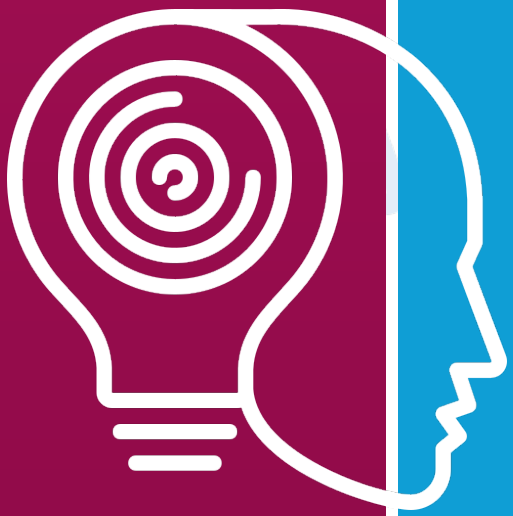

Intelligenza artificiale + garanzie + diritti

HIMMEL ADVISORS
www.himmeladvisors.it

I primi orientamenti del Garante Europeo della Protezione dei Dati (GEPD) per garantire la conformità alla protezione dei dati quando si utilizzano sistemi di *"Intelligenza artificiale"* generativa

V.1 - IT

EDPS
EUROPEAN DATA PROTECTION SUPERVISOR

Generative AI and the EUDPR
First EDPS Orientations for ensuring data protection compliance when using Generative AI systems

By European Data Protection Supervisor
The EU's independent data protection authority

Elaborazione, produzione e traduzione:

Himmel Advisors S.r.l
Il presente documento non è ufficiale

HIMMEL ADVISORS
www.himmeladvisors.it

Il testo originale è stato redatto dall'European Data Protection Supervisor (EDPS). La traduzione in italiano è stata curata da Francisco Garcia-Garrido, con l'intento di renderne comprensibile il contenuto a tutti i lettori. Le parti in blu sono state aggiunte dal traduttore per facilitare la comprensione del testo e contengono spiegazioni ulteriori rispetto ai contenuti degli orientamenti. La presente pubblicazione è distribuita a titolo gratuito e non persegue alcuno scopo economico. Si precisa che, in ogni caso, l'autore della traduzione declina ogni responsabilità in merito alla correttezza, completezza e aggiornamento dei contenuti tradotti. Per eventuali dubbi o necessità di chiarimenti, si fa riferimento unicamente alla versione originale redatta in lingua inglese dall'EDPS. Suggerimenti e proposte di miglioramento sono graditi e possono essere inviati alla nostra Amministrazione tramite e-mail: info [at] himmeladvisors.it

Il testo originale è stato redatto dall'European Data Protection Supervisor (EDPS). La traduzione in italiano è stata curata da Francisco -Garrido, con l'intento di renderne comprensibile il contenuto a tutti i lettori.

Premessa

Gli orientamenti del GEPD sull'intelligenza artificiale generativa (IA generativa) intendono fornire consigli e istruzioni pratiche alle istituzioni, agli organi e agli organismi dell'UE sul trattamento dei dati personali nell'utilizzo di sistemi di IA generativa, per agevolare l'adempimento dei loro obblighi in materia di protezione dei dati, come stabilito, in particolare, dal Regolamento (UE) 2018/1725 sulla tutela delle persone fisiche in relazione al trattamento dei dati personali da parte delle istituzioni, degli organi e degli organismi dell'Unione e sulla libera circolazione di tali dati, e che abroga il regolamento (CE) n. 45/2001 e la decisione n. 1247/2002/CE nonché dal Regolamento (UE) 2016/679.

Questi indirizzamenti sono stati redatti per coprire il maggior numero possibile di scenari e applicazioni e non prescrivono misure tecniche specifiche. Pongono invece l'accento sui principi generali della protezione dei dati che dovrebbero aiutare le istituzioni dell'UE (di seguito, per brevità, anche IUE o Istituzioni) a conformarsi ai requisiti di protezione dei dati previsti dal Regolamento (UE) 2018/1725.

Il presente documento è un primo passo verso orientamenti più dettagliati che terranno conto dell'evoluzione dei sistemi e delle tecnologie di IA generativa, del loro utilizzo da parte delle istituzioni dell'UE e dei risultati delle attività di monitoraggio e sorveglianza del GEPD. Il GEPD emette questi orientamenti nel suo ruolo di autorità di controllo della protezione dei dati e non nel suo nuovo ruolo di autorità di controllo dell'IA ai sensi della legge sull'IA. Questi orientamenti non pregiudicano la normativa in vigore sull'intelligenza artificiale.

Riferimenti normativi

Regolamento (UE) 2018/1725 sulla tutela delle persone fisiche in relazione al trattamento dei dati personali da parte delle istituzioni, degli organi e degli organismi dell'Unione e sulla libera circolazione di tali dati, e che abroga il regolamento (CE) n. 45/2001 e la decisione n. 1247/2002/CE	Link
Regolamento (UE) 2016/679 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati)	Link

Glossario

Introduzione e ambito di applicazione

1. Che cos'è l'IA generativa?
2. Le Istituzioni possono utilizzare l'IA generativa?
3. Come sapere se l'uso di un sistema di IA generativa comporta un trattamento di dati personali?
4. Qual è il ruolo del DPO nel processo di sviluppo o implementazione di sistemi di IA generativa?
5. L'Istituzione che vuole sviluppare o implementare sistemi di IA generativa, deve effettuare una DPIA?
6. In fase di progettazione, sviluppo e validazione di sistemi di IA generativa. Quando è lecito il trattamento dei dati personali?
7. Minimizzazione e sistemi di IA generativa?
8. I sistemi di intelligenza artificiale generativa rispettano il principio dell'esattezza dei dati?
9. Onere informativo nei sistemi di IA generativa
10. AI generativa e processi decisionali automatizzati?
11. Come si può garantire un trattamento equo ed evitare pregiudizi quando si utilizzano sistemi di AI generativa?
12. Esercizio dei diritti nell'ambito dei sistemi di AI generativa
13. Sistemi di AI generativa e sicurezza dei dati

Per saperne di più

Introduzione e ambito di applicazione

1. Gli orientamenti del GEPD intendono fornire alcuni indirizzamenti alle istituzioni, agli organi e agli organismi dell'UE sul trattamento dei dati personali nell'uso dei sistemi di IA generativa, per garantire che rispettino i loro obblighi in materia di protezione dei dati, in particolare quelli previsti dal Regolamento (UE) 2018/1725. Anche tale Regolamento non menziona esplicitamente il concetto di Intelligenza Artificiale (IA), la corretta interpretazione e applicazione dei principi di protezione dei dati è essenziale per ottenere un uso vantaggioso di questi sistemi che non danneggi i diritti e le libertà fondamentali delle persone.

Va tenuto in considerazione che “gli orientamenti del GEPD” sono considerati strumenti di "soft law", ovvero strumenti di "diritto mite" o "diritto flessibile", in quanto costituiscono un insieme di norme, linee guida, principi, raccomandazioni o dichiarazioni che, pur non avendo carattere giuridicamente vincolante come il diritto "hard" (diritto rigido o diritto forte), esercitano comunque un'influenza significativa sul comportamento degli attori nazionali e internazionali nonché soggetti privati.

2. Il GEPD emette questi orientamenti nel suo ruolo di autorità di controllo della protezione dei dati e non nel suo nuovo ruolo di autorità di controllo dell'IA ai sensi della normativa sull'IA (c.d. AI Act).

La proposta di AI Act, recentemente approvata dal Consiglio dell'Unione Europea, rappresenta un regolamento innovativo volto a disciplinare l'uso e lo sviluppo dell'intelligenza artificiale nell'UE. Questo regolamento mira a garantire un utilizzo etico e sicuro delle tecnologie di intelligenza artificiale, promuovendo al contempo l'innovazione. L'AI Act introduce un approccio basato sul rischio, classificando i sistemi di intelligenza artificiale in base al loro potenziale impatto sui diritti fondamentali e sulla sicurezza, e stabilendo requisiti specifici per ciascuna categoria di rischio. L'obiettivo principale di questo regolamento è creare un quadro normativo armonizzato che tuteli i cittadini europei, assicuri la trasparenza e la responsabilità nello sviluppo e nell'uso dell'IA e promuova la fiducia nelle tecnologie emergenti. Tuttavia, è importante sottolineare che si tratta ancora di una proposta e dovrà passare attraverso ulteriori fasi di approvazione prima di diventare legge.

3. Questi orientamenti non mirano a coprire in modo dettagliato tutte le questioni rilevanti relative al trattamento dei dati personali nell'uso dei sistemi di IA generativa che sono oggetto di analisi da parte delle autorità di protezione dei dati. Alcune di queste questioni sono ancora aperte ed è probabile che ne sorgano altre man mano che l'uso di questi sistemi aumenta e la tecnologia si evolve in modo da consentire una migliore comprensione del funzionamento dell'IA generativa.

4. Poiché la tecnologia dell'intelligenza artificiale si evolve rapidamente, gli strumenti e i mezzi specifici utilizzati per fornire questi tipi di servizi sono diversi e possono cambiare molto rapidamente. Pertanto, questi orientamenti sono stati redatti per coprire il maggior numero possibile di scenari e applicazioni.

5. Questi indirizzamenti sono strutturati come segue:

    - domande chiave, seguite da risposte iniziali con alcune conclusioni preliminari;
    - ulteriori chiarimenti o esempi.

6. Questi orientamenti iniziali rappresentano un passo preliminare verso lo sviluppo di uno strumento di guida più completo per le Istituzioni. Nel corso del tempo, questi orientamenti saranno aggiornati, perfezionati e ampliati per affrontare ulteriori elementi necessari a sostenere le Istituzioni nello

sviluppo e nell'attuazione di questi sistemi. Tale aggiornamento dovrebbe avvenire entro dodici mesi dalla pubblicazione del presente documento.

Il documento tradotto (V.1 IT) sarà soggetto a modifiche in linea con tali aggiornamenti e verrà fornito nella sua versione aggiornata non appena sarà disponibile.

Riferimenti normativi

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts

Link

Che cos'è l'IA generativa?

L'IA generativa è un tipo di intelligenza artificiale che utilizza modelli di apprendimento automatico per creare una varietà di output, come testo, immagini o audio, e può essere utilizzata per diversi compiti e applicazioni. Concretamente, si basa sull'uso dei cosiddetti modelli di base che servono come modelli di riferimento per altri sistemi di IA generativa che verranno "perfezionati" a partire da essi.

