoltre, deve comprendere un insieme di dati più ristretto rispetto al modello di informativa previsto dal GDPR.

Inoltre, il diritto di accesso, che si sostanzia sia nella conferma dell'esistenza di un trattamento in corso, sia nella possibilità di ottenere la comunicazione di una serie dettagliata di dati e informazioni, può essere ritardato, limitato o totalmente escluso, nella misura e per il tempo in cui ciò costituisca uno strumento necessario e proporzionato, tenuto conto dei diritti fondamentali e dei legittimi interessi della persona fisica interessata, per la tutela del buon esito dell'attività di prevenzione, indagine, accertamento e perseguimento di reati o dell'esecuzione di sanzioni penali, nonché per l'applicazione delle misure di prevenzione personali e patrimoniali e delle misure di sicurezza, per la tutela della sicurezza pubblica, della sicurezza nazionale e dei diritti e delle libertà altrui (art. 14, comma 2, D.Lgs. 51/18).

In relazione agli adempimenti, la direttiva 'LED' prevede che le autorità competenti devono tenere un *registro* delle operazioni di

---

63 L'applicazione 'flessibile' dei vari princìpi ha comportato che è stato stabilito nel corso degli anni dalla normativa di settore e dalla giurisprudenza che quando i dati personali sono raccolti all'insaputa dell'interessato, quest'ultimo deve esserne informato non appena la divulgazione dei dati non arrechi più pregiudizio alle indagini.
È stato, poi, chiarito che, all'atto di conservare dati personali, deve essere operata una chiara distinzione tra i dati amministrativi e i dati di polizia, i dati personali di diverse categorie di interessati, quali sospettati, detenuti, vittime e testimoni e fra i dati basati sui fatti reali e quelli basati su sospetti o deduzioni.

trattamento che effettuano nei sistemi di trattamento automatizzato (art. 25). Le registrazioni delle consultazioni e delle comunicazioni devono consentire di stabilire la data e l'ora di tali operazioni, la motivazione e, nella misura del possibile, l'identificazione della persona che ha consultato il sistema o comunicato i dati personali, e i destinatari dei dati personali in questione. Le registrazioni devono essere usate ai soli fini della verifica della liceità del trattamento, dell'autocontrollo, per garantire l'integrità e la sicurezza dei dati personali e nell'ambito di procedimenti penali.

Quando un trattamento può presentare un rischio elevato per i diritti delle persone, a causa, per esempio, dell'uso di nuove tecnologie, i titolari del trattamento devono effettuare una *valutazione dell'impatto* sulla protezione dei dati, prima di procedere al trattamento, come previsto nel GDPR.

Il trasferimento o la comunicazione a livello internazionale, infine, dovrebbero essere limitati alle autorità di polizia straniere ed essere basati su disposizioni giuridiche specifiche, possibilmente accordi internazionali, a meno che non siano necessari per prevenire un pericolo grave e imminente.

3.4. Rapporto tra *AI Act* e normativa in materia di trattamento dei dati personali

Lo sviluppo dell'intelligenza artificiale pone indubbiamente dei motivi di contrasto con le norme sopra tratteggiate in materia di protezione dei dati personali. Un aspetto particolarmente problematico è il procedimento che viene usato nella creazione e nell'addestramento dei sistemi in analisi: l'intelligenza artificiale elabora le risposte a partire dai dati che vengono forniti dal produttore e, in taluni casi, l'addestramento viene auto sviluppato dal sistema mediante l'acquisizione e l'elaborazione di una quantità significativa di dati, che costituisce il punto di partenza e di crescita di questo genere di tecnologia.

L'elemento, infatti, che distingue l'intelligenza artificiale dai cosiddetti sistemi domanda-risposta è il fatto che, mentre questi ultimi sono immediatamente pronti ad eseguire la loro funzione (l'esempio classico che viene fatto è quello della calcolatrice), l'IA nasce "stupida" e affinché sia in grado di elaborare risposte testuali o grafiche statisticamente corrette, deve essere addestrata su una grande quantità di dati e informazioni, vale a dire i *training data*. Questi possono com-

prendere qualsiasi tipo di informazione, dalle opere scritte, alle immagini, fino ai dati personali[64].

Tali modalità di funzionamento fanno sorgere un legittimo dubbio circa la compatibilità e il rispetto del quadro normativo in materia di protezione dei dati personali, soprattutto se rientranti nelle categorie particolari, quali, ad esempio, i dati biometrici. Da un lato, il ricorso a una quantità massiva di dati rappresenta un vantaggio, perché porta a un continuo miglioramento nelle prestazioni dei sistemi; dall'altro lato, i *training data* non vengono rivelati dal proprietario dell'intelligenza artificiale e questo causa frizioni legali in una moltitudine di ambiti, dal diritto d'autore alla protezione dei dati personali.

Per tale motivo, il Regolamento europeo sull'intelligenza artificiale e il GDPR si intersecano e si implementano a vicenda per garantire la protezione dei dati personali nell'uso e nello sviluppo dei sistemi di IA e ciò è ancora più vero quando il trattamento dei dati personali può avere un impatto significativo sui diritti e le libertà delle persone. Il Regolamento sull'IA non è, quindi, un atto legislativo *stand-alone*, ma si inserisce nell'ambito di una cornice normativa europea complessa e frastagliata ed è chiamato a coordinarsi con altre discipline proposte dalla Commissione in materia di dati e trasformazione digitale[65].

Nel caso di specie, il Regolamento sull'intelligenza artificiale non si occupa della raccolta e del trattamento dei dati, i quali rimangono, pertanto, regolati dalla normativa vigente e, in particolare, dal GDPR. Trascorso, dunque, il momento di raccolta dei dati, è fondamentale comprendere quale sia la normativa applicabile. Una risposta sembra pervenire dalla Commissione europea, che nella propria relazione esplicativa di accompagnamento del Regolamento sull'IA, chiarisce che quest'ultimo non pregiudica l'applicabilità del GDPR.

D'altronde, sia nella sistematica del GDPR sia in quella dell'*AI Act*, la protezione della dignità umana è l'elemento primario; da questo punto, poi, prendono avvio un insieme di similitudini, a partire dallo strumento normativo scelto dal legislatore europeo, vale a dire la forma del regolamento in luogo della direttiva. La

---

64 Cfr. Capone F. – Iaselli V., *La privacy nell'AI Act e nel DDL italiano sull'intelligenza artificiale: perché è un tema ineludibile*, in *Agenda digitale.eu*, https://www.agendadigitale.eu/sicurezza/privacy/la-privacy-nellai-act-e-nel-ddl-italiano-sullintelligenza-artificiale-perche-e-un-tema-ineludibile/, 2024.

65 Cfr. Falletta P. – Marsano A., *Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza artificiale e GDPR*, in *Rivista italiana di informatica e diritto*, n. 1, 2024, p. 126.

scelta della fonte utilizzata, come è evidente, esprime una vera e propria volontà di creare una base normativa unica per la disciplina della politica europea sul mondo digitale, in linea con la volontà di inaugurare una *governance* dell'innovazione incentrata sul principio del *"one continent, one law"*. Altro elemento comune è il ricorso all'approccio *"risk based"*: se nel GDPR i soggetti coinvolti devono dimostrare la propria conformità al dettato legislativo, secondo il principio dell'*accountability* (responsabilizzazione), nell'*AI Act* è il legislatore europeo che ha l'onere di individuare quali sistemi di intelligenza artificiale siano da considerare ad alto rischio e come tali rischi debbano essere mitigati. Conseguentemente, la staticità dell'approccio introdotto da ultimo pone come criticità quella di aver introdotto un sistema non abbastanza dinamico rispetto ai continui sviluppi dell'intelligenza artificiale.

L'effettività della sua tutela dipende, tuttavia, dall'azione di controllo e vigilanza delle Autorità competenti individuate dai Regolamenti. Con il GDPR il legislatore europeo ha affidato la competenza ad Autorità amministrative indipendenti[66], entità caratterizzate da indipendenza, imparzialità e terzietà nell'esercizio dei relativi compiti ed attività. Tale scelta regolatoria è andata nella direzione di istituire dei presidi reali di tutela del diritto alla protezione dei dati personali.

Nell'*AI Act*, invece, è stata imboccata una strada differente, problematica, nell'ottica di garantire l'effettività della tutela della dignità umana e dei diritti fondamentali dell'individuo, nonostante la potenziale, forte incidenza della materia in esame sulle situazioni giuridicamente inviolabili di cui sono titolari le persone. I rimedi riconosciuti, infatti, sono soltanto due: da un lato, vi è la possibilità di presentare un reclamo dinanzi all'autorità di sorveglianza del mercato per violazioni del Regolamento e, dall'altro, è possibile esercitare, nei confronti del produttore di sistemi di IA ad alto rischio, il diritto a ricevere spiegazioni circa eventuali decisioni assunte sulla base dei risultati del sistema che incidano sui diritti e le libertà del soggetto istante.

Il ruolo di vigilanza delle Autorità garanti per la protezione dei dati personali, dunque, appare critico e non chiaramente definito a livello operativo in alcuni settori strategici. Tra questi rientra una competenza cruciale come la supervisione sui sistemi di IA ad alto rischio usati a fini di attività di contrasto, gestione delle

66 Nel nostro ordinamento la competenza è stata attribuita al Garante per la protezione dei dati personali.

frontiere, giustizia e democrazia[67]. L'articolo 70, *AI Act*, infatti parla di istituzione o designazione di un'autorità di notifica e di un'autorità di vigilanza del mercato, per cui è lasciata agli Stati membri la scelta in merito, mentre in altre ipotesi, come quella appena menzionata, viene individuata l'*Autority* in materia di protezione dei dati personali come competente, ma non vengono definiti con chiarezza i compiti specifici[68].

Tra le ipotesi, in cui viene attribuita *expressis verbis* una competenza in materia di intelligenza artificiale all'Autorità nazionale di protezione dei dati personali, può essere richiamata quella relativa alla ricezione di notifiche sull'uso dei sistemi di identificazione biometrica da remoto *"real time"*: l'articolo 5, paragrafo 4, *AI Act*, stabilisce che «fatto salvo il paragrafo 3[69], ogni uso di un sistema di identificazione biometrica remota «in tempo reale» in spazi accessibili al pubblico a fini di attività di contrasto è notificato alla pertinente autorità di vigilanza del mercato e all'autorità nazionale per la protezione dei dati conformemente alle regole nazionali di cui al paragrafo 5. La notifica contiene almeno le informazioni di cui al paragrafo 6 e non include dati operativi sensibili».

In conclusione, sebbene l'*AI Act* preveda espressamente una clausola di salvaguardia a favore della disciplina in materia di protezione dati personali in relazione ai diritti e agli obblighi stabiliti dal predetto Regolamento 2024/1689[70], è evidente una mancanza di coordinamento tra le due discipline. I problemi di coordinamento normativo, infatti,

---

67 Cfr. art. 74, paragrafo 8, *AI Act*.

68 Si consideri, ad esempio, l'articolo 57, paragrafo 10, *AI Act*, che recita: «Le autorità nazionali competenti garantiscono che, nella misura in cui i sistemi di IA innovativi comportano il trattamento di dati personali o rientrano altrimenti nell'ambito di competenza di altre autorità nazionali o autorità competenti che forniscono o sostengono l'accesso ai dati, le autorità nazionali per la protezione dei dati e tali altre autorità nazionali o competenti siano associate al funzionamento dello spazio di sperimentazione normativa per l'IA e partecipino al controllo di tali aspetti nei limiti dei rispettivi compiti e poteri».

69 Secondo il quale «ogni uso di un sistema di identificazione biometrica remota «in tempo reale» in spazi accessibili al pubblico a fini di attività di contrasto è subordinato a un'autorizzazione preventiva rilasciata da un'autorità giudiziaria o da un'autorità amministrativa indipendente, la cui decisione è vincolante […]».

70 Cfr. art. 2, paragrafo 7, *AI Act*: « Il diritto dell'Unione in materia di protezione dei dati personali, della vita privata e della riservatezza delle comunicazioni si applica ai dati personali trattati in relazione ai diritti e agli obblighi stabiliti dal presente regolamento. Il presente regolamento lascia impregiudicati il regolamento (UE) 2016/679 o (UE) 2018/1725 o la direttiva 2002/58/CE o (UE) 2016/680, fatti salvi l'articolo 10, paragrafo 5, e l'articolo 59 del presente regolamento».

non si risolvono con la previsione di formule di salvaguardia, poiché sarebbe stato auspicabile prevedere nel testo dell'*AI Act* degli efficaci meccanismi di coordinamento operativo tra le diverse Autorità coinvolte nel ciclo di vita dei sistemi e dei modelli di IA, nel rispetto del principio di leale cooperazione. In loro assenza, emergono importanti lacune e discrasie nella complessiva struttura di *governance* dell'IA, che si traducono in una menomazione delle competenze affidate alle Autorità garanti di protezione dei dati personali.

Per questo motivo, sarebbe auspicabile affiancare alle Autorità di notifica e vigilanza del mercato, che debbono essere designate ai sensi dell'art. 70 dell'*AI Act*, l'Autorità garante in materia di protezione dei dati personali, che assumerebbe il compito di "Autorità garante dei diritti fondamentali" ai sensi dell'art. 77 dell'*AI Act*, affinché vigili sull'incidenza dell'IA sui diritti umani, in modo da essere messa nelle condizioni di esercitare i poteri riservatele da tale disposizione. Si tratta di poteri di richiesta informativa ed accesso documentale connessi all'uso di sistemi di IA ad alto rischio.

A titolo esemplificativo, l'Autorità garante in materia di protezione dati personali, se fosse designata ai sensi dell'art. 77 dell'*AI Act*, in sede di istruttoria per accertare una possibile violazione del GDPR, potrebbe accedere anche alla documentazione formata dal titolare del trattamento per rispettare gli obblighi dell'*AI Act*, come quella relativa alle misure di gestione dei rischi, ai *dataset* di addestramento, convalida e prova dei sistemi di IA, alla valutazione di impatto sui diritti fondamentali e così via. Diversamente, il diritto fondamentale alla protezione dei dati personali rischierebbe di essere privato della propria effettività e con esso tutti gli altri diritti essenziali che da quella tutela ottengono una protezione mediata, a detrimento dell'intangibilità della dignità personale.

3.5. Tensioni tra esigenze di salvaguardia della sicurezza e *privacy*. Casi di sorveglianza di massa

Nel nostro ordinamento giuridico il Garante per la protezione dei dati personali già da tempo ha avviato un'attività abbastanza intensa, incentrata sull'impiego dell'IA, come anche molte altre *Authority* europee, proprio perché questa tecnologia, per la sua stessa modalità operativa, viene già usata per trattare una grossa mole di dati personali, in sistemi automatizzati dall'impatto anche rilevante sulla vita di

tutti gli individui, come quelli di *credit scoring*[71], di modo che questa automatizzazione plasma in tempo reale la realtà, sulla mera base di dati ricevuti come *input*, elaborati senza alcun intervento umano.

Per quanto concerne, in particolare, le modalità di impiego nell'ambito della sicurezza e dell'ordine pubblico, degni di interesse sono gli impieghi, da parte di un numero sempre maggiore di autorità di contrasto (LEA, *Law Enforcement Authorities*), della tecnologia di riconoscimento facciale (FRT, *Facial Recognition Technology*), che spesso si avvale di componenti di intelligenza artificiale o di apprendimento automatico e che può essere utilizzata per autenticare o identificare una persona e può essere applicata nell'ambito dei video (ad esempio con le telecamere a circuito chiuso) o delle fotografie. Questa tecnologia può servire per varie finalità, tra cui la ricerca di persone nelle liste di controllo della polizia o il monitoraggio degli spostamenti di una persona nello spazio pubblico.

A tal proposito, il 26 aprile 2023 l'EDPB (*European data protection board* – Comitato europeo per la protezione dei dati) ha adottato delle linee guida sull'uso della tecnologia di riconoscimento facciale nel settore delle attività di contrasto e il Garante italiano ha adottato dei provvedimenti veramente determinanti sia nel contesto della videosorveglianza 'intelligente' in ambito urbano, sia a carico del Sistema automatico di riconoscimento delle immagini (S.A.R.I.) in uso alla Polizia di Stato[72].

L'attenzione che è stata posta sulle tecnologie di riconoscimento facciale deriva dal fatto che le stesse tendono a interferire con i diritti fondamentali (anche al di là del diritto alla protezione dei dati personali) e possono incidere sulla stabilità politica, sociale e democratica degli Stati.

In questo contesto, devono essere soddisfatti i requisiti della direttiva sulla protezione dei dati nelle attività di polizia e giudiziarie

---

71 Il *credit scoring* è un metodo statistico che consente di valutare l'affidabilità creditizia e la solvibilità di una determinata persona. È sostanzialmente un'attività di valutazione del rischio, sulla base del profilo di una persona, che nasconde però molti rischi di discriminazione. Viene effettuata mediante un processo trifasico: acquisizione dei dati, creazione del profilo utente, decisione automatizzata (accoglimento o rigetto della richiesta). La decisione dipende in gran parte dal profilo di rischio dell'utente, che viene valutato in modo automatico da *software* specifici, sulla base dei dati disponibili.

72 I provvedimenti del Garante per la protezione dei dati personali sul Sistema automatico di riconoscimento delle immagini (S.A.R.I.) verranno esaminati dettagliatamente nel capitolo IV.

(LED, *Law Enforcement Directive*), che prevede un determinato quadro normativo, in particolare l'articolo 3, punto 13 (che definisce il concetto di «dati biometrici»), l'articolo 4 (princìpi applicabili al trattamento di dati personali), l'articolo 8 (liceità del trattamento), l'articolo 10 (trattamento di categorie particolari di dati personali) e l'articolo 11 (processo decisionale automatizzato relativo alle persone fisiche).

In linea generale, il Comitato europeo per la protezione dei dati precisa, nelle proprie linee guida, che eventuali limitazioni all'esercizio dei diritti e delle libertà fondamentali devono essere previste dalla legge e rispettare il contenuto essenziale dei diritti e delle libertà fondamentali.

Nella sua formulazione, la base giuridica deve essere, inoltre, sufficientemente chiara, in modo da fornire ai cittadini un'indicazione adeguata in merito alle condizioni e alle circostanze in cui le autorità possono ricorrere a qualsiasi misura di raccolta di dati e di sorveglianza segreta; un mero recepimento nel diritto nazionale della clausola generale di cui all'articolo 10 della direttiva LED difetterebbe di precisione e prevedibilità. Viene precisato, infine, che «prima che il legislatore nazionale crei una nuova base giuridica per qualsiasi forma di trattamento dei dati biometrici che si avvalga del riconoscimento facciale, si dovrebbe consultare l'autorità di controllo competente per la protezione dei dati[73]».

Oltre a queste limitazioni procedurali, in verità, l'indirizzo, che viene espresso dall'organo europeo incaricato di fornire le linee di orientamento in materia di trattamento dei dati personali, è quello di garantire che i dati siano trattati in modo da assicurare l'applicabilità e l'efficacia delle norme e dei princìpi riconosciuti in materia dall'UE, affinché la valutazione della necessità e della proporzionalità individui e consideri tutte le possibili implicazioni per altri diritti fondamentali: se i dati vengono trattati sistematicamente all'insaputa degli interessati, è probabile che si generi un senso generale di sorveglianza costante, che può comportare effetti inibitori per quanto concerne alcuni o tutti i diritti fondamentali interessati, come la dignità umana, la libertà di pensiero, di coscienza e di religione, la libertà di espressione e la libertà di riunione e di associazione, pertanto anche le misure legislative devono essere appropriate per raggiungere gli obiettivi legittimi perseguiti dalla normativa in questione.

---

73 Cfr. *Linee guida 05/22 sull'uso della tecnologia di riconoscimento facciale nel settore delle attività di contrasto*, cit., p. 5.

L'orientamento che si può desumere, quindi, è che anche le misure legislative debbano essere appropriate, non essendo sufficiente una mera previsione normativa come copertura per l'impiego indiscriminato delle opportunità che la tecnologia offre alle investigazioni al giorno d'oggi.

In particolare, il trattamento di categorie particolari di dati, quali ad esempio i dati biometrici, si può considerare «strettamente necessario» (articolo 10 della direttiva LED) solo se l'ingerenza nella protezione dei dati personali e le sue limitazioni non eccedano la misura assolutamente necessaria, ossia indispensabile, escludendo qualsiasi trattamento di carattere generale o sistematico.

Relativamente agli impianti di videosorveglianza, è necessario premettere che il GDPR ha innovato la precedente disciplina, abolendo la notifica preventiva dei trattamenti all'Autorità di controllo, nota come *prior checking*, sostituendola con gli obblighi di tenuta di un registro dei trattamenti da parte del titolare/responsabile e di effettuazione di valutazioni di impatto, a causa della diffusione massiva degli impianti di videosorveglianza.

Lo sviluppo tecnologico, peraltro, ha aperto degli scenari veramente significativi, poiché «[…] dalla fine degli anni Novanta, i modelli di intelligenza artificiale ispirati all'attività neurale hanno consentito di svolgere compiti come il riconoscimento di immagini o di volti, contribuendo allo sviluppo di una nuova tecnologia come la videosorveglianza cd 'intelligente', basata sul riconoscimento di immagini e movimenti[74]».

La *video content analysis* (anche detta analisi video intelligente) costituisce, infatti, un insieme di tecniche dell'intelligenza artificiale che consente ad un calcolatore di analizzare un flusso video, allo scopo di comprenderne il contenuto e di annotarlo automaticamente, senza l'intervento umano; i sistemi di analisi video possono richiamare l'attenzione dell'operatore quando avviene qualche evento specifico nella scena inquadrata dalla telecamera e permettono di ridurre i tempi della ricerca, offrendo la possibilità di trovare solo quelle sequenze video che soddisfano alcuni criteri specificati *a priori*.

Il Garante per la protezione dei dati personali, con il Provvedimento generale in tema di videosorveglianza 8 aprile 2010, ha considerato tali sistemi come eccedenti rispetto alla normale attività di videosorveglianza, in quanto in grado di determinare effetti partico-

---

74 Cfr. Bernardi N., *Privacy. Protezione e trattamento dei dati*, IPSOA, Milano, 2019, p. 550.

larmente invasivi sulla sfera di autodeterminazione dell'interessato e, conseguentemente, sul suo comportamento; il relativo utilizzo, pertanto, è stato giustificato solo in casi particolari, tenendo conto delle finalità e del contesto in cui essi sono trattati, da verificare caso per caso, sul piano della conformità ai princìpi di necessità, proporzionalità, finalità e correttezza[75].

Prima di installare una qualsiasi tecnologia biometrica, pertanto, è necessario condurre un'attenta analisi preliminare, valutare l'impatto sulla dignità degli interessati e sulle implicazioni che tali sistemi comportano sui diritti fondamentali degli individui. I dati biometrici potranno essere oggetto di trattamento soltanto in presenza di una base giuridica e se il trattamento è adeguato, pertinente e non eccedente, rispetto alle finalità per le quali vengono utilizzati.

Per quanto concerne il trattamento tramite videosorveglianza effettuato per finalità di tutela dell'ordine e della sicurezza pubblica, prevenzione, accertamento o repressione dei reati, il citato Provvedimento generale dell'8 aprile 2010 ha stabilito che, sebbene anche per i soggetti pubblici sussista l'obbligo di fornire previamente l'informativa agli interessati (art. 13, Codice della *privacy*), la stessa può non essere resa quando i dati personali sono trattati per le ccdd finalità di polizia (tutela dell'ordine e della sicurezza pubblica e di polizia giudiziaria). Anche in questi casi, tuttavia, il trattamento deve essere effettuato in base ad espressa disposizione di legge, che lo preveda specificamente.

3.5.1. Il sistema di videosorveglianza 'intelligente' in ambito urbano installato dal Comune di Trento

Molto stimolante, per comprendere le ricadute pratiche della predetta disciplina normativa, è stato l'esame del provvedimento dell'11 gennaio 2024, che il Garante per la protezione dei dati personali ha adottato in merito a tre sistemi di intelligenza artificiale, denominati *"Marvel"*[76],

---

75 A scopo meramente esemplificativo, si può richiamare l'ipotesi di utilizzo di dati biometrici per presidiare accessi ad "aree sensibili", considerata la natura delle attività ivi svolte, come processi produttivi pericolosi o sottoposti a segreti di varia natura o al fatto che particolari locali sono destinati alla custodia di beni, documenti segreti o riservati o oggetti di valore.

76 In questo progetto, i dati, immediatamente anonimizzati dopo la raccolta, dovevano essere analizzati al fine di rilevare in maniera automatizzata, mediante tecniche di intelligenza artificiale, eventi rilevanti ai fini della salvaguardia della pubblica sicurezza (es. assembramenti, aggressioni, scippi, risse, ecc.).

"Protector"[77] e *"Precrisis"*[78], di cui si è dotato il Comune di Trento, finalizzati alla raccolta di informazioni in luoghi pubblici attraverso microfoni e telecamere di videosorveglianza per rilevare potenziali situazioni di pericolo per la pubblica sicurezza.

Nella documentazione prodotta, il Comune ha dichiarato che le molteplici norme nazionali in materia di sicurezza urbana, volte al perseguimento del bene pubblico alla sicurezza, sono state ritenute un adeguato fondamento giuridico legittimante il trattamento[79].

