il QG
Il Quotidiano Giuridico

ALTALEX

IP, IT E DATA PROTECTION

Il trattamento dei dati personali nel contesto dei modelli di IA: il parere dell'EDPB

Di Luca Tufarelli, Founding Partner, **Maria Lilia La Porta**, Senior associate e **Liliana Brindisi.** Associate - Studio Legale Ristuccia Tufarelli & Partners

L'integrazione sempre più diffusa di modelli di intelligenza artificiale (IA) nei servizi digitali di aziende pubbliche e private solleva interrogativi rilevanti sulla conformità del trattamento dei dati personali ai principi del GDPR. Il Parere 28/2024 dell'European Data Protection Board (EDPB affronta i principali nodi giuridici legati all'anonimità dei modelli, all'uso del legittimo interesse come base giuridica, e alle conseguenze dei trattamenti illeciti durante lo sviluppo dei modelli. Il documento rappresenta un primo tentativo di tracciamento normativo, ma evidenzia anche l'urgenza di armonizzare il quadro regolatorio tra GDPR e AI Act, per garantire un'efficace tutela dei diritti nell'ecosistema dell'IA.

Premesse

Un numero crescente di aziende pubbliche e private sta integrando nell'offerta di servizi digitali uno o più modelli di intelligenza artificiale (i "Modelli") che sono stati addestrati con dati personali o che potrebbero comportare l'elaborazione di dati personali, con conseguente incremento delle tematiche legate alla conformità del trattamento di tali dati alla normativa di riferimento.

Occorre stabilire quali modalità gli sviluppatori (c.d. **developer**) e i distributori (cd. **deployer**) debbano adottare al fine di garantire che i trattamenti di dati personali utilizzati da tali modelli siano conformi ai dettami del GDPR. In questa prospettiva, lo **European Data Protection Board** (l'"EDPB") - su richiesta dell'Autorità garante per la protezione dei dati personali irlandese – ha adottato il Parere n. 28 del 17 dicembre 2024 al fine di individuare le linee guida a cui tutte le Autorità Garanti degli stati membri devono adattarsi nel prendere in esame la legittimità dell'uso dei dati personali nel contesto dello sviluppo di tali Modelli.

Le questioni sottoposte all'EDPB riguardano i principi applicabili nel contesto delle fasi di sviluppo e di diffusione dei modelli di IA, compresi i Large Language Model (**LLM**). In particolare:

(i) la possibilità di considerare anonimo un modello di IA;

(ii) i limiti di utilizzo di dataset di training contenenti dati personali raccolti sulla base del legittimo interesse;

(iii) le conseguenze che un trattamento illecito di dati personali, nella fase di sviluppo e training di un modello di IA, possa comportare nella fase di utilizzo e diffusione del modello stesso.

Per formulare le risposte ai quesiti sottoposti dall'Autorità irlandese l'EDPB ha preso in esame solo **il sottoinsieme dei modelli di IA che sono stati addestrati con dati personali** e pertanto, i quesiti sono stati esaminati con riferimento quasi esclusivo all'applicazione del **GDPR**, senza quindi confrontarne le interazioni con quadri giuridici complementari, quale quello rappresentato dal recente Regolamento (UE) 2024/1689 in materia di intelligenza artificiale (c.d. "AI Act"). Per vero tale precisazione nasconde uno dei più grandi problemi che ci troveremo ad affrontare in un prossimo futuro giacché il Parere, come si vedrà, si sovrappone agli spazi di regolazione che l'AI Act sembrerebbe riservare all'European Artificial Intelligence Board (EAIB) e alle Autorità nazionali competenti per l'IA.

Il Parere rappresenta comunque il primo punto per tracciare il tema dei rapporti tra GDPR e IA con ampie note e ricostruzioni, di tipo anche fattuale, che contribuiscono ad anticipare gran parte degli atti di regolazione che scaturiranno da questo nuovo quadro normativo. Il parere, pur non risolvendo tutti i dubbi sollevati da vari commentatori sui vari strumenti in cui si declina la tecnologia dell'IA, fornisce comunque importanti indicazioni operative per gli attori della catena del valore dell'IA.

