# INTELLIGENZA ARTIFICIALE. LA DISCIPLINA

LUCIANO M. QUATTROCCHIO

## *1. Premessa.*

Il 21 maggio 2024, il Consiglio dell'UE ha approvato l'*Artificial intelligence Act* (AI Act), ovvero la legge europea sull'intelligenza artificiale, tesa ad armonizzare in tutta l'Unione le norme sull'intelligenza artificiale con un approccio "basato sul rischio": maggiore è il rischio di causare danni, più severe sono le regole.

Il Regolamento sull'IA si applica solo ad ambiti soggetti al diritto dell'UE e prevede esenzioni, ad esempio, per i sistemi utilizzati esclusivamente per scopi militari e di difesa, nonché per scopi di ricerca.

L'AI Act classifica diversi tipi di intelligenza artificiale. I sistemi di IA che presentano solo un rischio limitato sono soggetti a obblighi di trasparenza, mentre i sistemi di IA ad alto rischio sono ammessi, ma assoggettati a una serie di requisiti e obblighi per ottenere l'accesso al mercato dell'UE.

Al contrario altri sistemi di intelligenza artificiale come, ad esempio, i sistemi di manipolazione cognitiva comportamentale e i sistemi di attribuzione di un punteggio a fronte del comportamento (*social scoring*) sono vietati dall'UE perché il loro rischio è ritenuto inaccettabile. L'AI Act vieta inoltre l'uso dell'intelligenza artificiale per trattamenti predittivi basati sulla profilazione e sistemi che utilizzano dati biometrici per classificare le persone in base a categorie specifiche come razza, religione o orientamento sessuale.

Il Regolamento sull'IA disciplina anche l'uso di modelli di intelligenza artificiale per finalità generali: se non presentano rischi sistemici sono soggetti ad alcuni requisiti, ad esempio in materia di trasparenza, mentre quelli con rischi sistemici devono rispettare regole più severe.

Il regolamento UE prevede, inoltre, una maggiore trasparenza per quanto riguarda lo sviluppo e l'uso di sistemi di IA ad alto rischio: tali sistemi e anche gli utenti pubblici, che ne fanno uso, sono registrati in una apposita banca dati dell'UE per i sistemi di IA ad alto rischio.

Per garantire l'effettiva applicazione, sono stati istituiti diversi organi: un ufficio per l'IA presso la Commissione; un gruppo scientifico di esperti indipendenti a sostegno della Commissione per la stesura degli atti di esecuzione; un comitato per l'IA con

rappresentanti degli Stati membri per consigliare e assistere la Commissione e gli Stati membri in merito all'applicazione coerente ed efficace del Regolamento; un *forum* consultivo per i portatori di interessi, al fine di fornire consulenza tecnica al comitato per l'IA e alla Commissione.

Il regolamento prevede una scaletta progressiva di inizio di operatività per singoli capi o articoli dello stesso:

- la prima fase è collocata decorsi sei mesi dall'entrata in vigore, ovvero a fare data dal 2 febbraio 2025;
- la seconda fase è collocata decorsi dodici mesi dall'entrata in vigore, ovvero a fare data dal 2 agosto 2025;
- la terza fase è collocata decorsi ventiquattro mesi dall'entrata in vigore e a quella data (2 agosto 2026) il regolamento sarà a regime, salvo l'articolo 6;
- una quarta fase riguarda l'articolo, ovvero l'articolo 6, par. 1, dedicato alla classificazione dei sistemi di IA ad alto rischio ed i corrispondenti obblighi la data di decorrenza prevista è il 2 agosto 2027.

Per completare il quadro, è opportuno segnalare che il Governo italiano ha presentato un disegno di legge, all'esame del Senato (atto n. 1146), che contiene la delega legislativa per l'armonizzazione nell'ordinamento italiano dell'AI Act. Il disegno di legge rappresenta, peraltro, una legge quadro per la disciplina dell'intelligenza artificiale. Il 25 giugno 2025 la Camera dei Deputati ha approvato in seconda lettura il ddl 1146/2024.

### *2. Il Regolamento europeo 2024/1689 sull'intelligenza artificiale.*

Il 12 luglio 2024 è stato pubblicato sulla Gazzetta Ufficiale dell'Unione Europea il tanto atteso AI Act, ovvero la prima regolamentazione a livello europeo in materia di intelligenza artificiale.

Il Regolamento ha adottato un *risk-based approach*, un approccio fondato sul rischio, andando a creare una "piramide" dei singoli sistemi di intelligenza artificiale:

- applicazioni a rischio minimo, come ad esempio i filtri *anti-spam*, che non sono soggette a regole specifiche;
- applicazioni a rischio limitato, come ad esempio i *chatbot*, che prevedono obblighi di trasparenza;
- applicazioni ad alto rischio (es. quelle usate in sanità, nella giustizia, nella classificazione dei lavoratori), che possono avere un impatto significativo sui

diritti fondamentali e sono soggette a regole severe, con obblighi di trasparenza, valutazioni di conformità e monitoraggio umano.