Un modello di base funge da architettura centrale o da base su cui vengono costruiti altri modelli più specializzati. Questi modelli vengono allenati sulla base di serie di dati ampie e diversificate, comprese quelle contenenti informazioni disponibili pubblicamente. Possono rappresentare strutture complesse come immagini, audio, video o linguaggio e possono essere messi a punto per compiti o applicazioni specifiche.

I modelli linguistici di grandi dimensioni (anche LLM) sono un tipo specifico di modello di fondazione addestrato su enormi quantità di dati testuali (da milioni a miliardi di parole), in grado di generare risposte in linguaggio naturale a un'ampia gamma di input basati su modelli e relazioni tra parole e frasi. La grande quantità di testo utilizzata per addestrare il modello può essere tratta da Internet, da libri e da altre fonti disponibili. Alcune applicazioni già in uso sono sistemi di generazione di codici, assistenti virtuali, strumenti di creazione di contenuti, motori di traduzione linguistica, riconoscimento vocale automatizzato, sistemi di diagnosi medica, strumenti di ricerca scientifica, ecc.

"I modelli linguistici sono sistemi di intelligenza artificiale (IA) progettati per apprendere la grammatica, la sintassi e la semantica di una o più lingue al fine di generare linguaggio coerente e pertinente al contesto. I modelli linguistici sono stati sviluppati utilizzando reti neurali fin dagli anni '90, ma i risultati sono stati modesti. L'evoluzione verso modelli linguistici di grandi dimensioni (LLM) è stata resa possibile da sviluppi tecnici che hanno migliorato le prestazioni e l'efficienza dei sistemi di intelligenza artificiale. Questi sviluppi includono l'avvento di modelli pre-addestrati su larga scala, lo sviluppo di trasformatori (che apprendono il contesto e il significato tracciando le relazioni nei dati sequenziali) e meccanismi di auto-attenzione (che consentono ai modelli di ponderare l'importanza di diversi elementi in una sequenza in ingresso e di regolare dinamicamente la loro influenza sull'output)."

Autore: Xabier Lareo (link qui)

La relazione tra questi concetti è gerarchica. L'IA generativa è l'ampia categoria che comprende i modelli progettati per creare contenuti. Un modello di base, come un modello linguistico di grandi dimensioni, funge da architettura di base su cui vengono costruiti modelli più specializzati. I modelli specializzati, costruiti sul modello di base, si rivolgono a compiti o applicazioni specifiche, utilizzando le conoscenze e le capacità dell'architettura di base.

Il ciclo di vita di un modello di IA generativa comprende diverse fasi, a partire dalla definizione del caso d'uso e dell'ambito del modello. In alcuni casi, potrebbe essere possibile identificare un modello di base adeguato da cui partire, mentre in altri casi potrebbe essere necessario costruire un nuovo modello da zero. La fase successiva prevede l'addestramento del modello con un insieme di dati rilevanti per lo scopo del futuro sistema, compresa la messa a punto del sistema con insiemi di dati specifici e personalizzati necessari per soddisfare il caso d'uso del modello. Per completare l'addestramento, vengono utilizzate tecniche specifiche che richiedono l'intervento umano per garantire informazioni più accurate e un comportamento controllato. La fase successiva mira a valutare il modello e a stabilire delle metriche per valutare regolarmente fattori quali l'esattezza e l'allineamento del modello con il caso d'uso. Infine, i modelli vengono distribuiti e implementati, includendo il monitoraggio continuo e la valutazione periodica utilizzando le metriche stabilite nelle fasi precedenti.

I casi più rilevanti nell'ambito dell'IA generativa sono le applicazioni generali orientate al consumatore (es. ChatGPT ed i sistemi simili che si possono già trovare in diverse versioni e dimensioni[1], comprese quelle che possono essere eseguite direttamente dal cellulare). Esistono anche applicazioni aziendali pensate per essere utilizzate su settori specifici, modelli pre-addestrati, applicazioni basate su modelli allenati che vengono messi a punto per un uso specifico in un'area di attività e, infine, modelli in cui l'intero sviluppo, compreso il processo di addestramento, è realizzato dall'ente responsabile.

ChatGPT è una versione specializzata del modello GPT (Generative Pre-trained Transformer) che è stata addestrata specificamente per interagire e rispondere come un assistente conversazionale. Utilizza un'architettura di trasformatori per comprendere e generare testo in risposta a domande e messaggi degli utenti.

L'IA generativa, come altre nuove tecnologie, offre soluzioni in diversi campi destinate a sostenere e potenziare le capacità umane. Tuttavia, crea anche sfide con un potenziale impatto sui diritti e le libertà fondamentali che rischiano di passare inosservate, trascurate, non adeguatamente considerate e valutate.

L'addestramento di un Large Language Model (LLM) (e in generale di qualsiasi modello di apprendimento automatico) è un processo iterativo, complesso e ad alta intensità di risorse che prevede diverse fasi e tecniche finalizzate alla creazione di un modello in grado di generare testo simile a quello umano in reazione ai comandi (o alle richieste) forniti dagli utenti. Il processo inizia con l'addestramento del modello su enormi insiemi di dati, la maggior parte dei quali normalmente non etichettati e ottenuti da fonti pubbliche utilizzando tecnologie di webscraping (le autorità per la protezione dei dati hanno già espresso preoccupazione e delineato i principali rischi per la privacy e la protezione dei dati associati all'uso di dati personali pubblicamente accessibili). Successivamente, gli LLM vengono - non in tutti i casi - perfezionati utilizzando l'apprendimento supervisionato o tecniche che prevedono l'intervento umano (come il Reinforcement Learning with Human Feedback (RLHF) o l'Adversarial Testing via Domain experts) per aiutare il sistema a riconoscere ed elaborare meglio le informazioni e il contesto, nonché a determinare le risposte preferite, a limitare o meno l'output in risposta a domande sensibili e ad allinearlo con i valori degli sviluppatori (ad esempio, evitare di produrre output dannosi o tossici). Una volta in produzione, alcuni sistemi utilizzano i dati di input ottenuti attraverso l'interazione con gli utenti come un nuovo set di dati di addestramento per perfezionare il modello.

[1] The size of a Large Language Model is usually measured as the number of parameters (tokens it contains. The size of a LLM model is important since some capabilities only appear when the model grows beyond certain limits.

Le Istituzioni possono utilizzare l'IA generativa?

In linea di principio non vi è alcun ostacolo allo sviluppo, alla diffusione e all'utilizzo di sistemi di IA generativa nella fornitura di servizi pubblici, a condizione che le norme delle Istituzioni interessate lo consentano e che siano soddisfatti tutti i requisiti giuridici applicabili, in particolare considerando la speciale responsabilità del settore pubblico di garantire il pieno rispetto dei diritti e delle libertà fondamentali delle persone quando si fa uso di nuove tecnologie.

In sostanza, l'uso dei sistemi di intelligenza artificiale per i servizi pubblici dipende dalle normative delle Istituzioni coinvolte e dal rispetto dei requisiti legali, specialmente per proteggere i diritti degli interessati e le libertà fondamentali quando si adottano nuove tecnologie. È importante sottolineare la necessità di condurre una valutazione di impatto ai sensi dell'art. 35 del Regolamento (UE) 2016/679 e di fornire un'informativa ai soggetti interessati ai sensi dell'art. 13 del Regolamento (UE) 2016/679. Ciò garantisce il rispetto delle normative sulla protezione dei dati e assicura che le persone coinvolte siano pienamente informate sull'utilizzo delle loro informazioni personali.

In ogni caso, se l'uso di sistemi di intelligenza artificiale generativa comporta il trattamento di dati personali, il Regolamento (UE) 2018/1725 si applica integralmente. Si tratta di un atto giuridico tecnologicamente neutro per cui si applica a tutte le attività di trattamento dei dati personali, indipendentemente dalle tecnologie utilizzate e senza pregiudicare altre normative applicabili, in particolare l'AI Act. Il principio di *accountability* richiede che le responsabilità siano chiaramente identificate e rispettate tra i vari attori coinvolti nella catena di fornitura dei modelli di IA generativa.

In breve, se le Istituzioni intendono utilizzare sistemi di intelligenza artificiale che trattano dati personali, sono tenute ad applicare le disposizioni del Regolamento (UE) 2018/1725. Questo regolamento si applica a tutte le attività di trattamento dei dati personali, indipendentemente dalla tecnologia utilizzata e senza influenzare altre normative, come ad esempio l'AI Act. Il principio di accountability richiede che i ruoli e le responsabilità siano chiaramente definiti e rispettati da tutti coloro coinvolti nella fornitura dei modelli di intelligenza artificiale generativa (es. incaricati interni, responsabili esterni)

Le Istituzioni possono sviluppare e implementare le proprie soluzioni di IA generativa o, in alternativa, possono utilizzare per il proprio uso soluzioni disponibili sul mercato. In entrambi i casi, le Istituzioni possono ricorrere a fornitori per ottenere tutti o alcuni degli elementi che fanno parte del sistema di IA generativa. In questo contesto, le Istituzioni devono determinare chiaramente i ruoli specifici - responsabile del trattamento, incaricato del trattamento, contitolare del trattamento - per le specifiche operazioni di trattamento effettuate e le relative implicazioni in termini di obblighi e responsabilità ai sensi del Regolamento (UE) 2018/1725.

Con il rapido avanzamento delle tecnologie di IA, le Istituzioni europee devono considerare quando e come utilizzare l'IA generativa in modo responsabile e vantaggioso per l'interesse pubblico. Tutte le fasi del ciclo di vita di una soluzione di IA generativa dovrebbero operare in conformità agli ordinamenti giuridici e alle normative applicabili, compreso il Regolamento (UE) 2018/1725, quando il sistema comporta un potenziale trattamento di dati personali.

I termini “intelligenza artificiale affidabile” o “responsabile” si riferiscono alla necessità di garantire che i sistemi di IA siano sviluppati in modo etico e legittimo. Ciò implica la considerazione delle conseguenze indesiderate relative all'uso della tecnologia di IA e la necessità di seguire un approccio basato sul rischio che copra tutte le fasi del ciclo di vita del sistema. Implica inoltre la trasparenza sull'uso dei dati di addestramento e sulle loro fonti, su come vengono progettati e implementati gli algoritmi, sul tipo di pregiudizi che potrebbero essere presenti nel sistema e su come vengono affrontati i possibili impatti sui diritti e le libertà fondamentali dell'individuo. In questo contesto, i sistemi di IA generativa devono essere trasparenti, spiegabili, coerenti, verificabili e accessibili, in modo da garantire un trattamento equo dei dati personali.