Primo quesito: l'anonimità dei modelli di IA

Il primo quesito sottoposto al Board mira a chiarire **se i Modelli di IA addestrati utilizzando dati personali possano essere considerati come modelli che elaborano dati personali in ogni caso** oppure possano essere considerati anonimi al rispetto di certe condizioni.

L'EDPB ha innanzitutto precisato che i Modelli di IA sono progettati per fare previsioni o trarre conclusioni e che, se sono addestrati con dati personali, sono progettati generalmente per fare inferenze su individui diversi dai soggetti i cui dati sono stati utilizzati nella fase di addestramento. In alcuni casi, però, tali Modelli vengono addestrati appositamente per fornire proprio i dati personali relativi ai soggetti i cui dati sono stati utilizzati in fase di addestramento. Tali Modelli includono intrinsecamente dati personali e, comportando necessariamente un trattamento degli stessi, non possono essere considerati anonimi. **Il parere riguarda perciò solo i Modelli di IA che non sono progettati per fornire i dati personali utilizzati nella fase di addestramento**.

Tali Modelli possono comportare comunque il trattamento dei dati personali utilizzati durante l'addestramento, in quanto tali dati possono rimanere "assorbiti" nei parametri del modello o essere rappresentati tramite grandezze matematiche che, sebbene non

immediatamente intellegibili, conservano in una forma diversa le informazioni originali ad essi associate. In questo caso, i dati possono essere estratti, direttamente o indirettamente, attraverso tecniche avanzate di interrogazione o analisi e con mezzi ragionevolmente utilizzabili. L'EDPB ha, quindi, ritenuto che anche tali modelli di IA non possano essere considerati anonimi, in quanto sono soggetti a probabili rischi di estrazione e inferenza.

Sulla base di tali considerazioni, **l'EDPB ha ritenuto di dover indicare nel Parere due condizioni che devono sussistere al fine di considerare anonimo un modello di** IA. In particolare, tenuto conto di tutti i mezzi ragionevolmente utilizzabili: **(i)** le informazioni relative alle persone fisiche identificate o identificabili, i cui dati personali sono stati utilizzati per l'addestramento del modello, non devono poter essere estratte, nemmeno in modo probabilistico, dai parametri del modello stesso; **(ii)** qualsiasi output generato dal modello non deve, in alcun modo, riferirsi direttamente o indirettamente agli interessati i cui dati sono stati utilizzati nella fase di addestramento. In sintesi, sia la probabilità di estrazione diretta dei dati personali, sia la probabilità di ottenere tali dati da interrogazioni del modello devono essere "insignificanti", al fine di considerare il modello di IA anonimo. Su detti punti si veda anche il Parere 5/2014 WP29 sulle tecniche di anonimizzazione, dove si afferma che se non è possibile collegare o dedurre informazioni dall'insieme dei dati presumibilmente anonimi, i dati possono essere considerati anonimi

La valutazione circa l'anonimità dei modelli di IA addestrati con dati personali richiede un'analisi approfondita che deve essere condotta caso per caso, tenendo in considerazione le specificità del modello e del contesto di utilizzo, nonché applicando il criterio dei "**mezzi ragionevolmente utilizzabili**" da qualsiasi soggetto terzo per estrarre i dati personali e dunque identificare i soggetti ai quali essi si riferiscono.

Il criterio dei "mezzi ragionevolmente utilizzabili" menzionato è ripreso dal considerando 26 del GDPR, il quale lo utilizza per determinare, sulla base di elementi oggettivi, quando un dato può ritenersi anonimizzato oppure pseudonimizzato, ossia quando è ancora possibile identificare una persona fisica.

L'EDPB, richiamando il predetto criterio, non sembra aver tenuto conto della distinzione tra il concetto di pseudonimizzazione e di anonimizzazione che potrebbe avere diverse implicazioni rilevanti nella valutazione dell'anonimità o meno di un modello di IA. I concetti di pseudonimizzazione e anonimizzazione sono strettamente correlati in quanto si tratta di misure di sicurezza finalizzate al raggiungimento del medesimo scopo, ossia quello di ridurre - la prima - o neutralizzare - la seconda - il rischio per l'interessato che i suoi dati personali vengano rivelati durante un qualsiasi processo di analisi dei dati.