Inoltre, l'AI Act identifica specifiche pratiche di intelligenza artificiale a rischio inaccettabile, vietandone l'uso dal 2 febbraio 2025.

Tra queste pratiche rientrano:

a) Tecniche di manipolazione subliminale o ingannevole: Sistemi che influenzano il comportamento degli individui senza la loro consapevolezza, sfruttando vulnerabilità cognitive o psicologiche.
b) Sfruttamento delle vulnerabilità di gruppi specifici: Applicazioni che approfittano delle vulnerabilità di gruppi particolarmente vulnerabili, come minori o persone con disabilità, per influenzare il loro comportamento in modo dannoso.
c) Sistemi di *social scoring*: Valutazioni sistematiche della reputazione o dell'affidabilità delle persone basate sul loro comportamento sociale o sulle caratteristiche personali, che possono portare a discriminazioni ingiustificate.
d) Identificazione biometrica remota in tempo reale in spazi pubblici: L'uso di sistemi di riconoscimento facciale o altre tecnologie biometriche per identificare persone in tempo reale in luoghi pubblici, salvo specifiche eccezioni.
e) Riconoscimento delle emozioni in ambiti sensibili: Applicazioni che cercano di determinare le emozioni degli individui in contesti come il lavoro o l'istruzione, dove ciò potrebbe portare a discriminazioni o violazioni della *privacy*.
f) Creazione o ampliamento di banche dati di riconoscimento facciale tramite *scraping* non mirato: La raccolta massiva di immagini o dati biometrici da fonti online senza il consenso degli individui coinvolti, per creare database utilizzati in sistemi di riconoscimento facciale.

Dunque, i soggetti che utilizzano o sviluppano sistemi di IA devono garantire che nessuna di queste pratiche vietate sia stata sviluppata nei loro prodotti o servizi. La violazione di questi divieti può comportare sanzioni significative, che possono raggiungere i 35 milioni di euro o il 7% del fatturato annuo globale della società, a seconda di quale importo sia maggiore.

Da tale data, inoltre, in tutti gli Stati membri UE valgono le definizioni contenute nell'AI Act – che, vale la pena ricordare, non necessita di recepimento negli ordinamenti nazionali, trattandosi di regolamento – ed è scattato l'obbligo, a carico di fornitori e *deployer* (quindi, gli *utilizzatori)* di sistemi di IA, di garantire "un *livello sufficiente di*

*alfabetizzazione in materia di IA del loro personale, nonché di qualsiasi altra persona che si occupa del funzionamento e dell'utilizzo dei sistemi di IA per loro conto".*

Il legislatore ha, dunque, adottato un duplice approccio per la sicurezza delle persone.

Da un lato, ha previsto una tutela *ex lege* attraverso un regime di divieto generale di quei sistemi di IA utilizzabili a fini di manipolazione, sfruttamento e controllo sociale e, dunque, in grado di mettere maggiormente a repentaglio i diritti e i valori su cui si fonda l'Unione europea, come il *rispetto della dignità umana, la libertà, l'uguaglianza, la democrazia, il diritto alla non discriminazione, alla protezione dei dati e alla vita privata e i diritti dei minori.*

Dall'altro lato, ha fornito uno strumento di *self-defense,* vale a dire l'alfabetizzazione in materia di IA (o *AI Literacy)* cioè la promozione e la diffusione di competenze, conoscenze e comprensione dei sistemi di IA, in modo da generare consapevolezza in merito alle opportunità e ai rischi dell'AA e ai possibili danni che essa può causare. Dunque, insieme ad autorità e Stati membri, sono chiamati in prima linea gli stessi fornitori e utilizzatori di sistema di IA, su cui incombe questo obbligo di garantire la corretta e sufficiente formazione in materia di IA. D'altronde, come nel contesto della *cybersecurity* e, prima ancora, della protezione dei dati personali, tra i presidi non possono che esservi anche l'educazione, la sensibilizzazione e l'*awareness* dei singoli.

Come si è detto, l'obbligo di alfabetizzazione riguarda tutti i soggetti che utilizzano sistemi e modelli di intelligenza artificiale.

Per quanto riguarda i fornitori, l'applicazione del Regolamento abbraccia anche i soggetti non stabiliti nell'UE, atteso che l'ambito di applicazione oggettiva del Regolamento coinvolge anche scenari in cui l'*output* prodotto dai sistemi di IA è utilizzato all'interno del territorio europeo, seppur prodotto al di fuori del suo territorio.

Successivamente, a partire dal mese di agosto 2025, entreranno in vigore le norme sulla governance dell'IA e gli obblighi specifici per i modelli di IA di uso generale ("*General Purpose AI*" o "*GPAI*").

I destinatari della normativa dovranno:

a) mantenere documentazione dettagliata sui *test* e sullo sviluppo dei loro sistemi di IA;
b) seguire procedure standardizzate per garantire la sicurezza dei sistemi di IA durante tutto il loro ciclo di vita;

c) effettuare valutazioni periodiche per assicurarsi che i sistemi di IA rispettino le normative vigenti.