È stato, inoltre, rappresentato che è stata garantita l'anonimizzazione dei volti delle persone e delle targhe dei veicoli (mediante sfocatura o alterazione), di modo che non fosse in concreto possibile riconoscere delle caratteristiche personali sufficienti per permettere di identificare univocamente i soggetti ritratti; inoltre, l'acquisizione dei video e degli audio era destinata al solo scopo di "addestrare" il *software* al riconoscimento di situazioni di pericolo. L'Ente territoriale, pertanto, ha dichiarato che la persuasione, cui si era addivenuti, era che le circostanze oggettive e gli accorgimenti tecnici fossero tali da escludere un sostanziale trattamento di dati riferibili a persone identificate o identificabili.

I microfoni, inoltre, erano stati tarati in modo da rilevare solo suoni molto forti, come un'esplosione o una sirena d'emergenza e, per la rara eventualità di registrazione delle voci di persone, era stata comunque prevista un'anonimizzazione alla fonte dei segmenti contenenti parlato, con alterazione delle caratteristiche della voce in modo da rendere l'identità del parlatore non più riconoscibile.

Il Garante, in verità, ha ritenuto che il Comune abbia posto in essere un vero e proprio trattamento di dati personali in particolare nella

---

77 Il progetto prevedeva l'acquisizione di filmati da telecamere di videosorveglianza (senza segnale audio), ma anche la raccolta e l'analisi, mediante tecniche di intelligenza artificiale, di messaggi d'odio pubblicati sulla piattaforma *"Twitter"* (ora denominata *"X"*) e commenti pubblicati sulla piattaforma *"YouTube"*, al fine di rilevare eventuali emozioni negative (aggressività, rabbia o altre emozioni negative sul tema della religione), elaborando informazioni, al fine di identificare rischi e minacce per la sicurezza dei luoghi di culto.

78 In questo caso il Comune ha dichiarato che non è stato posto in essere alcun trattamento di dati personali, essendo il progetto ancora in fase di programmazione e analisi di contesto.

79 Cfr. parere citato, p. 5: «la base giuridica per la normale attività di videosorveglianza posta in essere dal Comune, tramite il Corpo di polizia locale, è rinvenibile […] [nell'] art. 6, comma 1, lettera e, [del Regolamento]), ai sensi in particolare di quanto previsto dal [d.l.] n. 11/2009, convertito in [legge] n. 38/2009, il quale attribuisce ai Comuni specifici compiti in materia di sicurezza urbana stabilendo, all'art. 6, comma 7, che "per la tutela della sicurezza urbana, i comuni possono utilizzare sistemi di videosorveglianza in luoghi pubblici o aperti al pubblico"».

fase di raccolta delle informazioni ritenute d'interesse (filmati di videosorveglianza, audio proveniente dai microfoni, messaggi, commenti, profili ottenuti da reti sociali), perché l'acquisizione e la temporanea memorizzazione di dati personali, ancorché per una ridotta frazione temporale, costituisce un trattamento di dati personali, spesso anche appartenenti alle ccdd categorie particolari.

Ha concluso anche nel senso che le tecniche, impiegate successivamente alla raccolta dei dati, non potevano considerarsi idonee a realizzare un'effettiva anonimizzazione, poiché la sola sostituzione della voce del soggetto parlante non è in alcun modo idonea ad anonimizzare i dati personali correlati a una conversazione, atteso che dal contenuto della stessa è possibile ricavare informazioni relative sia al soggetto parlante, sia a terzi e che tali informazioni possono rendere identificabile il parlante, i suoi interlocutori o i soggetti terzi a cui si fa rifermento nel discorso; allo stesso modo per le immagini, anche la tecnica di offuscamento dei visi non può ritenersi idonea ad assicurare l'effettiva anonimizzazione dei dati, atteso che gli interessati sono comunque potenzialmente identificabili tramite altre caratteristiche fisiche (corporatura, abbigliamento, caratteristiche fisiche particolari, ecc.) o informazioni detenute da terzi (notizie di stampa relative a fatti di cronaca ecc.) o ancora informazioni desumibili dalla localizzazione della telecamera.

Premesso, pertanto, che i progetti in parola costituivano trattamento a tutti gli effetti, il Garante, oltre ad un insieme di violazioni di regole procedurali e al mancato adempimento di svariati obblighi gravanti a carico del titolare del trattamento, fondamentalmente non ha riconosciuto la liceità e ha imposto il fermo totale dell'operatività degli stessi, essendo stato effettuato un trattamento in assenza di una previsione normativa di autorizzazione ed essendo stata potenzialmente lesa non soltanto la riservatezza dei cittadini, ma anche la libera manifestazione del pensiero, la partecipazione alla vita politica e sociale, la libertà di riunione e la libertà di manifestare la propria fede religiosa, poiché simili forme di sorveglianza negli spazi pubblici possono modificare il comportamento delle persone e condizionare finanche l'esercizio delle libertà democratiche.

Il principio imprescindibile, in definitiva ricavabile da questo provvedimento, è quello secondo il quale eventuali limitazioni ai diritti fondamentali, al rispetto della vita privata e alla protezione dei dati personali possono essere apportate soltanto se esse siano previste dalla legge e rispettino il contenuto essenziale dei diritti fondamentali,

nonché il principio di proporzionalità, per cui possono essere apportate limitazioni solo laddove siano necessarie e rispondano effettivamente a obiettivi di interesse generale riconosciuti dall'Unione o all'esigenza di proteggere i diritti e le libertà altrui.

Il d.l. 20 febbraio 2017, n. 14, inoltre, ha consentito ai Comuni l'installazione di sistemi di videosorveglianza ai soli fini di "prevenzione e contrasto dei fenomeni di criminalità diffusa e predatoria", previa stipula di un accordo per l'attuazione della sicurezza urbana con la Prefettura territorialmente competente; in queste ipotesi non è comunque contemplato l'utilizzo di microfoni per l'acquisizione del segnale audio.

Capitolo IV
L'intelligenza artificiale al servizio della sicurezza

di Gennaro Corrado*, Daniela Gregorio**
e Marialuisa Simona Gregalli***

4.1. Il ricorso all'intelligenza artificiale nell'attività di *law enforcement*

L'intelligenza artificiale costituisce una risorsa ed una sfida anche per le Forze di polizia, la cui efficienza è da sempre inscindibilmente connessa alla capacità di adattarsi all'evoluzione dei fenomeni umani ed all'avanzamento della scienza e della tecnica.

Ancor più che nel passaggio dalle macchine da scrivere ai *personal computer*, come nel superamento delle buste da lettere in favore delle comunicazioni via posta elettronica, l'implementazione dell'intelligenza artificiale costituisce un radicale punto di svolta nel costante processo di evoluzione tecnologica dell'attività di polizia.

Le elevatissime potenzialità che contraddistinguono tali applicativi informatici, per parte già implementati dagli apparati di *law enforcement* delle nazioni all'avanguardia tra cui l'Italia, consentono di sfruttare le enormi capacità computazionali dei moderni *hardware*, così da incrementare sensibilmente l'efficienza e l'efficacia dell'attività svolta a tutela della collettività.

L'applicazione dell'intelligenza artificiale all'attività delle Forze di polizia, più che un'opportunità per risparmiare od ottimizzare il tempo e le risorse, rappresenta un'inevitabile necessità per contrastare

* Vice Questore della Polizia di Stato, già frequentatore del XL Corso di Alta Formazione presso la Scuola di Perfezionamento per le Forze di Polizia.
** Vice Questore della Polizia di Stato, già frequentatrice del XL Corso di Alta Formazione presso la Scuola di Perfezionamento per le Forze di Polizia.
*** Dirigente della Polizia Penitenziaria, già frequentatrice del XL Corso di Alta Formazione presso la Scuola di Perfezionamento per le Forze di Polizia.

efficacemente fenomenologie criminali, suscettibili di incidere sull'ordine e la sicurezza pubblica, sempre maggiormente evolute da un punto di vista tecnologico. Basti pensare, nel complesso della moltitudine di esempi cui potrebbe farsi ricorso, alle nuove sfide rappresentate dall'uso delle criptovalute da parte delle organizzazioni criminali su scala mondiale, oppure alla complessità di muoversi nei meandri della rete, in particolare nel *deep web*, per prevenire e reprimere la radicalizzazione fondamentalista.

Parallelamente, va posto in evidenza che, allo stato attuale, nel nostro ordinamento giuridico esiste un vero e proprio obbligo normativo, per gli apparati pubblici, di procedere alla digitalizzazione ed all'informatizzazione dei procedimenti amministrativi, nell'ottica di migliorarne l'efficienza.

Difatti, oltre a quanto stabilito dal Codice dell'amministrazione digitale, l'art. 3-*bis* della Legge nr. 241/1990, come novellato per effetto del d.l. nr. 76/2020, stabilisce che «per conseguire maggiore efficienza nella loro attività, le amministrazioni pubbliche agiscono mediante strumenti informatici e telematici, nei rapporti interni, tra le diverse amministrazioni e tra queste e i privati»[80].

A tali irrinunciabili potenzialità fa tuttavia da contraltare l'imprescindibile necessità di guidare il cambiamento, senza dunque subirlo, in considerazione dei profili di rischio e dei margini di responsabilità connessi al ricorso a tale tecnologia.

Oltre a quanto rappresentato nel precedente capitolo, con puntuale riferimento alle implicazioni riguardanti il trattamento dei dati personali, l'implementazione dell'intelligenza artificiale nell'attività di polizia pone in evidenza rilevanti ulteriori profili di rischio, tra cui quelli relativi alla sicurezza cibernetica ed alla possibilità che azioni o fenomeni di disinformazione possano ripercuotersi negativamente, andando ad influenzare i processi svolti in forma automatizzata[81].

---

80 Nella precedente formulazione l'articolo stabiliva: «per conseguire maggiore efficienza nella loro attività, le amministrazioni pubbliche incentivano l'uso della telematica, nei rapporti interni, tra le diverse amministrazioni e tra queste e i privati». La novella introduce un obbligo/dovere rispetto alla precedente disposizione volta semplicemente ad incentivare l'uso della telematica.

81 Pollicino O., *Disinformazione e intelligenza artificiale: un cocktail esplosivo?*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024, nel muovere da una critica all'incondizionato ricorso alla locuzione inglese facente riferimento alle ccdd *"fake news"*, evidenzia i rischi connessi ai tre distinti fenomeni della disinformazione, misinformazione e malinformazione. Sul punto si tornerà dettagliatamente nel prosieguo del presente capitolo.

Inoltre, l'utilizzo dell'intelligenza artificiale a tutela della pubblica sicurezza impone di affrontare il tema relativo all'imputazione delle responsabilità, di natura civile, penale, amministrativa e contabile, tenendo conto del fatto che ricorrere a tali strumenti significa a tutti gli effetti coinvolgere nell'attività di polizia, sia pur indirettamente, gli sviluppatori di tali tecnologie come pure, se non prodotte *in house*, i rispettivi produttori e venditori.

Individuate nei termini suddetti le principali questioni da affrontare, occorre muovere dall'analisi dei principali utilizzi dell'intelligenza artificiale da parte delle Forze di polizia, facendo accenno anche a quelli di cui risulta la sperimentazione in ambito internazionale.

4.1.1. Analisi di video e immagini. Riconoscimento facciale

Una delle principali applicazioni dell'intelligenza artificiale all'attività di polizia, anche nel contesto italiano, è rappresentata dall'utilizzo di tale tecnologia nell'ambito dell'analisi di contenuti multimediali ai fini del riconoscimento di persone, località o elementi di interesse ai fini info-investigativi.

Tali strumentazioni consentono innanzitutto di procedere alla categorizzazione delle immagini, determinando con precisione se, all'interno delle stesse, siano ritratte persone, animali, edifici, armi, esplosivi od altri oggetti di rilievo. Permettono inoltre l'analisi automatizzata dei fotogrammi ai fini della ricerca di elementi specifici come, ad esempio, codici alfanumerici come le targhe di automobili o di altre tipologie di veicoli.

Attraverso detti applicativi, in particolare, è possibile procedere alla comparazione di immagini della medesima tipologia, ottenendo come risultato un punteggio di similitudine. È questo il caso del S.A.R.I., Sistema Riconoscimento Automatico Immagini, sviluppato dal Dipartimento della pubblica sicurezza, che consente di ricercare eventuali somiglianze tra persone ritratte in foto – o fotogrammi estratti da video – rispetto a quelle già sottoposte ai rilievi fotodattiloscopici da parte delle Forze di polizia. L'applicativo, grazie a due algoritmi di riconoscimento facciale, partendo da un'immagine fotografica di un "soggetto ignoto" fornisce un elenco di immagini, ordinato secondo un grado di similarità, così fornendo un rilevantissimo ausilio all'attività di identificazione che, tuttavia, permane in capo all'operatore di polizia[82].

---

82 Si rinvia al prossimo paragrafo per un'analisi dettagliata.

4.1.2. Identikit e age progression

Nel quadro delle applicazioni all'attività di polizia giudiziaria, tali algoritmi sono impiegati ai fini della ricostruzione di *identikit* e per la *age progression*, ovverosia per ricostruire le credibili fattezze attuali di soggetti, partendo da immagini datate. I principali utilizzi di tali dotazioni tecniche sono riconducibili alle investigazioni finalizzate alla cattura dei latitanti, come pure alla ricerca delle persone scomparse a distanza di margini di tempo rilevanti.

Detti applicativi sono altresì utilizzati anche in relazione alle tracce audio, consentendo di estrarre e campionare sequenze di dialogo anche ai fini della ricostruzione, con ampio margine di verosimiglianza, dei tratti vocali di un soggetto da ricercare[83].

4.1.3. Riconoscimento delle emozioni

Uno dei profili di maggiore complessità e delicatezza, è rappresentato dal possibile utilizzo dell'intelligenza artificiale per il riconoscimento delle emozioni, *real time* o tramite contenuti audiovisivi, attraverso l'analisi delle espressioni facciali, del tono di voce come da altri segnali biometrici.

Queste tecnologie si basano sulla *Basic Emotion Theory*, nota con l'acronimo BET, elaborata dallo psicologo Paul Ekman negli anni Sessanta dello scorso secolo, secondo cui dall'analisi delle micro espressioni facciali sarebbe possibile comprendere le emozioni effettivamente provate da ciascun individuo.

La validità di tale teoria è stata contestata nell'ambito di studi tesi a dimostrare la non veridicità del suo assunto di fondo, ossia l'universalità del modo in cui l'essere umano esprimerebbe le sue emozioni. All'opposto, è stato rilevato che profili di natura culturale e di contesto, ovvero condizioni patologiche, incidono sensibilmente nel modo in cui sono espresse le emozioni, anche attraverso i movimenti del volto.

Oltre alle criticità che dunque scaturiscono dagli stessi dubbi relativi alla validità della BET, il possibile impiego di *software* in tal sen-

---

83 Pistilli C., *L'utilizzo dell'intelligenza artificiale nel campo delle attività investigative delle forze dell'ordine: tra prospettive di sviluppo ed esigenze di coordinamento*, in Balbi G. – De Simone F. – Esposito A. – Manacorda S. (a cura di), *Diritto Penale e Intelligenza Artificiale "Nuovi scenari"*, Torino, 2022.

so pone evidenti interrogativi etici e giuridici, soprattutto in relazione alla possibilità che tali tecnologie possano essere concretamente applicate nell'ambito dell'attività di polizia o giudiziaria.

In particolare, come osservato nell'ambito degli approfondimenti sul punto, un rilevante fattore di criticità sotto il profilo tecnico giuridico afferisce al fatto che l'effettivo ricorso a tali algoritmi, soprattutto nel momento della formazione della prova, renderebbe estremamente complesso confutarne la validità, rimettendo dunque all'intelligenza artificiale un ruolo dirimente nell'attribuzione della responsabilità[84].

Sul punto il cd *AI Act*, ovverosia il Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024, ha innanzitutto stabilito cosa debba intendersi come "sistema di riconoscimento delle emozioni", definendolo, all'art. 3, nr. 39, come «un sistema di IA finalizzato all'identificazione o all'inferenza di emozioni o intenzioni di persone fisiche sulla base dei loro dati biometrici»[85]. Tali applicativi sono stati inclusi tra i "sistemi di IA ad alto rischio".

In particolare, in linea con quanto premesso in merito ai dubbi scientifici sulla teoria di fondo, il legislatore unionale, al considerando nr. 44, ha rilevato come «sussistono serie preoccupazioni in merito alla base scientifica dei sistemi di IA volti a identificare o inferire emozioni, in particolare perché l'espressione delle emozioni varia notevolmente in base alle culture e alle situazioni e persino in relazione a una stessa persona.

---

84 Martorana M., *IA e riconoscimento delle emozioni: rischi e possibili vantaggi*, articolo pubblicato su *Agendadigitale.eu* il 21 settembre 2023, evidenzia che «un altro aspetto molto interessante legato all'uso delle IA per decifrare le emozioni provate dagli individui è che questi non hanno modo di confutare il risultato dell'algoritmo: se una Intelligenza Artificiale analizza le mie espressioni (anche, magari, in combinazione con i miei movimenti o tono di voce, se il sistema è particolarmente sofisticato) e ne deduce che sono impaurito, come posso dimostrare che c'è stato un errore? È la parola di un algoritmo contro la mia; a quale delle due dare retta? Pensiamo alle derive problematiche di questo discorso quando le IA vengono utilizzate nel settore giudiziario o di polizia, ad esempio (come già succede in alcune parti del mondo) per determinare se una persona sta dicendo la verità o no durante un interrogatorio».

85 La definizione è approfondita nel considerando nr. 18, in cui si evidenzia che «la nozione si riferisce a emozioni o intenzioni quali felicità, tristezza, rabbia, sorpresa, disgusto, imbarazzo, eccitazione, vergogna, disprezzo, soddisfazione e divertimento. Non comprende stati fisici, quali dolore o affaticamento, compresi, ad esempio, i sistemi utilizzati per rilevare lo stato di affaticamento dei piloti o dei conducenti professionisti al fine di prevenire gli incidenti. Non comprende neppure la semplice individuazione di espressioni, gesti o movimenti immediatamente evidenti, a meno che non siano utilizzati per identificare o inferire emozioni. Tali espressioni possono essere espressioni facciali di base quali un aggrottamento delle sopracciglia o un sorriso, gesti quali il movimento di mani, braccia o testa, o caratteristiche della voce di una persona, ad esempio una voce alta o un sussurro».

Tra le principali carenze di tali sistemi figurano la limitata affidabilità, la mancanza di specificità e la limitata generalizzabilità. Pertanto, i sistemi di IA che identificano o inferiscono emozioni o intenzioni di persone fisiche sulla base dei loro dati biometrici possono portare a risultati discriminatori e possono essere invasivi dei diritti e delle libertà delle persone interessate. Considerando lo squilibrio di potere nel contesto del lavoro o dell'istruzione, combinato con la natura invasiva di tali sistemi, questi ultimi potrebbero determinare un trattamento pregiudizievole o sfavorevole di talune persone fisiche o di interi gruppi di persone fisiche. È pertanto opportuno vietare l'immissione sul mercato, la messa in servizio o l'uso di sistemi di IA destinati a essere utilizzati per rilevare lo stato emotivo delle persone in situazioni relative al luogo di lavoro e all'istruzione. Tale divieto non dovrebbe riguardare i sistemi di IA immessi sul mercato esclusivamente per motivi medici o di sicurezza, come i sistemi destinati all'uso terapeutico».

4.1.4. Analisi di database

Le elevate capacità computazionali dei *software* di intelligenza artificiale assumono una particolare rilevanza nell'ambito delle potenzialità espresse nell'analisi di *database* e grosse raccolte di documenti informatici, sia in riferimento alle indagini finalizzate all'individuazione degli autori di reati, sia nell'ambito dello svolgimento di attività di natura preventiva.

Tali applicazioni consentono, ad esempio, di sottoporre ad analisi automatizzata i numerosi *file* – dati non strutturati – acquisiti nell'ambito di attività tecniche di intercettazione o nel corso dell'esecuzione di perquisizioni informatiche di *device* come *personal computer*, *notebook*, *tablet*, *smartphone* o dispositivi di memorizzazione, agevolando l'attività investigativa attraverso l'indicizzazione dei documenti, oppure procedendo all'estrazione di specifici contenuti, in relazione ai parametri di ricerca stabiliti dall'operatore.

Parallelamente, gli applicativi di intelligenza artificiale permettono di processare gli innumerevoli elementi contenuti nei *database* – dati semi strutturati – presenti in *file* XML, ricavati da siti *web*, presenti nei metadati o ricavati da *file* di *log* delle applicazioni installate sui citati dispositivi; un esempio su tutti può essere riferito all'utilità di tali strumenti nell'analisi dei tabulati telefonici, attività che richiederebbe tempistiche incalcolabili senza l'ausilio di strumenti in grado di

normalizzare i contenuti, che sono di regola forniti dagli operatori telefonici mediante *file* complessi ed eterogenei.

In particolare, tali algoritmi assumono un rilievo fondamentale nell'ambito dell'analisi della *blockchain*, consentendo di tracciare ed approfondire i movimenti di *cryptocurrency*.

4.1.5. OSINT e SOCMINT. Il CRAIM del Dipartimento della pubblica sicurezza

Le tecnologie di intelligenza artificiale sin qui riepilogate assumono un ruolo cruciale in relazione all'attività info-investigativa svolta nell'ambito della rete, orientata segnatamente a prevenire e contrastare fenomeni che possono costituire un pericolo per l'ordine e la sicurezza pubblica, tra cui in particolare la commissione di reati con finalità di terrorismo. In tale quadro gli algoritmi impiegati per il riconoscimento di immagini e audio e l'analisi testuale rappresentano una potenzialità straordinaria nell'ambito delle attività note con gli acronimi di OSINT e SOCMINT, ossia la *open source intelligence* e la *social media intelligence*, sottocategoria della prima.

Tali sono, in particolare, le strumentazioni tecnologiche che il CRAIM, Centro di Ricerca per l'Analisi delle Informazioni Multimediali del Dipartimento della pubblica sicurezza, mette a disposizione delle D.I.G.O.S. delle Questure, ossia delle articolazioni della Polizia di Stato, funzionalmente dipendenti dalla Direzione Centrale della Polizia di Prevenzione, che hanno come principali funzioni l'attività informativa per la tutela dell'ordine e della sicurezza pubblica ed il contrasto del terrorismo[86].

I servizi del CRAIM comprendono diverse applicazioni, che si compongono come segue:

- OSINT generalizzato. La piattaforma WebLive è dedicata alla raccolta di informazioni generalizzate da fonti aperte. Il meccanismo di ricerca si basa essenzialmente su parole chiave, che vengono ricercate in una pluralità di fonti, quali *news*, *blog*, pa-

86 I servizi forniti dal CRAIM sono stati descritti, nei termini seguenti, nell'ambito di una scheda riepilogativa fornita, come contributo al presente lavoro, dal Direttore Tecnico Superiore della Polizia di Stato Dr. Tommaso Fornaciari, membro del Gruppo di lavoro per lo Sviluppo, l'Innovazione, la Ricerca e i Progetti Speciali della Polizia di Stato, che si ringrazia per la cortese collaborazione.

gine *web* generiche. WebLive mette poi a disposizione diversi strumenti di analisi, che permettono di filtrare e raggruppare i dati secondo le logiche degli utenti, e di rappresentare i dati raccolti attraverso diverse metriche e mediante vari strumenti di visualizzazione. Si tratta di uno strumento che ha lo scopo di analizzare fenomeni macro, raccogliendo grandi quantità di dati e offrendo un quadro di sintesi.

- OSINT "targetizzato". Per il monitoraggio di una vasta pluralità di *social media*, è stata adottata la soluzione Medusa. Il motore interno di tale applicazione è dato da una batteria di *avatar* accreditati su diversi *social media*, che permettono lo *scraping* dei dati, senza interazione diretta con gli utenti. Scopo della soluzione è monitorare conversazioni pubbliche e fenomeni di carattere generale, sempre all'interno di specifici *social media*. Medusa permette poi, analogamente a WebLive, di filtrare i dati, rappresentando i contenuti di interesse, tra cui le reti di relazione tra i vari *account*.
- *Broadcast Monitoring*. Il CRAIM ha poi acquisito un'applicazione, Iride, che permette di registrare programmi televisivi da digitale terrestre e satellitari, così come da canali *streaming* come YouTube, in base alle indicazioni degli uffici richiedenti. I filmati raccolti vengono poi trascritti, ne vengono estratti entità e relazioni e la base di conoscenza creata viene rappresentata graficamente, con la possibilità di navigare gli stessi filmati, raggiungendo rapidamente l'informazione di interesse.

Inoltre, il CRAIM mette a disposizione prototipi sviluppati internamente, per lo svolgimento di una serie di attività specifiche.

- *Image analysis*. Per l'analisi delle immagini è disponibile una soluzione che permette l'estrazione dei volti da foto e video, valutando il livello di similarità tra volti di interesse, senza tuttavia permettere l'identificazione forense. Inoltre è possibile effettuare ricerche specifiche in linguaggio naturale, con la possibilità di estrarre immagini pertinenti sia per oggetti concreti, che per concetti astratti.
- *Smart* OCR. Documenti *pdf* e immagini possono essere elaborati da sistemi di intelligenza artificiale specifici per l'estrazione dei testi.
- Servizi di trascrizione e traduzione. Il CRAIM dispone di applicazioni, in parte internalizzate, in parte legate a motori esterni, che vengono utilizzati in modalità anonima, sia pure

con condivisione del dato, che permettono di trascrivere file audio e video e di tradurre testi in una pluralità di lingue, particolarmente vasta nel caso dei motori esterni.
- Navigazione anonima. Infine, il CRAIM permette di navigare in modo sicuro e anonimo sia sul *web* superficiale, sia sul *dark web*.