Come sapere se l'uso di un sistema di IA comporta il trattamento di dati personali?

Il trattamento dei dati personali all'interno un sistema di IA generativa può avvenire a vari livelli e fasi del suo ciclo di vita, senza essere necessariamente evidente. Ciò include la creazione dei dataset di addestramento, la fase di addestramento stesso, l'inferenza di informazioni nuove o aggiuntive una volta che il modello è stato creato e utilizzato, o semplicemente attraverso gli input e gli output del sistema una volta in funzione.

In sostanza, i dati personali possono essere utilizzati in diverse fasi del ciclo di vita di un sistema di intelligenza artificiale, come durante la creazione dei dati, l'addestramento del sistema o l'uso dei risultati ottenuti.

Quando uno sviluppatore o il fornitore di un sistema di IA generativa afferma che il suo sistema non tratta dati personali (per ragioni quali il presunto uso di set di dati anonimizzati o di dati sintetici durante la progettazione, lo sviluppo e i test), è fondamentale accertarsi di quali siano i controlli specifici messi in atto per garantirlo. In sostanza, le Istituzioni potrebbero voler sapere quali sono le fasi o le procedure utilizzate dal fornitore per garantire che il modello non tratti effettivamente dati personali.

Un problema potrebbe emergere quando il creatore di un sistema di intelligenza artificiale dichiara che il sistema non manipola dati personali, magari utilizzando dati anonimizzati o aggregati. È cruciale capire quali precauzioni vengano effettivamente adottate dallo sviluppatore (es. misure di sicurezza tecniche). In tal caso, le Istituzioni sono tenute a verificare le specifiche procedure seguite dal fornitore per assicurare che il sistema non trattenga dati personali.

Il GEPD si è già messo in guardia[2] relativamente all'uso di tecniche di web scraping per la raccolta di dati personali, attraverso le quali gli interessati possono perdere il controllo dei loro dati personali quando queste vengono raccolte a loro insaputa, contro le loro aspettative e per scopi diversi da quelli della raccolta originale. Il GEPD ha inoltre sottolineato che il trattamento dei dati personali disponibili al pubblico è regolamentato dal GDPR. A questo proposito, l'uso di tecniche di web scraping per raccogliere dati da siti web e il loro utilizzo a fini di formazione potrebbe non essere conforme ai principi di protezione dei dati pertinenti, tra cui la minimizzazione dei dati e il principio di esattezza, nella misura in cui non vi è alcuna valutazione sull'affidabilità delle fonti.

Le tecniche di web scraping sono metodi utilizzati per estrarre informazioni da pagine web in automatico. Questo processo permette di recuperare dati strutturati o non strutturati da siti web per vari scopi come ricerca, monitoraggio dei prezzi, analisi di mercato e molto altro. Generalmente, il web scraping coinvolge l'automazione di richieste HTTP per scaricare il contenuto delle pagine web e l'estrazione specifica delle informazioni desiderate da esse. Il problema principale legato alle tecniche di web scraping è che talvolta possono violare le normative sulla privacy e i diritti di proprietà intellettuale dei siti web. Se non autorizzato, il web scraping può comportare la raccolta non autorizzata di dati personali o sensibili, violare i termini di servizio dei siti web e causare sovraccarichi ai server. Inoltre, può creare una concorrenza sleale se i dati sono utilizzati per fini commerciali senza consenso. Pertanto, è importante esercitare il web scraping in modo etico e legale, rispettando le normative vigenti e i termini di utilizzo dei siti web.

[2] Opinion 41/2023, of 25 September 2023, on the Proposal for a Regulation on European Union labour market statistics on businesses

Un monitoraggio regolare e l'implementazione di controlli in tutte le fasi possono aiutare a verificare che non vi sia alcun trattamento di dati personali, nei casi in cui il modello non è previsto.

EUI-X, un'istituzione immaginaria dell'UE, sta valutando l'acquisto di un prodotto per il riconoscimento e la trascrizione automatica del parlato. Dopo aver studiato le opzioni disponibili, si è concentrata sulla possibilità di utilizzare un sistema di intelligenza artificiale generativa per facilitare questa funzione. In questo caso particolare, si tratta di un sistema che offre un modello pre-addestrato per il riconoscimento e la traduzione del parlato. Poiché questo modello verrà utilizzato per la trascrizione di riunioni utilizzando file vocali registrati, è stato stabilito che l'uso di questo modello richiede il trattamento di dati personali e quindi deve garantire la conformità al Regolamento.

Qual è il ruolo del DPO nel processo di sviluppo o di implementazione di un sistema di IA generativa?

L'art. 45 del Regolamento (UE) 2018/1725 stabilisce i compiti del Responsabile della Protezione dei Dati Personali (RPDP) o Data Protection Officer (DPO). I DPO informano e consigliano le Istituzioni sugli obblighi di protezione dei dati, assistono i titolari del trattamento nel monitoraggio della conformità interna, forniscono consulenza, se richiesta, in merito alle DPIA e fungono da punto di contatto per gli interessati e l'Autorità Garante per la Protezione dei Dati Personali nazionale competente.

Il DPO, acronimo di Data Protection Officer, è una figura professionale incaricata della supervisione e del monitoraggio della conformità alle normative sulla protezione dei dati all'interno di un'organizzazione. Il ruolo del DPO include la gestione delle questioni relative alla privacy dei dati, il monitoraggio della conformità al Regolamento (UE) 2016/679 e la cooperazione con le autorità di controllo nazionale competente in materia di protezione dei dati.

Nel contesto dell'implementazione da parte delle Istituzioni di sistemi di IA generativa che trattano dati personali, è importante garantire che il DPO, nell'ambito del proprio ruolo, conoscano – sin dall'inizio in virtù del principio di *privacy by design* e *by default* – l'intero ciclo di vita del sistema di IA generativa che l'Istituzione sta valutando di acquistare, progettare o implementare e del suo funzionamento. Ciò significa ottenere informazioni su quando e come questi sistemi elaborano i dati personali, su come funzionano i meccanismi di input e output e sui processi decisionali attuati attraverso il modello. È importante, come sottolinea il regolamento[3], fornire consulenza ai titolari del trattamento quando sono tenuti a condurre una valutazione d'impatto sulla protezione dei dati ai sensi dell'art. 35 del GDPR. Le Istituzioni, quali titolari del trattamento, devono assicurarsi che tutti i processi siano adeguatamente documentati e che sia garantita la trasparenza, compreso l'aggiornamento dei registri dei trattamenti e, come migliore prassi, l'esecuzione di un inventario specifico sui sistemi e le applicazioni generativi guidati dall'intelligenza artificiale. Infine, il DPO dovrebbe essere sempre coinvolto.

Dal punto di vista organizzativo, l'implementazione di sistemi di IA generativa in conformità al regolamento dovrebbe essere un lavoro di squadra. Dovrebbe esserci un dialogo continuo tra tutte le parti interessate coinvolte nel ciclo di vita del prodotto. Pertanto, i titolari del trattamento dovrebbero collaborare con tutte le figure competenti all'interno della propria organizzazione, in particolare con il proprio DPO, l'ufficio legale, l'IT e, ove presente, il responsabile della sicurezza informatica (LISO), al fine di garantire che l'Istituzione operi secondo i parametri di un'IA generativa affidabile, di una buona governance dei dati ed in conformità alla normativa nazionale / europea applicabile. La creazione di una task force sull'IA, che includa il DPO, e la preparazione di un piano d'azione, che comprenda azioni di sensibilizzazione a tutti i livelli dell'organizzazione e la preparazione di una guida interna, possono contribuire al raggiungimento di questi obiettivi.

[3] Article 39(2) of the Regulation

I titolari del trattamento dei dati devono collaborare con diverse figure interne, incluso il DPO, per assicurare una corretta gestione dei dati nell'uso dell'intelligenza artificiale, rispettando le normative nazionali ed europee. La creazione di una squadra dedicata all'intelligenza artificiale, con il coinvolgimento del DPO, e la definizione di un piano d'azione con attività di sensibilizzazione e linee guida interne sono cruciali per il successo dell'Istituzione.

Come esempio di clausole contrattuali, la Commissione europea, attraverso l'iniziativa "Procurement of AI Community", ha riunito le parti interessate all'acquisto di soluzioni di intelligenza artificiale per sviluppare ampi modelli di clausole contrattuali per l'acquisto di intelligenza artificiale da parte di organizzazioni pubbliche. È inoltre importante considerare le clausole contrattuali standard tra titolari del trattamento e incaricati o responsabili del trattamento ai sensi del Regolamento.

L'Istituzione che vuole sviluppare o implementare sistemi di IA generativa, deve effettuare una DPIA?

I principi di *privacy by design* e *by default*[4] mirano a proteggere i dati personali durante l'intero ciclo di vita di un trattamento di dati personali sin dalla progettazione. Rispettando questo principio del Regolamento – basato su un approccio orientato al rischio – le minacce e i rischi che l'IA generativa può generare possono essere considerati e, ove possibile, mitigati in anticipo. Gli sviluppatori e gli implementatori sono tenuti ad effettuare le proprie valutazioni dei rischi e documentare qualsiasi azione di mitigazione intrapresa.

Gli sviluppatori di un sistema di AI sono tenuti a mettere a disposizione dei Titolari ovvero delle Istituzioni tutti i documenti relativi alle valutazioni dei rischi effettuate e alle misure di sicurezza adottate per l'impiego dei sistemi di intelligenza artificiale. Questo garantisce il rispetto dei principi di privacy by design e by default, che mirano a proteggere i dati personali sin dalla fase di progettazione e a prevenire potenziali minacce e rischi derivanti dall'utilizzo di intelligenza artificiale generativa. La comunicazione dettagliata di queste valutazioni e misure di sicurezza consente ai titolari del trattamento di valutare in modo adeguato l'opportunità di adottare di tali sistemi nonché a valutare l'adeguatezza dei medesimi rispetto delle normative sulla protezione dei dati.

Il regolamento prevede che una DPIA[5] debba essere effettuata prima di qualsiasi trattamento che possa comportare un rischio elevato[6] per i diritti e le libertà fondamentali delle persone. Il regolamento sottolinea l'importanza di effettuare tale valutazione nel caso in cui debbano essere utilizzate nuove tecnologie o siano di tipo nuovo in relazione alle quali il titolare del trattamento non ha mai effettuato una valutazione in precedenza, ad esempio nel caso dei sistemi di intelligenza artificiale generativa. Inoltre, l'art. 35 del Regolamento (UE) 2016/679 stabilisce che "quando un tipo di trattamento, allorché prevede 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 effettua, prima di procedere al trattamento, una valutazione dell'impatto dei trattamenti previsti sulla protezione dei dati personali". Di conseguenza, il titolare del trattamento è tenuto a chiedere il parere del DPO quando effettua una DPIA. A seguito della valutazione, devono essere adottate misure tecniche e organizzative adeguate a mitigare i rischi identificati, ove presenti, tenuto conto del contesto e delle misure disponibili allo stato dell'arte.