Entro il 2 agosto 2026 è prevista l'applicazione completa dell'AI Act per tutti i sistemi di IA, inclusi quelli classificati come ad alto rischio, per i quali sono previsti adempimenti più onerosi.

Le organizzazioni che sviluppano o utilizzano sistemi di AI ad alto rischio sono elencate nell'Allegato III dell'AI Act, e comprendono – tra le altre – soluzioni per il riconoscimento biometrico o ancora sistemi valutativi o di selezione in ambito lavorativo, scolastico o per l'accesso a servizi essenziali.

Il termine è, invece, di 36 mesi (sino al 2 agosto 2027) per altre tipologie si sistemi di AI ad alto rischio – non ricompresi nell'Allegato III – destinati ad essere utilizzati come componenti di sicurezza di un prodotto o che sono essi stessi un prodotto, per il quale è prescritta una valutazione di conformità ai sensi della normativa UE (ad esempio macchinari industriali, giocattoli o dispositivi medici).

Va, inoltre, rilevato che il Regolamento disciplina anche i sistemi di intelligenza artificiale già immessi sul mercato o che saranno immessi sul mercato prima della data di applicazione delle relative disposizioni. In particolare, i fornitori di modelli di GPAI immessi sul mercato prima del 2 agosto 2025, dovranno adottare le misure necessarie per conformarsi agli obblighi stabiliti dal Regolamento entro il 2 agosto 2027.

In particolare, sarà applicabile dal 2 agosto 2027 la classificazione del sistema ad alto rischio in presenza delle condizioni delineate all'art. 6, 1° comma, mentre una disciplina speciale è riservata ai sistemi e ai modelli di IA già immessi sul mercato o messi in servizio prima dell'applicazione del Regolamento. Più specificamente:

- i sistemi di IA ad alto rischio, già in uso prima del 2 agosto 2026, devono osservare il Regolamento soltanto qualora, a decorrere da tale data, siano stati soggetti a modifiche significative della loro progettazione;
- i fornitori di modelli di IA per finalità generali che sono stati immessi sul mercato prima del 2 agosto 2025 devono conformarsi agli obblighi del Regolamento entro il 2 agosto 2027;
- i sistemi di IA facenti parte di sistemi IT su larga scala (si intendono i sistemi informatici operanti nello spazio di libertà, sicurezza e giustizia, come ad esempio i sistemi deputati alla gestione delle frontiere nei Paesi Schengen o i sistemi di informazione sui visti) devono essere resi conformi al Regolamento, ove messi in

servizio o immessi sul mercato prima del 2 agosto 2027.

Per i sistemi di AI – che siano componenti di sistemi IT su larga scala stabiliti dal diritto dell'UE nei settori della sicurezza e della giustizia, specificamente individuati nel Regolamento (ad esempio, lo *Schengen Information System*), e che siano immessi sul mercato prima del 2 agosto 2027 – il termine ultimo di adeguamento è indicato al 31 dicembre 2030.

Con riguardo, invece, ai sistemi di AI ad alto rischio – diversi da quelli sopra menzionati, immessi sul mercato o messi in servizio prima del 2 agosto 2026 – le regole si applicheranno solo se tali sistemi saranno soggetti a modifiche significative della loro progettazione. Ciò significa che, in mancanza di modifiche significative, le disposizioni dell'AI Act destinate a tale tipologia di sistemi non saranno suscettibili di applicazione, con l'eccezione delle soluzioni di AI ad alto rischio destinate all'uso da parte delle autorità pubbliche; per questi sistemi, gli sviluppatori e gli utilizzatori dovranno, in ogni caso, adottare le misure necessarie per conformarsi ai requisiti e agli obblighi del Regolamento entro il 2 agosto 2030.

Sempre nel quadro del processo di implementazione dell'AI Act, va segnalato che la Commissione Europea ha già provveduto ad istituire l'AI Office. L'AI Office svolgerà un ruolo fondamentale nell'attuazione dell'AI Act, soprattutto in relazione ai modelli di GPAI. L'Ufficio garantirà l'attuazione coerente della nuova normativa, prestando il proprio supporto in favore degli Stati membri. Avrà, inoltre, il compito di monitorare l'effettiva attuazione degli obblighi prescritti dal Regolamento da parte dei fornitori di modelli di GPAI. In collaborazione con gli sviluppatori di AI, la comunità scientifica e le altre parti interessate, l'AI Office coordinerà altresì la stesura di codici di condotta.

### *3. L'AI Pact.*

È interessante, in tale contesto, oltre alla citata previsione di strumenti di *soft law* ed autoregolamentazione come i codici di condotta (art. 56), l'iniziativa della Commissione denominata *AI Pact*.

La Commissione ha, già da tempo, promosso l'impegno volontario delle aziende del settore ad anticipare gli effetti dell'AI Act e ad avviare l'attuazione dei suoi requisiti prima del termine, all'evidente scopo di creare *standard* di riferimento per il mercato.