4.1.6. Polizia predittiva

Altro rilevante ambito in cui l'intelligenza artificiale può trovare applicazione, quanto all'attività di polizia, riguarda l'analisi predittiva finalizzata alla pianificazione dell'attività di controllo del territorio. È questo l'ambito della cd "polizia predittiva", definibile come l'insieme delle attività di studio e analisi statistica avente l'obiettivo di prevedere dove e quando potrà essere commesso un reato, al fine di prevenirne la commissione.

In tale ambito si manifesta, in modo del tutto evidente, l'essenza dell'ausilio che la tecnologia più avanzata può fornire all'attività di polizia, atteso che ragionare in termini di prevenzione dei reati, sulla base degli elementi informativi e dei precedenti, rappresenta un'attività di polizia per così dire "tradizionale". L'intelligenza artificiale in tale ambito non altera la sostanza ma ne rivoluziona la portata, permettendo di analizzare e mettere in correlazione, se opportunamente raccolto, un patrimonio informativo fuori dalla portata cognitiva dell'essere umano, fatto di luoghi, orari, condizioni meteorologiche, periodi dell'anno, azioni criminali ricorrenti.

Tuttavia, l'impiego di tali applicativi ha posto all'attenzione di chi ne ha approfondito l'utilizzo una serie di evidenti fattori di criticità[87]:
- essi possono fornire adeguate previsioni solo in relazione a limitate, determinate categorie di reati, come ad esempio quelli ccdd "predatori" o relativi allo spaccio di stupefacenti, che non rappresentano necessariamente quelli più pericolosi per la sicurezza pubblica;
- l'uso di tali strumenti, oltre ai profili inerenti alla *privacy*, potrebbe collidere con il divieto di discriminazione laddove l'a-

87 Basile F., *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in Balbi G. – De Simone F. – Esposito A. – Manacorda S. (a cura di), *Diritto Penale e Intelligenza Artificiale "Nuovi scenari"*, Torino, 2022.

nalisi predittiva riguardasse anche profili connessi a caratteristiche etniche, religiose o sociali[88];
- tali sistemi si autoalimentano con dati prodotti dal loro stesso utilizzo, creando il rischio di circoli viziosi, dando origine al fenomeno della "profezia che si auto-avvera"[89].

4.1.7. Valutazione della pericolosità sociale

Altro profilo applicativo, concettualmente collegato all'analisi predittiva di scenario, riguarda il possibile ricorso a tali algoritmi ai fini della valutazione della pericolosità sociale, per l'applicazione di misure di prevenzione, sicurezza o alternative alla detenzione.

In proposito occorre evidenziare quanto recentemente stabilito nell'ambito dell'*AI Act* che, all'art. 5, rubricato "pratiche di IA vietate", al punto 1, lettera d), stabilisce il divieto di commercializzazione e uso di sistemi di intelligenza artificiale per effettuare valutazioni del rischio di commissione di reati laddove siano basati sulla profilazione ovvero sulla valutazione dei tratti e delle caratteristiche della personalità. Lo stesso articolo precisa che tale divieto non si applica ai *software* utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, a patto che si basi su fatti oggettivi, verificabili direttamente e connessi a crimini.

Appare dunque evidente come il legislatore europeo, nel ribadire il divieto di discriminazione in relazione a caratteristiche personali, consenta la possibilità che la magistratura e gli apparati di *law enforcement* possano far uso di tali applicativi, ai fini della valutazione del pericolo di commissione di reati, a condizione che gli elementi a base della valutazione riguardino profili oggettivi e verificabili e, soprattutto, che tali sistemi rappresentino esclusivamente strumenti

---

88 Si ritornerà a breve sul punto in relazione all'utilizzo di tali *software* ai fini della valutazione della pericolosità sociale.

89 Basile F., op. cit., così descrive tale rischio: «se, ad esempio, un *software* predittivo individua una determinata "zona calda", i controlli e i pattugliamenti della polizia in quella zona si intensificheranno, con inevitabile conseguente crescita del tasso dei reati rilevati dalla polizia in quella zona, che diventerà, quindi, ancora più "calda", mentre altre zone, originariamente non ricondotte nelle "zone calde", e quindi non presidiate dalla polizia, rischiano di rimanere, o di diventare, per anni "zone fredde", ove la commissione di reati non viene adeguatamente monitorata».

a supporto della decisione umana e, dunque, dell'assunzione di responsabilità di chi di dovere[90].

4.1.8. Controllo automatizzato del territorio

Per quanto attiene al possibile ricorso all'intelligenza artificiale nell'attività di prevenzione dei reati, si rileva come in talune nazioni siano in corso di sperimentazione ipotesi di pattugliamento delle città attraverso droni a guida autonoma, finanche dotati di strumenti per contrastare ipotesi di aggressione, potenzialmente offensivi se non addirittura letali[91].

Tali applicazioni, non riguardanti le Forze di polizia italiane, risultano, come di palese evidenza, tra le più critiche quanto ai profili di rischio ed in relazione agli interessi e beni giuridici coinvolti. Tuttavia, è opportuno rilevare come l'ipotesi in discorso sia fortemente connessa all'utilizzo dei dispositivi a guida autonoma, che – come si dirà nel prosieguo – costituisce, allo stato, il principale ambito di uso dell'intelligenza artificiale analizzato, da dottrina e giurisprudenza, nel quadro delle implicazioni di responsabilità civile.

4.1.9. Compilazione automatizzata della documentazione di servizio

Un ulteriore campo di applicazione dell'intelligenza artificiale all'attività di polizia è costituito dalla possibilità di avvalersi di tale tecnologia ai fini della compilazione di atti, relazioni e documenti di servizio.

Il principale e forse più emblematico caso risulta essere quello del *software Draft One*, prodotto dalla Axon per supportare e potenziare gli agenti, con il dichiarato fine di «ridurre le scartoffie e produrre rapporti di qualità superiore»[92].

---

90 E questo un primo riferimento a quel principio di "riserva di umanità" descritto in dottrina, su cui diffusamente si tornerà.

91 Basile F., op. cit.

92 L'uso dell'applicazione in alcune città statunitensi è oggetto dell'articolo, pubblicato sul portale *www.rivista.ai* il 5 ottobre 2024, dal titolo *L'uso dell'Intelligenza Artificiale nei rapporti di polizia: svolta tecnologica o rischio per la giustizia?*.
La *Axon Enterprise, Inc.* è un'azienda statunitense nota per la produzione e distribuzione dell'arma a impulsi elettrici TASER (dal precedente nome della società *TASER Internatio-*

Come pubblicizzato nell'ambito della pagina *web* italiana del produttore, l'applicativo *Draft One* consente agli agenti di risparmiare tempo utilizzando l'audio della videocamera indossabile per produrre bozze di rapporti di polizia in pochi secondi.

Tra le misure di sicurezza, il *software* «richiede che siano agenti in carne e ossa a prendere le decisioni nei momenti chiave»:

- una volta modificato il resoconto del rapporto e aggiunte le informazioni chiave, gli agenti sono tenuti a confermarne l'accuratezza prima di sottoporlo al successivo ciclo di revisione umana;
- i resoconti dei rapporti sono redatti rigorosamente a partire dalla trascrizione audio della registrazione della videocamera indossabile;
- l'esperienza predefinita limita le bozze dei rapporti ai casi minori, escludendo in particolare quelli in cui si siano verificati arresti e quelli per illeciti penali.

Lo sviluppo e l'adozione di tecnologie di intelligenza artificiale ai fini della predisposizione di bozze di atti, relazioni o documenti di servizio in generale costituisce un'interessante soluzione tecnologica, utile a consentire che l'intelligenza umana possa essere valorizzata in tutti quei momenti dell'attività di servizio in cui il fattore umano è – e credibilmente resterà sempre – necessario ed imprescindibile.

In altri termini, a convinto avviso di chi scrive, tali tecnologie possono – anzi, devono – essere sviluppate e diffuse per ottimizzare tutte quelle operazioni meccaniche che possono fare a meno dell'esperienza e della professionalità dell'operatore di polizia. Nello scrivere l'immaginazione rimanda alla produzione, in occasione di un arresto in flagranza, di tutti gli atti in ciclostile; documenti la cui predisposizione in bozza potrebbe essere affidata ad una intelligenza artificiale, sviluppata *in house* che, partendo naturalmente dagli *input* necessari e fermo restando il controllo e l'assunzione di responsabilità finale, potrebbe curarsi del "copia e incolla", lasciando all'operatore di polizia più tempo per fare il suo mestiere: tutelare l'ordine e la sicurezza pubblica e, nel redigere i verbali e gli atti prescritti, concentrarsi su ciò che la macchina non può conoscere, ossia i fatti e la loro umana percezione.

Parallelamente, non può che pienamente condividersi la posizione di chi ha sostenuto che «se l'AI viene utilizzata come "scorciatoia"

---

nal, *Inc.*). Ha sede a Scottsdale, Arizona. La citazione è dalla pagina italiana del produttore, raggiungibile all'*url* https://it.axon.com/caso-di-studio/approfondimenti/draft-one/.

per evitare di lavorare cerebralmente alla stesura di una sentenza, di una comparsa difensiva, di una conciliazione, di un rogito notarile, di un provvedimento amministrativo, proseguirà, anche sul lavoro, il crescente abbassamento del livello culturale di categorie "colte" e "pensanti", quali magistrati, avvocati, notai, dirigenti pubblici, in perfetta sintonia con il generalizzato abbassamento culturale e linguistico dei giovani nascente dall'uso sistematico di cellulari e PC per redigere temi e ricerche di scuola, risolvere problemi matematici, preparare esami universitari, rinunciando al peso della "fatica dello studio"»[93].

4.2. Il Sistema automatico di riconoscimento immagini (S.A.R.I)

Nel contesto specifico dell'attività di prevenzione e repressione dei reati e di gestione dell'ordine e della sicurezza pubblica, la Polizia di Stato ha dotato la propria articolazione competente in materia di polizia scientifica di un *software* in grado di identificare, in poco tempo, l'autore di un reato da un *frame* sfocato di una telecamera di sicurezza: il portale *Sari* (Sistema automatico di riconoscimento immagini), presentato nel 2019, è capace di ricercare, grazie ad algoritmi di intelligenza artificiale, la foto di un soggetto ignoto all'interno della banca dati dei soggetti fotosegnalati "AFIS[94]" (la stessa in cui vengono salvate le impronte digitali), che contiene le immagini di circa 20 milioni di persone, realizzando un'attività che sarebbe impossibile svolgere manualmente[95].

In tempi brevissimi il Sari produce una lista di candidati, ordinata sulla base di un criterio di similarità e di una percentuale di *'matching'*,

---

93 Tenore V., *Una riflessione sistematica: l'intelligenza artificiale può sostituire un giudice o trasformarlo da homo sapiens in homo numericus?*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024.

94 Acronimo di *"Automated Fingerprint Identification System"*, in italiano "Sistema Automatizzato di Identificazione delle Impronte".

95 Sul punto, si pone in evidenza la descrizione dell'applicativo riportata sul portale istituzionale del Ministero dell'Interno, alla pagina *web* https://www.interno.gov.it/it/notizie/sistema-automatico-riconoscimento-immagini-futuro-diventa-realta: «S.A.R.I, questo il suo nome derivato dall'acronimo, riesce da un'immagine fotografica di un soggetto ignoto a effettuare una ricerca computerizzata nella banca dati A.F.I.S., e grazie a due algoritmi di riconoscimento facciale, è in grado di fornire un elenco di immagini ordinato secondo un grado di similarità. Sono però sempre gli operatori specializzati della Polizia scientifica ad effettuare la necessaria comparazione fisionomica. Questo – in caso di *match* – consente di integrare l'utilità investigativa del risultato anche con un accertamento tecnico a valenza dibattimentale».

con associazione di un punteggio che rappresenta la misura di somiglianza fornita; la lista deve essere, poi, revisionata a mano da un operatore specializzato per la ricerca di un possibile candidato su cui concentrare ulteriori investigazioni. Il sistema, inoltre, non esegue soltanto analisi di volti, ma, da un fotogramma di un soggetto ripreso di spalle o con il casco indossato, riesce ad effettuare una stima dell'altezza.

I risultati ottenuti sono certamente di tipo "investigativo" e non "forense" e servono ad evidenziare un sospettato e a farlo sottoporre ad altri accertamenti come un pedinamento o un confronto con testimoni; tuttavia, sono in grado di orientare in modo significativo lo sviluppo delle attività di indagine.

L'introduzione di questo sistema ha indotto il Garante per la protezione dei dati personali ad attivare il procedimento di indagini, previsto dall'art. 37, D.Lgs. 51/2018, con una richiesta di informazioni al Ministero dell'Interno, in particolare sul sistema *"Sari Enterprise"*; il Ministero ha rappresentato che il sistema è destinato ad affiancare il sistema AFIS-SSA, per fornire all'operatore un supporto informatico in modo da agevolare l'attività di indagine. Il sistema AFIS-SSA, infatti, già consentiva una selezione tra le immagini presenti, caricate in occasione di fotosegnalamenti, mediante la selezione di connotati e contrassegni (ad es. colore degli occhi, tatuaggi, ecc.), inoltre tale applicativo risultava già incluso nel d.m. 24 maggio 2017[96] tra i trattamenti effettuati dal Centro elaborazione dati del Dipartimento della pubblica sicurezza.

"Sari Enterprise", pertanto, è stato presentato come un sistema che è in grado di effettuare le medesime elaborazioni di AFIS-SSA, con l'aggiunta dell'automatizzazione di alcune operazioni, che prima richiedevano l'inserimento manuale dei connotati fisici, di conseguenza non comporterebbe il trattamento di nuovi dati, ma una nuova modalità del trattamento dei medesimi dati biometrici.

Il Garante si è espresso con provvedimento del 26 luglio 2018, richiamando la disciplina posta dall'art. 7, D.Lgs. 51/2018, che consente il trattamento di categorie particolari di dati personali se specifica-

---

96 D.m. Ministero dell'Interno 24 maggio 2017, recante norme in materia di «Individuazione dei trattamenti di dati personali effettuati dal Centro elaborazione dati del Dipartimento della pubblica sicurezza o da Forze di polizia sui dati destinati a confluirvi, ovvero da organi di pubblica sicurezza o altri soggetti pubblici nell'esercizio delle attribuzioni conferite da disposizioni di legge o di regolamento, effettuati con strumenti elettronici e i relativi titolari, in attuazione dell'articolo 53, comma 3, del decreto legislativo 30 giugno 2003, n. 196».

mente previsto dal diritto dell'Unione europea o da legge o, nei casi previsti dalla legge, da regolamento; nel caso di specie, l'inclusione di AFIS-SSA tra i trattamenti disciplinati dal d.m. 24 maggio 2017 sarebbe elemento necessario e sufficiente, oltre al fatto che il sistema in parola risponde pienamente alla finalità di consentire l'attività di identificazione, che è funzionale ai compiti di polizia. È stato valutato positivamente, in definitiva, che *"Sari Enterprise"* velocizza e semplifica ciò che era già gestito legittimamente dalla Polizia, per cui è stato giudicato *non critico sotto il profilo della protezione dei dati personali*.

Esito differente ha avuto la valutazione della seconda infrastruttura *Sari*, che è stata denominata *"real time"*: lo scopo, in questo caso, è il confronto, automatico e in tempo reale, dei volti inquadrati dalle telecamere con una lista contenente quelli delle persone da ricercare (*watch list*), la cui grandezza è di massimo 10.000 volti. Se il viso è presente nella lista, il sistema lancia un *alert* e lo segnala all'operatore addetto al controllo. Il sistema, progettato e sviluppato come soluzione mobile, prevede l'installazione di una serie di telecamere in un'area geografica predeterminata e delimitata, laddove sorga l'esigenza di disporre di una tecnologia di riconoscimento facciale per coadiuvare le Forze di polizia nella gestione dell'ordine e della sicurezza pubblica o in relazione a specifiche esigenze di polizia giudiziaria. Il *"Sari real time"* consente, inoltre, di registrare i flussi video delle telecamere, fungendo, in tal senso, quale attività di videosorveglianza.

Considerata la maggiore "invasività" del sistema, prima di rilasciare la propria autorizzazione, il Garante ha richiesto che fosse redatta una valutazione d'impatto sulla protezione dei dati personali. La valutazione è stata inviata al Garante che ha rilevato come l'identificazione di una persona in un luogo pubblico comporta il trattamento biometrico di tutte le persone che circolano nello spazio pubblico monitorato; pertanto, si determinerebbe, con questa tecnologia, un'evoluzione della natura stessa dell'attività di sorveglianza, passando dalla sorveglianza mirata di alcuni individui alla possibilità di sorveglianza universale allo scopo di identificare alcuni individui.

È stato eccepito che il trattamento in parola rientrerebbe tra quelli disciplinati dalla direttiva 'LED' e decreto di attuazione interno e avrebbe ad oggetto dati rientranti tra le categorie particolari, in quanto relativi all'immagine degli individui, ma potenzialmente anche alle loro opinioni politiche o all'appartenenza a determinati sindacati, ecc., qualora il sistema sia utilizzato per monitorare manifestazioni in luogo pubblico, che hanno ricadute sulla gestione dell'ordine pubblico.

Per queste categorie di dati, l'art. 7 del D.Lgs. 51/2018 richiede una previsione del diritto dell'Unione europea o di legge o, nei casi previsti dalla legge, di regolamento. Nel proprio parere del 25 marzo 2021, il Garante, pur passando in rassegna le varie disposizioni normative citate dal Ministero nella propria valutazione, non ha trovato una base normativa adeguata, che consentisse esplicitamente l'utilizzo del sistema; pertanto, ha espresso parere negativo alla sua messa in opera. Il *"Sari real time"* non è, quindi, mai entrato in uso.

Come ulteriore prescrizione, il Garante ha chiesto anche di verificare che il sistema sia pienamente adeguato anche rispetto alle minoranze etniche. Secondo alcuni studi, infatti, tra cui quelli condotti dal NIST (*National Institute of Standards and Technologies*)[97], gli algoritmi che regolano il funzionamento dei sistemi di riconoscimento facciale avrebbero percentuali di precisione variabili: mentre nel caso di uomini di carnagione chiara si registrerebbe una precisione del 99%, quando si analizzano i volti di donne o di persone di colore, le percentuali scenderebbero, arrivando ad una precisione del 35% per le donne di pelle scura. La causa sarebbe da ricercare nei *dataset* utilizzati per alimentare le intelligenze artificiali, composte per l'80% da soggetti bianchi e per il 75% da uomini.

Successivamente all'adozione del predetto parere, è entrato in vigore il d.l. 8 ottobre 2021, n. 139, così come convertito dalla legge 3 dicembre 2021, n. 205, che all'art. 9, comma 12 ha introdotto una deroga per i trattamenti effettuati dalle autorità competenti a fini di prevenzione e repressione dei reati o di esecuzione di sanzioni penali alla regola generale prevista dal comma 9, secondo la quale «l'installazione e l'utilizzazione di impianti di videosorveglianza con sistemi di riconoscimento facciale operanti attraverso l'uso dei dati biometrici […] in luoghi pubblici o aperti al pubblico, da parte delle autorità pubbliche o di soggetti privati, sono sospese fino all'entrata in vigore di una disciplina legislativa della materia e comunque non oltre il 31 dicembre 2025». Nell'ipotesi di trattamento "di polizia" il comma 12 stabilisce, tuttavia, che è possibile l'installazione e l'utilizzazione di impianti di videosorveglianza con sistemi di riconoscimento facciale operanti attraverso l'uso dei dati

97 Agenzia del governo degli Stati Uniti d'America, che si occupa della gestione delle tecnologie di diverse discipline. Tra le proprie iniziative, nel 2024 ha lanciato un programma avanzato per *test* e valutazioni sociotecniche per l'intelligenza artificiale, denominato "ARIA" (*Assessing Risks and Impacts of AI*).

biometrici in luoghi pubblici o aperti al pubblico, essendo allo scopo sufficiente un parere preventivo favorevole del Garante[98].

Nonostante quest'apertura da parte del legislatore, allo stato, il sistema non è stato oggetto di rivalutazione, in attesa dell'approvazione da parte del legislatore nazionale di un atto normativo di adattamento dell'ordinamento giuridico interno all'*AI Act* unionale, il quale all'art. 5, paragrafo 5 demanda a ciascuno Stato membro la decisione sulla possibilità di autorizzare in tutto o in parte l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico per finalità di polizia nei limiti previsti, disciplinando tra l'altro condizioni e procedure per l'autorizzazione e per la notifica dell'uso di tali sistemi all'Autorità di protezione dei dati[99].

4.3. L'uso dell'intelligenza artificiale nelle strutture penitenziarie

In linea generale, le aree di potenziale significativo impiego dell'intelligenza artificiale nelle strutture adibite all'esecuzione della privazione della libertà personale possono raggrupparsi in:

1. Sicurezza e Sorveglianza. Uno degli impieghi più evidenti dell'IA nel contesto penitenziario riguarda il potenziamento della sicurezza. Sistemi intelligenti di videosorveglianza

98 Cfr. art. 9, comma 12, d.l. 8 ottobre 2021, n. 139: «i commi 9, 10 e 11 non si applicano ai trattamenti effettuati dalle autorità competenti a fini di prevenzione e repressione dei reati o di esecuzione di sanzioni penali di cui al decreto legislativo 18 maggio 2018, n. 51, in presenza, salvo che si tratti di trattamenti effettuati dall'autorità giudiziaria nell'esercizio delle funzioni giurisdizionali nonché di quelle giudiziarie del pubblico ministero, di parere favorevole del Garante reso ai sensi dell'articolo 24, comma 1, lettera b), del medesimo decreto legislativo n. 51 del 2018».

99 Cfr. art. 5, paragrafo 5, *AI Act*: «Uno Stato membro può decidere di prevedere la possibilità di autorizzare in tutto o in parte l'uso di sistemi di identificazione biometrica remota «in tempo reale» in spazi accessibili al pubblico a fini di attività di contrasto, entro i limiti e alle condizioni di cui al paragrafo 1, primo comma, lettera h), e ai paragrafi 2 e 3. Gli Stati membri interessati stabiliscono nel proprio diritto nazionale le necessarie regole dettagliate per la richiesta, il rilascio, l'esercizio delle autorizzazioni di cui al paragrafo 3, nonché per le attività di controllo e comunicazione ad esse relative. Tali regole specificano inoltre per quali degli obiettivi elencati al paragrafo 1, primo comma, lettera h), compresi i reati di cui alla lettera h), punto iii), le autorità competenti possono essere autorizzate ad utilizzare tali sistemi a fini di attività di contrasto. Gli Stati membri notificano tali regole alla Commissione al più tardi 30 giorni dopo la loro adozione. Gli Stati membri possono introdurre, in conformità del diritto dell'Unione, disposizioni più restrittive sull'uso dei sistemi di identificazione biometrica remota».

possono monitorare in tempo reale le aree interne ed esterne degli istituti penitenziari, analizzando i comportamenti dei detenuti e segnalando automaticamente situazioni anomale. Tali tecnologie sono in grado di rilevare atteggiamenti sospetti, identificare oggetti proibiti (come armi o telefoni cellulari) e persino anticipare potenziali conflitti o atti di violenza.

2. Gestione e Monitoraggio dei Detenuti. Tecnologie come il riconoscimento facciale, vocale, biometrico e l'analisi comportamentale consentono di identificare, verificare e monitorare i detenuti in modo automatico e con un alto grado di precisione che può migliorare anche la loro gestione all'interno del carcere.
   L'IA può offre ulteriori sistemi avanzati per la gestione dei detenuti, quali l'utilizzo di algoritmi predittivi capaci di valutare il rischio individuale di recidiva o di coinvolgimento in episodi critici. Queste analisi possono influenzare le decisioni relative alla sicurezza, alla classificazione dei ristretti e alla personalizzazione dei percorsi detentivi e riabilitativi. Tuttavia, l'utilizzo di tali strumenti solleva interrogativi importanti sul piano etico e giuridico, soprattutto in relazione alla trasparenza e all'affidabilità dei criteri algoritmici.
3. Riabilitazione e Reinserimento Sociale. Un altro ambito di utilizzo dell'IA nelle carceri riguarda i percorsi riabilitativi dei detenuti. Tecnologie intelligenti possono essere utilizzate per sviluppare programmi educativi, di formazione professionale e di supporto psicologico.
   Merita una menzione progetto *Cognify*, ideato da Hashem Al-Ghali, biologo molecolare e comunicatore scientifico yemenita; si tratta di una "prigione tecnologica" pensata per manipolare i ricordi degli atti criminosi commessi da una persona, trasportando il soggetto a vivere l'esperienza del suo reato con lo sguardo della vittima, per stimolare rimorso e pentimento nel pregiudicato. La sperimentazione di *Cognify* è iniziata sugli animali nel 2018. L'obiettivo è una riabilitazione più rapida ed economica. Una pena tecnologico-mentale che probabilmente va molto oltre la nostra idea di giustizia, condanna e riabilitazione, perché cerca di immettere nel cervello del condannato, memorie artificiali. Un'idea che sfida le nostre concezioni di giustizia, punizione e riabilitazione, sollevando al contempo profonde questioni etiche e filosofiche.