Nel contesto dell'uso dell'IA generativa, può essere opportuno chiedere il parere delle persone interessate dal sistema, ovvero gli stessi interessati o i loro rappresentanti nell'area di trattamento prevista. Oltre alle verifiche per valutare se la DPIA è stata correttamente attuata, è necessario effettuare un monitoraggio e una revisione regolari delle valutazioni dei rischi, poiché il funzionamento del modello può aggravare i rischi identificati o crearne di nuovi. Tali rischi sono legati alla protezione dei dati personali, ma anche ad altri diritti e libertà fondamentali.

[4] Article 27 of the Regulation
[5] Articles 39 and 89 Regulation (EU) 2016/679
[6] The classification of an AI system as posing "high-risk" due to its impact on fundamental rights according to the AI Act, does trigger a presumption of "high-risk" under the GDPR, the EUDPR and the LED to the extent that personal data is processed.

Tutti gli attori coinvolti nella DPIA devono garantire che ogni decisione e azione sia adeguatamente documentata, coprendo l'intero ciclo di vita del sistema di IA generativa, comprese le azioni intraprese per gestire i rischi e le successive revisioni da effettuare.

È responsabilità dell'Istituzione gestire in modo appropriato i rischi connessi all'uso dei sistemi di IA generativa. I rischi per la protezione dei dati devono essere identificati e affrontati durante l'intero ciclo di vita del sistema di IA generativa. Ciò include un monitoraggio regolare e sistematico per determinare, man mano che il sistema si evolve, se i rischi già identificati si aggravano o se ne compaiono di nuovi. La comprensione dei rischi legati all'uso dell'IA generativa è ancora in corso; pertanto, è necessario mantenere un approccio vigile nei confronti dei rischi emergenti non identificati. Se vengono identificati rischi che non possono essere mitigati con mezzi ragionevoli, è il momento di consultare il GEPD.

Il GEPD ha elaborato un modello che consente ai responsabili del trattamento di valutare se devono effettuare una DPIA [allegato sei alla parte I del kit di strumenti per la responsabilizzazione]. Inoltre, il GEPD ha stabilito un elenco aperto di trattamenti soggetti all'obbligo di una DPIA. Se necessario, il responsabile del trattamento effettua un riesame per valutare se il trattamento dei dati viene eseguito in conformità alla valutazione d'impatto sulla protezione dei dati, almeno quando si verifica una modifica dei rischi rappresentati dalle operazioni di trattamento. Se a seguito della DPIA i responsabili del trattamento non sono sicuri che i rischi siano adeguatamente attenuati, devono procedere a una consultazione preliminare con il GEPD.

In fase di progettazione, sviluppo e validazione di sistemi di IA generativa. Quando è lecito il trattamento dei dati personali?

Il trattamento dei dati personali nei sistemi di IA generativa può riguardare l'intero ciclo di vita del sistema, comprendendo tutte le attività di trattamento relative alla raccolta dei dati, all'addestramento, all'interazione con il sistema e alla generazione dei contenuti del sistema. Le attività di trattamento relative alla raccolta e all'addestramento includono l'ottenimento di dati da fonti disponibili liberamente su Internet e accessibili direttamente, da terzi o dagli archivi delle Istituzioni. I dati personali possono anche essere ottenuti dal modello di IA generativa direttamente dagli utenti, attraverso gli input del sistema o l'inferenza di nuove informazioni. Nel contesto dei sistemi di IA generativa, l'addestramento e l'uso dei sistemi si basano normalmente sul trattamento sistematico e su larga scala di dati personali, in molti casi senza la consapevolezza delle persone i cui dati sono trattati.

È di fondamentale importanza limitare al minimo il trattamento dei dati personali pubblicamente (es. presenti sui siti web istituzionali) come nelle sezioni di amministrazione trasparente, al fine di evitare la diffusione eccessiva di informazioni nonché la possibilità di reperire tali dati ai sistemi di intelligenza artificiale. Questo approccio mira a tutelare la sicurezza dei dati personali degli interessati nel rispetto della normativa nazionale ed europea applicabili nonché dei provvedimenti delle autorità nazionali di competenza.

Il trattamento di qualsiasi dato personale da parte delle Istituzioni è lecito se è applicabile almeno uno dei motivi di liceità[7] previsti dal Regolamento. Inoltre, affinché il trattamento di categorie particolari di dati personali sia lecito, deve applicarsi una delle eccezioni[8] elencate nel regolamento. Quando il trattamento è effettuato per l'esecuzione di un compito di interesse pubblico[9] o è necessario per l'adempimento di un obbligo legale[10] a cui è soggetto il titolare del trattamento, il motivo del trattamento deve essere stabilito dal diritto dell'UE. Inoltre, il diritto dell'UE a cui si fa riferimento deve essere chiaro e preciso e la sua applicazione deve essere prevedibile per le persone che vi sono soggette, in conformità con i requisiti stabiliti nella Carta dei diritti fondamentali dell'Unione Europea e nella Convenzione europea per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali.

Inoltre, quando una base giuridica comporta una grave interferenza con i diritti fondamentali alla protezione dei dati e della vita privata, vi è una maggiore necessità di norme chiare e precise che disciplinino la portata e l'applicazione della misura, nonché le relative garanzie. Pertanto, maggiore è l'interferenza, più solide e dettagliate devono essere le norme e le garanzie. Quando ci si affida a norme interne, queste ultime devono definire con precisione la portata dell'interferenza con il diritto alla protezione dei dati personali, attraverso l'identificazione delle finalità del trattamento, delle categorie di soggetti interessati, delle categorie di dati

[7] Article 5 of the Regulation
[8] Article 10(2) of the Regulation
[9] Article 5(1)(a) of the Regulation
[10] Article 5(1)(b) of the Regulation

personali che verrebbero trattati, dell'eventuale responsabile e/o incaricato del trattamento, dei periodi di conservazione, insieme a una descrizione delle garanzie e delle misure minime concrete per la protezione dei diritti delle persone.

L'uso del consenso[11] come base giuridica può essere applicato in alcune circostanze nel contesto dell'uso di sistemi di IA generativi. L'ottenimento del consenso[12] ai sensi del regolamento, e affinché tale consenso sia valido, deve soddisfare tutti i requisiti previsti dalla norma, tra cui la necessità di una chiara azione affermativa da parte dell'individuo, essere dato liberamente, specifico, informato ed inequivocabile. Considerato il modo in cui vengono addestrati i sistemi di IA generativa e le fonti di dati per l'addestramento, comprese le informazioni disponibili al pubblico, l'uso del consenso in quanto tale deve essere attentamente riconsiderato, anche nel contesto del suo utilizzo da parte di organismi pubblici, come le Istituzioni. Inoltre, è da considerare che, se il consenso viene revocato, tutte le operazioni di trattamento dei dati che si basavano su tale consenso e che hanno avuto luogo prima della revoca - e in conformità al regolamento – comporterebbe una necessaria interruzione delle operazioni di trattamento in questione. In tal caso, senza il consenso il Titolare non è legittimato a continuare ad effettuare il già menzionato trattamento.

Nel contesto dei trattamenti effettuati dalle istituzioni o dai soggetti che svolgono un compito di interesse pubblico, l'uso del consenso come base giuridica per il trattamento dei dati personali può presentare delle sfide alla luce del GDPR. Poiché le istituzioni pubbliche sono spesso tenute a trattare i dati per adempiere a compiti pubblici e a specifici obblighi normativi, il consenso potrebbe non essere sempre la base giuridica più appropriata anche nel contesto dell'utilizzo di sistemi di AI. Il considerando 43 del GDPR stabilisce che "per assicurare la libertà di espressione del consenso, è opportuno che il consenso non costituisca un valido presupposto per il trattamento dei dati personali in un caso specifico, qualora esista un evidente squilibrio tra l'interessato e il titolare del trattamento, specie quando il titolare del trattamento è un'autorità pubblica e ciò rende pertanto improbabile che il consenso sia stato espresso liberamente in tutte le circostanze di tale situazione specifica. Si presume che il consenso non sia stato liberamente espresso se non è possibile esprimere un consenso separato a distinti trattamenti di dati personali, nonostante sia appropriato nel singolo caso, o se l'esecuzione di un contratto, compresa la prestazione di un servizio, è subordinata al consenso sebbene esso non sia necessario per tale esecuzione". La questione andrebbe affrontata più approfonditamente.

I fornitori di servizi di modelli di IA generativa utilizzano il legittimo interesse ai sensi del GDPR come base giuridica applicabile al trattamento dei dati personali, in particolare per quanto riguarda la raccolta dei dati utilizzati per sviluppare i propri sistemi, compresi i processi di formazione e convalida. La Corte di giustizia dell'Unione europea (CGUE) ha stabilito[13] che l'uso del legittimo interesse stabilisce tre condizioni cumulative affinché il trattamento dei dati personali coperto da tale base giuridica sia legittimo:

A. il perseguimento di un interesse legittimo da parte del titolare del trattamento o di un terzo;
B. la necessità di trattare i dati personali ai fini degli interessi legittimi perseguiti;
C. che gli interessi o le libertà fondamentali e i diritti della persona interessata dalla protezione dei dati non prevalgano sull'interesse legittimo del titolare del trattamento o di un terzo.

[11] Article 5(1)(d) and 7 of the Regulation
[12] EDPB Guidelines 05/2020 on consent under Regulation 2016/679, available at https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf
[13] Judgment of 4 July 2023, Meta Platforms and Others (General terms of use of a social network), C-252/21, EU:C:2023:537, paragraph 106 and the case-law cited.

La domanda di pronuncia pregiudiziale verte sull'interpretazione dell'articolo 4, paragrafo 3, TUE nonché dell'articolo 6, paragrafo 1, dell'articolo 9, paragrafi 1 e 2, dell'articolo 51, paragrafo 1, e dell'articolo 56, paragrafo 1, del regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati) (GU 2016, L 119, pag. 1, e rettifiche in GU 2016, L 314, pag. 72, GU 2018, L 127, pag. 3 e GU 2021, L 74, pag. 35). Tale domanda è stata presentata nell'ambito di una controversia tra Meta Platforms Inc., già Facebook Inc., Meta Platforms Ireland Ltd, già Facebook Ireland Ltd, e Facebook Deutschland GmbH, da un lato, e il Bundeskartellamt (autorità federale garante della concorrenza, Germania), dall'altro, in merito alla decisione di quest'ultimo di vietare a tali società di procedere al trattamento di taluni dati personali previsto dalle condizioni generali di utilizzo del social network Facebook.