In particolare, a fine 2023 è stato pubblicato il primo invito a manifestare interesse, che ha ottenuto risposte da oltre 550 organizzazioni di varie dimensioni, settori e paesi.

L'AI Office ha fondato lo sviluppo dell'*AI Pact* su due pilastri:

- il primo pilastro prevede il coinvolgimento delle organizzazioni che hanno espresso interesse per il patto, incoraggiando lo scambio di pratiche e fornendo informazioni operative sul processo di attuazione dell'AI Act;
- il secondo pilastro contempla l'invito ai fornitori ed operatori di sistemi di IA ad adottare in anticipo misure per conformarsi ai requisiti e agli obblighi stabiliti dal Regolamento. Nello specifico, le imprese che forniscono o distribuiscono sistemi di IA possono dimostrare e condividere i loro impegni volontari in materia di trasparenza e gestione dei prodotti ad alto rischio e prepararsi alla loro attuazione, formulando "dichiarazioni di impegno".

### *4. AI: le Linee Guida della Commissione europea sulle pratiche proibite previste dall'AI Act.*

Il 4 febbraio 2025 la Commissione europea ha pubblicato un primo *set* di linee guida volte a meglio chiarire alcune disposizioni del Regolamento (UE) 2024/1689 ("AI Act") e, in particolare, le pratiche di intelligenza artificiale proibite previste dall'articolo 5, uno schema (non ancora ufficialmente adottato) di un documento che illustra gli "*Orientamenti della Commissione sulle pratiche vietate di intelligenza artificiale istituiti dal regolamento (UE) 2024/1689 (AI Act)*" ("*Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689*").

Tali linee guida – volte a garantire un'applicazione uniforme delle norme da parte delle autorità competenti e degli stakeholders – sono particolarmente rilevanti in quanto fanno riferimento ad un articolo dell'AI Act, dallo scorso 2 febbraio già pienamente applicabile.

Le linee guida non sono vincolanti per gli operatori, ma rappresentano uno strumento interpretativo per l'implementazione dell'art. 5, e chiariscono come l'applicazione delle proibizioni debba essere basata su una valutazione caso per caso, che tenga conto del contesto e delle circostanze specifiche.

In particolare, le Linee Guida sulle Pratiche Proibite forniscono una panoramica delle pratiche proibite dall'art. 5, chiarendo alcuni punti controversi attraverso esempi pratici.

Come si è detto, sono vietati i sistemi di intelligenza artificiale che usano tecniche subliminali o manipolative per distorcere il comportamento di una persona, inducendola a prendere decisioni dannose. È, altresì, vietato l'uso dell'IA per sfruttare vulnerabilità dovute a età, disabilità o condizioni socio-economiche.

Tra gli esempi forniti, un videogioco basato su neurotecnologie che utilizza interfacce cervello-macchina per raccogliere dati, anche di natura sensibile, senza che il giocatore ne sia consapevole. L'intelligenza artificiale potrebbe, infatti, sfruttare questi dati per influenzare le decisioni di gioco in modo che il giocatore spenda più denaro negli acquisti *in-app*. Il divieto riguarda i soli casi di manipolazione subliminale significativamente dannosa e non, in generale, le applicazioni di interfaccia macchina-cervello, se progettate in modo sicuro e rispettoso della *privacy* e dell'autonomia individuale.

È, inoltre, proibito qualsiasi sistema di classificazione sociale basato su comportamenti o caratteristiche personali che porti a trattamenti discriminatori o sproporzionati.

Tra i casi di pratiche vietate delineati dalla Commissione, l'utilizzo da parte di un'agenzia per l'assistenza sociale di un sistema di IA per valutare la probabilità di frode da parte dei beneficiari di assegni familiari che si basa su caratteristiche raccolte o desunte da contesti sociali, senza alcun collegamento apparente con la frode, come il fatto di avere un coniuge di una certa nazionalità o origine etnica, di avere una connessione a *Internet*, il comportamento sulle piattaforme sociali o le prestazioni sul posto di lavoro. Al contrario, i dati rilevanti per l'assegnazione delle prestazioni e raccolti legalmente potrebbero essere utilizzati per determinare il rischio di frode, poiché le autorità pubbliche perseguono un obiettivo legittimo nel verificare se le prestazioni sociali sono assegnate correttamente.

Sono, altresì, vietati i sistemi di IA che valutano o prevedono il rischio di crimini basandosi unicamente su *profiling* o caratteristiche personali. Sono, invece, consentiti i sistemi che supportano valutazioni umane basate su fatti obiettivi e verificabili.

Per la Commissione è, ad esempio, vietato un *software AI based* utilizzato dalla polizia per prevedere il rischio che bambini e adolescenti siano coinvolti in "*futuri reati violenti e contro la proprietà*", basandosi esclusivamente sulle loro relazioni con altre persone e ai loro presunti livelli di rischio; il che significa che i bambini possono essere considerati a rischio di reato più elevato semplicemente per il fatto di essere collegati a un altro individuo con una valutazione ad alto rischio, come un fratello o un amico. Anche i livelli di rischio dei genitori possono influire sul livello di rischio del bambino. Un simile sistema può portare a discriminazioni razziali e sociali.