La differenza risiede nella possibilità di re-identificazione o meno dell'interessato tramite l'utilizzo di informazioni aggiuntive, da valutare sulla base del predetto criterio di cui al considerando 26 del GDPR, nonché nella differente natura del dato protetto dalla misura di sicurezza.

Il criterio dei "mezzi ragionevolmente utilizzabili" implica, infatti, che **occorre verificare la natura giuridica del dato alla luce degli strumenti che il soggetto detiene** o meno per poter re-indentificare la persona fisica da un punto di vista concreto. Le misure di sicurezza adottate per pseudonimizzare il dato possono rendere il dato anonimo per un altro soggetto, a cui il titolare trasmetta i dati, quando il ricevente non sia in possesso degli

strumenti idonei a rendere quel dato identificabile, tenendo conto dell'insieme dei fattori obiettivi, tra cui i costi e il tempo necessario per l'identificazione nonché delle tecnologie impiegate per separare le informazioni dai dati identificativi.

Anche per un Modello di IA ci si potrebbe chiedere, quindi, se esso possa essere considerato anonimo solo per il soggetto che lo utilizza, seppur non lo sia per un altro soggetto dotato delle informazioni aggiuntive. Domanda che nel contesto dell'IA si potrebbe declinare se un Modello debba essere considerato anonimo per il solo soggetto deployer/utente finale, oppure anche (o solo) per il soggetto (developer) dotato di quelle informazioni aggiuntive grazie alle quali re-identificare l'origine "personale" del dato di addestramento (o di utilizzo); non è chiaro, in sostanza, a che livello della catena del ciclo di vita del Modello di IA debba essere verificata l'anonimità del modello.

Nel Parere **l'unico riferimento alla pseudonimizzazione è presente nell'indicazione delle misure specifiche che i titolari del trattamento possono implementare nella fase di sviluppo di un modello di IA** al fine "di **ridurre significativamente la probabilità di identificazione** o re-identificazione dei dati personali" e rendere il Modello di IA anonimo. Tra queste misure, infatti, sono inclusi l'uso di dati anonimi e l'uso di dati pseudonimizzati, nonché l'impiego di tecniche di minimizzazione volte a ridurre la raccolta di dati personali a quelli strettamente necessari per il funzionamento del modello.

In conclusione, per effettuare la valutazione circa l'anonimità del modello, l'EDPB richiama il criterio dei "mezzi ragionevolmente utilizzabili", idoneo, ai sensi del GDPR, a distinguere il dato anonimizzato da quello pseudonimizzato.

Vale ricordare che l'EDPB con le "Linee Guida 1/2025 sulla pseudonimizzazione", ha precisato che i dati pseudonimizzati **rimangono tali pure "se i dati pseudonimizzati e le informazioni aggiuntive non sono nelle mani della stessa persona"**. Genera, quindi, qualche dubbio il fatto che ai fini della tematica in esame l'EDPB non abbia preso in considerazione né il concetto di pseudonimizzazione alla luce del GDPR, né i concetti dallo stesso poi esposti nelle Linee Guida di recente pubblicazione. Ritenere che il dato pseudonimizzato sia tale anche qualora i dati pseudonimizzati e le informazioni aggiuntive non sono nelle mani della stessa persona complica ulteriormente la verifica dell'anonimità del Modello IA. Una tale conclusione, infatti, sembra non tenere in considerazione né alcuni aspetti della distinzione tra dato anonimo e dato pseudonimizzato, né la complessità, anche tecnica, dei modelli di IA.

L'EDPB procede poi nella sua analisi precisando che occorre valutare, caso per caso, anche tutte le altre misure di sicurezza implementate dal titolare per garantire e dimostrare che un Modello di IA sia anonimo e che attengono alla: (i) **progettazione del modello**, (ii) **analisi del modello**, (iii) **test del modello e resistenza agli attacchi** e (iv) **documentazione**.