Nel caso del trattamento dei dati da parte di sistemi di IA generativi, molte circostanze possono influenzare il processo di bilanciamento inerente alla disposizione, portando a effetti quali l'imprevedibilità per gli interessati, nonché l'incertezza giuridica per i titolari del trattamento. A questo proposito, le Istituzioni hanno la responsabilità di verificare che i propri fornitori di sistemi di IA generativi abbiano rispettato le condizioni di applicazione di questa base giuridica, tenendo conto delle condizioni specifiche del trattamento effettuato.

In qualità di titolari del trattamento dei dati personali, le Istituzioni sono responsabili dei trasferimenti di dati personali che avviano e di quelli che vengono effettuati per loro conto anche oltre lo Spazio Economico Europeo (SEE). Questi trasferimenti possono avvenire solo se l'Istituzione in questione ha consentito tale trasferimento oppure se tali trasferimenti sono richiesti dalla legislazione dell'UE o degli Stati membri dell'UE. I trasferimenti possono avvenire a diversi livelli nel contesto dello sviluppo o dell'utilizzo di sistemi di IA generativa, anche quando le Istituzioni utilizzano sistemi basati su servizi cloud o quando devono fornire, in alcuni casi, dati personali da utilizzare per addestrare, testare o convalidare un modello. In entrambi i casi, questi trasferimenti di dati devono essere conformi alle disposizioni di cui al capo V[14] del regolamento, nonché alle altre disposizioni del regolamento, e devono essere coerenti con la finalità originaria del trattamento dei dati.

Nello specifico, l'art. 45 del GDPR stabilisce che "il trasferimento di dati personali verso un paese terzo o un'organizzazione internazionale è ammesso se la Commissione ha deciso che il paese terzo, un territorio o uno o più settori specifici all'interno del paese terzo, o l'organizzazione internazionale in questione garantiscono un livello di protezione adeguato. In tal caso il trasferimento non necessita di autorizzazioni specifiche". Il comma 3 del medesimo art. prevede che "la Commissione, previa valutazione dell'adeguatezza del livello di protezione, può decidere, mediante atti di esecuzione, che un paese terzo, un territorio o uno o più settori specifici all'interno di un paese terzo, o un'organizzazione internazionale garantiscono un livello di protezione adeguato ai sensi del paragrafo 2 del presente articolo. L'atto di esecuzione prevede un meccanismo di riesame periodico, almeno ogni quattro anni, che tenga conto di tutti gli sviluppi pertinenti nel paese terzo o nell'organizzazione internazionale. L'atto di esecuzione specifica il proprio ambito di applicazione geografico e settoriale e, ove applicabile, identifica la o le autorità di controllo di cui al paragrafo 2, lettera b), del presente articolo. L'atto di esecuzione è adottato secondo la procedura d'esame di cui all'articolo 93, paragrafo 2." La condivisione dei dati personali elaborati e trattati all'interno di un sistema di IA oltre lo SEE apre uno scenario di opportunità ma anche di potenziali rischi per le libertà e i diritti dei cittadini europei.

Il trattamento dei dati personali nel contesto dei sistemi di intelligenza artificiale generativa richiede una base giuridica in linea con il regolamento. Se il trattamento dei dati si basa su un obbligo legale o sull'esercizio di un'autorità pubblica, tale base giuridica deve essere stabilita in modo chiaro e preciso dal diritto dell'UE. L'uso del consenso come base giuridica richiede un'attenta considerazione per garantire che soddisfi i requisiti del Regolamento, per essere valido.

Ad esempio, la risoluzione dell'AAP sui sistemi di intelligenza artificiale generativa stabilisce che, laddove richiesto dalla legislazione applicabile, gli sviluppatori, i fornitori e i distributori di sistemi di intelligenza artificiale devono identificare fin dall'inizio la base giuridica per il trattamento dei dati personali relativi a: a) la raccolta dei dati utilizzati per sviluppare sistemi di intelligenza artificiale generativa; b) i set di dati di formazione, convalida e test utilizzati per sviluppare o migliorare i sistemi di intelligenza artificiale generativa; c) le interazioni degli interessati con i sistemi di intelligenza artificiale generativa; d) i contenuti generati dai sistemi di intelligenza artificiale generativa.

[14] Articles 46 to 51 of the Regulation

Minimizzazione e sistemi di IA generativa

Il principio di minimizzazione dei dati significa che i titolari del trattamento devono garantire che i dati personali oggetto di trattamento siano adeguati, pertinenti e limitati a quanto necessario in relazione alle finalità per cui sono trattati. Esiste l'idea errata che il principio di minimizzazione dei dati[15] non trovi spazio nel contesto dell'intelligenza artificiale. Tuttavia, i titolari del trattamento hanno l'obbligo di limitare la raccolta e il trattamento dei dati personali a quanto necessario per le finalità del trattamento, evitando il trattamento indiscriminato dei dati personali. Tale obbligo riguarda l'intero ciclo di vita del sistema, comprese le fasi di test, accettazione e messa in produzione. I dati personali non devono essere raccolti e trattati indiscriminatamente. Le Istituzioni devono garantire che il personale coinvolto nello sviluppo di modelli generativi di IA sia a conoscenza delle diverse procedure tecniche disponibili per ridurre al minimo l'uso dei dati personali e che queste siano debitamente prese in considerazione in tutte le fasi dello sviluppo.

Anche nell'ambito dell'intelligenza artificiale, i titolari del trattamento devono rispettare questo principio, assicurandosi di utilizzare solo i dati strettamente indispensabili per adempiere agli scopi previsti. Ciò implica la necessità di valutare attentamente i dati raccolti e trattati, evitando la raccolta eccessiva o indiscriminata di informazioni personali. Il principio di minimizzazione sottolinea l'importanza di garantire che il trattamento dei dati sia proporzionato, limitato e in linea con le finalità per cui i dati sono stati originariamente raccolti. Mediante una pratica attenta e finalizzata, i titolari del trattamento devono assicurarsi di rispettare i diritti fondamentali degli interessati in relazione alla protezione dei loro dati personali, conformemente ai principi normativi e alla giurisprudenza in materia di protezione dei dati personali. Tale principio, però, nell'ambito dei sistemi di AI dovrebbe essere estensibile anche ai responsabili ovvero agli sviluppatori ed implementatori di tali sistemi. Dovrebbe essere indispensabile limitare l'utilizzo di qualunque dato od informazione e stabilire, all'interno di una policy o vademecum ad hoc le informazioni oggetto di trattamento tramite sistemi di IA.

Le Istituzioni dovrebbero sviluppare e utilizzare modelli addestrati con serie di dati di alta qualità, limitati ai dati personali necessari per raggiungere lo scopo del trattamento. In tal modo, questi set di dati dovrebbero essere ben etichettati e curati, nell'ambito di adeguate procedure di governance dei dati, compresa la revisione periodica e sistematica del contenuto. I dataset e i modelli devono essere accompagnati da una documentazione sulla loro struttura, sulla manutenzione e sull'uso previsto. Quando utilizzano sistemi progettati o gestiti da fornitori di servizi terzi, le Istituzioni devono includere nelle loro valutazioni considerazioni relative al principio della minimizzazione dei dati.

Il problema nell'utilizzo di sistemi di intelligenza artificiale si estende anche all'impiego di dati anonimi o aggregati. Per quanto riguarda i dati anonimi, sebbene non possano identificare direttamente o indirettamente una persona fisica, potrebbero ancora essere utilizzati per ricavare informazioni che potrebbero essere ricondotte a un individuo. Questo rischio sottolinea la sfida nel garantire un effettivo anonimato dei dati e la necessità di considerare attentamente le possibili implicazioni per la privacy anche quando si trattano dati che sembrano essere stati resi anonimi. Si evidenzia dunque la complessità nel gestire i dati anonimi e l'importanza di adottare misure adeguate a proteggere la privacy e la riservatezza delle informazioni, anche quando si utilizzano dati che sembrano essere stati depersonalizzati.

L'utilizzo di una grande molle di dati per addestrare i sistemi di IA generativa non implica necessariamente una maggiore efficacia o risultati migliori. La progettazione accurata di insiemi di dati ben strutturati, da utilizzare in sistemi che privilegiano la "qualità" rispetto alla "quantità",

[15] In accordance with Article 4(1)(c) of the Regulation, personal data undergoing processing shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

seguendo un processo di addestramento adeguatamente supervisionato e sottoposto a un monitoraggio regolare, è essenziale per ottenere i risultati attesi, non solo in termini di minimizzazione dei dati, ma anche per quanto riguarda la qualità dell'output e la sicurezza dei dati.

EUI-X intende addestrare un sistema di intelligenza artificiale che sia in grado di assistere nelle attività di sviluppo e programmazione del software. A tal fine, vorrebbe utilizzare uno strumento di generazione di contenuti che sarà disponibile attraverso gli account dei singoli membri del personale IT. L'EUI-X deve riflettere prima di addestrare l'algoritmo per assicurarsi di non trattare dati personali che non sarebbero utili per lo scopo previsto. Ad esempio, può effettuare un'analisi statistica per dimostrare che è necessaria una quantità minima di dati per ottenere il risultato. Inoltre, dovranno verificare e giustificare se tratteranno categorie speciali di dati personali. Inoltre, dovranno esaminare la tipologia dei dati (cioè sintetizzati, anonimizzati o pseudonimizzati). Infine, dovranno verificare tutti gli elementi tecnici e legali pertinenti delle fonti di dati utilizzate, tra cui la loro legittimità, trasparenza ed esattezza.