È, poi, vietata la creazione di database di riconoscimento facciale tramite *scraping* non mirato di immagini da Internet o telecamere di sorveglianza.

Le linee guida individuano come vietata la condotta di una società che raccoglie fotografie dai social media attraverso un sistema di *scraping* automatico, che individua le immagini

contenenti volti umani e raccoglie queste immagini con tutte le informazioni associate (come la fonte dell'immagine (URL), la geolocalizzazione e talvolta i nomi delle persone), per consentire agli utenti di ricercare attraverso l'immagine di un individuo se quest'ultimo è presente sul sistema di intelligenza artificiale: un caso molto simile a quello già sanzionato dal Garante *Privacy* per violazione del *GDPR* e che ha visto coinvolta la società statunitense *ClearView AI*.

È, inoltre, vietato l'uso di sistemi IA per dedurre emozioni sul luogo di lavoro o in contesti educativi.

Tra gli esempi di pratiche vietate fornite dalla Commissione: l'uso di *webcam* e sistemi di riconoscimento vocale da parte di un *call center* per tracciare le emozioni dei propri dipendenti, come la rabbia. Se utilizzati solo a scopo di formazione del personale, i sistemi di riconoscimento delle emozioni sono consentiti, solo nella misura in cui i risultati non siano condivisi con i responsabili delle risorse umane e non possano influire sulle valutazioni, sugli avanzamenti di carriera e, più in generale, sul rapporto di lavoro.

Sono, altresì, proibiti i sistemi che classificano le persone in base a dati biometrici per dedurne origine etnica, opinioni politiche, religione o orientamento sessuale.

La Commissione considera vietato un sistema di intelligenza artificiale che classifica le persone attive su una piattaforma *social* in base al loro presunto orientamento sessuale, analizzando i dati biometrici delle foto condivise su tale piattaforma, e su tale base propone a tali persone annunci pubblicitari.

È, poi, vietato l'uso di sistemi di identificazione biometrica in spazi pubblici per finalità di polizia, salvo eccezioni strettamente regolamentate: ricerche mirate per vittime di gravi crimini, prevenzione di attacchi terroristici, localizzazione di sospetti per reati gravi.

Al riguardo, le linee guida chiariscono che possono considerarsi spazi accessibili al pubblico anche luoghi che possono essere utilizzati per il commercio, come negozi, ristoranti e caffè; per i servizi, come banche, attività professionali (uno studio medico così come uno studio contabile) o *hotel*; per il trasporto, come stazioni di autobus, metropolitana e ferrovie e aeroporti; per l'intrattenimento, come cinema, teatri, musei, sale per concerti e conferenze.

Le linee guida esplorano anche le aree di sovrapposizione tra AI Act ed altre normative europee, l'*enforcement* dell'articolo 5 dell'AI Act nonché le sanzioni previste per il caso di violazione (fino a 35.000.000 di euro o, se il trasgressore è un'impresa, fino al 7 % del suo fatturato mondiale totale annuo dell'esercizio precedente, se superiore); nonché il

perimetro delle esclusioni previste dall'ambito di applicazione dell'AI Act: ad esempio, con riguardo ai sistemi di IA sviluppati per la sicurezza nazionale, difesa o scopi militari; nell'ambito di attività di ricerca e sviluppo; o per l'uso personale e non professionale di sistemi di IA da parte di privati.

### *5. Al e diritto d'autore nel report della Global Partnership on Artificial Intelligence (GPAI).*

La GPAI ha pubblicato un *report* che analizza le possibili soluzioni alle criticità che caratterizzano il rapporto tra lo sviluppo dell'intelligenza artificiale e la tutela del diritto d'autore

La GPAI è un'iniziativa internazionale che si compone allo stato di 44 paesi membri, tra cui l'Italia, e che promuove lo sviluppo e l'uso responsabile dell'intelligenza artificiale.

Anche la GPAI si è occupata del difficile rapporto tra Al generativa e *copyright* e, più nello specifico, delle note preoccupazioni in merito alla tutela del diritto d'autore legate all'aumento della domanda di dati per l'addestramento degli algoritmi di intelligenza artificiale.

Il *report*, del febbraio di quest'anno, ha analizzato in particolare il fenomeno del c.d. *data scraping*, ossia il metodo principale per ottenere i dati utili all'addestramento dell'Al, che fa ricorso in buona sostanza all'estrazione automatizzata di informazioni da siti, *database* e piattaforme di social media di terze parti.

Il *report* evidenzia, peraltro, come il *data scraping* venga utilizzato non solo per fini commerciali, ma anche ad altri fini, ad esempio per il sostegno alla ricerca scientifica. Da cui l'opportunità di disporre di strumenti di regolamentazione differenti a seconda dei diversi utilizzi.

La GPAI suggerisce codici di condotta volontari, strumenti tecnici e clausole contrattuali standardizzati, oltre ad iniziative di sensibilizzazione. Secondo la relazione, gli strumenti tecnici standardizzati potrebbero includere meccanismi di controllo dell'accesso ai dati, un monitoraggio automatizzato dei contratti e sistemi di pagamento diretto.