In particolare, con riferimento alla fase di progettazione del Modello, nel Parere è presente un elenco esemplificativo e non esaustivo di elementi la cui adozione, secondo l'EDPB, potrebbe determinare l'anonimato del Modello, suddivisi in **quattro aree chiave di valutazione**: (i) selezione delle fonti utilizzate per addestrarlo; (ii) preparazione e minimizzazione dei dati; (iii) scelte metodologiche relative allo sviluppo del Modello e (iv) misure relative agli output del Modello.

Inoltre, l'adozione di misure tecniche durante l'addestramento, come la regolarizzazione, contribuisce a migliorare la capacità di generalizzazione del Modello e a prevenire il fenomeno del sovra-adattamento ("**overfitting**"). L'integrazione di tecniche di privacy differenziale permette, invece, di anonimizzare i dati già nella fase di acquisizione degli stessi mediante l'utilizzo di appositi algoritmi e l'inserimento di elementi di disturbo. Con il termine privacy differenziale si fa riferimento a quell'insieme di tecniche volte a consentire di effettuare analisi statistiche sui dati senza rivelare le informazioni personali o quelle identificabili in merito agli individui cui i dati personali si riferiscono (sul punto si vedano le "Guidelines for Evaluating Differential Privacy Guarantees, dicembre 2023, emesse dal NIST)

Infine, con riferimento alla fase di documentazione, l'EDPB, in osservanza al **principio di accountability**, suggerisce ai titolari del trattamento della fase di addestramento di fornire ai titolari del trattamento che utilizzano il modello la documentazione dettagliata relativa alle misure di sicurezza implementate nel Modello al fine di ridurre la probabilità di identificazione e i possibili rischi residui.

Secondo quesito: l'adeguatezza del legittimo interesse come base giuridica del trattamento dei dati personali utilizzati nelle fasi di sviluppo e diffusione del Modello di IA

Il tema è **verificare la sussistenza del legittimo interesse** (art. 6 par.1 lett. f del GDPR) quale base giuridica del trattamento dei dati personali utilizzati nelle fasi di sviluppo e diffusione dei modelli di IA.

L'EDPB, facendo riferimento alle sue Linee Guida 1/2024, ribadisce che anche nell'ambito dei modelli di IA l'applicazione della base giuridica del legittimo interesse richiede **un'analisi rigorosa che si articola in tre fasi**:

i) individuare l'interesse del titolare del trattamento o del terzo che, nel contesto dell'IA, potrebbe includere obiettivi di innovazione tecnologica, miglioramento dell'efficienza operativa o prevenzione di frodi;

ii) determinare se il trattamento è necessario per perseguire l'interesse identificato e verificare che non esistono alternative meno invasive ma ugualmente idonee al raggiungimento dello stesso obiettivo (c.d. "test di necessità"). Ad esempio, il trattamento dei dati non è necessario quando è possibile conseguire la finalità anche attraverso un modello di IA che non comporta il trattamento degli stessi;

iii) individuare e descrivere, da un lato, l'interesse del titolare del trattamento e, dall'altro, gli interessi, i diritti e le libertà fondamentali degli interessati (c.d. "test di bilanciamento"). Questa valutazione deve tener conto della natura dei dati trattati, del contesto del trattamento e delle ulteriori conseguenze che il trattamento può avere e la probabilità che si verifichino e della ragionevole aspettativa degli interessati.

Nell'ambito dei Modelli di IA, riveste un ruolo dirimente la valutazione circa la **ragionevole aspettativa dell'interessato che il trattamento basato sul legittimo interesse e per una diversa finalità possa aver luogo**, così come previsto dal considerando 47 del GDPR.

Nell'ambito dei Modelli di IA, infatti, le ragionevoli aspettative degli interessati possono variare caso per caso e sono strettamente legate alle varietà di utilizzo del modello e alla consapevolezza dell'interessato in merito alle capacità specifiche dello stesso.

Nel caso in cui dal bilanciamento di interessi dovesse emergere la prevalenza dei diritti e delle libertà degli interessati rispetto al legittimo interesse del titolare del trattamento, l'EDPB ricorda la possibilità per il titolare di implementare misure di mitigazione dei rischi al fine di consentire allo stesso di ricorrere alla base giuridica del legittimo interesse per il trattamento dei dati utilizzati dal modello di IA.