Sistemi di IA generativa e principio di esattezza

I sistemi di IA generativa possono utilizzare in tutte le fasi del loro ciclo di vita, in particolare durante la fase di addestramento, enormi quantità di informazioni, compresi i dati personali. Il principio dell'esattezza dei dati[16] richiede che i dati siano corretti e aggiornati, mentre il titolare del trattamento è tenuto ad aggiornare o cancellare i dati inesatti. I titolari del trattamento devono garantire l'esattezza dei dati in tutte le fasi dello sviluppo e dell'utilizzo di un sistema di IA generativa. In effetti, devono implementare le misure necessarie per integrare la protezione dei dati dalla progettazione che contribuirà ad aumentare l'esattezza dei dati in tutte le fasi. Ciò implica la verifica della struttura e del contenuto dei set di dati utilizzati per l'addestramento dei modelli, compresi quelli provenienti od ottenuti da soggetti terzi. È altrettanto importante avere il controllo sui dati di output, comprese le inferenze fatte dal modello, il che richiede un monitoraggio regolare di tali informazioni, compresa la supervisione umana. Gli sviluppatori dovrebbero utilizzare set di validazione[17] durante l'addestramento e set di test separati per la valutazione finale, per ottenere una stima delle prestazioni del sistema. Sebbene in genere non siano orientate alla protezione dei dati personali, le metriche sull'esattezza statistica (la capacità dei modelli di produrre output o previsioni corrette in base ai dati su cui sono stati addestrati), quando disponibili, possono offrire un indicatore dell'esattezza dei dati utilizzati dal modello e delle prestazioni previste.

Questo significa che le organizzazioni che elaborano e gestiscono dati avvalendosi di sistemi di IA generativa hanno l'onere di assicurare che le informazioni utilizzate in tutte le fasi di creazione e impiego di un sistema di intelligenza artificiale generativa siano corrette, aggiornate e attendibili. Questo impegno per garantire l'esattezza dei dati è estremamente importante per il corretto funzionamento e l'affidabilità del sistema di IA generativa. Il rischio legato alla mancanza di garanzia sull'esattezza dei dati in tutte le fasi dello sviluppo e dell'utilizzo di un sistema di intelligenza artificiale generativa potrebbe portare a decisioni errate o inefficaci prese dal sistema basandosi su informazioni inesatte o obsolete. Questo potrebbe compromettere l'affidabilità, l'efficacia e la sicurezza del sistema di IA, con possibili conseguenze negative sugli utenti, sui dati stessi e sull'ambiente in cui il sistema viene utilizzato.

Quando le Istituzioni utilizzano un sistema di IA generativa o set di dati, test o convalida forniti da terzi, è necessario ottenere garanzie contrattuali e documentazione sulle procedure utilizzate per garantire l'esattezza dei dati utilizzati per lo sviluppo del sistema. Ciò include le procedure di raccolta dei dati, le procedure di preparazione, come l'annotazione, l'etichettatura, la pulizia, l'arricchimento e l'aggregazione, nonché l'identificazione di eventuali lacune e problemi che possono influire sull'esattezza. La documentazione tecnica e d'uso del sistema, comprese le schede dei modelli, deve consentire al *controller* del sistema di effettuare regolarmente controlli e azioni appropriate per garantire il principio di esattezza. Ciò è tanto più importante in quanto i modelli, anche se addestrati con dati rappresentativi di alta qualità, possono generare output contenenti informazioni imprecise o false, compresi i dati personali, le cosiddette "allucinazioni".

Nonostante gli sforzi per garantire l'esattezza dei dati, i sistemi di IA generativa sono ancora soggetti a risultati imprecisi che possono avere un potenziale impatto sui diritti e le libertà fondamentali delle persone. Mentre i fornitori stanno implementando sistemi di formazione avanzati per garantire che i modelli utilizzino e generino dati esatti e corretti, le Istituzioni dovrebbero valutare attentamente

[16] Article 4(1) of the Regulation
[17] Validation sets are used to fine-tune the parameters of a model and to assess its performance.

l'esattezza dei dati durante l'intero ciclo di vita dei sistemi di IA generativa che vengono utilizzati e prendere in considerazione l'uso di tali sistemi se l'esattezza non può essere mantenuta.

L'EUI-X, seguendo il consiglio del DPO, ha deciso che i risultati del modello ASR, se utilizzati per la trascrizione di riunioni e udienze ufficiali, saranno soggetti a convalida da parte di personale qualificato dell'Istituzione. Nei casi in cui il modello venga utilizzato per altre riunioni meno sensibili, la trascrizione sarà sempre accompagnata da una chiara indicazione che si tratta di un documento generato da un sistema di IA. L'EUI-X ha preparato e approvato a livello di top management una politica per l'uso del modello, oltre a informative sulla protezione dei dati conformi al regolamento che richiedono il consenso delle persone, sia per la registrazione della loro voce durante le riunioni sia per il loro trattamento da parte del sistema di trascrizione. Prima dell'implementazione del sistema di IA da parte dell'Istituzione è stata inoltre effettuata una DPIA.

Onere informativo nei sistemi di IA generativa

L'implementazione di politiche di trasparenza e la messa a disposizione degli interessati di informative privacy può contribuire a ridurre i rischi per gli interessati e a garantire la conformità ai requisiti del regolamento, in particolare fornendo informazioni dettagliate su come, quando e perché le Istituzioni hanno deciso di trattare i dati personali nei propri sistemi di IA generativa. Ciò implica la disponibilità di informazioni esaustive - che devono essere fornite dagli sviluppatori o dai fornitori, a seconda dei casi - sulle attività di trattamento svolte nelle diverse fasi di sviluppo, tra cui l'origine dei set di dati, la procedura di catalogazione/tagging, nonché qualsiasi trattamento associato. In particolare, le Istituzioni devono assicurarsi di ottenere informazioni adeguate e pertinenti sugli insiemi di dati utilizzati dai loro fornitori o prestatori e che tali informazioni siano affidabili e regolarmente aggiornate. Alcuni sistemi (es. le chatbot) possono richiedere specifici requisiti di trasparenza, tra cui informare gli interessati che stanno interagendo con un sistema di intelligenza artificiale senza intervento umano.

La trasparenza e la divulgazione di informazioni sulla privacy sono importanti perché aiutano a proteggere le persone riducendo i rischi legati al trattamento dei dati personali nei sistemi di intelligenza artificiale. Fornire dettagli su come, quando e perché i dati personali sono utilizzati nelle tecnologie di intelligenza artificiale consente agli interessati di capire e controllare meglio l'utilizzo delle proprie informazioni, creando fiducia e rispettando le normative sulla privacy. In pratica, ciò significa che gli interessati possono sapere come vengono gestiti i loro dati e quali rischi potrebbero comportare, garantendo un ambiente più sicuro e trasparente per tutti gli interessati. Il modo in cui è possibile farlo è con la

Poiché il diritto all'informazione[18] prevede l'obbligo di fornire agli interessati, nei casi di profilazione e di decisioni automatizzate, informazioni significative sulla logica di tali decisioni, nonché sul loro significato e sulle possibili conseguenze sulle persone, è importante che l'Istituzione sia in grado di mantenere le informazioni aggiornate, non solo sul funzionamento degli algoritmi utilizzati, ma anche sugli insiemi di dati trattati. Questo obbligo dovrebbe essere generalmente esteso ai casi in cui la procedura decisionale, pur non essendo interamente automatizzata, comprende atti preparatori basati su un trattamento automatizzato.

Per garantire la trasparenza e il rispetto della privacy nei sistemi di intelligenza artificiale generativa, i titolari del trattamento dovrebbero regolamentare il suo utilizzo, ad esempio tramite l'adozione di apposita policy o regolamento aziendale (es. "Vademecum sull'Intelligenza Artificiale" oppure "Artificial Intelligence Policy"). Questo strumento di regolamentazione procedimentalizzata potrebbe formalizzare i processi aziendali per il trattamento dei dati personali attraverso i sistemi di intelligenza artificiale, fornendo una chiara spiegazione su come verranno gestite queste informazioni. Inoltre, tali strumenti potrebbero dettagliare in che modo gli interessati possono, ove possibile, esercitare i propri diritti in termini di accesso, rettifica, cancellazione e limitazione del trattamento dei loro dati personali. Attraverso questa "Artificial Intelligence Policy", le Istituzioni possono assicurare – ai sensi del principio della responsabilizzazione – un approccio coerente e trasparente nel trattamento dei dati personali all'interno dei sistemi di intelligenza artificiale generativa, promuovendo la fiducia e garantendo il rispetto delle normative sulla privacy.

Le Istituzioni devono fornire agli interessati tutte le informazioni richieste dal regolamento quando utilizzano sistemi di IA generativa e tramite tali sistemi vengono gestiti, elaborati e trattati dati personali. Le informazioni fornite agli interessati devono essere aggiornate quando necessario per mantenerle adeguatamente informate e in grado di controllare i propri dati.

[18] Article 14 of the Regulation.

EU-X sta preparando una chatbot che assisterà gli utenti nell'accesso ad alcune aree del suo sito web. I titolari del trattamento, con la consulenza del DPO, hanno predisposto un'informativa (ex art. 13 GDPR), disponibile sul sito web di EU-X. L'informativa comprende informazioni sulle finalità del trattamento, la base giuridica, l'identificazione del titolare del trattamento e i dettagli di contatto del DPO, i destinatari dei dati, le categorie di dati personali oggetto di trattamento, la conservazione dei medesimi e le modalità di esercizio dei diritti che possono far valere gli interessati. L'informativa comprende anche informazioni sul funzionamento del sistema e sul possibile utilizzo dei dati forniti dall'utente per perfezionare la funzione fornita dalla chat. EU-X utilizza il consenso come base giuridica, ma gli utenti possono revocare il proprio consenso in qualsiasi momento. L'avviso chiarisce inoltre che l'uso della chatbot non è consentito ai minori. Prima di utilizzare la chatbot dell'Istituzione, gli utenti forniscono il proprio consenso dopo aver letto l'informativa sulla protezione dei dati.

AI generativa e processi decisionali automatizzati

L'uso di un sistema di intelligenza artificiale generativa non implica necessariamente un processo decisionale automatizzato[19] ai sensi del regolamento. Tuttavia, esistono sistemi di IA generativa che forniscono informazioni decisionali ottenute con mezzi automatizzati che comportano la profilazione e/o valutazioni individuali. A seconda dell'utilizzo di tali informazioni per l'adozione della decisione finale da parte di un servizio pubblico, le Istituzioni possono rientrare nell'ambito di applicazione dell'art. 24 del regolamento, per cui è necessario garantire garanzie individuali, tra cui almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere il proprio punto di vista e di contestare la decisione.

Un processo decisionale automatizzato è un processo in cui le decisioni sono prese da un sistema informatico o da un algoritmo senza l'intervento diretto di un essere umano. Questi sistemi utilizzano dati, regole e algoritmi per analizzare le informazioni e produrre risultati o decisioni in maniera automatizzata. In contesti come l'intelligenza artificiale e il machine learning, i processi decisionali automatizzati vengono utilizzati per compiti come la classificazione di dati, la previsione di risultati futuri o la raccomandazione di azioni da intraprendere, senza la necessità di un intervento umano diretto durante ciascuna decisione presa. L'art. 24 del Regolamento (UE) 2018/1725 stabilisce che "l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona". Tuttavia, vi sono selle eccezioni in quanto tale disposizione non è applicabile qualora la decisione a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e il titolare del trattamento; b) sia autorizzata dal diritto dell'Unione, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e degli interessi legittimi dell'interessato; oppure c) si basi sul consenso esplicito dell'interessato.