### *6. Antiriciclaggio e AI: l'evoluzione dell'obbligo di segnalazione di operazione sospetta*

In data 4 luglio 2025, è stato posto in consultazione il provvedimento con cui l'UIF detta le nuove istruzioni per la rilevazione e la segnalazione delle operazioni sospette. L'UIF ricorda che i destinatari dell'obbligo segnaletico possono avvalersi di strumenti, anche

informatici, per la selezione delle operatività anomale e basati su regole e parametri quantitativi e qualitativi. Nell'ambito di questi strumenti – cui è opportuno ricorrere, soprattutto in presenza di attività caratterizzate da operazioni frequenti o della stessa tipologia, in funzione delle esigenze di contenimento del rischio – rientrano anche quelli basati sull'intelligenza artificiale che, laddove utilizzati, devono essere conformi alle disposizioni a essi eventualmente applicabili e basarsi su dati oggettivi e verificabili e su adeguate valutazioni svolte con l'intervento umano, al fine di controllare ed eventualmente validare le anomalie da essi evidenziate.

Il motivo alla base della riedizione delle istruzioni va ricercato nel più ampio processo di riforma della disciplina unionale cui si associa, indissolubilmente, la precisa volontà di "*agevolare la rilevazione delle operazioni sospette e assicurare tempestività, completezza e riservatezza della segnalazione nonché di accrescerne la qualità*".

Ben si comprende quindi il richiamo a "*strumenti, anche informatici, per la selezione delle operatività anomale e basati su regole e parametri quantitativi e qualitativi*" invogliando, se del caso, anche al ricorso all'intelligenza artificiale per l'emersione di comportamenti e schemi operativi anomali.

***7. Protezione dei dati personali.***

In data 17 dicembre 2024, il Comitato Europeo per la Protezione dei dati personali ("EDPB") ha adottato l'Opinion 28/2024 on "*certain data protection aspects related to the processing of personal data in the context of AI models*", su richiesta della *Data Protection Commission* ("DPC") irlandese. Il Parere affronta temi chiave fondamentali legati al trattamento dei dati personali nelle fasi di *development* e *deployment* dei modelli di IA.

L'apertura del parere opera una precisione di rilievo secondo la quale a prescindere dall'uso dei dati personali per il loro addestramento, i modelli di IA sono progettati per fare previsioni o deduzioni, spesso generando inferenze anche su soggetti diversi da quelli i cui dati personali sono stati utilizzati nel *training*. Alcuni esempi sono i modelli generativi perfezionati sulle registrazioni vocali di una persona per imitarne la voce o anche i modelli progettati per rispondere con dati personali contenuti nel *set* di addestramento quando richiesti. Nel caso specifico, rispondendo alla prima domanda posta dalla DPC, il Comitato si concentra sui modelli di IA non progettati per fornire dati personali riconducibili ai dati di *training*. Sul punto, l'EDPB assume una posizione tanto

pragmatica quanto prudente: secondo il Parere, infatti, non tutti i modelli di IA possono essere considerati anonimi in maniera automatica. Tale considerazione discende dal fatto che i dati personali potrebbero rimanere "assorbiti" nei parametri del modello; pertanto, anche se gli stessi differiscono dai punti originari del *set* di addestramento, potrebbero mantenere informazioni riconducibili ai dati personali di origine. L'anonimato deve essere dunque valutato caso per caso attraverso il ricorso a criteri specifici che analizzino le caratteristiche tecniche del modello e la possibilità di estrarre, direttamente o indirettamente, informazioni personali dai parametri stessi.

Affinché un modello di IA possa essere considerato anonimo è necessario in particolare che:

- la probabilità di estrazione diretta, anche in termini probabilistici, di dati personali relativi agli individui i cui dati sono stati utilizzati per l'addestramento del modello sia insignificante per qualsiasi interessato;
- e che la probabilità di ottenere i dati personali dell'addestramento, intenzionalmente o accidentalmente, attraverso interrogazioni al modello utilizzando tutti i mezzi ragionevolmente prevedibili da parte del titolare del trattamento o di terzi, sia altrettanto insignificante.

L'*Opinion* distingue inoltre tra modelli pubblici e privati e riconosce che i modelli pubblici poiché accessibili ad una vasta platea di persone fisiche, presentano rischi di identificazione più elevati rispetto ai modelli privati, utilizzati in contesti ristretti. La valutazione della probabilità di identificazione o reidentificazione deve essere effettuata dall'autorità di controllo tenendo conto di fattori specifici, come le caratteristiche progettuali del modello di IA, le quali dovranno tener conto:

- delle modalità di selezione delle fonti di dati, di preparazione e minimizzazione dei dati di addestramento;
- delle scelte metodologiche effettuate durante l'addestramento del modello e delle misure adottate per garantire che gli output generati dal modello non contengano informazioni personali.