4. Supporto al Personale Penitenziario. L'intelligenza artificiale può essere impiegata per supportare il personale nelle sue funzioni operative quotidiane. L'uso di sistemi intelligenti può ridurre il carico di lavoro manuale, migliorando l'efficienza e consentendo agli operatori di concentrarsi su compiti a più alto valore relazionale e umano. Può inoltre ottimizzare la gestione delle risorse all'interno dell'istituto automatizzando compiti amministrativi, assistendo nell'identificazione di situazioni a rischio e monitorando il benessere del personale.

Sebbene esempi specifici e ampiamente pubblicizzati siano limitati per ragioni di sicurezza, l'indagine sull'uso di sistemi potenziati dall'IA nelle carceri rivela una crescente attenzione verso l'impiego di tecnologie sempre più avanzate. Emerge, altresì, che l'impiego attuale nel sistema penitenziario presenta un panorama internazionale eterogeneo; l'utilizzo concreto di sistemi basati su IA all'interno delle amministrazioni penitenziarie, sia europee che extraeuropee, evidenzia infatti un quadro estremamente variegato, caratterizzato da differenze significative in termini di obiettivi, livello di implementazione e approccio etico.

Nei Paesi *extra*-UE, dove l'IA nelle carceri è già in fase avanzata di sperimentazione e, in alcuni casi, di applicazione sistematica, le tecnologie adottate sono prevalentemente orientate a finalità di sicurezza, sorveglianza e gestione del rischio, con un impiego massiccio di strumenti digitali finalizzati al controllo dei detenuti.

In Cina, ad esempio, sono stati introdotti sistemi sofisticati di videosorveglianza potenziati da riconoscimento facciale e analisi predittiva del comportamento, capaci di rilevare movimenti anomali, segnali di agitazione o tentativi di autolesionismo. Telecamere e sensori intelligenti vengono spesso integrati con braccialetti biometrici che monitorano i parametri vitali dei detenuti, trasmettendo dati in tempo reale agli operatori penitenziari, garantendo un rapido intervento in caso di problemi di salute.

Negli Stati Uniti, l'IA viene utilizzata per monitorare le comunicazioni telefoniche e i contenuti digitali dei detenuti. Sistemi di elaborazione del linguaggio naturale analizzano le conversazioni alla ricerca di parole chiave sospette che potrebbero indicare la pianificazione di attività illecite. In alcuni istituti penitenziari si sperimentano inoltre programmi di realtà virtuale per supportare la riabilitazione, la gestione dello stress e il reinserimento dei detenuti nella società. Il Dipartimento di correzione della Pennsylvania, ad esempio, ha acquistato diversi vi-

sori e fatto sviluppare una serie di esperienze in realtà virtuale per far incontrare genitori detenuti e figli nel metaverso, così da creare una forma di interazione tra le due parti più intima rispetto ai rari incontri nelle sale colloqui degli istituti penitenziari. In Maryland invece i detenuti reclusi da molti anni e in procinto di lasciare le carceri per la fine della condanna sono stati coinvolti in sedute di *training* in realtà virtuale per riprendere confidenza con il mondo esterno, con un *focus* particolare sul lavoro. Un modo per favorire il reinserimento sociale una volta liberi, secondo gli ideatori del progetto, e abbattere la recidiva.

Ad Hong Kong sono stati sperimentati *robot* di sorveglianza dotati di microfoni, telecamere e sensori ambientali, che pattugliano autonomamente le sezioni detentive. L'obiettivo è duplice: rafforzare la capacità di controllo da remoto e ridurre il carico psicofisico del personale penitenziario. Tecnologie simili sono state introdotte anche in Corea del Sud già a partire dal 2012.

Le carceri di Singapore impiegano una sorveglianza basata sull'IA nelle celle, il riconoscimento facciale per i controlli del numero dei detenuti e un sistema di rilevamento del comportamento, con l'obiettivo di migliorare l'efficienza e la sicurezza, ma sollevando preoccupazioni in merito alla *privacy* e a potenziali pregiudizi.

Nel contesto europeo, l'adozione dell'IA in ambito penitenziario è più prudente e in genere maggiormente orientata al rispetto dei diritti umani, alla trasparenza, e al supporto nei percorsi riabilitativi. L'utilizzo dell'intelligenza artificiale tende ad affiancarsi a strategie educative e sociali, piuttosto che sostituirsi ad esse.

Paesi come la Finlandia si sono distinti per approcci innovativi. Nel progetto pilota *"Smart Prison"*[100], i detenuti vengono coinvolti in attività di *data labeling* e classificazione per l'addestramento di algoritmi, acquisendo così competenze digitali richieste e spendibili nel mondo del lavoro. L'obiettivo non è solo l'efficienza, ma la formazione professionale e il reinserimento sociale, in linea con il principio di normalità che ispira il sistema penitenziario nordico.

Nel Regno Unito, alcune carceri hanno introdotto sistemi di videosorveglianza intelligenti per il rilevamento di oggetti proibiti e comportamenti sospetti. Un esempio emblematico è la prigione di Liverpool, dove le telecamere monitorate da IA vengono impiegate per

---

100 Progetto avviato in collaborazione con la Metroc AI, azienda finlandese fondata nel 2019 e con sede a Helsinki, che sviluppa soluzioni *software* basate sull'intelligenza artificiale per i settori delle costruzioni e dell'immobiliare.

prevenire il contrabbando di droghe, telefoni e armi e per individuare comportamenti sospetti. Inoltre, sono in corso ricerche sull'utilizzo del *machine learning* per valutare la probabilità e prevenire episodi di violenza all'interno degli istituti.

In Spagna, la Catalogna ha sviluppato e implementato da oltre quindici anni il sistema algoritmico *RisCanvi*, uno strumento predittivo volto a valutare il rischio di recidiva o violenza da parte dei detenuti, che influisce direttamente sul trattamento penitenziario e sulla concessione dei benefici. Tuttavia, questo sistema ha sollevato dibattiti etici, in particolare per il fatto che si basa su fattori statici legati al passato del detenuto, sollevando dubbi in merito al rispetto dei diritti fondamentali e al principio del trattamento individualizzato[101].

Infine, in merito alla situazione italiana, il Dipartimento dell'Amministrazione Penitenziaria (DAP) ha avviato diverse iniziative volte all'innovazione tecnologica del sistema penitenziario nazionale, con l'obiettivo di potenziare la sicurezza[102], ottimizzare la gestione dei detenuti e promuovere efficaci percorsi di reinserimento sociale[103]. Tra le

---

101 Il *RisCanvi* ha come obiettivo quello di prevedere la probabilità con cui un detenuto può mettere in atto un comportamento violento; in particolare mira a prevedere il rischio di violenza etero-diretta, di violenza intramuraria e di recidiva violenta. Il livello di rischio individuato determina il programma di trattamento del singolo detenuto. L'algoritmo si divide in *RisCanvi screening* e *RisCanvicomplet*, rispettivamente relativi alla valutazione e alla gestione del rischio. La valutazione del rischio viene effettuata sulla base di dieci fattori, alcuni dei quali strettamente connessi alle caratteristiche socio-demografiche, biografiche o penali della persona detenuta, ovvero elementi statici pertanto non modificabili da parte dell'individuo. Negli ultimi anni sono state sollevate preoccupazioni in merito al fatto che il livello di rischio individuato attraverso la scala predittiva *RisCanvi* ha delle implicazioni dirette circa le condizioni di vita all'interno del carcere e, quindi, anche sui diritti fondamentali dei soggetti detenuti. In particolare, il fatto che il passato individuale della persona detenuta e determinate variabili soggettive possano comportare che una persona debba passare più tempo in carcere in quanto non idonea a ottenere benefici penitenziari, oppure che debba essere soggetta a un regime di vita più duro, risulta assai problematico visto che la Costituzione spagnola (art 25.2) pone come fine ultimo di qualsiasi condanna il reinserimento sociale.

102 Nel luglio 2023, il DAP ha siglato un protocollo d'intesa con l'Ente Nazionale Aviazione Civile (ENAC) per l'utilizzo di droni negli istituti penitenziari. Questa collaborazione mira a potenziare la sorveglianza aerea dei perimetri esterni e interni delle carceri, facilitando il monitoraggio durante rivolte o manifestazioni di protesta. I droni saranno impiegati anche nella ricerca di detenuti evasi e nel contrasto all'introduzione illecita di oggetti proibiti tramite dispositivi aerei.

103 Nel dicembre 2024, il DAP ha avviato un'iniziativa congiunta con il Fondo per la Repubblica Digitale per promuovere la formazione digitale dei detenuti. L'obiettivo è facilitare il loro reinserimento sociale e lavorativo attraverso lo sviluppo delle competenze digitali. Inoltre, nel luglio 2023, è stato firmato un protocollo d'intesa con la *Cyber Security Italy*

priorità del DAP vi è la digitalizzazione delle strutture detentive, attraverso progetti mirati all'aggiornamento tecnologico delle infrastrutture e dei processi amministrativi. Questi interventi mirano a rendere il sistema più efficiente e trasparente, ma anche a creare condizioni più adeguate alla formazione[104], l'accesso ai servizi e la riabilitazione dei detenuti. In quest'ottica, l'impiego di tecnologie avanzate si sta facendo progressivamente strada anche negli istituti penitenziari italiani, comprese alcune soluzioni basate sull'intelligenza artificiale, sebbene queste in fase ancora progettuale o in via sperimentale.

Una sperimentazione dell'intelligenza artificiale nel contesto penitenziario italiano è rappresentata dal progetto pilota, promosso dall'Azienda Sociosanitaria Ligure, che prevede la realizzazione di visite mediche nel metaverso presso la casa di reclusione di Chiavari (Genova). Sebbene ancora in fase di sviluppo, questa iniziativa rappresenta una novità per l'Italia, ispirata a progetti esteri simili. Nello specifico, grazie all'uso di visori VR, i detenuti selezionati hanno partecipato a consultazioni mediche in un ambiente virtuale ricostruito come uno spazio sanitario più accogliente rispetto alle infermerie degli istituti. Dotati di questi appositi visori, i ristretti hanno quindi svolto teleconsulti medici interagendo con l'*avatar* del medico, alcuni seduti virtualmente a un tavolo, altri esplorando l'ambiente esterno per una temporanea sensazione di libertà durante il colloquio. La sperimentazione è stata apprezzata da detenuti e azienda sanitaria, che intende acquistare i visori (finora in prestito) per lanciare ufficialmente il progetto di telemedicina con IA.

Gli esempi citati e, soprattutto, i numerosi interrogativi riguardanti il funzionamento degli strumenti di intelligenza artificiale in contesto detentivo, mostrano come il loro utilizzo rappresenti, prima di tutto, una sfida etica. Pur offrendo un potenziale supporto alle amministrazioni penitenziarie, tali tecnologie sollevano importanti questioni legate alla tutela della *privacy* e alla salvaguardia della dignità delle persone private della libertà. Inoltre, come già anticipato, l'affidamento di decisioni cruciali, come la valutazione del rischio di recidiva

---

Foundation e la Camera Penale di Roma per introdurre competenze digitali e cultura della cybersicurezza negli istituti penitenziari, sensibilizzando i detenuti sui rischi delle tecnologie e promuovendo comportamenti responsabili *online*.

104 A luglio 2023, è stata inaugurata una nuova aula digitalizzata presso la Casa Circondariale Rebibbia Nuovo Complesso "Raffaele Cinotti". Questo progetto, frutto della collaborazione tra l'Università di Roma Tor Vergata e il DAP, offre ai detenuti l'opportunità di seguire lezioni a distanza, accedere a risorse educative *online* e migliorare la preparazione per gli esami, facilitando il loro percorso di formazione e reinserimento sociale.

o l'assegnazione a determinate misure penitenziarie, a un algoritmo solleva interrogativi sulla giustizia e sulla responsabilità. Gli algoritmi possono essere influenzati da *bias* impliciti o generare risultati imprecisi, con il rischio di dar luogo a discriminazioni nei confronti dei detenuti e di compromettere il rispetto dei loro diritti fondamentali.

Le opportunità e le criticità derivanti dall'utilizzo dell'intelligenza artificiale nei contesti di detenzione e libertà vigilata costituiscono il fulcro della Raccomandazione CM/Rec(2024)5 del Comitato dei Ministri del Consiglio d'Europa, adottata il 9 ottobre 2024. Il documento affronta gli aspetti etici e organizzativi dell'uso dell'IA e delle tecnologie digitali nei servizi penitenziari e di *probation*, delineando princìpi guida per un'implementazione responsabile e rispettosa dei diritti umani, con l'obiettivo di bilanciare i benefici dell'IA e la tutela dei diritti individuali[105]. Il testo definisce nove princìpi fondamentali che devono

---

105 Tra le disposizioni generali della Raccomandazione viene esplicitato chi sono i destinatari della raccomandazione (le autorità pubbliche responsabili dei servizi penitenziari e di libertà vigilata e le aziende private che si occupano di tali tecnologie). Si pone poi l'accento sulle esigenze specifiche dei minori inseriti nel circuito penale e si introducono i contesti in cui si ritiene pertinente l'utilizzo di IA e delle tecnologie digitali correlate. Vengono anche esposti i princìpi base dell'utilizzo di strumenti di IA in ambito penitenziario, nell'ordine: il rispetto della dignità umana e dei diritti fondamentali; il principio di legalità, della certezza del diritto e di responsabilità; il principio di uguaglianza e non discriminazione; quello di proporzionalità, efficacia e necessità dell'IA; il principio di buona *governance*, trasparenza, tracciabilità e spiegabilità; quello del diritto a una verifica umana delle decisioni; il principio di qualità, affidabilità e sicurezza; il principio dell'uso incentrato sull'uomo dell'IA e delle tecnologie digitali correlate; infine il principio dell'IA e dell'alfabetizzazione digitale, che si riferisce al fatto che le modalità di utilizzo, le finalità e le regole etiche di queste tecnologie devono essere comprensibili agli utenti. Nel quarto paragrafo del testo della raccomandazione, dedicato alla protezione dei dati e della *privacy*, si legge che, nel caso di utilizzo dell'IA nei confronti di autori di reato, i dati raccolti non devono essere conservati in modo che sia consentita l'identificazione delle persone coinvolte oltre un periodo di tempo strettamente funzionale al loro fine. Allo stesso modo, devono essere utilizzati solamente i dati personali di quantità e tipo necessari allo scopo per cui vengono impiegati e, quando possibile, devono essere anonimizzati. L'utilizzo di categorie speciali di dati deve avvenire in ambienti controllati.
La Raccomandazione tratta poi dell'uso dell'intelligenza artificiale in relazione alle esigenze di sicurezza, protezione e buon ordine; in particolare nella formazione del personale penitenziario e dell'utilizzo di queste tecnologie nel monitoraggio elettronico, che dovrebbe essere comunque regolato da controllo umano, proporzionato allo scopo e teso a favorire il reinserimento dei detenuti. Successivamente viene trattato l'argomento dell'uso di IA nel caso di gestione di fascicoli dei detenuti per generare avvisi automatici; questo non comporta una sostituzione del personale; infatti, non va tralasciata l'importanza della figura professionale nella decisione finale, che va a verificare quanto stabilito dall'intelligenza artificiale. La Raccomandazione si occupa anche delle possibili distorsioni algoritmiche e della rappresentatività dei dati, a cui si dovrebbe porre particolare attenzione al fine di evitare ogni tipo di discriminazione, tenendo in considerazione la prospettiva di

guidare la progettazione, lo sviluppo, l'uso e la dismissione dell'IA e delle tecnologie digitali correlate. Questi princìpi includono legalità, certezza del diritto, responsabilità, buona *governance*, trasparenza, tracciabilità, esplicabilità, diritto a un riesame umano delle decisioni e un uso dell'IA centrato sull'essere umano, fornendo così un quadro concreto per un'implementazione etica. Un principio cardine è che l'impiego dell'IA nelle diverse aree applicative all'interno degli istituti penitenziari (come la sicurezza, la gestione dei detenuti, la valutazione del rischio, i percorsi riabilitativi e il supporto al personale) deve sempre rispettare i diritti fondamentali e preservare la dignità della persona detenuta. In questa ottica, il Consiglio d'Europa sollecita gli Stati membri ad adottare sistemi IA conformi ai princìpi di legittimità, proporzionalità e finalità rieducativa, invitando a integrare questi criteri nelle normative e politiche nazionali.

La Raccomandazione sottolinea anche l'importanza dell'interazione umana e della supervisione, precisando che l'IA non dovrebbe sostituire il personale penitenziario e di *probation*, ma piuttosto assisterlo nelle sue funzioni quotidiane, contribuendo così al sistema di giustizia penale, all'esecuzione delle sanzioni e alla riduzione della recidiva.

In conclusione, l'intelligenza artificiale ha il potenziale di trasformare profondamente il sistema penitenziario, migliorando la gestione della sicurezza, la personalizzazione dei percorsi riabilitativi e il supporto al personale. Tuttavia, affinché il suo impiego non sfoci in abusi, è fondamentale che venga regolato da princìpi chiari, ispirati a giustizia, trasparenza, equità e proporzionalità. L'obiettivo non deve essere il semplice rafforzamento del controllo, ma la promozione di una riforma penitenziaria orientata alla riabilitazione e al reinserimento sociale.

---

genere e il multiculturalismo. Si parla poi del possibile utilizzo dell'IA al fine di elaborare diagnosi mediche per le persone detenute. Tuttavia, l'utilizzo di strumenti di questo tipo connessi alla telemedicina non possono mai sostituire la figura professionale sanitaria; il contatto umano resta imprescindibile.

Una delle ultime questioni affrontate dalla raccomandazione riguarda l'uso dell'IA e delle tecnologie digitali correlate per la selezione, la gestione, la formazione e lo sviluppo del personale penitenziario; l'intelligenza artificiale, infatti, può configurarsi anche come un'importante risorsa per la formazione dello *staff* e anche per prendere decisioni manageriali, sempre rispettando i diritti del personale. Nel paragrafo dedicato alla ricerca, allo sviluppo, alla valutazione e alla revisione periodica, si solleva il tema del finanziamento e del sostegno a progettazione, sviluppo e ricerca sull'intelligenza artificiale. Infine, il Comitato dei Ministri esprime la necessità di un aggiornamento e di una revisione periodica del testo della Raccomandazione CM/Rec(2024)5, affinché continui a tutelare i diritti umani e le libertà fondamentali degli utenti e la sicurezza sociale.

Capitolo V
La gestione del rischio ed i profili di responsabilità legati all'uso dell'intelligenza artificiale

di Gennaro Corrado*

5.1. I profili di rischio connessi all'uso dell'intelligenza artificiale nei sistemi a supporto delle decisioni

Venendo a trattare dei profili di rischio connessi all'impiego dell'intelligenza artificiale nei sistemi a supporto delle decisioni, oltre a quanto già illustrato in merito agli aspetti riguardanti il trattamento dei dati personali, occorre focalizzare i fattori di criticità ricollegabili a due ulteriori temi: i rischi in termini sicurezza cibernetica e il pericolo che tali sistemi automatizzati possano risentire di attività di disinformazione.

Per quanto attiene al primo aspetto si rileva che l'Agenzia per l'Italia Digitale, ha fornito un dettagliato e schematico quadro di insieme in relazione alla gestione dei rischi di sicurezza cibernetica derivanti dall'adozione dell'intelligenza artificiale, dedicandovi ampio spazio nell'ambito della bozza delle «Linee guida per l'adozione dell'Intelligenza Artificiale nella Pubblica Amministrazione», resa accessibile in consultazione pubblica dal 18 febbraio al 20 marzo 2025[106].

Sul punto si reputa opportuno soffermarsi sugli aspetti salienti, ripercorrendo i contenuti del citato documento, che, per chiarezza argomentativa, si propongono suddividendoli schematicamente come segue:

---

* Vice Questore della Polizia di Stato, già frequentatore del XL Corso di Alta Formazione presso la Scuola di Perfezionamento per le Forze di Polizia.

106 Determinazione del Direttore Generale del 17.02.2025, avente per oggetto «Avvio dell'*iter* di consultazione e informazione delle Linee guida per l'adozione dell'Intelligenza Artificiale nella Pubblica Amministrazione».

- classificazione dei sistemi di intelligenza artificiale sulla base del rischio;
- tipologie di attacchi informatici specificamente collegati ai *software* di IA;
- gestione del rischio cibernetico.

5.1.1. Classificazione dei sistemi di IA sulla base del rischio

Come evidenziato nel capitolo II, in linea con quanto definito dal legislatore unionale nell'*AI Act* i sistemi di intelligenza artificiale sono oggetto della seguente classificazione in base ai livelli di rischio:

- Sistemi di IA vietati: sono i sistemi dal rischio inaccettabile e, pertanto, vietati. Vi rientrano i *software* che agiscono senza che una persona ne sia consapevole o che adottino tecniche volutamente manipolative o ingannevoli, volte a distorcere il comportamento delle persone; meccanismi di sfruttamento delle persone vulnerabili; sistemi che introducono meccanismi di "*scoring* sociale";
- Sistemi di IA ad alto rischio: sono i sistemi che pongono elevati rischi poiché, in caso di malfunzionamento, rappresentano una potenziale rilevante dannosità, con elevata compromissione di interessi pubblici e diritti fondamentali[107];
- Sistemi di IA a rischio limitato: sono sistemi dal rischio minore che, in caso di malfunzionamento, si contraddistinguono per una potenziale dannosità limitata, con un impatto ridotto su interessi pubblici e diritti fondamentali;
- Sistemi di IA a rischio minimo o nullo: sono sistemi dal rischio irrilevante, la cui dannosità, in caso di malfunzionamento, è molto limitata.

È di fondamentale rilevanza evidenziare che, per quanto concerne i sistemi di IA ad alto rischio, l'art. 14 dell'*AI Act* definisce un obbligo di sorveglianza umana, prevedendo in particolare che tali applicativi «sono progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui sono in uso. La sorveglianza umana mira a prevenire o ridurre al minimo i rischi per la salute, la sicu-

107 Per qualificare un sistema come ad alto rischio le pubbliche amministrazioni devono fare riferimento alla classificazione delineata dall'*AI Act* all'art. 6 e all'Allegato III.

rezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile».

5.1.2. Tipologie di attacco informatico specificamente connesse all'IA

Nell'evidenziare che la conoscenza delle specifiche tassonomie di attacco informatico ai sistemi di intelligenza artificiale è necessaria ai fini di attuare azioni di protezione e contenimento del rischio, l'AGID ha ricalcato la distinzione sviluppata dal NIST[108] tra le seguenti macro-categorie di attacchi:

- *Evasion attack*: hanno come obiettivo quello di generare un errore nella classificazione del modello, introducendo perturbazioni negli *input*, spesso impercettibili all'uomo, denominate *adversarial example*;
- *Poisoning attack*: hanno come obiettivo degradare le prestazioni di un modello o fargli generare uno specifico risultato alterando i dati di addestramento; è il caso del *label flipping* nel quale l'attaccante cambia l'etichetta dei dati di addestramento con l'obiettivo di addestrare il modello sulla base dell'etichetta da lui scelta;
- *Privacy attack*: hanno come obiettivo compromettere le informazioni degli utenti ricostruendole a partire dai dati di addestramento;
- *Abuse attack*: hanno come obiettivo alterare il comportamento di un sistema di IA generativa per adattarlo ai propri scopi come, ad esempio, perpetrare frodi, distribuire *malware* e manipolare informazioni.

5.1.3. Gestione integrata della sicurezza dei sistemi di IA

La sicurezza dei sistemi di intelligenza artificiale è definita come l'insieme di strumenti, strategie e processi implementati per identificare e prevenire le minacce e gli attacchi che potrebbero compromettere la

---

108 *National Institute of Standards and Technology*, agenzia governativa degli Stati Uniti d'America.

riservatezza, l'integrità o la disponibilità di un modello di IA o di un sistema abilitato all'IA. Rappresenta, dunque, un requisito essenziale per l'adozione, l'acquisizione e lo sviluppo dell'intelligenza artificiale nella pubblica amministrazione poiché, se da un lato tali applicativi offrono strumenti utili per rispondere alla crescente necessità di migliorare l'efficienza e l'efficacia nella gestione e nell'erogazione dei servizi pubblici, dall'altro tale tecnologia introduce nuovi rischi, minacce e vulnerabilità che devono essere presi in considerazione nella sua implementazione.

Per perseguire efficacemente gli obiettivi di protezione dai rischi di sicurezza cibernetica, l'AGID ha evidenziato la necessità di assicurare una «gestione integrata della sicurezza dei sistemi di IA», messa a fuoco attraverso l'immagine che segue[109].

La sicurezza di tali applicativi si incentra, dunque, in una costante attività di contenimento del rischio nell'arco di tutto il ciclo di vita dei modelli di intelligenza artificiale[110], articolata nelle seguenti attività.

1. Adottare l'IA in modo responsabile: modelli, applicazioni e sistemi di intelligenza artificiale devono essere adottati dopo essere stati sottoposti a valutazione di sicurezza; vanno indi-

---

109 Immagine tratta dalla citata bozza delle «Linee guida per l'adozione dell'Intelligenza Artificiale nella Pubblica Amministrazione».