Nella gestione degli strumenti decisionali di IA, le Istituzioni devono considerare attentamente come garantire che il diritto di ottenere l'intervento umano sia attuato correttamente. Ciò è di fondamentale importanza nel caso in cui le Istituzioni impieghino agenti di IA autonomi in grado di svolgere compiti e prendere decisioni senza l'intervento o la guida dell'uomo. Le Istituzioni devono prestare molta attenzione al peso che le informazioni fornite dal sistema hanno nelle fasi finali della procedura decisionale e se hanno un'influenza decisiva sulla decisione finale presa dal titolare del trattamento. È importante riconoscere i rischi unici e i potenziali danni dei sistemi di IA generativa nel contesto del processo decisionale automatizzato, in particolare per le popolazioni vulnerabili e i bambini[20].

In occasione della Global Privacy Assembly (GPA) (2023) è stata sottolineata l'importanza della trasparenza nel trattamento dei dati personali utilizzati nei sistemi di intelligenza artificiale generativa. I fornitori di tali sistemi devono implementare non solo misure di sicurezza tecnica ed organizzativa ma anche di trasparenza. Tali fornitori sono tenuti, a sua volta, ad informare i clienti sui potenziali rischi per la protezione dei dati legati all'utilizzo di questi sistemi, illustrando come affrontano tali questioni attraverso politiche e pratiche adeguate. Queste informazioni devono essere chiare, facilmente comprensibili e accessibili agli utenti prima e durante l'utilizzo del sistema. Inoltre, se un sistema di intelligenza artificiale generativa viene impiegato per supportare processi decisionali, gli sviluppatori ed i fornitori devono comunicare in modo trasparente tali pratiche alle parti interessate. È inoltre richiesto che venisse fornita una politica di trasparenza riguardo ai set di dati utilizzati, comprese le fonti, le licenze e le pratiche di modifica o filtraggio dei dati, al fine di garantire una maggiore chiarezza e sicurezza nell'utilizzo di queste tecnologie. Come si evince anche dalla Resolution della GPA, il principio di trasparenza è applicabile non solo ai Titolari del trattamento (siano esse Istituzioni od organizzazioni private) ma anche ai Responsabili (fornitori).

[19] Article 24 of the Regulation.
[20] Article 24 of the Regulation.

Quando i sistemi di IA generativa sono previsti per supportare le procedure decisionali, le Istituzioni devono valutare attentamente se metterli in funzione se il loro uso solleva dubbi sulla loro legittimità o sul loro potenziale di essere decisioni ingiuste, non etiche o discriminatorie.

I fornitori di sistemi di IA dovrebbero essere soggetti ad audit periodici da parte dei titolari del trattamento, cioè le Istituzioni che intendono utilizzare tali sistemi, al fine di valutare se sono state implementate correttamente le misure di sicurezza tecniche, organizzative e di trasparenza richieste. Questi audit periodici consentirebbero alle Istituzioni di verificare che i fornitori stiano rispettando gli standard e le linee guida di sicurezza, organizzazione e trasparenza concordati. In tal modo, le Istituzioni possono garantire che i sistemi di intelligenza artificiale utilizzati rispettino le normative sulla protezione dei dati e siano conformi alle politiche interne e alle normative di settore. L'audit periodico aiuta a rafforzare la fiducia nell'utilizzo di tali sistemi e a garantire un trattamento sicuro e responsabile dei dati all'interno delle Istituzioni.

EUI-X sta valutando la possibilità di utilizzare un sistema di intelligenza artificiale per lo screening iniziale e il filtraggio delle candidature. Il fornitore di servizi C ha offerto un sistema di intelligenza artificiale generativa che esegue un'analisi dei requisiti formali e una valutazione automatica delle candidature, fornendo punteggi e suggerimenti su quali candidati intervistare nella fase successiva. Dopo aver consultato la documentazione sul modello, comprese le misure disponibili sull'esattezza statistica (misure sulla precisione e sulla sensibilità del modello) e in considerazione della possibile presenza di bias nel modello, EUI-X ha deciso che non utilizzerà il sistema almeno fino a quando non ci saranno chiare indicazioni che il rischio di bias è stato eliminato e le misure sulla precisione migliorano, all'analisi dei requisiti formali. In ogni caso, se tale sistema è considerato "adatto allo scopo" (ossia lo screening dei candidati) e conforme a tutti i regolamenti applicabili all'Istituzione, quest'ultima dovrebbe essere in grado di dimostrare di potersi validamente avvalere di una delle eccezioni di cui all'articolo 24, paragrafo 2, del regolamento; che l'Istituzione ha attuato misure adeguate per salvaguardare i diritti delle persone, compreso il diritto di ottenere un intervento umano da parte dell'Istituzione, di esprimere il proprio punto di vista e di contestare la decisione (ad esempio, la non ammissibilità). L'Istituzione deve fornire informazioni, ai sensi dell'articolo 15, paragrafo 2, lettera f), del regolamento, se i dati sono raccolti presso l'individuo, sulla logica coinvolta dal sistema di IA, nonché sulle conseguenze previste di tale trattamento per l'individuo. Una DPIA deve essere effettuata anche prima dell'implementazione del sistema di IA da parte dell'Istituzione. L'IUE-X può decidere di utilizzare, invece di un sistema di IA generativa, uno strumento automatizzato online più "semplice" per lo screening delle candidature (ad esempio, uno strumento informatico che verifichi automaticamente il numero di anni di esperienza professionale o di istruzione).

Come si può garantire un trattamento equo ed evitare pregiudizi quando si utilizzano sistemi di AI generativa?

In generale, le soluzioni di intelligenza artificiale tendono ad amplificare i pregiudizi umani esistenti ed eventualmente a incorporarne di nuovi, il che può creare nuove sfide etiche e rischi di conformità. I pregiudizi possono insorgere in qualsiasi fase dello sviluppo di un sistema di IA generativa, attraverso l'addestramento dei set di dati, gli algoritmi o le persone che sviluppano o utilizzano tale sistema. Le distorsioni nei sistemi di IA generativa possono portare a conseguenze negative significative per i diritti e le libertà fondamentali delle persone, tra cui il trattamento poco equo o la discriminazione, in particolare in settori quali la gestione delle risorse umane, l'assistenza medica pubblica e la fornitura di servizi sociali, le pratiche scientifiche e ingegneristiche, i processi politici e culturali, il settore finanziario, l'ambiente e gli ecosistemi, nonché la pubblica amministrazione.

Uno dei rischi maggiori per la protezione dei dati personali derivante dalle distorsioni nei sistemi di intelligenza artificiale generativa è la possibilità di discriminazione o trattamenti poco equi. Se tali sistemi non vengono adeguatamente bilanciati e tengono conto di potenziali distorsioni, possono perpetuare o addirittura accentuare pregiudizi o disparità già presenti nei dati di addestramento. Ciò potrebbe portare a decisioni ingiuste o discriminatorie che violerebbero i diritti fondamentali delle persone, come il diritto alla privacy e alla non discriminazione. In settori critici come la gestione delle risorse umane, l'assistenza medica, i servizi sociali, la finanza o la pubblica amministrazione, tali distorsioni potrebbero avere un impatto significativo sulla vita delle persone e sulla società nel suo complesso. Pertanto, assicurare la corretta gestione delle distorsioni nei sistemi di intelligenza artificiale generativa è cruciale per proteggere i dati personali e garantire decisioni equilibrate e non discriminatorie.

Le principali fonti possono derivare, tra l'altro, da modelli esistenti nei dati di formazione, dalla mancanza di informazioni (totali o parziali) sulla popolazione colpita, dall'inclusione o dall'omissione di variabili e dati che non dovrebbero o dovrebbero far parte dei set di dati, da errori metodologici o anche da distorsioni introdotte attraverso il monitoraggio.

È essenziale che gli insiemi di dati utilizzati per creare e addestrare i modelli garantiscano una rappresentazione adeguata ed equa del mondo reale - senza pregiudizi che possano aumentare il danno potenziale per individui o collettivi non ben rappresentati negli insiemi di dati per l'addestramento - implementando al contempo meccanismi di responsabilità e di supervisione che consentano un monitoraggio continuo per prevenire l'insorgere di pregiudizi che hanno un effetto sugli individui, nonché per correggere tali comportamenti. Ciò include la garanzia che le attività di trattamento siano tracciabili e verificabili[21] e che le Istituzioni conservino la documentazione di supporto. A questo proposito, è importante che le Istituzioni adottino e implementino modelli di documentazione tecnica, che possono essere particolarmente importanti quando i modelli utilizzano diversi set di dati e/o combinano diverse fonti di dati.

[21] Global Privacy Assembly (GPA) (2023). Resolution on Generative Artificial Intelligence Systems.

I fornitori di sistemi di IA generativa cercano di individuare e attenuare le distorsioni nei loro sistemi. Tuttavia, le Istituzioni, conoscendo meglio la propria realtà dovrebbero verificare e monitorare regolarmente se i risultati del sistema sono distorti utilizzando dati di input adattati alle loro esigenze aziendali. Le Istituzioni, in quanto autorità pubbliche, dovrebbero mettere in atto misure di salvaguardia per evitare di fare eccessivo affidamento sui risultati forniti dai sistemi, che possono portare a pregiudizi di automazione e di conferma.

Il rischio per le istituzioni pubbliche di fare un eccessivo affidamento sui risultati forniti dai sistemi di intelligenza artificiale, che potrebbero portare a pregiudizi di automazione e di conferma, è quello di prendere decisioni errate, discriminatorie o fuorvianti basate su tali risultati automatizzati. L'automazione e la conferma dei pregiudizi possono creare un circolo vizioso in cui i sistemi di intelligenza artificiale generano o amplificano discriminazioni esistenti o pregiudizi, rendendo le decisioni più sbagliate o distorti rispetto alla realtà. Questo potrebbe danneggiare la reputazione delle istituzioni, minare la fiducia del pubblico nell'equità e nell'obiettività delle decisioni prese e potenzialmente violare i diritti e le libertà fondamentali degli interessati coinvolti. Pertanto, per evitare tali rischi, le istituzioni pubbliche devono adottare misure di salvaguardia per garantire un uso responsabile e consapevole dei risultati forniti dai sistemi di intelligenza artificiale.