Alla documentazione è assegnato un ruolo centrale: infatti, secondo il Comitato il titolare del trattamento è tenuto a dimostrare la natura anonima di un modello di IA attraverso qualsiasi informazione relative alla DPIA comprese:

- le valutazioni e le decisioni che hanno determinato che la valutazione di impatto di cui all'art. 35 del GDPR non fosse necessaria;

- il parere del DPO;
- esiti di *test*;
- l'implementazione di misure specifiche come filtri di *output*, architetture *privacy by design* (ad esempio, l'uso della *differential privacy*) e procedure di validazione.

Se il titolare riesce a fornire evidenze solide che escludano la possibilità di reidentificazione, il modello può essere considerato escluso dall'ambito di applicazione del GDPR.

In questo contesto, la documentazione svolge un ruolo essenziale sia nella verifica della natura anonima del modello sia nella dimostrazione del rispetto degli obblighi di responsabilizzazione previsti dal GDPR. Qualora l'autorità di controllo, sulla base della documentazione fornita, non rilevasse misure efficaci per garantire l'anonimizzazione del modello, tale situazione potrebbe configurare una violazione degli obblighi di *accountability*.

Un tema centrale dell'*Opinion* riguarda le questioni poste dalla DPC nelle domande due e tre, relative alla possibilità di invocare il legittimo interesse come base giuridica ai sensi dell'art. 6(1)(f) GDPR nelle fasi di *development* e *deployment* dei modelli di IA.

Premesso che non è possibile individuare una gerarchia tra le basi giuridiche per il trattamento dei dati e dovendosi sottolineare l'importanza di un'analisi e documentata per ogni ipotesi di trattamento, nel caso del legittimo interesse, il titolare del trattamento deve sempre dimostrare di soddisfare le tre condizioni del cosiddetto "*three-step test*" (Cfr. Guidelines 1/2024 sull'art. 6(1)(f) GD PR):

- la prima condizione richiede che il titolare identifichi un interesse legittimo, lecito e concretamente perseguito, allo scopo di evitare finalità vaghe, speculative o illecite;
- la seconda condizione riguarda la necessità del trattamento: il *processing* dei dati deve, infatti, essere indispensabile per raggiungere la finalità identificata e il titolare deve dimostrare che non esistano alternative meno invasive che possano ottenere lo stesso risultato;
- la terza condizione prevede un'analisi di bilanciamento (*balancing test*), in cui il titolare deve valutare i potenziali rischi per i diritti e le libertà fondamentali degli interessati rispetto ai benefici, pubblici o privati, derivanti dal trattamento. Tale valutazione deve tener conto delle ragionevoli aspettative degli interessati, che potrebbero anche variare in base alla natura dei dati trattati, al contesto del

trattamento e alle possibili conseguenze, nonché alla probabilità che queste si verifichino. Elementi quali l'origine dei dati e il metodo di raccolta assumono un ruolo di particolare importanza soprattutto nei casi in cui il trattamento si colleghi al *web scraping* o si realizzi in assenza di un rapporto diretto con gli interessati e tali aspetti determinano un impatto significativo sul bilanciamento tra i diritti degli interessati e il legittimo interesse del titolare.

Il Comitato riconosce in ogni caso che i titolari, per controbilanciare eventuali esiti sfavorevoli del *balancing test*, possono adottare misure di mitigazione ulteriori, le stesse devono essere progettate per minimizzare l'impatto del trattamento sui diritti e sulle libertà degli interessati, tenendo conto delle specificità e delle criticità che emergono nelle fasi di *development* e *deployment* dei modelli di IA. Il Parere individua a titolo esemplificativo alcune di queste misure, per la fase di sviluppo, ad esempio, si raccomandano tecniche come il *data masking* e la pseudonimizzazione, oltre all'utilizzo di dati fittizi, laddove i dati personali non siano indispensabili.

Nella fase di *deployment*, invece, l'Opinion suggerisce l'introduzione di filtri sugli *output* per evitare la generazione di dati personali, specialmente nei modelli generativi, e l'uso di *watermarking* digitale per ridurre il rischio di riutilizzo illecito.

L'EDPB evidenzia la necessità di facilitare l'esercizio dei diritti degli interessati attraverso misure come il ritardo temporale tra la raccolta e l'utilizzo dei dati, l'introduzione di un *opt-out* incondizionato per l'esclusione dal *dataset* e l'ampliamento del diritto alla cancellazione anche nei casi in cui non ricorrano i requisiti previsti dall'art. 17(1) GDPR.

Il Parere propone una trasparenza rafforzata al fine di garantire agli interessati informazioni chiare sulle modalità di raccolta, sulle fonti e sulle finalità dei dati attraverso strumenti come campagne di sensibilizzazione e FAQ.

Relativamente al *web scraping*, il Parere raccomanda di escludere la raccolta di dati particolari o invasivi, di rispettare i meccanismi di esclusione previsti da strumenti come *robots.txt* o ai *.txt*, di limitare rigorosamente l'ambito della raccolta e di predisporre una *opt-out list* che consenta agli interessati di opporsi alla raccolta dei propri dati.