110 Il ciclo di vita di tali *software* è stato che è stato ricondotto a dieci fasi: 1) definizione dell'obiettivo; 2) raccolta, esplorazione e pre-elaborazione dei dati; 3) selezione delle caratteristiche; 4) selezione del modello; 5) addestramento del modello; 6) *tuning* del modello; 7) trasferimento dell'apprendimento; 8) distribuzione del modello; 9) manutenzione del modello; 10) valutazione del raggiungimento dell'obiettivo.

cati chiaramente agli utenti le eventuali limitazioni note, i potenziali guasti, gli aspetti di sicurezza di cui sono responsabili e le modalità relative alla gestione dei dati personali;
2. Identificare, tracciare, mantenere e proteggere gli *asset*[111]: devono essere previsti processi e strumenti per identificare, tracciare, mantenere e proteggere gli *asset* dei sistemi di intelligenza artificiale prevedendo, ad esempio, specifici controlli per la gestione e protezione dei dati e dei contenuti generati dai sistemi; devono essere inoltre implementati controlli per proteggere la riservatezza, l'integrità e la disponibilità dei *log*;
3. Proteggere la catena di approvvigionamento: la sicurezza della *supply chain* dei vari componenti deve essere valutata e monitorata durante l'intero ciclo di vita dei sistemi; i componenti devono essere acquisiti da fornitori verificati e devono essere adeguatamente protetti e documentati; è fondamentale analizzare la gestione dei dati di terze parti, assicurandosi che siano implementate misure di sicurezza adeguate, come la crittografia e i controlli di accesso; è necessario richiedere ai fornitori che le loro misure di sicurezza siano allineate con le politiche di sicurezza adottate e con l'analisi dei rischi effettuata;
4. Proteggere il modello e i dati: i modelli e i dati dei sistemi di intelligenza artificiale devono essere protetti da accessi impropri o manomissioni. Un attaccante può essere in grado di ricostruire la funzionalità di un modello o i dati su cui è stato addestrato, accedendo direttamente ad essi (acquisendo i pesi del modello) o indirettamente (interrogando il modello tramite un'applicazione o un servizio); può inoltre manomettere i modelli, i dati o le richieste durante o dopo la fase di addestramento del modello, rendendone il risultato non affidabile. È possibile proteggere il modello e i dati dall'accesso adottando le *best practice* di *cybersecurity* e implementando controlli sull'interfaccia di interrogazione del modello per rilevare e prevenire i tentativi di accesso, modifica ed esfiltrazione delle informazioni;
5. Monitorare il comportamento del sistema e degli *input*: devono essere monitorati i risultati e le prestazioni del modello e del sistema di IA in modo da poter osservare e rispondere a

111 Un sistema di intelligenza artificiale è costituito da: dati; modelli; attori; processi; strumenti; artefatti.

cambiamenti improvvisi e graduali del comportamento che influiscono sulla sicurezza come, ad esempio, potenziali intrusioni e compromissioni. In linea con i requisiti di protezione dei dati, anche personali, gli *input* devono essere monitorati e registrati per consentire verifiche di conformità, *audit*, analisi e ripristino in caso di compromissione o uso improprio;

6. Sviluppare un piano di risposta agli incidenti: deve essere definito e attuato un piano di risposta agli incidenti che riflette i diversi scenari di minaccia; il piano deve essere revisionato periodicamente e al verificarsi di eventi interni (come, ad esempio, l'aggiornamento di piani strategici o modifiche organizzative), eventi esterni (come l'evoluzione del contesto normativo e legislativo) o mutamenti dell'esposizione alle minacce e ai relativi rischi regolarmente valutati con l'evoluzione del sistema e della ricerca in generale;
7. Formare e sensibilizzare il personale sulle minacce e sui rischi: gli attori responsabili dell'adozione dell'IA devono comprendere le minacce e le relative mitigazioni; i *data scientist* e gli sviluppatori devono essere formati allo sviluppo sicuro del *software* e alle pratiche di sistemi di IA sicuri e responsabili;
8. Proteggere l'infrastruttura ICT: le infrastrutture che ospitano i sistemi devono essere adeguatamente protette; le relative vulnerabilità possono essere infatti sfruttate da un attaccante per perpetrare attacchi al sistema come, ad esempio, la compromissione del modello o la degradazione delle sue prestazioni; pertanto, devono essere adottate misure di *cybersecurity* adeguate ai connessi rischi sulle infrastrutture ICT utilizzate in ogni parte del ciclo di vita dei sistemi di IA;
9. Proteggere le identità e gli accessi: la gestione delle identità e degli accessi è fondamentale per proteggere i sistemi di intelligenza artificiale, segnatamente quando si trattano categorie particolari di dati. I principali rischi associati includono accessi non autorizzati, furti di dati e manipolazioni dei modelli, che possono compromettere l'integrità e la riservatezza delle informazioni; al fine di mitigare questi rischi e mantenere la sicurezza complessiva dei sistemi di IA, è fondamentale implementare controlli centralizzati che consentano una supervisione efficace degli accessi, definire chiaramente ruoli e permessi per garantire un accesso appropriato e utilizzare l'autenticazione multi-fattore.

5.1.4. Rischi di disinformazione

L'implementazione dell'intelligenza artificiale nell'attività di polizia pone in risalto un ulteriore sensibile profilo di rischio, rappresentato dall'eventualità che fenomeni di disinformazione possano ripercuotersi negativamente sull'attività di *open source intelligence* e *social media intelligence*, svolta attraverso processi automatizzati mediante tali applicativi[112].

---

112 Pollicino O., op. cit., evidenzia la distinzione tra i tre fenomeni della disinformazione, misinformazione e malinformazione nei termini seguenti: «nonostante sia invalsa sempre più nel linguaggio comune la locuzione inglese facente riferimento alle cosiddette *"fake news"*, l'utilizzo di tale espressione è in anni recenti andata sempre più incontro a critiche […] In tal senso si è espresso, in particolare, l'*High Level Group on fake news and online disinformation* (HLEG), istituito dalla Commissione europea, il quale, nel suo Rapporto finale pubblicato nel 2018, ha rigettato la locuzione *"fake news"*. L'HLEG ha argomentato, evidenziando, essenzialmente, due motivi: in primo luogo, si è rilevato come il concetto di "notizie false" rischi di essere riduttivo in quanto esclude tutta una serie di manifestazioni del fenomeno; in secondo luogo, l'utilizzo mediatico e popolare della stessa espressione *"fake news"* ha nel tempo condotto a una connotazione fuorviante del fenomeno, soprattutto alla luce dell'uso strumentale che ne è stato fatto da parte di alcune personalità politiche e dei loro sostenitori, al fine di screditare e respingere notizie, informazioni e opinioni non volute. A fronte di ciò, appare preferibile il ricorso al concetto di "disinformazione", che, secondo l'HLEG, ricomprende tutte quelle forme di manifestazione della libertà che si sostanziano di quelle informazioni false, inaccurate o fuorvianti, sovente artificiosamente create, le quali siano presentate e diffuse con lo scopo precipuo di trarre un profitto di carattere economico, politico o ideologico e/o di provocare un danno a livello pubblico, ivi inclusa l'ingerenza nei processi elettorali e democratici. […] Occorre, peraltro, precisare che la nozione di disinformazione, così intesa, include non tanto quelle condotte che l'ordinamento riconosce come inerentemente illecite (si pensi, per esempio, al caso della diffamazione o a quello della calunnia), ma, piuttosto, tutti quei casi ove la produzione e la diffusione di contenuto falso non integra, di per sé, una fattispecie illecita, ma produce, comunque, un potenziale danno a princìpi e valori democratici. Sono del resto proprio questi ultimi casi a rappresentare la maggiore sfida per il legislatore nazionale e per quello europeo, in quanto impongono la necessità di contemperare l'interesse pubblico connesso al corretto funzionamento dello stato democratico con la necessità di non arrecare uno sproporzionato nocumento alla libertà individuale dei consociati, che pongono in essere condotte che l'ordinamento ritiene, di per sé, lecite. L'elemento della volontarietà della creazione e diffusione di un contenuto falso, così come la volontà di causare un danno o di ottenere un profitto, rappresenta, peraltro, l'elemento di discrimine tra il fenomeno della disinformazione e quello, connesso ma ben diverso, della "misinformazione". Quest'ultimo "disordine informazionale" (*information disorder*) si caratterizza, infatti, per un elemento di carattere soggettivo, in quanto implica precisamente l'assenza della volontà di diffondere informazioni false e, pertanto, si sostanzia nella diffusione di materiale considerato genuino […] Fenomeno ancora differente è quello della malinformazione, che consta della diffusione di informazioni rispondenti al vero, o comunque basate su elementi fattuali reali, le quali vengono tuttavia comunicate e diffuse al preciso fine di provocare un danno».

Va infatti sottolineata la diffusione di sistemi di IA generativa in grado di creare immagini, video e testi altamente realistici, come pure il fatto che le applicazioni di intelligenza artificiale contribuiscono ad incrementare esponenzialmente la diffusione di informazioni false, inaccurate o fuorvianti. Tali contenuti, captati a loro volta dagli applicativi di intelligenza artificiale utilizzati nella SOCMINT, possono falsarne i risultati e, pertanto, rendono necessaria l'adozione e la costante implementazione delle contromisure già analizzate.

5.1.5. Deepfake e altri contenuti irreali

Il riferimento è in particolare ai *deepfake*, immagini e video dal contenuto estremamente realistico, realizzati tramite tecniche avanzate che consentono di manipolare gli attributi facciali, modificando caratteristiche riferibili al volto della persona ritratta, ad esempio riproducendo fattezze ringiovanite o invecchiate (*face attribute manipulation*); sostituzione del volto presente nell'immagine o nel video originale con un viso differente (*face swap*); modifica delle espressioni e del movimento del volto così da sincronizzare l'immagine con un contenuto audio a sua volta falso (*face reenactment* e *lip synching*).

Come accennato, inoltre, la produzione della disinformazione attraverso applicativi di intelligenza artificiale riguarda la realizzazione di contenuti audio in cui, partendo da registrazioni o *input* testuali, sono riprodotte particolarità vocali, incluso il timbro, di specifiche persone.

Analogamente, tali applicativi possono essere impiegati anche per la realizzazione di contenuti testuali o articoli corredati da immagini estremamente realistici ma del tutto irreali. Come posto in evidenza nell'ambito degli approfondimenti sul fenomeno, a riprova del crescente ruolo dell'intelligenza artificiale nella produzione di disinformazione e manipolazione dell'informazione, negli ultimi anni si è registrato un rilevante incremento percentuale delle notizie false generate – o comunque manipolate – attraverso l'uso dell'intelligenza artificiale[113].

---

113 Pollicino O., op. cit.: «a puro titolo indicativo del crescente ruolo dell'IA nella produzione di disinformazione e manipolazione dell'informazione, appare di particolare interesse un dato che emerge dalla banca dati di *Newsguard*, dedicata alla raccolta e catalogazione di narrative false (e virali) in rete. Nel periodo tra il 1° gennaio 2021 e il 30 maggio 2024, il *database* contiene 930 voci di false narrative: di queste 930 voci, 51 (ovverosia, il 5,48%) sono catalogate come notizie generate attraverso l'uso dell'IA (*AI-Generated Media*) o, comunque, come media manipolati o artificiosamente fabbricati (*Manipulated or Fabricated*

5.1.6. Contrasto alla disinformazione nel disegno di legge nr. 1146/2024

In proposito appare opportuno evidenziare i contenuti del disegno di legge nr. 1146/2024, di iniziativa governativa, che prevede, in particolare, l'introduzione di modifiche al codice penale relative all'utilizzo dell'intelligenza artificiale e, soprattutto, l'introduzione dell'art. 612-*quater* del codice penale, con cui si prevede la punibilità per il reato di «illecita diffusione di contenuti generati o manipolati artificialmente».

Nel quadro di tale atto di iniziativa legislativa, si prevede l'adozione di disposizioni di natura penale riguardanti l'utilizzo degli applicativi di intelligenza artificiale, definendo un'aggravante comune all'art. 61, nonché specifiche circostanze aggravanti in diversi reati, tra cui quelli di sostituzione di persona, truffa, frode informatica, riciclaggio e autoriciclaggio.

In particolare, il disegno di legge prevede l'introduzione dell'art. 612-*quater* nel codice penale, nella cui previsione «chiunque cagiona ad altri un danno ingiusto, mediante invio, consegna, cessione, pubblicazione o comunque diffusione di immagini o video di persone o di cose ovvero di voci o suoni in tutto o in parte falsi, generati o manipolati mediante l'impiego di sistemi di intelligenza artificiale, atti a indurre in inganno sulla loro genuinità o provenienza, è punito con la reclusione da uno a cinque anni».

Tra i principali profili di novità di tale previsione normativa si evidenzia come si punti a punire la diffusione di immagini o video realizzati specificamente attraverso sistemi di intelligenza artificiale, a patto che non si tratti di falsi macroscopici o grossolani, dovendo essere comunque idonei a «indurre in inganno sulla loro genuinità o provenienza».

Si rileva, infatti, che l'articolo così proposto individua un profilo di responsabilità nella condotta di chi cagioni un danno ingiusto mediante «invio, consegna, cessione, pubblicazione o comunque diffusione», non prevedendo dunque una responsabilità penale per l'autore del contenuto tramite l'IA – credibilmente, comunque, riconducibile ad altri illeciti – ma puntando a sanzionare penalmente la condotta di chi arrechi nocumento alla persona offesa tramite la diffusione del contenuto artefatto.

Media). È, tuttavia, da notare come la percentuale di tali categorie di contenuti generati attraverso l'uso di IA (o comunque manipolati, o fabbricati) sulla quantità delle "bufale" raccolte, sia andata crescendo in modo alquanto significativo nel (breve) lasso di tempo considerato: 1 su 127 (0,78%) nel 2021; 6 su 253 (2,37%) nel 2022; 22 su 293 (7,5%) nel 2023; 22 su 257 (8,5%) nella prima parte del 2024».

5.1.7. Obblighi specifici di trasparenza nell'AI Act

In riferimento ai profili di rischio così posti in evidenza, appare opportuno rilevare quanto sancito dall'art. 50 dell'*AI Act*, che prevede specifici obblighi di trasparenza per i fornitori e i *deployer* di determinati sistemi di IA[114].

In particolare, è stabilito che i fornitori di sistemi di intelligenza artificiale che generano contenuti audio, immagine, video o testuali sintetici, sono tenuti a garantire che gli *output* siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente; devono garantire che le loro soluzioni tecniche siano efficaci, interoperabili, solide e affidabili nella misura in cui ciò sia tecnicamente possibile, tenendo conto delle specificità e dei limiti dei vari tipi di contenuti, dei costi di attuazione e dello stato dell'arte generalmente riconosciuto, come eventualmente indicato nelle pertinenti norme tecniche. Tale obbligo, tuttavia, non si applica se i sistemi di intelligenza artificiale svolgono una funzione di assistenza per l'*editing standard* o non modificano in modo sostanziale i dati di *input* o la rispettiva semantica.

I *deployer* di sistemi che generano o manipolano immagini o contenuti audio o video che costituiscono un *deepfake* devono rendere noto che il contenuto è stato generato o manipolato artificialmente. Qualora il contenuto faccia parte di un'analoga opera o di un programma che siano manifestamente artistici, creativi, satirici o fittizi, gli obblighi di trasparenza si limitano al dovere di rivelare l'esistenza di tali contenuti generati o manipolati in modo adeguato, senza ostacolare l'esposizione o il godimento dell'opera.

Analogamente, i *deployer* di *software* di intelligenza artificiale che generano o manipolano testo pubblicato allo scopo di informare su questioni di interesse pubblico sono tenuti a rendere noto che il testo è stato generato o manipolato artificialmente. Tale obbligo non si applica se il contenuto generato dall'IA è stato sottoposto a un processo di revisione umana o di controllo editoriale e una persona fisica o giuridica detiene la responsabilità editoriale della pubblicazione del contenuto.

Infine, si evidenzia che tali obblighi non si applicano se l'uso è autorizzato dalla legge per accertare, prevenire, indagare o perseguire reati; cosa che dà ulteriore riprova della possibile implementazione di

---

114 L'inosservanza degli obblighi stabiliti dall'articolo 50 è punita con le sanzioni di cui all'art. 99.

tali applicativi nell'attività di *law enforcement*, finanche di quelli attraverso cui sarà possibile realizzare *deepfake* per finalità di servizio.

5.2. L'attribuzione della responsabilità nei casi di ricorso all'intelligenza artificiale da parte delle Forze di polizia e nell'attività giudiziaria

5.2.1. Machina delinquere non potest

L'ultimo profilo da chiarire, in linea con i propositi di cui in premessa, è rappresentato dunque dall'inquadramento della responsabilità nell'eventualità in cui sia fatto ricorso ad applicativi di intelligenza artificiale.

La questione, oggetto di rilevante attenzione in dottrina e nell'ambito della letteratura scientifica, si incentra sulla possibilità o meno di individuare una "personalità giuridica elettronica" in capo all'IA, specie se generativa. In altri termini, si riflette circa la possibilità di poter giungere o meno a ritenere che una "macchina intelligente" possa essere soggetto di diritti e centro di imputazione e riferimento di interessi e di rapporti giuridici.

Adattando l'antico brocardo latino all'esigenze dei tempi, la dottrina risulta orientata, in via quasi unanime, a sostenere che *machina delinquere non potest*, sebbene non siano mancate riflessioni di giuristi che hanno ipotizzato paradigmi di responsabilità penale riferibili a entità dotate di intelligenza artificiale, come pure posizioni di tecnici informatici che sono giunti a sostenere che i *software* più evoluti avrebbero oramai raggiunto la dimensione di vere e proprie entità senzienti[115].

---

115 Tenore V., op. cit., riporta sul punto gli assunti del giurista israeliano Gabriel Hallevy che «ha ipotizzato tre astratti paradigmi di responsabilità penale applicabili a un'entità dotata di intelligenza artificiale: a) *Perpetration through another*, ovvero il programmatore o l'utente finale potrebbero essere ritenuti responsabili di aver istruito direttamente l'entità IA a commettere il crimine. La responsabilità dolosa è imputabile esclusivamente all'essere umano che ha agito attraverso un'entità dotata di apprendimento automatico o per mezzo di un algoritmo preimpostato; b) *Natural probable consequence*, ovvero il programmatore o l'utente finale non hanno pianificato la commissione di alcun crimine. Secondo lo schema della colpa cosciente, il reato commesso materialmente dalla macchina costituiva una conseguenza probabile, non presa in considerazione dall'uomo che l'ha progettata; c) *Direct liability*, ovvero il programmatore o l'utente finale non hanno nessuna responsabilità, che ricade esclusivamente in capo all'entità IA. Solo il primo paradigma, *perpetration through*

Ancorché i contenuti elaborati da tali applicativi informatici siano divenuti fortemente somiglianti e, in taluni casi, addirittura indistinguibili da quelli realizzati dagli esseri umani, non può che condividersi che una macchina, allo stato della scienza e della tecnica, non possa in alcun caso essere ritenuta autrice di azioni e comportamenti, né di ciò responsabile, mancando di umana coscienza.

5.2.2. "Riserva di umanità"

È questo il presupposto concettuale che, per quanto apparentemente pleonastico, appare necessario focalizzare adeguatamente nell'ottica di definire quel "principio del servizio" o "riserva di umanità" che, così enunciato in dottrina, si manifesta appieno nell'evoluzione del quadro normativo in materia[116].

Difatti, oltre a quanto già considerato in relazione all'*AI Act*, il citato disegno di legge nr. 1146/2024, di iniziativa governativa – approvato dal Senato della Repubblica il 20 marzo 2025 – definisce chiaramente la funzione eminentemente strumentale dell'intelligenza artificiale, ribadendo il mantenimento della responsabilità in capo alla persona competente all'adozione dell'atto o provvedimento.

Nello specifico, l'art. 13 del testo proposto, in tema di «uso dell'intelligenza artificiale nella pubblica amministrazione», prevede, al secondo comma, che «l'utilizzo dell'intelligenza artificiale avviene in funzione strumentale e di supporto all'attività provvedimentale, nel rispetto dell'autonomia e del potere decisionale della persona che resta l'unica responsabile dei provvedimenti e dei procedimenti in cui sia stata utilizzata l'intelligenza artificiale»[117]. Ana-

---

another, prevede una responsabilità penale unicamente umana; mentre il secondo e il terzo paradigma presuppongono il riconoscimento di una personalità giuridica con una conseguente responsabilità penale delle entità IA». In altro punto del medesimo contributo, l'autore illustra che «ben oltre il riconoscimento di una soggettività giuridica va l'impostazione culturale tesa ad una sorta di "umanizzazione" dell'AI, che rasenta il fanatismo, propugnata da alcuni "scienziati", che sembrano riconoscerla in macchine come LaMDA, il generatore di *chatbot* artificialmente intelligente di Google».

116 Fava P., *La c.d. "intelligenza artificiale": personalità elettronica, imputazione e responsabilità tra diritto civile, penale ed amministrativo*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024.

117 Si riporta il testo integrale dell'articolo: «Art. 13. (Uso dell'intelligenza artificiale nella pubblica amministrazione). 1. Le pubbliche amministrazioni utilizzano l'intelligenza artificiale allo scopo di incrementare l'efficienza della propria attività, di ridurre i tempi di definizione dei procedimenti e di aumentare la qualità e la quantità dei servizi erogati ai citta-

logamente, l'art. 14, avente per oggetto «uso dell'intelligenza artificiale nell'attività giudiziaria» prevede, che «i sistemi di intelligenza artificiale sono utilizzati esclusivamente per l'organizzazione e la semplificazione del lavoro giudiziario, nonché per la ricerca giurisprudenziale e dottrinale»; inoltre «è sempre riservata al magistrato la decisione sulla interpretazione della legge, sulla valutazione dei fatti e delle prove e sulla adozione di ogni provvedimento»[118].

Sul punto è opportuno evidenziare che la giurisprudenza amministrativa, sviluppatasi prevalentemente in relazione all'utilizzo di applicativi informatici nell'ambito delle procedure concorsuali ed ai fini della gestione delle istanze di trasferimento, ha più volte ribadito che il ricorso ad algoritmi nell'ambito del procedimento amministrativo non è di per sé vietato, neppure nell'ambito di quelli caratterizzati da discrezionalità, anche tecnica.

Come evidenziato da chi ha approfondito il tema, «il ricorso all'algoritmo in funzione istruttoria, integrativa e servente della decisione umana, ovvero anche in funzione parzialmente decisionale nei procedimenti a basso tasso di discrezionalità, non può mai comportare un abbassamento del livello delle tutele garantite dalla legge sul procedimento amministrativo, e in particolare di quelle sulla individuazione del responsabile del procedimento, sull'obbligo di motivazione, sulle garanzie partecipative e sulla cd non esclusività della decisione algoritmica»[119].

In particolare, il fatto che il provvedimento venga emanato ricorrendo all'utilizzo di algoritmi produce l'effetto di rafforzare l'obbligo

---

dini e alle imprese, assicurando agli interessati la conoscibilità del suo funzionamento e la tracciabilità del suo utilizzo. 2. L'utilizzo dell'intelligenza artificiale avviene in funzione strumentale e di supporto all'attività provvedimentale, nel rispetto dell'autonomia e del potere decisionale della persona che resta l'unica responsabile dei provvedimenti e dei procedimenti in cui sia stata utilizzata l'intelligenza artificiale. 3. Le pubbliche amministrazioni adottano misure tecniche, organizzative e formative finalizzate a garantire un utilizzo dell'intelligenza artificiale responsabile e a sviluppare le capacità trasversali degli utilizzatori. 4. Le pubbliche amministrazioni provvedono agli adempimenti previsti dal presente articolo con le risorse finanziarie, umane e strumentali disponibili a legislazione vigente».

118 Si riporta il testo integrale dell'articolo: «Art. 14. (Uso dell'intelligenza artificiale nell'attività giudiziaria). 1. I sistemi di intelligenza artificiale sono utilizzati esclusivamente per l'organizzazione e la semplificazione del lavoro giudiziario, nonché per la ricerca giurisprudenziale e dottrinale. Il Ministero della giustizia disciplina l'impiego dei sistemi di intelligenza artificiale da parte degli uffici giudiziari ordinari. Per le altre giurisdizioni l'impiego è disciplinato in conformità ai rispettivi ordinamenti. 2. È sempre riservata al magistrato la decisione sulla interpretazione della legge, sulla valutazione dei fatti e delle prove e sulla adozione di ogni provvedimento».

119 Tenore V., op. cit.

di motivazione in capo all'amministrazione, chiamata a rendere la propria decisione finale «non solo conoscibile, ma anche comprensibile».

Nel costante orientamento della giurisprudenza, allo scopo di assicurare il fondamentale principio di trasparenza dell'azione amministrativa, risulta dunque assolutamente necessario che sia assicurata la piena conoscibilità dell'utilizzo dell'algoritmo e del suo stesso funzionamento; un dovere che, come visto, è stato opportunamente tipizzato dal legislatore europeo che, con l'*AI Act*, ha definito specifici obblighi di trasparenza per i sistemi di intelligenza artificiale.