Il Parere fornisce un **elenco non esaustivo di misure di mitigazione**, che possono essere adottate quando gli interessi, i diritti e le libertà degli interessati prevalgano sui legittimi interessi perseguiti dal titolare del trattamento o da terzi. Tali misure 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 sviluppo e diffusione dei Modelli di IA.

Durante la fase di sviluppo, vengono raccomandate tecniche come il **data masking**, la pseudonimizzazione e l'impiego di dati fittizi quando i dati personali non sono strettamente necessari per il funzionamento del Modello. Nella fase di utilizzo e diffusione, è suggerita l'adozione di filtri sugli output per evitare la memorizzazione, il rigurgito o la generazione accidentale di dati personali, in particolare nei Modelli generativi, e l'implementazione di tecniche come il **watermarking digitale** per ridurre il rischio di riutilizzo illecito dei dati.

L'EDPB **sottolinea**, inoltre, **l'importanza di facilitare l'esercizio dei diritti degli interessati**. A tal fine, vengono proposte misure come l'introduzione di un intervallo di tempo ragionevole tra la raccolta e l'utilizzo dei dati, l'adozione di un'opzione di opt-out per consentire l'esclusione di dati da un dataset, e l'ampliamento del diritto alla cancellazione anche nei casi in cui non ricorrono i requisiti specifici previsti dall' art. 17, par. 1 del GDPR.

In relazione alla raccolta dei dati, la stessa può avvenire da fonti interne all'organizzazione del titolare (data base aziendali) o tramite consultazione di fonti esterne e particolarmente diffusa è la tecnica del web scraping che consiste nell'estrazione di dati da fonti online pubblicamente accessibili, come siti web, social media e forum. Proprio con riferimento al web scraping, il Parere suggerisce di escludere dalla raccolta mediante tale tecnica i dati personali che potrebbero comportare rischi per particolari persone o gruppi di persone e di rispettare i meccanismi di esclusione dallo scraping automatizzato previsti da file come robots.txt o ai.txt, e di introdurre una lista di opt-out che permetta agli interessati di opporsi alla raccolta dei propri dati.

Terzo quesito: conseguenze del trattamento illecito di dati personali nella fase di sviluppo di un modello di IA sul successivo trattamento nella fase di utilizzo e diffusione dello stesso

L'EDPB ha esaminato, infine, le **conseguenze di un trattamento illecito effettuato durante la fase di sviluppo di un modello di IA** e il suo impatto sui trattamenti operati nella successiva fase di utilizzo e diffusione. Il Parere ha suddiviso l'analisi in tre scenari differenziati in base al fatto che il Modello di IA conservi i dati personali e che il trattamento venga effettuato dallo stesso titolare del trattamento o da un altro titolare:

Primo scenario: il titolare del trattamento tratta illecitamente i dati personali nel contesto dello sviluppo del Modello di IA. I dati personali vengono conservati dal Modello di IA e successivamente trattati dallo stesso titolare nella fase di utilizzo e diffusione. In questo caso l'illegittimità del trattamento iniziale influisce sulla legittimità di quello successivo.

Secondo scenario: il titolare del trattamento tratta illecitamente i dati personali nel contesto dello sviluppo del Modello di IA. I dati personali vengono conservati dal Modello di IA e successivamente trattati da un diverso titolare nella fase di utilizzo e diffusione. In base al principio secondo cui ciascun titolare è tenuto a garantire la liceità del trattamento dei dati, il titolare che utilizza il Modello in fase di utilizzo e diffusione è tenuto a condurre una valutazione per assicurarsi che il modello non sia stato sviluppato sulla base di un trattamento illecito di dati. È il caso relativo ad OpenAI, sanzionata dal Garante (Provvedimento del 2 novembre 2024 [10085455]) per aver addestrato in maniera illegittima il modello su cui si basa ChatGPT tramite dati personali di utenti non informati.