Il quarto quesito dell'Autorità irlandese esamina le conseguenze legali di un trattamento illecito (ex art. 5 e 6 del GD PR) durante la fase di sviluppo di un modello di IA e il suo impatto sui trattamenti successivi nella fase di implementazione. Il Parere analizza due

situazioni distinte: una in cui il modello conserva dati personali e un'altra in cui i dati sono stati anonimizzati, sviluppando l'analisi in diversi scenari.

Nel primo scenario, se un titolare del trattamento effettua un trattamento illecito di dati personali per sviluppare un modello di IA, con i dati personali conservati nel modello stesso e successivamente trattati dallo stesso titolare nella fase di *deployment*, il Comitato ritiene che la base giuridica per il trattamento successivo potrebbe essere compromessa.

Nel secondo scenario, invece, se un modello sviluppato in violazione del GDPR da un titolare viene adottato da un diverso titolare nella fase di *deployment*, resta fermo anzitutto il principio che ciascun titolare deve garantire la liceità del trattamento dei dati. Il titolare che utilizza il modello è sempre tenuto a condurre una valutazione per assicurarsi che il modello non sia stato sviluppato utilizzando dati personali trattati illecitamente.

Aspetto rilevante resta altresì il fatto che i titolari della fase di *deployment* potrebbero non avere accesso alle stesse informazioni del titolare della fase di sviluppo. Pertanto, viene raccomandato un approccio proporzionato da parte delle autorità di controllo, la profondità delle verifiche in base al livello di rischio e alla complessità del modello di IA.

In ogni caso, è necessario tenere in debito conto anche il quadro normativo nazionale applicabile, come ad esempio in Italia, con l'inutilizzabilità dei dati ai sensi dell'articolo 2-decies del Codice della *Privacy*.

Nel terzo scenario, se un titolare del trattamento effettua un trattamento illecito di dati personali per sviluppare un modello di IA, ma successivamente anonimizza i dati ivi contenuti prima che lo stesso titolare o un diverso titolare avvii un successivo trattamento durante la fase di deployment, l'illiceità del trattamento iniziale non dovrebbe influire sull'utilizzo successivo del modello. Resta fermo che una mera dichiarazione di anonimizzazione non soddisfa i requisiti richiesti, rendendosi necessaria la produzione di evidenze tecniche e documentali idonee a dimostrare l'effettiva rimozione di qualsiasi elemento qualificabile come dato personale.

Nel quarto scenario, in cui – nella fase di *deployment* – vengano effettuati trattamenti di nuovi dati personali, l'applicabilità del GDPR si limiterà esclusivamente a tali ulteriori operazioni di trattamento.

In relazione al tema dell'illiceità, l'EDPB richiama l'attenzione sul fatto che, in situazioni di particolare gravità, le autorità di controllo possono esercitare poteri correttivi di natura incisiva, tra cui l'irrogazione di sanzioni amministrative, la limitazione temporanea del trattamento oppure, nei casi più gravi, la cancellazione di una parte del *dataset* trattato in

modo illecito. Nel caso in cui non risulti possibile effettuare una cancellazione parziale, l'autorità di controllo, tenendo conto della gravità del caso e in conformità al principio di proporzionalità, può ordinare la cancellazione integrale del *dataset* utilizzato per lo sviluppo del modello di IA e/o la rimozione completa del modello stesso.

L'EDPB rileva che alcune questioni rilevanti in materia di protezione dei dati e intelligenza artificiale non sono state approfondite nel Parere, poiché escluse dall'ambito della richiesta formulata dall'Autorità irlandese.

Tra queste rientra il trattamento dei dati particolari ai sensi dell'art. 9 GDPR, per il quale si dispone il divieto generale, salvo le deroghe tassativamente previste dal paragrafo 2. L'EDPB ricorda il disposto di cui alla sentenza della Corte di Giustizia dell'Unione Europea del 4 luglio 2023 (C-252/21, Meta contro Bundeskartellamt), secondo la quale, qualora un *dataset* contenga dati particolari non suscettibili di separazione da altre categorie di dati, il relativo trattamento deve essere qualificato come trattamento di dati particolari. Ai fini dell'applicazione dell'eccezione prevista dall'art. 9, par. 2, lett. e) GDPR, è richiesto che i dati siano stati resi pubblici dall'interessato in modo esplicito e inequivocabile, attraverso un atto che manifesti in modo chiaro e volontario la volontà dell'interessato stesso.

Non sono state altresì affrontate, nel dettaglio, le questioni relative alle decisioni automatizzate e al *profiling* ai sensi dell'art. 22 GDPR, né fornite indicazioni su garanzie aggiuntive quali l'intervento umano o la trasparenza per i sistemi di IA che producono effetti significativi sugli interessati. Analogamente, tematiche come la compatibilità delle finalità (art. 6(4) GDPR), la protezione dei dati *by design* (art. 25GDPR) e le *DPIA* (art. 35GDPR), sebbene richiamate, non sono state specificamente declinate per i modelli di IA.