5.2.3. Responsabilità per danni da IA

Allo stato attuale, dunque, il nostro ordinamento, pur prevedendo l'utilizzo dell'intelligenza artificiale in diversi settori, tra cui quello propriamente relativo all'attività di polizia e giudiziaria, non ammette alcuna ipotesi in cui possa definirsi una responsabilità non ascrivibile, direttamente o indirettamente, alla persona fisica incaricata dell'adozione dell'atto o provvedimento per la cui realizzazione si avvale del *software*.

Tuttavia, l'impiego di applicativi di intelligenza artificiale significa, come anticipato, coinvolgere sviluppatori, venditori o fornitori di tali *software* nell'ambito di azioni che possono determinare eventuali danni, come tali suscettibili di determinare responsabilità a vario titolo.

In ambito mondiale, segnatamente negli Stati Uniti d'America, la questione è stata approfondita, anche nei tribunali, con riferimento ai casi di danni o lesioni verificatisi in relazione all'uso di auto a guida autonoma, ossia di autoveicoli completamente manovrati da applicativi di intelligenza artificiale.

Sebbene, allo stato attuale, la regolamentazione europea ponga un divieto alla circolazione di tali mezzi, consentendo esclusivamente dispositivi di assistenza alla guida – e, dunque, si ispiri agli stessi princìpi già richiamati, secondo cui l'IA può costituire esclusivamente un supporto alla decisione o azione umana – il tema evidenzia uno dei principali aspetti critici nel quadro dell'attribuzione della responsabilità per un eventuale danno causato attraverso l'utilizzo dell'intelligenza artificiale: l'effettiva difficoltà di provare che sia effettivamente dipeso dal *software* e non dall'azione umana.

Il tema si ricollega, come evidente, a quelle esigenze di trasparenza sopra argomentate, che rendono necessario, nell'ipotesi in cui tali algo-

ritmi siano impiegati nell'attività di polizia od in quella giudiziaria, l'assoluta conoscibilità e comprensibilità dei rispettivi processi operativi.

Sul punto, occorre evidenziare che la Commissione europea aveva proposto l'adozione di una direttiva volta a definire, in maniera uniforme, il quadro giuridico per i danni causati da sistemi di intelligenza artificiale; tale direttiva aveva per oggetto le azioni civili di responsabilità extracontrattuale per colpa per il risarcimento dei danni causati da *output* generato o non prodotto ancorché atteso.

La direttiva, recentemente ritirata per la riferita impossibilità di giungere ad un accordo[120], prevedeva in particolare degli specifici oneri di divulgazione degli elementi di prova a carico degli sviluppatori di tali sistemi, al fine di limitare l'asimmetria informativa e consentire l'effettivo esperimento delle azioni risarcitorie; stabiliva, inoltre, casi di presunzione del nesso causale tra la colpa dello sviluppatore e l'*output* generato[121].

5.2.4. Software come prodotto nella Direttiva (UE) 2024/2853

In proposito occorre tuttavia rilevare che lo scorso 8 dicembre 2024 è entrata in vigore la nuova Direttiva (UE) 2024/2853 sulla responsabilità per danno da prodotti difettosi, che ha abrogato la precedente Direttiva 85/374/CEE.

La nuova direttiva, quanto alla tematica in oggetto, amplia il previsto regime di responsabilità oggettiva per danno da prodotto difettoso ai *software*, ivi compresi i *file* per la fabbricazione digitale, le applicazioni o i sistemi di intelligenza artificiale, immessi sul mercato o messi in servizio a partire dal 9 dicembre 2026; termine altresì fissato per il recepimento da parte degli Stati membri.

La direttiva in questione, tuttavia, riguarda il risarcimento dei danni generati a causa di sistemi di intelligenza artificiale sviluppati o forniti nell'ambito di attività commerciali e, difatti, non si applica al *software* libero e *open source*.

Sulla stessa linea della proposta cui si è fatto accenno, in tema di generale regolamentazione dei danni da IA, la direttiva prevede un meccanismo di divulgazione degli elementi di prova che ne alleggeri-

---

120 Si veda, tra i vari, l'articolo *La Commissione Ue ritira la direttiva sulla responsabilità da intelligenza artificiale*, pubblicato il 12 febbraio 2025 sul portale dell'agenzia ANSA.

121 Fava P., op. cit.

sce l'onere per gli attori nei casi maggiormente complessi, prevedendo la presunzione della difettosità del prodotto – tra cui i *software* in oggetto – in caso di mancata condivisione da parte del convenuto delle informazioni ritenute pertinenti[122].

Tra i difetti, in particolare, la direttiva annovera anche le vulnerabilità in termini di sicurezza cibernetica.

Nello specifico, all'art. 10, in tema di «Onere della prova», si prevede che «si presume il carattere difettoso di un prodotto qualora sia soddisfatta una delle seguenti condizioni:

a) il convenuto omette di divulgare i pertinenti elementi di prova a norma dell'articolo 9, paragrafo 1;
b) l'attore dimostra che il prodotto non rispetta i requisiti obbligatori di sicurezza del prodotto stabiliti dal diritto dell'Unione o nazionale intesi a proteggere dal rischio del danno subito dal danneggiato;
c) l'attore dimostra che il danno è stato causato da un malfunzionamento evidente del prodotto durante l'uso ragionevolmente prevedibile o in circostanze ordinarie».

Sempre lo stesso articolo prevede che «si presume l'esistenza del nesso di causalità tra il carattere difettoso del prodotto e il danno nel caso in cui sia stato provato che il prodotto è difettoso e che la natura del danno cagionato è compatibile con il difetto in questione».

Inoltre, la norma stabilisce che «l'organo giurisdizionale nazionale presume il carattere difettoso del prodotto o il nesso di causalità tra il carattere difettoso e il danno, o entrambi, qualora, nonostante la divulgazione di prove a norma dell'articolo 9 e tenuto conto di tutte le circostanze pertinenti del caso: l'attore incontri difficoltà eccessive, in particolare a causa della complessità tecnica o scientifica, nel provare il carattere difettoso del prodotto o il nesso di causalità tra il carattere

---

122 Nel considerando nr. 3 si rileva che: «La direttiva 85/374/CEE ha rappresentato uno strumento efficace e importante, ma dovrebbe essere rivista alla luce degli sviluppi legati alle nuove tecnologie, compresa l'intelligenza artificiale (IA), ai nuovi modelli imprenditoriali dell'economia circolare e alle nuove catene di approvvigionamento globali, che sono fonte di incoerenze e di incertezza giuridica, specialmente in relazione al significato del termine "prodotto". L'esperienza maturata con l'applicazione di tale direttiva ha dimostrato inoltre che per il danneggiato è difficile ottenere il risarcimento del danno a causa sia delle limitazioni alla possibilità di presentare domande di risarcimento sia della difficoltà di raccogliere elementi di prova per dimostrare la responsabilità, soprattutto alla luce della crescente complessità tecnica e scientifica. Ciò riguarda anche le domande di risarcimento del danno in relazione alle nuove tecnologie. La revisione di tale direttiva promuoverebbe pertanto la diffusione e l'adozione di tali nuove tecnologie, compresa l'IA».

difettoso e il danno o entrambi; l'attore dimostri che è probabile che il prodotto sia difettoso o che esista un nesso di causalità tra il carattere difettoso del prodotto e il danno, o entrambi»[123].

5.2.5. Responsabilità amministrativa-contabile per colpa (non) dell'IA

Venendo in conclusione ai profili concernenti la responsabilità amministrativa-contabile in relazione all'impiego dell'intelligenza artificiale nell'ambito dell'attività della pubblica amministrazione e, nello specifico, al *law enforcement*, appaiono evidenti le implicazioni scaturenti dai richiami già rivolti, nel quadro della "riserva di umanità" sopra descritta.

Gli atti ed i provvedimenti amministrativi adottati attraverso tale strumento, come richiamato, soggiacciono ad un cd "obbligo di motivazione rafforzata", frutto dell'applicazione del principio di trasparenza ad un agire amministrativo in cui viene a coinvolgersi un sistema di algoritmi di difficile accessibilità al destinatario del provvedimento che, tuttavia, deve essere garantito nella sua piena capacità di conoscere, comprendere ed eventualmente agire contro le determinazioni assunte.

Le principali difficoltà sul punto derivano, come palese, dalla complessità tecnica di tali strumenti, dal possibile coinvolgimento nel procedimento amministrativo di più soggetti che facciano ricorso all'intelligenza artificiale, come pure dal fatto che, salvi i casi di gravi negligenze, non sarebbe corretto contestare profili di responsabilità in capo ad appartenenti alla pubblica amministrazione non in grado di poter rilevare eventuali errori o vizi del procedimento riferibili all'attività automatizzata demandata ad algoritmi[124].

---

123 L'articolo prevede, infine, che il convenuto abbia il diritto di confutare tali presunzioni.

124 Evangelista P., *Intelligenza artificiale e responsabilità amministrativo-contabile*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024, rileva che: «tali criticità si rinvengono per l'opacità tecnica del *software* o del metodo algoritmico (si parla di *black box*) nelle ipotesi di IA generativa, dove è possibile conoscere i dati immessi (*input*) ed il risultato finale (*output*) ma non sempre è possibile comprendere il percorso 'logico' seguito dal *software* che ha portato alla determinazione conclusiva della macchina/calcolatrice. Difficoltà sono riscontrabili anche nell'individuazione della persona a cui imputare la responsabilità amministrativa-contabile; si pensi alle ipotesi di istruttoria di un procedimento amministrativo incompleta o viziata. Sul punto è stato affermato che "ove l'istruttoria nella quale si esprime l'IA sia

5.2.6. Conclusione. Un'intelligenza senza coscienza

Dunque, soprattutto nell'attività di polizia come in quella giudiziaria, l'implementazione degli applicativi di intelligenza artificiale non può prescindere da quella costante "sorveglianza umana" che l'*AI Act* prescrive per ciascun "sistema di IA ad alto rischio", a riprova del principio della "riserva di umanità".

Oggi più che mai l'avanzamento tecnologico e la rivoluzione della rete impongono alle Forze di polizia di stare al passo, di guidare il cambiamento, ma evidenziano che, nonostante rappresenti una risorsa straordinaria ed irrinunciabile, «l'intelligenza artificiale senza coscienza è niente»[125].

Nel quotidiano dell'operatore di polizia, prima di qualsiasi formazione, cultura o intelligenza, è il buon senso della coscienza a fare la differenza.

---

affidata a responsabile che non è legittimato ad adottare il provvedimento finale, si può configurare una lettura adattata dell'art. 6, lett. e) della legge n. 241 del 1990, con aggravamento dell'onere motivazionale del dirigente che intenderà discostarsi dalle risultanze dell'istruttoria. Tale facoltà, come noto immaginata per disgiungere i profili di responsabilità discendenti da istruttoria viziata o incompleta dalla produzione di effetti giuridici verso terzi del provvedimento, risulta ancor più complessa da esercitare ove l'istruttoria riposi su un sistema di IA. Il responsabile del procedimento predispone in tali casi uno schema di provvedimento costruito in ragione dell'*output* algoritmico, del quale dovrà aver controllato il corretto esercizio; il dirigente per potersi discostare dovrà affrontare un onere motivazionale non solo limitato al valutato difetto istruttorio, ma esteso a vizi di funzionamento, *lato sensu* intesi, del sistema di IA". Altra criticità è rinvenibile nell'assenza di adeguate competenze e figure professionali 'tecniche' nelle piante organiche delle pubbliche amministrazioni, per cui non è semplice né corretto, pur richiamando la cd colpa professionale *ex* art. 1176, c. 2, Codice civile, imputare una condotta gravemente negligente al dipendente pubblico che non è in possesso di conoscenze per intercettare l'errore e l'illegittimità della 'decisione algoritmica'. Ed allora possiamo più propriamente contestare una responsabilità amministrativa contabile nelle ipotesi di gravi negligenze nelle fasi endoprocedimentali, propedeutiche alla decisione amministrativa automatizzata, tra cui la selezione e approvazione del *software*, dove è rilevante l'intervento 'umano' del dipendente pubblico che peraltro in tale ambito non può agire in violazione sia di norme interne che sovranazionali […] Dalle premesse considerazioni e tenendo conto, come abbiamo visto, che deve essere sempre assicurato un controllo umano del procedimento algoritmico in funzione di garanzia (il cd *human in the loop*) possiamo pervenire alla conclusione che è possibile imputare ad un funzionario pubblico una responsabilità amministrativo-contabile per il danno erariale – diretto o indiretto – conseguente ad un non corretto utilizzo dell'IA, dovendo verificare a monte l'esattezza dei dati da elaborare e mantenere il costante controllo del procedimento 'algoritmico'».

125 Così il noto fisico Federico Faggin, dall'articolo *L'intelligenza artificiale senza coscienza è niente*, pubblicato il 28 aprile 2025 sul portale dell'agenzia ANSA.

Diamole però una possibilità. Chiediamole come si porrebbe, adesso, se fosse di fronte ad una delle situazioni più difficili vissute in servizio.

Sono le 22.49 del 28 aprile 2025 e ho appena chiesto a *ChatGPT*: sei un Funzionario di polizia. Sei il responsabile di turno delle volanti della Polizia di Stato di Napoli. La sala operativa ti informa che, su un cavalcavia, un uomo di mezza età minaccia di lanciarsi nel vuoto. Non hai mai vissuto un caso simile. La prima pattuglia arrivata sul posto ha chiesto il tuo aiuto. Tu arrivi in fretta e quell'uomo è lì, sul ciglio. Che fai?

Se fossi un Funzionario di polizia responsabile delle volanti a Napoli in una situazione simile, agirei in modo immediato ma **molto attento**, tenendo presente che ogni gesto o parola può fare la differenza tra la vita e la morte.
Ecco come procederei:

1. **Mi avvicinerei con estrema calma e senza movimenti bruschi**, evitando assolutamente di correre o gridare: potrei innescare una reazione impulsiva.
2. **Farei indietreggiare i colleghi troppo vicini** o in posizione visibile, per non farlo sentire accerchiato o minacciato. I presenti devono essere pochi e scelti con attenzione.
3. **Mi presenterei**, in modo tranquillo e rassicurante: ad esempio, direi il mio nome, specificando che sono lì per **aiutarlo** e **ascoltarlo**, non per giudicarlo né forzarlo.
4. **Cercherei un contatto verbale**, iniziando con frasi semplici e non invasive, ad esempio:
   "*Capisco che stai vivendo un momento molto difficile. Sono qui per parlarti. Possiamo parlarne insieme?*"
5. **Cercherei di guadagnare tempo**, che è fondamentale: più passa il tempo, più la tensione emotiva può ridursi.
6. **Contatterei subito la sala operativa** chiedendo il supporto di:
   - Un **negoziatore** specializzato (se disponibile nella Questura o nel reparto mobile);
   - Un'**ambulanza** e i **Vigili del Fuoco**, pronti sotto il cavalcavia con i mezzi di soccorso (rete o materasso gonfiabile, se disponibile).
7. **Osserverei attentamente** il linguaggio del corpo dell'uomo: ogni gesto, movimento o parola può dare indizi sul suo stato mentale.
8. **Non tenterei mai di avvicinarmi di scatto o afferrarlo** senza una chiara apertura da parte sua: un intervento improvviso può precipitare la situazione.
9. **Mi preparerei** anche a gestire l'emergenza se l'uomo scivolasse o si lanciasse, coordinandomi con i soccorritori per un'azione di recupero rapida.

In sostanza: sangue freddo, empatia, comunicazione continua, e collaborazione con le unità specialistiche.
Obiettivo: mettere al sicuro la persona senza forzature, salvaguardando la sua dignità e la sua vita.
Se vuoi, posso anche simularti un dialogo tipo che potrei intavolare con lui in quei primi minuti. Vuoi?

Un intervento da manuale! Dritte che, a dire il vero, mi avrebbero fatto parecchio comodo prima di arrivare sul posto. Consigli preziosi di una meravigliosa invenzione senza coscienza, che sceglie le parole giuste ma non ne è consapevole.

Perché il fatto non è cosa fare in teoria, ma metterlo in pratica. Non è "cercare un contatto verbale", ma trovare le parole. È agire guidati sì dalla propria intelligenza, ma soprattutto con coscienza.

La coscienza, poi, è non averci dormito la notte dopo. È aver rivissuto mille volte quel momento. È la gioia nel raccontarlo durante i corsi, ai futuri Agenti, provando a trasmettergli, non tanto con le parole quanto con le emozioni, la consapevolezza di aver scelto il lavoro più bello.

Capitolo VI
La sperimentazione e l'uso dei sistemi di intelligenza artificiale per finalità securitarie in Unione europea e negli Emirati Arabi Uniti

di Filippo Vanni* e Ibrahim Nasser Al Mandoos**

6.1. Francia: tra sperimentazione e controversie. Il caso PREDICT e le Olimpiadi di Parigi 2024

Negli ultimi anni, la Francia ha intrapreso un percorso ambizioso per integrare tecnologie di intelligenza artificiale nel campo della sicurezza pubblica, suscitando accesi dibattiti tra innovazione tecnologica, tutela dei diritti fondamentali e rischi di discriminazione sistemica. Due progetti emblematici – PREDICT (2015-2020) e l'IA per il riconoscimento facciale durante le Olimpiadi 2024 – incarnano questa dualità, diventando casi studio di rilevanza internazionale.

PREDICT (acronimo per *Prévision des Événements et Déploiement des Interventions Correspondantes*), sviluppato dal Ministero dell'interno francese, rappresentava un sistema di polizia predittiva volto a identificare aree a rischio di crimini violenti attraverso l'analisi di dati storici (luoghi, orari, tipologie di reati). L'algoritmo, basato su modelli statistici, mirava a ottimizzare il dispiegamento delle forze dell'ordine, concentrandosi su "zone calde" identificate come ad alta criticità.

Il progetto è stato immediatamente contestato da organizzazioni per i diritti umani, accademici e attivisti. Le principali accuse includevano il rafforzamento della sorveglianza selettiva (l'algoritmo tendeva a concentrarsi su quartieri già marginalizzati, come le *banlieue*, perpe-

* Colonnello dell'Arma dei Carabinieri, già frequentatore del XL Corso di Alta Formazione presso la Scuola di Perfezionamento per le Forze di Polizia.

** *First Ltn.* EAU, già frequentatore del XL Corso di Alta Formazione presso la Scuola di Perfezionamento per le Forze di Polizia.

tuando cicli di stigmatizzazione e controllo poliziesco, alimentando il sospetto di un "razzismo algoritmico") e la violazione del principio di proporzionalità (il trattamento massivo di dati personali, incluso lo storico di individui non condannati, sollevava dubbi sulla conformità al GDPR e alla Convenzione Europea dei Diritti dell'Uomo).

Nel 2020, dopo anni di pressioni da parte di ONG come *"La Quadrature du Net"* e *"Amnesty International"*, il *Conseil d'État* – massimo organo giurisdizionale amministrativo francese – ha dichiarato PREDICT illegittimo, giudicandolo sproporzionato rispetto agli obiettivi di sicurezza. La sentenza ha sottolineato l'assenza di garanzie sufficienti contro abusi e discriminazioni, decretando la fine del progetto.

Per quanto riguarda il secondo recente esperimento, occorre evidenziare che in vista delle Olimpiadi di Parigi 2024 il Parlamento francese ha approvato una legge che autorizza l'uso sperimentale di sistemi di riconoscimento facciale basati su IA per monitorare gli spazi pubblici "in tempo reale". Il sistema, giustificato come misura eccezionale per prevenire atti terroristici, permetterebbe di identificare individui sospetti incrociando volti con *database* preesistenti (es. liste di ricercati).

Le critiche a tale sistema hanno per lo più riguardato i seguenti aspetti:

- sorveglianza di massa: la ONG *La Quadrature du Net* ha denunciato il rischio di creare un precedente per la normalizzazione della sorveglianza biometrica, violando il diritto alla *privacy* (art. 8 CEDU) e il principio di minimizzazione dei dati (GDPR, art. 5)[126];
- opacità algoritmica: l'assenza di trasparenza sui criteri di funzionamento dell'IA ha sollevato interrogativi su possibili errori di identificazione e *bias* razziali, già documentati in sistemi analoghi[127];
- tensioni con l'*Artificial intelligence act* europeo: la legislazione UE, allora in fase di approvazione, avrebbe di lì a poco vietato l'uso di IA per il riconoscimento biometrico "in tempo reale" in spazi pubblici, salvo eccezioni per emergenze.

Nel 2023, il *Conseil d'État* ha emanato una pronuncia cardine sull'uso di algoritmi predittivi nella sicurezza pubblica, richiamando

---

126 https://contropiano.org/news/internazionale-news/2025/01/30/intelligenza-artificiale-la-francia-apre-la-strada-alla-sorveglianza-di-massa-in-europa-0179744. Intelligenza artificiale: la Francia apre la strada alla sorveglianza di massa in Europa.

127 Ad esempio gli studi sul riconoscimento facciale del *MIT Media Lab*.

due princìpi fondamentali: il controllo umano effettivo, secondo cui ogni decisione basata su IA deve prevedere un'intermediazione umana, e il principio di precauzione, in base al quale le autorità devono dimostrare l'assenza di rischi sproporzionati per i diritti fondamentali, garantendo *audit* indipendenti e meccanismi di correzione degli errori.

Questa sentenza, allineata allo spirito del regolamento UE *"AI Act"*, segna un tentativo di bilanciare innovazione e garanzie giuridiche, imponendo limiti stringenti all'uso di tecnologie intrusive. Tuttavia, permangono tensioni tra il governo francese – determinato a potenziare gli strumenti di sicurezza – e le istituzioni sovranazionali, chiamate a vigilare sul rispetto dello Stato di diritto.

6.2. Paesi Bassi: il caso SyRI. Algoritmi, discriminazione e diritti fondamentali nell'era dell'intelligenza artificiale

Il caso SyRI (*Systeem Risico Indicatie*) rappresenta un precedente fondamentale nel panorama europeo sull'utilizzo dell'intelligenza artificiale nel settore della sicurezza interna e della lotta alle frodi nei sistemi di *welfare*. Questo controverso sistema algoritmico, implementato dal governo olandese nel 2014, ha portato a una serie di conseguenze drammatiche che culminarono in una sentenza storica della Corte Distrettuale dell'Aja nel 2020 e, in seguito, nelle dimissioni del governo olandese. Quello di SyRI rappresenta il primo caso in cui un governo è caduto a causa di un algoritmo accusato di essere discriminatorio e invasivo nella vita dei cittadini, evidenziando le profonde implicazioni sociali, etiche e politiche dell'uso di tecnologie algoritmiche nella gestione pubblica e nella sorveglianza sociale.

Il sistema SyRI venne introdotto nei Paesi Bassi nel 2014 come strumento tecnologico avanzato per combattere le frodi nel sistema di *welfare* dello Stato. In un contesto politico ed economico caratterizzato da crescenti pressioni sui sistemi di protezione sociale, il governo olandese cercava soluzioni per ottimizzare l'allocazione delle risorse e identificare potenziali casi di frode fiscale e abuso di sussidi sociali[128].

---

128 https://spremutedigitali.com/intelligenza-artificiale-approccio-olandese. "L'Intelligenza Artificiale (AI) rivoluzionerà il mondo come tecnologia chiave, proprio come la rivoluzione industriale ha fatto nel XVIII e XIX secolo. Cinque anni fa era inconcepibile che l'AI potesse fare diagnosi più affidabili di un medico specialista. Era anche inconcepibile che l'AI ci conoscesse meglio del nostro *partner*, guardando solo ai *Like* su Facebook. Allo stesso modo, è

L'algoritmo, sviluppato come parte di un più ampio progetto predisposto da un *team* antifrode, mirava a individuare rapidamente e con maggiore precisione comportamenti fraudolenti nei sistemi di sussidio, in particolare quelli relativi all'infanzia. Le autorità olandesi consideravano questo sistema come un passo avanti nella digitalizzazione dei controlli fiscali e nell'efficientamento delle procedure di monitoraggio, riducendo la necessità di controlli manuali e aumentando la precisione nell'identificazione di potenziali irregolarità.

L'implementazione di SyRI si inseriva in un più ampio *trend* europeo di digitalizzazione della pubblica amministrazione e di adozione di soluzioni tecnologiche per migliorare la *governance* pubblica. Sistemi simili erano già operativi in altri paesi europei come il Regno Unito, mentre Stati Uniti e Australia avevano iniziato a elaborare dati per l'accesso ai servizi sociali tramite algoritmi. La peculiarità di SyRI, tuttavia, risiedeva nella sua pervasività e nella mancanza di trasparenza con cui veniva applicato.

Infatti, il sistema non veniva applicato uniformemente su tutto il territorio nazionale, ma era prevalentemente utilizzato in quartieri svantaggiati di città come Rotterdam, Eindhoven e Haarlem. Questa scelta geografica non era casuale: le autorità avevano deliberatamente selezionato aree urbane caratterizzate da alti tassi di povertà e significativa presenza di comunità di immigrati. Già questa selezione territoriale ha sollevato i primi dubbi sulla possibile discriminazione insita nel sistema. Inoltre, un altro aspetto particolarmente controverso dell'applicazione di SyRI, era che il monitoraggio non si limitava ai soli beneficiari di sussidi, ma coinvolgeva tutti i residenti delle aree interessate, creando di fatto una sorveglianza generalizzata in specifici contesti urbani. Questa caratteristica avrebbe successivamente alimentato le critiche sul sistema, accusato di implementare una forma di profilazione sociale indiscriminata.