Terzo scenario: il titolare del trattamento tratta illecitamente i dati personali nel contesto dello sviluppo del Modello di IA, ma successivamente li anonimizza prima che egli stesso o un altro titolare avvii un trattamento durante la fase di utilizzo e diffusione del modello. In questo caso, secondo l'EDPB, l'illiceità del trattamento iniziale non dovrebbe influire sull'utilizzo successivo del modello di IA, in quanto non comportando il trattamento di dati personali non troverebbe applicazione il GDPR.

Per tutti e tre gli scenari, l'EDPB sottolinea che **è competenza delle Autorità garanti valutare la liceità dei trattamenti ed eventualmente imporre misure correttive**, tenendo conto delle circostanze concrete caso per caso.

Nei tre scenari analizzati l'EDPB sembra aver attribuito poca rilevanza ad un momento che al contrario appare nodale nell'ambito dei trattamenti di dati personali e cioè il momento della raccolta del dato personale.

L'EDPB, infatti, ha ritenuto **che l'illegittimità di un trattamento di dati personali effettuato da un titolare nell'ambito dello sviluppo e training di un modello di IA possa essere considerata irrilevante in un successivo momento di utilizzo del Modello di IA**, qualora i dati siano stati anonimizzati, e quindi resi non più personali. **Da tale assunto si potrebbe desumere che** (i) il trattamento operato nella fase di utilizzo non si possa ritenere collegato o derivante dal trattamento effettuato nella fase di sviluppo e (ii) l'adozione della misura di sicurezza della anonimizzazione dei dati personali oggetto del trattamento illecito possa far venir meno l'illegittimità del trattamento iniziale. In tal caso, il titolare del trattamento della fase di utilizzo del Modello non dovrebbe quindi porsi il problema della legittimità dei dati utilizzati dal titolare sviluppatore.

Tale impostazione appare, tuttavia, **in contrasto con il principio generale per cui il titolare del trattamento**, anche ai fini dell'accountability e in base al risk based approach, **debba verificare la provenienza dei dati e accertare la liceità del loro trattamento prima di iniziare il nuovo trattamento**. In conclusione, **l'anonimizzazione dei dati non dovrebbe far venire meno la illiceità del trattamento iniziale**, perché l'esecuzione illecita di tale trattamento è comunque avvenuta e non è possibile escludere la ragionevole probabilità di una regressione e re-identificazione delle persone fisiche cui i dati di training appartengono.

L’EDPB, peraltro, si è espresso solo con riferimento alla anonimità del Modello, affermando che lo stesso possa essere considerato anonimo sulla base delle considerazioni già effettuate nella risposta al primo quesito e all’applicazione di adeguate misure di sicurezza quali la anonimizzazione dei dati. Tuttavia, il profilo problematico di tale scenario non attiene all’anonimità del Modello, quanto alla illegittimità del trattamento dei dati personali nella fase di sviluppo, a prescindere dal fatto che poi sia stata adottata la misura di sicurezza dell’anonimizzazione, che di per sé potrebbe essere adeguata e aver efficacemente reso il modello anonimo.

In conclusione, **sarebbe forse stato opportuno approfondire ulteriormente la questione relativa alla presunta legittimità del trattamento dei dati anonimizzati utilizzati dal titolare del trattamento nella fase di utilizzo del modello di IA** e, eventualmente, prevedere l’adozione di eventuali ulteriori misure di mitigazione volte a ridurre la ragionevole probabilità di regressione e re-identificazione delle persone fisiche cui i dati si riferiscono.

Conclusioni

Il Parere 28/2024 dell’EDPB offre un contributo significativo alla comprensione delle implicazioni della protezione dei dati personali nel contesto dello sviluppo e della diffusione dei Modelli di IA. Attraverso un’analisi dettagliata delle questioni più rilevanti – dall’anonimità dei Modelli alla validità della base giuridica del legittimo interesse, fino alle conseguenze del trattamento illecito nella fase di sviluppo – l’EDPB fornisce indicazioni utili per gli sviluppatori e i distributori di IA. Tuttavia, sebbene il Parere delinei alcune linee guida operative, lascia aperte diverse aree di approfondimento e non risolve completamente le incertezze che emergono nell’applicazione del GDPR alla materia delle tecnologie di intelligenza artificiale.