Il cuore del sistema SyRI consisteva nella capacità di incrociare dati personali e sensibili provenienti da ben 17 diversi *database* governativi. Questa massiccia raccolta e elaborazione di dati includeva informazioni sul fisco, sui servizi sociali, sulla storia medica dei cittadi-

---

difficile prevedere cosa significherà l'intelligenza artificiale per noi in cinque, dieci o quindici anni. Sundar Pichai, amministratore delegato di Google, ha detto: "AI è una delle questioni più importanti per l'umanità e avrà un impatto maggiore rispetto all'elettricità." Così inizia il *report* dell'AINED, la rete avviata dal governo olandese con l'obiettivo di fornire una risposta alla domanda: cosa dovrebbe fare l'Olanda nel campo dell'intelligenza artificiale?

ni, sulle utenze elettriche e telefoniche e su numerosi altri dati sensibili relativi alla vita privata degli individui.

L'algoritmo, sulla base della profilazione dei dati presenti nelle dichiarazioni dei redditi e in altri documenti ufficiali registrati nel sistema, assegnava a ciascun cittadino un "punteggio di rischio". Quando questo punteggio raggiungeva una determinata soglia, il sistema generava automaticamente un "sospetto di frode" che giustificava l'avvio di controlli più approfonditi da parte delle autorità fiscali.

Un elemento chiave del funzionamento di SyRI, che avrebbe successivamente contribuito alla sua delegittimazione, era la completa mancanza di trasparenza: i cittadini non venivano informati del fatto che i loro dati personali fossero oggetto di analisi algoritmica, né avevano modo di conoscere i criteri in base ai quali veniva calcolato il loro punteggio di rischio.

Una volta che un individuo veniva identificato come potenziale frodatore veniva automaticamente inserito in una specifica lista di controllo. A questo punto, si verificava un'inversione dell'onere della prova: i cittadini inseriti nella lista dovevano dimostrare la correttezza dei loro adempimenti fiscali.

Ad aggravare ulteriormente la situazione, gli utenti destinatari dell'avvio di un procedimento di accertamento fiscale non venivano informati tempestivamente, e ciò limitava fortemente gli interessati rispetto alla possibilità di prepararsi adeguatamente a difendere la propria posizione. In sintesi, secondo quanto emerso dalle indagini successive, circa 20.000 famiglie olandesi si sono viste ingiustamente privare di sussidi sociali poiché l'algoritmo le aveva erroneamente sospettate di frode fiscale. Le conseguenze sociali di queste false accuse sono state devastanti per molte famiglie, specialmente quelle già in condizioni di vulnerabilità economica.

Questa discriminazione algoritmica non era necessariamente intenzionale, ma risultava dalla selezione dei dati e dai parametri utilizzati per addestrare il sistema. Il *focus* su quartieri a basso reddito e ad alta concentrazione di immigrati ha inevitabilmente portato a una sorveglianza sproporzionata di queste comunità, contribuendo a rafforzare pregiudizi sociali esistenti anziché combatterli. L'opacità del sistema costituiva un ulteriore problema: l'algoritmo funzionava come una "scatola nera", i cui meccanismi decisionali rimanevano sconosciuti non solo ai cittadini ma spesso anche agli stessi operatori pubblici incaricati di implementarne le decisioni, il che impediva qualsiasi forma di controllo democratico sul suo funzionamento.

Conseguentemente, nel 2018 una coalizione di associazioni che si occupavano di *welfare* e diritti digitali ha deciso di portare il governo olandese davanti al giudice. Questa coalizione comprendeva il Comitato di giuristi per i diritti umani dei Paesi Bassi, la fondazione *Privacy First*, un sindacato e un'associazione di consumatori. Le associazioni sostenevano che SyRI violava norme regionali e internazionali per la protezione dei diritti umani, in particolare il diritto alla *privacy* garantito dall'articolo 8 della Convenzione europea per la salvaguardia dei diritti dell'uomo (CEDU).

Il 5 febbraio 2020, la Corte distrettuale dell'Aja ha emesso una sentenza storica, dichiarando illegittimo l'impiego del sistema algoritmico SyRI da parte del governo dei Paesi Bassi per violazione dell'articolo 8 CEDU, sottolineando la violazione del principio di proporzionalità richiesto dalla Convenzione: l'interesse legittimo dello Stato di combattere le frodi non giustificava infatti un'intrusione così massiccia nella vita privata dei cittadini. Inoltre, la mancanza di trasparenza e di adeguate garanzie procedurali rendeva il sistema incompatibile con i princìpi fondamentali dello stato di diritto.

Questa sentenza ha rappresentato un precedente importante nel panorama giuridico europeo, stabilendo che anche gli strumenti tecnologici avanzati utilizzati dalle pubbliche amministrazioni devono rispettare i diritti fondamentali dei cittadini e sottostare a princìpi di trasparenza, proporzionalità e non discriminazione.

Nonostante la sentenza della Corte Distrettuale dell'Aja avesse già nel 2020 dichiarato illegittimo il sistema SyRI, le implicazioni politiche più profonde si sono manifestate solo successivamente. Inizialmente, il governo olandese ha tentato di limitare i danni, ma la crescente pressione dell'opinione pubblica e le indagini parlamentari hanno progressivamente portato alla luce la gravità della situazione.

Un rapporto parlamentare predisposto per far valere la responsabilità politica dell'esecutivo ha formalizzato gli errori commessi nell'adozione del sistema di accertamento, mettendo in luce sia i problemi tecnici dell'algoritmo come anche le responsabilità politiche nella sua implementazione e nel mancato controllo sui suoi effetti discriminatori. Il governo olandese si è visto quindi costretto a rassegnare ufficialmente le dimissioni.

Una delle principali lezioni apprese dal caso SyRI riguarda l'importanza della trasparenza nei sistemi algoritmici utilizzati dalle pubbliche amministrazioni; la sua mancanza non solo viola i diritti fondamentali dei cittadini, ma mina anche la fiducia nelle istituzioni

pubbliche. La sentenza della Corte distrettuale ha ribadito che l'innovazione tecnologica non può avvenire a scapito dei diritti umani, in particolare del diritto alla *privacy* e del diritto alla non discriminazione. La tecnologia, infatti, non è intrinsecamente né buona né cattiva: i suoi effetti dipendono da come viene utilizzata.

6.3. Germania: l'intelligenza artificiale nella sicurezza interna. Un modello di cautela e regolamentazione. Il caso PRECOBS

Il panorama dell'intelligenza artificiale applicata alla sicurezza interna in Germania è caratterizzato da un approccio notevolmente cauto, bilanciato da una forte regolamentazione che riflette la sensibilità tedesca verso le questioni di *privacy* e diritti civili. L'utilizzo di sistemi predittivi come PRECOBS rappresenta un caso emblematico di come la Germania stia integrando le tecnologie avanzate nelle strategie di sicurezza pubblica, mantenendo al contempo un forte impianto normativo e di vigilanza. Le autorità tedesche hanno implementato un sistema di controlli e bilanciamenti che cerca di sfruttare i benefici dell'IA mitigandone i rischi potenziali, ponendosi come modello di riferimento nel contesto europeo per un approccio equilibrato tra innovazione tecnologica e tutela dei diritti fondamentali.

Ben prima dell'approvazione dell'*AI Act* unionale, la Germania ha infatti sviluppato un quadro normativo particolarmente articolato che regola l'applicazione dell'intelligenza artificiale nei settori della sicurezza pubblica, nel solco di una tradizione giuridica fortemente orientata alla protezione dei diritti civili, radicata nell'esperienza storica e giuridica del Paese. L'atteggiamento dei cittadini tedeschi verso l'IA è generalmente caratterizzato da un cauto ottimismo, con un crescente riconoscimento del potenziale di questa tecnologia. Secondo recenti rilevazioni, il 72% dei tedeschi ritiene che l'IA sarà la tecnologia dominante del prossimo decennio, un incremento significativo rispetto al 42% registrato solo due anni fa.

Un elemento cruciale del modello tedesco è rappresentato dal robusto sistema di controlli istituzionali sulle attività di *intelligence* e sicurezza. Il *Bundesamt für Verfassungsschutz* (BfV), l'agenzia di *intelligence* interna della Germania, è sottoposta a una rigorosa autorità di vigilanza, similmente a quanto avviene per il servizio di *intelligence* esterna (BND). Questa struttura di supervisione è stata rafforzata a seguito di una sentenza storica della Corte costituzionale tedesca, che ha

dichiarato illegali le metodologie impiegate dal servizio di sicurezza interna dello stato baverese (LFV), in quanto prevedevano il tracciamento dei telefoni cellulari e l'uso di strumenti di infiltrazione informatica senza adeguate autorizzazioni.

Con riferimento al sistema PRECOBS (*Pre Crime Observation System*), si può affermare che esso rappresenti uno dei casi più significativi dell'applicazione dell'IA alla sicurezza interna. Questo sistema di polizia predittiva utilizza dati sui crimini avvenuti nel recente passato per elaborare previsioni precise su potenziali attività criminali future in aree geografiche definite. Sviluppato originariamente dall'Istituto di Oberhausen per la Tecnologia di Previsione basata su *Pattern* (IfmPt) e successivamente portato avanti da LogObject Deutschland GmbH, PRECOBS è diventato un pioniere nel campo del *"predictive policing"*.

Il sistema opera attraverso un monitoraggio quotidiano delle attività criminali, analizzando i dati dei casi e calcolando le probabilità di futuri reati. Gli algoritmi di PRECOBS sono in grado di determinare dove è più probabile che si verifichino specifiche tipologie di crimini, come furti con scasso, reati contro i veicoli, rapine e incendi dolosi. Questa capacità predittiva consente alle Forze di polizia di essere presenti nei luoghi giusti con tempistiche adeguate, ottimizzando l'allocazione delle risorse e potenziando l'efficacia degli interventi preventivi.

Un aspetto distintivo di PRECOBS è la sua solida base scientifica e criminologica. Il *software* si fonda infatti su evidenze criminologiche consolidate e utilizza algoritmi sofisticati per elaborare previsioni precise. Questo approccio scientifico è fondamentale per garantire l'affidabilità del sistema e per giustificarne l'adozione in un contesto normativo rigido come quello tedesco.

In tale quadro, un aspetto fondamentale dell'applicazione dell'IA nella sicurezza interna in Germania è la chiara distinzione tra le funzioni di *polizia di sicurezza*, che ha carattere preventivo, e *polizia giudiziaria*, con funzione repressiva. Questa distinzione non è meramente teorica, ma ha importanti conseguenze pratiche e giuridiche nell'implementazione di sistemi come PRECOBS. I sistemi di polizia predittiva si collocano principalmente nell'ambito della funzione preventiva, mirando a impedire la violazione dell'ordine sociale. Tuttavia, la loro integrazione con le attività repressive tradizionali deve avvenire nel rispetto di confini giuridici ben definiti. Questa necessità di chiarezza normativa ha portato a un'articolata riflessione sulle modalità di utilizzo dei dati generati dai sistemi predittivi e sulla loro ammissibilità in contesti giudiziari.

La Germania non ha agito isolatamente nella regolamentazione dell'IA applicata alla sicurezza, ma ha preso attivamente parte agli sforzi congiunti a livello europeo. Mesi prima della proposta e dell'approvazione dell'*Artificial Intelligence Act* dell'UE, Italia, Germania e Francia hanno raggiunto un importante accordo sulla regolamentazione dell'intelligenza artificiale, proponendo – tra l'altro – norme di condotta e trasparenza obbligatorie per tutti i fornitori di IA, indipendentemente dalle loro dimensioni, nonché di affidare la supervisione del rispetto degli *standard* a un'autorità europea dedicata. Questo approccio è stato poi fatto proprio dalla Commissione europea all'atto della scrittura dell'*AI Act*, che infatti ha previsto la costituzione di un Ufficio europeo per l'Intelligenza artificiale (*European AI Office*) per una supervisione generale dei sistemi in campo a livello unionale rispetto agli *standard* fissati dalla norma e per l'esame degli effetti concreti della regolazione nella ricerca di settore e nella pratica operativa.

6.4. Emirati Arabi Uniti: *best practices* e sistemi in uso

6.4.1. Breve introduzione sulle peculiarità di quello Stato, dalla forte caratterizzazione multietnica e multireligiosa

Nel corso di una visita presso l'Ambasciata degli Emirati Arabi Uniti (EAU) in Roma, organizzata dalla Scuola di perfezionamento alla fine del mese di aprile u.s., ai frequentatori del corso Alta formazione è stato fatto omaggio di un piccolo volume recante la strategia degli EAU per convivere e coesistere con la diversità[129]. Questa premessa vuole partire da qui per evidenziare come un Paese con una simile caratterizzazione multietnica riesca a garantire forme di coesistenza e tolleranza tra persone con radici a volte anche profondamento diverse, attuando per questo scopo importanti forme di controllo per la repressione – tra gli altri – di ogni crimine che sia connesso a odio, razzismo o xenofobia.

Nella prefazione al richiamato volume si legge che "l'accettazione dell'altro ha costituito da sempre una delle più grandi sfide che la società abbia mai dovuto affrontare. Le differenze hanno fomentato

129 "Coesistere con la diversità e accettare la differenza", a cura di Ali Rashid al-Nuaimi.

innumerevoli conflitti e guerre in tutto il mondo, giacché l'essere umano si sente più a suo agio tra coloro che gli sono simili e prova verso chi è diverso una sorta di timore". In questo, il termine "tolleranza" viene spesso impiegato in luogo di "coesistenza", mentre nell'idioma arabo deriva dalla medesima radice del termine che indica il "perdono" ed è quindi naturalmente associato con qualità quali la gentilezza e la generosità. La tolleranza non è sinonimo di compromesso, sottomissione o debolezza, ma rappresenta piuttosto la *decisione* di riconoscere i diritti altrui e valutare la loro diversità, non rimanendovi indifferenti. Purtroppo, l'evoluzione del significato di questo termine nella lingua inglese (*tolerance*) si è sempre più riferita all'accettazione dell'altro da una posizione di superiorità, volto a rapportarsi all'esistenza in modo da "tollerare" la sua esistenza.

Tre significativi documenti citati nel volume[130] costituiscono i fondamenti ideologici delle modalità secondo cui le comunità musulmane hanno coesistito con la diversità e la differenza negli ultimi 1.400 anni, rivelando che tali idee erano dedotte dalle regole basilari, dai fondamenti e dai valori dell'Islam[131]. Una società capace di crescita e creatività continua può nascere solo dal pluralismo religioso, etnico e culturale. Gli EAU hanno scelto questa via nell'interesse del progresso e della prosperità: il mantenimento del successo degli EAU ha bisogno di ampie interazioni culturali con persone provenienti dal tutto il mondo, giacché sono un Paese giovane con una ricca eredità e radici profonde nella storia. Nel volume si legge ancora che nei suoi 50 anni di storia gli Emirati non hanno mai registrato un singolo crimine legato al razzismo e nessun tipo di xenofobia diretta verso quanti professano religioni diverse o provengono da un diverso entroterra etnico, linguistico e culturale. In tale quadro, il dialogo interreligioso e interculturale diviene irrilevante all'interno di una società veramente tollerante, perché una simile società ha superato il bisogno di un tale dialogo. Tra gli elementi necessari per raggiungere questa realtà vengono richiamati lo stato di diritto e un sistema giudiziario equo, a cui da alcuni anni si aggiunge un *Ministero della tolleranza e della coesi-*

---

130 Il documento fondativo sulla costituzione della Medina, l'*ashtiname* (ovvero un patto di protezione) di Muhammad col monastero di S. Caterina in Egitto, la garanzia di sicurezza di 'Umar agli abitanti di Gerusalemme.

131 «Non disputate con le genti del Libro se non nel modo migliore, tranne contro coloro che commettono ingiustizia, ma dite: *"Crediamo in ciò che è stato rivelato a noi e in ciò che è stato rivelato a voi. Il nostro Dio e il vostro Dio sono il medesimo"*» (S. Corano 29:46).

stenza, col compito di rendere la tolleranza un elemento radicato nel tessuto sociale, protetta e preservata dalla legge[132].

Il volume di cui si tratta conclude evidenziando due concetti fondamentali:

- da un lato definisce questo modello di società come "*post*-tolleranza", giacché garantisce la continuità della diversità e della differenza religiosa, etnica e culturale e le considera una risorsa vitale della comunità;
- dall'altro richiama l'esigenza di una criminalizzazione severa di ogni discorso d'odio e di razzismo (in generale, degli *hate speech*), anche attraverso un controllo accurato dei *social media*.

6.4.2. Oyoon (Eyes) Project[133]

As such, the Oyoon project, which Dubai Police launched back in 2018, is one of the most ambitious and high profile of AI powered surveillance systems in the UAE. The principal purpose of the project is to set up an extensive network of AI enabled cameras in Dubai where these cameras, are fixed at various places of public spaces across Dubai to continuously monitor the activities around. It incorporates multiple AI capacities, for example, live facial perception, conduct investigation, auto recognisance, and grouping the group. These technologies seek to make UAE public safety better, and speedier in responding to incidents. One of the key features of Oyoon project is real time facial recognition. The system can compare people on the camera against individuals known in a database of faces, like those of people of interest or on a wanted poster. This will allow the authorities to very quickly identify suspects even in packed settings. The system is further strengthened by AI driven behavioural analysis of the same, which detects suspicious activities in public places with deviations from normal. Using AI system, the movement patterns can be monitored in order to flag any unusual behaviour which could hint at a threat[134].

One significant feature of the Oyoon project is vehicle recognition. The AI system uses the licence plate recognition (LPR) technology to track and

---

132 Ogni discriminazione su base religiosa dovrebbe essere rifiutata e proibita nel rispetto del versetto coranico "A voi la vostra religione, a me la mia" (109:6).

133 Il presente paragrafo, al pari dei seguenti, è stato redatto in lingua inglese da Ibrahim Nasser Al Mandoos. Segue traduzione in lingua italiana.

134 Lewis *et al.*, 2023.

monitor vehicles in a real time. This is a very importance tool which is used to track stolen cars, also cars in conduction of criminal activities as well ensuring that traffic rules are observed. In addition, the system can be used to analyse the type of vehicle, e.g., to differentiate commercial from private vehicles, which can be used by the law enforcement to detect the criminal activities connected with particular types of vehicles. Furthermore, the Oyoon project uses AI in order to monitor and manage crowds in big gatherings in public places or occasions. Crowd movements can be analysed by the system to detect any anomalies or patterns suggesting a security risk, for example, a stampede or disorderly movement[135]*. The system helps law enforcement take fast action if potential threats are found. The Oyoon project is more than an automated surveillance network – it includes human supervision. AI analytics work with real time human intervention through a central command centre that makes decisions a combination of data driven insights and resultant human judgement*[136]*. This has further shown to be a synergy that has reduced crime rates in monitored areas and increasing time it takes to respond in case of an incident. Reportedly, the project has a highly positive effect on enhancing public safety and reduction of crime in Dubai.*

Il progetto *Oyoon*, lanciato dalla Polizia di Dubai nel 2018, è uno dei sistemi di sorveglianza alimentati dall'intelligenza artificiale più ambiziosi e di alto profilo degli Emirati Arabi Uniti. Lo scopo principale del progetto è quello di creare una vasta rete di telecamere abilitate all'intelligenza artificiale a Dubai, fissate in vari spazi pubblici per monitorare continuamente le attività circostanti. Il sistema incorpora diverse capacità di intelligenza artificiale, come ad esempio la percezione facciale in *real-time*, l'investigazione e il riconoscimento automatico. Queste tecnologie mirano a migliorare la sicurezza pubblica degli Emirati Arabi Uniti e a rispondere più rapidamente agli incidenti. Una delle caratteristiche principali del progetto *Oyoon* è il riconoscimento facciale in tempo reale. Il sistema è in grado di confrontare le persone riprese dalla telecamera con quelle conosciute in un *database* di volti di interesse o ricercati. Ciò consentirà alle autorità di identificare molto rapidamente i sospetti anche in ambienti affollati. Il sistema è ulteriormente rafforzato dall'analisi comportamentale guidata dall'intelligenza artificiale, che rileva attività sospette in luoghi pubblici con deviazioni dalla norma. Grazie al sistema di intelligenza artificiale, i

135 Raghav *et al.*, 2025.
136 Lewis *et al.*, 2023.

modelli di movimento possono essere monitorati per segnalare qualsiasi comportamento insolito che possa far pensare a una minaccia.

Una caratteristica significativa del progetto *Oyoon* è il riconoscimento dei veicoli. Il sistema AI utilizza la tecnologia di riconoscimento delle targhe (LPR) per tracciare e monitorare i veicoli in tempo reale. Si tratta di uno strumento molto importante che viene utilizzato per rintracciare le auto rubate, le auto coinvolte in attività criminali e per garantire il rispetto delle regole del traffico. Inoltre, il sistema può essere utilizzato per analizzare il tipo di veicolo, ad esempio per differenziare i veicoli commerciali da quelli privati, il che può essere utilizzato dalle forze dell'ordine per individuare le attività criminali legate a particolari tipi di veicoli. Inoltre, il progetto *Oyoon* utilizza l'intelligenza artificiale per monitorare e gestire le folle nei grandi assembramenti in luoghi o occasioni pubbliche. I movimenti della folla possono essere analizzati dal sistema per rilevare eventuali anomalie o schemi che suggeriscono un rischio per la sicurezza, ad esempio un assembramento o un movimento disordinato. Il sistema aiuta le forze dell'ordine a intervenire rapidamente in caso di potenziali minacce. Il progetto *Oyoon* non è solo una rete di sorveglianza automatizzata, ma anche una supervisione umana. Le analisi dell'intelligenza artificiale lavorano con l'intervento umano in tempo reale attraverso un centro di comando centrale che prende le decisioni grazie a una combinazione di intuizioni basate sui dati e al conseguente giudizio umano. Questa sinergia si è dimostrata in grado di ridurre il tasso di criminalità nelle aree monitorate e di aumentare il tempo di risposta in caso di incidente. Secondo quanto riferito, il progetto ha un effetto altamente positivo sul miglioramento della sicurezza pubblica e sulla riduzione della criminalità a Dubai.

6.4.3. National Security Operations Center (NSOC)

Another indispensable element of the UAE's AI strategy for crime prevention is the National Security Operations Centre (NSOC). The purpose of this centre is to gather, analyse, and act on the security intelligence as an integrated hub. Nicer than the NSOC uses AI to process huge amounts of data coming from different sources, like social media, surveillance systems, and cybersecurity networks, to be able to identify and predict potential dangers. The centre uses machine learning algorithms to catch incoming security risks before they completely develop so that it can take

pro active control. Threat intelligence is one of the main functions of the NSOC. At the centre, the AI systems spend time combing through datasets as large as WMDDB itself to find patterns that could indicate a security risk, like terrorist activity, organised crime, or cyberattacks. The predictive analytics also seeks out trends and behaviours that have preceded criminal acts in order to predict future threats. As per Intelligence Online (2023), this provides the UAE's law enforcement agencies with a powerful tool to anticipate and tackle potential perils before they escalate.

Third, the NSOC also uses AI for social media monitoring, in addition to threat intelligence. As the social platforms are becoming used for communication and coordination become more common, security agencies must monitor these channels for evidence of criminal activity. Social media posts, hashtags and user behaviour are scanned by AI algorithms to detect potential threats or to somehow indicate that criminality may be unfolding. Authorities can take action whenever they notice suspicious activity as it is real time monitoring. Furthermore, pertaining to cybersecurity the NSOC is also very significant. UAE's increasing digitization of infrastructure will certainly pose a rising threat of cyberattacking its critical systems. At the NSOC, AI tools are used to identify and neutralise cyber threats prior to them affecting national infrastructure. AI systems can recognise malicious activity and create an alert for the security team to act upon whether it's government networks being attacked, or private sector targets[137].

Un altro elemento indispensabile della strategia di AI degli Emirati Arabi Uniti per la prevenzione del crimine è il *National Security Operations Centre* (NSOC). Lo scopo di questo centro è raccogliere, analizzare e agire sulle informazioni di sicurezza come un *hub* integrato. L'NSOC utilizza l'intelligenza artificiale per elaborare enormi quantità di dati provenienti da fonti diverse, come i *social media*, i sistemi di sorveglianza e le reti di sicurezza informatica, per essere in grado di identificare e prevedere potenziali pericoli. Il centro utilizza algoritmi di apprendimento automatico per cogliere i rischi di sicurezza in arrivo prima che si sviluppino completamente, in modo da poter assumere un controllo attivo. L'*intelligence* delle minacce è una delle funzioni principali del NSOC. Presso il centro, i sistemi di intelligenza artificiale setacciano grandi insiemi di dati per trovare modelli che potrebbero indicare un rischio per la sicurezza, come

137 *Intelligence Online*, 2023.

attività terroristiche, criminalità organizzata o attacchi informatici. L'analisi predittiva cerca anche le tendenze e i comportamenti che hanno preceduto gli atti criminali per prevedere le minacce future. Secondo *Intelligence Online* (2023), ciò fornisce alle forze dell'ordine degli Emirati Arabi Uniti un potente strumento per anticipare e affrontare potenziali pericoli prima che si aggravino.

In terzo luogo, il NSOC utilizza l'IA anche per il monitoraggio dei *social media*, oltre che per l'*intelligence* sulle minacce. Poiché le piattaforme sociali vengono utilizzate per la comunicazione e il coordinamento diventano sempre più comuni, le agenzie di sicurezza devono monitorare questi canali alla ricerca di prove di attività criminali. I *post* sui *social media*, gli *hashtag* e il comportamento degli utenti vengono analizzati dagli algoritmi di intelligenza artificiale per individuare potenziali minacce o per indicare in qualche modo che si sta svolgendo un'attività criminale. Le autorità possono intervenire ogni volta che notano attività sospette, poiché si tratta di un monitoraggio in tempo reale. Inoltre, per quanto riguarda la sicurezza informatica, l'NSOC è molto importante. La crescente digitalizzazione delle infrastrutture degli Emirati Arabi Uniti comporterà sicuramente una crescente minaccia di attacco informatico ai sistemi critici. Presso l'NSOC, gli strumenti di intelligenza artificiale vengono utilizzati per identificare e neutralizzare le minacce informatiche prima che colpiscano le infrastrutture nazionali. I sistemi di intelligenza artificiale sono in grado di riconoscere le attività dannose e di creare un allarme a favore del *team* di sicurezza, sia che si tratti di reti governative attaccate, sia che si tratti di obiettivi del settore privato.

6.4.4. Falcon Eye System

Another instance of AI being implemented in the UAE for the purpose of crime prevention is through Abu Dhabi's Falcon Eye system. Spanning across the emirate, the system makes use of thousands of cameras and sensors armed with AI powered analytics to monitor urban areas as well as road and critical infrastructure networks. The design of Falcon Eye will allow for constant surveillance of all major city areas. The incident detection ability of the system can be described as excellent. With AI powered analysis, Falcon Eye can detect emergencies in traffic such as accidents, criminal activities etc in real time. The system automatically generates alerts once a potential incident is detected, alerting security personnel for a quick reaction. An automated

alert system is also in place that makes no time is wasted in dealing with the issue, something that is essential in traffic accidents or crimes in progress[138].

In addition to being a security tool, the Falcon Eye system is a tool of environmental monitoring. It monitors air quality, temperature, weather pattern, etc. And it is useful data when there is a natural tragedy or emergency, authorities can make decisions and take steps to keep the public safe[139]*. Other than its usual purpose of tackling crime, the Falcon eye system shows a great application of how artificial intelligence can go beyond crime prevention. It provides a holistic approach of urban management that helps ensure the safety and well being of Abu Dhabi's residents. Credited with reducing crime rates and helping manage traffic, the system has proven the broad scope of how AI can make urban life better.*

Un altro esempio di applicazione dell'intelligenza artificiale negli Emirati Arabi Uniti per la prevenzione del crimine è il sistema *Falcon Eye* di Abu Dhabi. Il sistema, che si estende in tutto l'emirato, si avvale di migliaia di telecamere e sensori dotati di analisi basate sull'intelligenza artificiale per monitorare le aree urbane e le reti stradali e di infrastrutture critiche. Il progetto di *Falcon Eye* consentirà una sorveglianza costante di tutte le principali aree cittadine. La capacità di rilevamento degli incidenti del sistema può essere definita eccellente. Grazie all'analisi dell'intelligenza artificiale, *Falcon Eye* è in grado di rilevare in tempo reale le emergenze nel traffico, come incidenti, attività criminali, ecc. Il sistema genera automaticamente degli avvisi una volta rilevato un potenziale incidente, allertando il personale di sicurezza per una rapida reazione.

Oltre a essere uno strumento di sicurezza, il sistema *Falcon Eye* è uno strumento di monitoraggio ambientale. Monitora la qualità dell'aria, la temperatura e le condizioni meteorologiche. Oltre allo scopo abituale di contrastare il crimine, il sistema *Falcon Eye* mostra una grande applicazione di come l'intelligenza artificiale possa andare oltre la prevenzione del crimine. Fornisce un approccio olistico alla gestione urbana che contribuisce a garantire la sicurezza e il benessere dei residenti di Abu Dhabi. Accreditato di aver ridotto i tassi di criminalità e di aver aiutato a gestire il traffico, il sistema ha ampiamente dimostrato come l'intelligenza artificiale possa migliorare la vita della comunità urbana.

---

138 *The Gulf News*, 2018.
139 *The Gulf News*, 2018.

6.4.5. Smart Police Stations (SPS)

Smart Police Stations are Dubai's brainchild and a pioneering initiative that brings an altogether new facet in how citizens come in contact with law enforcement. There are no human police staff at these police stations that run 24/7, it is the advanced artificial intelligence (AI) systems that deliver police services. The idea started off in Dubai but has evolved to other emirates where the police provide public services with a modern touch of technology. Automated reporting is one of SPS's core features. Citizens needing to file a reported report for a stolen item, a traffic violation or a minor crime would in the past visit a physical police station face to face with an officer. The SPS eliminates this requirement. Rather, citizens can utilise the AI interfaces to file reports and acquire police service at any time. It reduces significantly the load on human officers and gives them a chance to deal with more sophisticated tasks. The system can easily handle and categorise reports so that they are linked with the department or officer to be followed as Aldhaheri (2023) points out.

Another function that SPS provides is identity verification. Biometric systems such as facial and finger scanning are used by the stations to authenticate the identity of persons. This is important to ensure that services are only provided to authorised people and also prevent fraud or misuse of the system. This facilitates facial recognition, in particular, ensuring that the person filing the report, or requesting a service can be verified officially and accurately in less than a second[140]*. Additionally, the Smart Police Stations run on case management systems that are powered by AI. As the citizens file the different reports, this system automatically processes, categorises same, ensuring each report is kept up. Using AI's capacity to look at the information rapidly, the SPS can verify that no report is missed, and all important movement is taken accordingly and rapidly*[141]*. Also, thanks to real time language translation capabilities, the (SPS) is one of the most exciting one whose services can be offered in multiple languages. It guarantees that the police services are accessible by citizens speaking different linguistic backgrounds. This feature is of great benefit to Dubai's diverse population, both residents and tourists, for the reasons that it makes tagging interpreters unnecessary and quickens the process of writing reports and seeking help*[142]*.*

Le stazioni di polizia intelligenti sono un'idea di Dubai e un'iniziativa pionieristica che introduce un aspetto completamente nuovo

---

140 Al Darmaki, 2022.
141 Aldhaheri, 2023.
142 Zia *et al.*, 2023.

nel modo in cui i cittadini entrano in contatto con le forze dell'ordine. In queste stazioni di polizia, attive 24 ore su 24 e 7 giorni su 7, non c'è personale di polizia umano, ma sono i sistemi avanzati di intelligenza artificiale che forniscono servizi di polizia. L'idea è nata a Dubai, ma si è evoluta in altri emirati. La segnalazione automatica è una delle caratteristiche principali delle SPS. In passato, i cittadini che avevano bisogno di sporgere una denuncia per un oggetto rubato, una violazione del codice della strada o un reato minore si recavano in una stazione di polizia, faccia a faccia con un agente. L'SPS elimina questo requisito. I cittadini possono utilizzare le interfacce dell'intelligenza artificiale per sporgere denuncia e richiedere il servizio di polizia in qualsiasi momento. Questo riduce in modo significativo il carico di lavoro degli agenti umani e dà loro la possibilità di occuparsi di compiti più sofisticati. Il sistema è in grado di gestire e classificare facilmente le segnalazioni in modo da collegarle al dipartimento competente.

Un'altra funzione fornita da SPS è la verifica dell'identità. I sistemi biometrici, come la scansione del volto e delle dita, sono utilizzati dalle stazioni per autenticare l'identità delle persone. Questo è importante per garantire che i servizi siano forniti solo alle persone autorizzate e per prevenire frodi o usi impropri del sistema. In particolare, il riconoscimento facciale facilita la verifica ufficiale e accurata della persona che presenta una denuncia. Inoltre, le stazioni di polizia intelligenti utilizzano sistemi di gestione dei casi basati sull'intelligenza artificiale. Man mano che i cittadini presentano le diverse segnalazioni, il sistema le elabora e le categorizza automaticamente, assicurando che ogni segnalazione venga tenuta aggiornata. Grazie alla capacità dell'intelligenza artificiale di esaminare rapidamente le informazioni, le SPS sono in grado di verificare che nessuna segnalazione venga tralasciata. Inoltre, grazie alle capacità di traduzione linguistica in tempo reale, il sistema SPS è in grado di offrire servizi in più lingue. Ciò garantisce che i servizi di polizia siano accessibili a cittadini che parlano idiomi diversi. Questa funzione è di grande utilità per la popolazione eterogenea di Dubai e per i suoi numerosi turisti, perché rende superfluo il ricorso a interpreti e velocizza il processo di stesura dei rapporti e le richieste di aiuto.

6.4.6. Predictive Policing Systems

Predictive policing is the emerging law enforcement concept where crime can be forecasted using machine learning algorithms at which place

and when crime will next take place. In fulfilment of its commitment to enhance public safety, the UAE has taken to adopting predictive policing systems. These systems use AI to examine past crime data and extract patterns that indicate future criminal behaviour. Next to predictive policing, crime mapping is at the heart of predictive policing in the UAE. By analysing huge amounts of historical crime data such as the time, location, and type of crime, AI systems can determine the high-risk times and locations for crimes. Identification of these patterns would help law enforcement agencies to efficiently deploy the resources they have available. Just for example, if the system can determine that some neighbourhoods or places have a surge in thefts at certain hours or days, police officers can escalate their patrols in those spots during that anticipated time. Therefore, this approach is proactive, and it prevents the occurrence of the crimes, instead of reacting to the crimes after occurring[143].

Predictive policing also includes offender analysis. AI systems can extract patterns in criminal behaviour to determine the chance of an offender committing another offence. It is particularly useful to predict recidivism of previously convicted criminals and their tendency to re offend. Machine learning algorithms can consider numerous factors like what type of crimes are being committed, whether the offender has a history of committing a crime, and behavioural patterns to predict the possibility of the offenders being involved in future criminal acts. It is information that helps law enforcement allocate resources better, like monitoring of high-risk people or places[144]*. Resource allocation is also an important application for predictive policing. AI systems have the power to predict where and when crimes are more likely to happen and recommend the best place to deploy the resources. It pertains to regulating where to distribute police officers, patrol cars, and even other resources to places found to be on high-risk. It might involve for example if the system is predicting a surge of petty theft in a certain district, then additional officers can be deployed there so that there would be an immediate response to any case of potential incident*[145]*.*

It is also subject to the integration of AI based early warning systems into predictive policing strategies. Such systems can help identify patterns related to the planning or occurrence of criminal activities like setting up a protest, a robbery or any other illegal activity. This makes it easy for authorities to treat these patterns before hand and by doing so take preventive measures through increased surveillance or intervention before the actual crime is committed[146]*.*

---

143 Al Shamsi and Safei, 2023.
144 Zia *et al.*, 2023.
145 Al Darmaki, 2022.
146 Aldhaheri, 2023.

La polizia predittiva è un concetto emergente nelle attività di *law enforcement*, consentendo di ipotizzare il tempo e il luogo di un'attività criminale utilizzando algoritmi di apprendimento automatico. Per adempiere all'impegno di migliorare la sicurezza pubblica, gli Emirati Arabi Uniti hanno da tempo adottato sistemi di polizia predittiva. Questi sistemi utilizzano l'intelligenza artificiale per esaminare i dati sui crimini passati ed estrarre modelli che indicano possibili comportamenti criminali futuri. Oltre alla polizia predittiva, la mappatura dei crimini è il cuore della polizia predittiva negli Emirati Arabi Uniti. Analizzando enormi quantità di dati storici sulla criminalità, come l'ora, il luogo e il tipo di crimine, i sistemi di intelligenza artificiale possono determinare i momenti e i luoghi ad alto rischio, aiutando le forze dell'ordine a impiegare in modo efficiente le risorse disponibili.

La polizia predittiva comprende anche l'analisi delle caratteristiche dei criminali. I sistemi di intelligenza artificiale sono in grado di estrarre modelli di comportamento criminale per determinare la possibilità che un criminale commetta un altro reato. È particolarmente utile per prevedere la recidiva di persone già condannate e la loro tendenza a delinquere nuovamente. Si tratta di informazioni che aiutano le forze dell'ordine ad allocare meglio le risorse e monitorare persone e luoghi ad alto rischio.

L'integrazione di sistemi di allerta precoce basati sull'intelligenza artificiale nelle strategie di polizia predittiva è altrettanto importante. Tali sistemi possono aiutare a identificare modelli relativi alla pianificazione o all'imminente verificarsi di attività criminali come l'organizzazione di una protesta, una rapina o qualsiasi altra attività illegale. In questo modo è più semplice per le autorità agire in anticipo e adottare misure preventive attraverso una maggiore sorveglianza o un intervento prima che venga commesso il reato vero e proprio.

6.4.7. Border Security and Immigration Systems

At UAE's borders and airports, it has many AI powered systems to improve security as well as facilitating the interaction of the immigrants. Border security in the UAE is one of the most advanced in the world thanks to these systems, which are based on face recognition, biometric verification, and behaviour analysis. Smart Gate is one of the most clicked up AI systems at UAE's airports as it features a facial recognition technology that verifies the identity of travellers within seconds. It has cut down the waiting time at

immigration and this has enhanced the efficiency of the process not only to the citizens but the visitors as well. As a result, travellers are no longer required to go through the process of manual identity cheques as the AI system scans the face of the travellers instantly and compares it with the passport information to quickly grant them entry into the country[147].

Additionally, the border security is strengthened by use of the AI based risk assessment algorithms which assess the travellers on several factors including their travel history, the purpose of the visit and any security risks. In Detail, these algorithms successfully discover candidates with potential to undermine national security or public safety. For instance, if someone in the system has a past of criminal offences, suspicious travel patterns, or missing documentation, the system can flag the individual for further investigation or detention[148]*. Document verification is another critical AI application at the borders. AI tools validating travel document, including passport, visa, and ticket are not forged and authentic. By using these systems, counterfeit documents and any alteration can be immediately recognised, preventing illegal immigration, human trafficking, and other border related crimes*[149].

Alle frontiere e negli aeroporti degli Emirati Arabi Uniti sono presenti molti sistemi basati sull'intelligenza artificiale per migliorare la sicurezza e facilitare l'interazione con gli immigrati. La sicurezza alle frontiere degli EAU è una delle più avanzate al mondo grazie a questi sistemi, che si basano sul riconoscimento facciale, sulla verifica biometrica e sull'analisi del comportamento. *Smart Gate* è uno dei sistemi di intelligenza artificiale più usati negli aeroporti degli Emirati Arabi Uniti, poiché è dotato di una tecnologia di riconoscimento facciale che verifica l'identità dei viaggiatori in pochi secondi. Ha ridotto i tempi di attesa all'immigrazione, migliorando l'efficienza del processo non solo per i cittadini ma anche per i visitatori. Di conseguenza, i viaggiatori non devono più sottoporsi al processo di verifica manuale dell'identità, poiché il sistema di intelligenza artificiale scansiona istantaneamente il volto dei viaggiatori e lo confronta con le informazioni del passaporto per concedere rapidamente l'ingresso nel Paese.

Inoltre, la sicurezza delle frontiere è rafforzata dall'uso di algoritmi di valutazione del rischio basati sull'intelligenza artificiale che valutano i viaggiatori in base a diversi fattori, tra cui la loro storia di

147 Lori, 2022.
148 Kabir *et al.*, 2023.
149 Lori, 2022.

viaggio, lo scopo della visita ed eventuali rischi per la sicurezza. In dettaglio, questi algoritmi riescono a scoprire i candidati che possono potenzialmente minare la sicurezza nazionale o la sicurezza pubblica. Ad esempio, se qualcuno nel sistema ha un passato criminale, modelli di viaggio sospetti o documentazione mancante, il sistema può segnalare l'individuo per ulteriori indagini o per l'adozione di un provvedimento restrittivo. La verifica dei documenti è un'altra applicazione critica dell'intelligenza artificiale alle frontiere. Gli strumenti di intelligenza artificiale convalidano i documenti di viaggio, tra cui passaporto, visto e biglietto, che non sono contraffatti e autentici. Utilizzando questi sistemi, i documenti contraffatti o alterati possono essere immediatamente riconosciuti, impedendo l'immigrazione clandestina, il traffico di esseri umani e altri crimini legati alle frontiere.

A guisa di conclusioni. *We have a dream*.

Giunti alla fine di questo nostro studio, possiamo considerare come sia emerso, con sufficiente chiarezza, che l'intelligenza artificiale rappresenti uno strumento fondamentale per liberare l'uomo dalle mansioni più ripetitive e routinarie, quelle che spesso assorbono tempo ed energie senza valorizzare appieno le sue capacità distintive.

Delegando all'IA questi compiti, l'essere umano può finalmente riappropriarsi di spazi di tempo preziosi da dedicare a ciò che lo rende davvero unico: la creatività, l'intelligenza emotiva, la riflessione, la produzione artistica e musicale. Queste attività, che fanno vibrare la coscienza e arricchiscono la sfera emozionale, sono il vero patrimonio dell'umanità e costituiscono il tratto distintivo rispetto a qualsiasi altra forma di intelligenza, naturale o artificiale. In questo scenario, l'IA non si pone come antagonista, ma come alleato dell'uomo, permettendogli di esprimere al massimo il proprio potenziale e di contribuire in modo significativo al progresso culturale e sociale.

Forse è questa la forma di "intelligenza complementare" che sogniamo.

Bibliografia

- Ali Rashid al-Nuaimi, *Coesistere con la diversità e accettare la differenza*, Tawasul International, 2023.
- Aresu A., *Geopolitica dell'intelligenza*, Feltrinelli 2024.
- Basile F., *Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione*, in Balbi G. – De Simone F. – Esposito A. – Manacorda S. (a cura di), *Diritto Penale e Intelligenza Artificiale "Nuovi scenari"*, Torino, 2022.
- Bernardi N., *Privacy. Protezione e trattamento dei dati*, IPSOA, Milano, 2019.
- Bolognesi F., *Intelligenza Artificiale, istruzioni per l'uso*, EPC, 2024.
- Capone F. – Iaselli V., *La privacy nell'AI Act e nel DDL italiano sull'intelligenza artificiale: perché è un tema ineludibile*, in *Agenda digitale.eu*, 2024.
- Colamedici A., *L'algoritmo di Babele. Storie e miti dell'intelligenza artificiale*, Solferino, 2024.
- Evangelista P., *Intelligenza artificiale e responsabilità amministrativo-contabile*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024.
- Falletta P. – Marsano A., *Intelligenza artificiale e protezione dei dati personali: il rapporto tra Regolamento europeo sull'intelligenza artificiale e GDPR*, in *Rivista italiana di informatica e diritto*, n. 1, 2024.
- Fava P., *La c.d. "intelligenza artificiale": personalità elettronica, imputazione e responsabilità tra diritto civile, penale ed amministrativo*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024.
- Ferrina Ceroni G., *La dignità della persona: tra GDPR E AI ACT*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024, *"Intelligenza artificiale: impiego e implicazioni sotto il profilo tecnologico, etico e giuridico"*, 2024.
- Floridi L., *Etica dell'intelligenza artificiale*, R.Cortina Editore, 2022.
- Galletti M. – Caiani Zipoli S., *Filosofia dell'Intelligenza Artificiale*, Il Mulino, 2024.
- *Manuale del diritto europeo in materia di protezione dei dati*, *Publications Office of the European Union*, 2018.

- Moriggi S. – Pireddu M., *L'intelligenza Artificiale e I suoi fantasmi*, Il Margine, 2024.
- Pezzolo Giacaglia G., *Making Things Think: How AI and Deep Learning Power the Products We Use*, Hollowey, 2021.
- Pistilli C., *L'utilizzo dell'intelligenza artificiale nel campo delle attività investigative delle forze dell'ordine: tra prospettive di sviluppo ed esigenze di coordinamento*, in Balbi G. – De Simone F. – Esposito A. – Manacorda S. (a cura di), *Diritto Penale e Intelligenza Artificiale "Nuovi scenari"*, Torino, 2022.
- Pollicino O., *Disinformazione e intelligenza artificiale: un cocktail esplosivo?*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024.
- Quintarelli S., *Intelligenza artificiale, cos'è davvero, come funziona, che effetti avrà*, B. Boringhieri, 2020.
- Tenore V., *Una riflessione sistematica: l'intelligenza artificiale può sostituire un giudice o trasformarlo da homo sapiens in homo numericus?*, in *Rivista della Corte dei Conti*, Quaderno n. 2/2024.
- Tesconi C., *Intelligenza Artificiale, Esplorando il mondo del Machine Learning e Deep Learning*, Autoedizione, 2023.
- Turing A., *"Intelligent Machinery"*, a cura di Marconi D., Einaudi, 2025.
- Uria-Recio P., *Il futuro che si aspetta*, Autoedizione, 2024.
- Williams S., *Storia dell'intelligenza artificiale. La battaglia per la conquista della scienza del XXI secolo*, Garzanti, 2003.

Sitografia

- Discorso del Santo Padre Leone XIV al Collegio Cardinalizio del 10 maggio 2025, pubblicato sul portale vatican.va.
- *AI History - Artificial Intelligence (AI)* – Library Guides at Ohio University, libguides.library.ohio.edu.
- *Artificial Intelligence – Questions and Answers*, in ec.europa.eu, portale della Commissione europea.
- *La storia dell'IA in 7 esperimenti* – Giornalist, generalist.com/briefing/the-history-of-ai.
- *The History of Artificial Intelligence* – IBM, ibm.com.
- *History of AI: Timeline and the Future* – Maryville University Online, online.maryville.edu.
- *What is Artificial Intelligence?* – UST, ust.com.
- *Strong AI vs. Weak AI: What's the Difference?* – Built In, builtin.com.
- *Types of AI* – Artificial Intelligence and Libraries – Subject Guides, subjectguides.library.american.edu.
- *4 Types of AI: Getting to Know Artificial Intelligence* – Coursera, coursera.org.
- *Ethical Considerations of the Trolley Problem in Autonomous Driving: A Philosophical and Technological Analysis* – MDPI, mdpi.com.
- *AI & Ethics - Artificial Intelligence (AI) in Health Sciences* – LibGuides at University of New Mexico, libguides.health.unm.edu.
- *The Frontlines of Artificial Intelligence Ethics: Human-Centric Perspectives on Technology's Advance* – Routledge, routledge.com.
- *What is the history of artificial intelligence (AI)?* – Tableau, tableau.com.
- *The Most Significant AI Milestones So Far* – Bernard Marr, bernardmarr.com.
- *A short history of AI in 10 landmark moments* – The World Economic Forum, weforum.org.
- *Advancements in Artificial Intelligence and Machine Learning* – Online Master's in Engineering | CWRU, online-engineering.case.edu.

- *Special Issue: Advancements in Deep Learning and Its Applications* – MDPI, mdpi.com.
- *The birth of Artificial Intelligence (AI) research* – Science and Technology, st.llnl.gov.
- *What is an AI winter and is one coming?* – Search Engine Land, searchengineland.com.
- *History of Machine Learning: How We Got Here* – Akkio, akkio.com.
- *History of Artificial Intelligence - ChatGPT* – Research Guides at Black Hawk College, bhc.libguides.com.
- *History of Machine Learning* – A Journey through the Timeline – Clickworker, clickworker.com.
- *The historic Dartmouth Conference of 1956* - Setting the stage for AI – roboticsbiz.com.
- *A Comparative Analysis of the First and Second AI Winters* – Demandify Media, demandifymedia.com.
- *What is Deep Learning? Its history and place in our future* – Thematic, getthematic.com.
- *The AI Boom (1980–1987)* – Making Things Think – Holloway books, holloway.com.
- *What Is Deep Learning and How Does It Work?* – Built In, builtin.com.
- *Intelligenza artificiale in carcere* – rm.coe.int.
- *Technology in Prisons: The Impact of Artificial Intelligence* – International Network for Criminal Justice, criminaljusticenetwork.net.
- *The use and future of artificial intelligence monitoring in prisons* – Reason Foundation, reason.org.
- *Intelligenza artificiale nei servizi penitenziari e di libertà vigilata: la nuova raccomandazione del Consiglio d'Europa* – Consiglio d'Europa, coe.int.
- *IA e riconoscimento delle emozioni: rischi e possibili vantaggi*, articolo del 21 settembre 2023, pubblicato su Agendadigitale.eu.
- *L'uso dell'Intelligenza Artificiale nei rapporti di polizia: svolta tecnologica o rischio per la giustizia?*, articolo del 5 ottobre 2024, pubblicato su www.rivista.ai.

- *La Commissione Ue ritira la direttiva sulla responsabilità da intelligenza artificiale*, articolo del 12 febbraio 2025, pubblicato sul portale dell'agenzia ANSA.
- *L'intelligenza artificiale senza coscienza è niente*, articolo del 28 aprile 2025, pubblicato sul portale dell'agenzia ANSA.
- *Niente limiti all'intelligenza artificiale: Italia, Germania e Francia bloccano la legge Ue*, in europa.today.it.
- *Intelligenza artificiale: la Francia apre la strada alla sorveglianza di massa in Europa*, in contropiano.org.
- *L'Intelligenza Artificiale: l'approccio olandese*, in spremutedigitali.com.