L'applicazione di procedure e best practices per la minimizzazione e la mitigazione dei pregiudizi dovrebbe essere una priorità in tutte le fasi del ciclo di vita dei sistemi di IA generativa, per garantire un trattamento equo ed evitare pratiche discriminatorie. A tal fine, è necessaria una supervisione e una comprensione del funzionamento degli algoritmi e dei dati utilizzati per l'addestramento del modello.

Il mantenere procedure e best practice per minimizzare e mitigare i pregiudizi dovrebbe essere una priorità in tutte le fasi del ciclo di vita dei sistemi di intelligenza artificiale generativa, al fine di garantire un trattamento equo ed evitare pratiche discriminatorie. Per raggiungere questo obiettivo, è essenziale che all'interno delle istituzioni venga fornita formazione agli individui che useranno tali strumenti di IA. Inoltre, è importante designare un responsabile incaricato di supervisionare regolarmente la qualità – anzitutto rispetto alla quantità – delle informazioni generate, verificandone l'esattezza e la completezza. Questa figura (es. Designato/a AI) sarebbe incaricata di garantire che le informazioni prodotte siano affidabili, privi di pregiudizi e rispettino le normative sulla privacy e la sicurezza dei dati. La supervisione e la formazione costanti sono fondamentali per assicurare un utilizzo responsabile ed efficace dei sistemi di intelligenza artificiale all'interno delle istituzioni.

L'EU-X sta valutando l'esistenza di un bias di campionamento nel sistema di riconoscimento vocale automatizzato. I servizi di traduzione hanno riportato tassi di errore di parola significativamente più elevati per alcuni oratori rispetto ad altri. Sembra che il sistema abbia difficoltà a gestire alcuni accenti inglesi. Dopo aver consultato lo sviluppatore, si è giunti alla conclusione che i dati di addestramento sono carenti per alcuni accenti, in particolare quando i parlanti non sono nativi. Poiché si tratta di un problema sistematico, EU-X sta valutando la possibilità di perfezionare il modello utilizzando set di dati generati in proprio.

Esercizio dei diritti nell'ambito dei sistemi di AI generativa

Le particolari caratteristiche dei sistemi di IA generativa fanno sì che l'esercizio dei diritti individuali[22] possa presentare sfide particolari, non solo nell'ambito del diritto di accesso, ma anche in relazione ai diritti di rettifica, cancellazione e opposizione al trattamento dei dati. Ad esempio, uno degli elementi più rilevanti è la difficoltà di identificare e accedere ai dati personali memorizzati dal sistema. Nei modelli linguistici di grandi dimensioni, ad esempio, singole parole come "gatto" o "cane" non vengono memorizzate come stringhe di testo. Vengono invece rappresentate come vettori numerici attraverso un processo chiamato word *embedding*. Questi vettori derivano dall'addestramento del modello su grandi quantità di dati testuali. La conseguenza è che l'accesso, l'aggiornamento o la cancellazione dei dati memorizzati in questi modelli, se possibile, è molto difficile. In questo senso, una corretta gestione degli insiemi di dati può facilitare l'accesso alle informazioni, cosa difficile nel caso di un addestramento non supervisionato basato su fonti pubblicamente disponibili che incorporano dati personali. Altrettanto complessa è la gestione della produzione di dati personali ottenuti per inferenza. Infine, l'esercizio di alcuni diritti, come il diritto alla cancellazione, può avere un impatto sull'efficacia del modello.

Come sopra accennato, per garantire la trasparenza e il rispetto della privacy nei sistemi di intelligenza artificiale generativa, i titolari del trattamento dovrebbero regolamentare il suo utilizzo, ad esempio tramite l'adozione di apposita policy o regolamento aziendale (es. "Vademedum sull'Intelligenza Artificiale" oppure "Artificial Intelligence Policy"). Questo strumento di regolamentazione procedimentalizzata potrebbero formalizzare i processi aziendali per il trattamento dei dati personali attraverso i sistemi di intelligenza artificiale, fornendo una chiara spiegazione su come verranno gestite queste informazioni. Inoltre, tali strumenti potrebbero dettagliare in che modo gli interessati possono, ove possibile, esercitare i propri diritti in termini di accesso, rettifica, cancellazione e limitazione del trattamento dei loro dati personali. Attraverso questa "Artificial Intelligence Policy", le Istituzioni possono assicurare – ai sensi del principio della responsabilizzazione – un approccio coerente e trasparente nel trattamento dei dati personali all'interno dei sistemi di intelligenza artificiale generativa, promuovendo la fiducia e garantendo il rispetto delle normative sulla privacy.

La conservazione di un registro tracciabile del trattamento dei dati personali e la gestione degli insiemi di dati in modo da consentire la tracciabilità del loro uso possono favorire l'esercizio dei diritti individuali. Le tecniche di minimizzazione dei dati possono anche contribuire a mitigare i rischi legati all'impossibilità di garantire il corretto esercizio dei diritti individuali in conformità al Regolamento.

Le Istituzioni, in qualità di titolari del trattamento, sono responsabili e tenute ad attuare misure tecniche, organizzative e procedurali adeguate a garantire l'effettivo esercizio dei diritti individuali. Tali misure devono essere progettate e attuate fin dalle prime fasi del ciclo di vita del sistema, consentendo la registrazione dettagliata e la tracciabilità delle attività di trattamento.

EU-X ha incluso nell'informativa sulla protezione dei dati per la chatbot un riferimento all'esercizio dei diritti individuali, tra cui l'accesso, la rettifica, la cancellazione, l'obiezione e la limitazione del trattamento in conformità all'EUDPR. L'avviso include i dettagli di contatto del titolare del trattamento e dell'EU-X DPO, nonché un riferimento alla possibilità di presentare un reclamo al GEPD. A seguito di una richiesta di accesso da parte di un interessato in merito al contenuto delle sue conversazioni con la chatbot, EU-X ha risposto, dopo aver effettuato i controlli del caso, che non è stato conservato alcun contenuto di tali conversazioni oltre il periodo di conservazione stabilito, 30 giorni. Le conversazioni, come indicato al singolo, non sono state utilizzate per addestrare il modello della chatbot.

[22] Chapter III of the Regulation.

Sistemi di AI generativa e sicurezza dei dati

L'uso di sistemi di IA generativa può amplificare i rischi di sicurezza esistenti o crearne di nuovi, creando anche nuove fonti e canali di trasmissione di rischi sistemici nel caso di modelli ampiamente utilizzati. Rispetto ai sistemi tradizionali, i rischi specifici per la sicurezza dell'IA generativa possono derivare dall'inaffidabilità dei dati utilizzati, dalla complessità dei sistemi, dall'opacità, dai problemi nell'esecuzione di test adeguati, dalle vulnerabilità nelle protezioni del sistema, ecc. L'offerta limitata di modelli in settori critici per la fornitura di servizi pubblici come la sanità può amplificare l'impatto delle vulnerabilità in questi sistemi. Il regolamento impone alle Istituzioni di attuare misure tecniche e organizzative appropriate per garantire un livello di sicurezza[23] adeguato al rischio per i diritti e le libertà delle persone fisiche.

Oltre ai tradizionali controlli di sicurezza per i sistemi informatici, i controllori dovrebbero integrare controlli specifici per le vulnerabilità già note di questi sistemi - attacchi di inversione di modello[24], *prompt injection*[25], *jailbreak*[26] - in modo da facilitare il monitoraggio continuo e la valutazione della loro efficacia. Si consiglia ai controllori di utilizzare solo set di dati forniti da fonti affidabili e di eseguire regolarmente procedure di verifica e convalida, anche per i set di dati interni.

Gli attacchi di inversione del modello, noti anche come "Model Inversion Attacks", sono un tipo di attacco in cui un aggressore cerca di invertire o estrarre informazioni sensibili dal modello stesso. Questo tipo di attacco sfrutta le vulnerabilità nei modelli di intelligenza artificiale per recuperare, ad esempio, dati sensibili o informazioni riservate utilizzate per addestrare il modello. In pratica, l'aggressore cerca di ottenere informazioni specifiche sulle istanze di addestramento utilizzate per costruire il modello, compromettendo così la riservatezza dei dati. Tali attacchi sono una preoccupazione per la sicurezza e la privacy dei dati nell'ambito dell'intelligenza artificiale.

Le Istituzioni dovrebbero formare il proprio personale su come identificare e gestire i rischi per la sicurezza legati all'uso dei sistemi di IA generativa. Poiché i rischi evolvono rapidamente, è necessario monitorare e aggiornare regolarmente la valutazione dei rischi. Allo stesso modo, poiché le modalità degli attacchi possono cambiare, è necessario garantire un accesso adeguato a conoscenze e competenze avanzate. Un modo possibile per affrontare i rischi sconosciuti è quello di utilizzare tecniche di "*red teaming*[27] " per cercare di trovare ed esporre le vulnerabilità. Quando si utilizza la *Retrieval Augmented Generation*[28] con i sistemi di IA generativa, è necessario verificare che il sistema di IA generativa non faccia trapelare dati personali che potrebbero essere presenti nella base di conoscenza del sistema.

[23] Article 33 of the Regulation.
[24] A Model inversion attacks takes place when an attacker extracts information from it through reverse-engineering.
[25] Malicious actors use prompt injection attacks to introduce malicious instructions as if they were harmless.
[26] Malicious actors use jailbreaking techniques to disregard the model safeguards.
[27] A red team uses attacking techniques aiming at finding vulnerabilities in the system.
[28] AI systems in which a Large Language Model bases its answers in a knowledge base prepared by the generative AI system owner (e.g. an EUI) with internal sources and not in the knowledge stored by the LLM itself.

La mancanza di informazioni sui rischi per la sicurezza legati all'uso dei sistemi di IA generativa e sulla loro possibile evoluzione impone alle Istituzioni di esercitare estrema cautela e di effettuare una pianificazione dettagliata di tutti gli aspetti legati alla sicurezza informatica, compresi il monitoraggio continuo e il supporto tecnico specializzato. Le Istituzioni devono essere consapevoli dei rischi derivanti da attacchi da parte di terzi malintenzionati e degli strumenti disponibili per mitigarli.

EU-X, a seguito di una valutazione della sicurezza, ha deciso di implementare il sistema ASR in loco, invece di utilizzare i servizi API forniti per lo sviluppatore del modello. EU-X formerà il proprio personale IT sull'uso e l'ulteriore sviluppo del sistema, in stretta collaborazione con il fornitore. Ciò può includere la formazione su come perfezionare il modello. Inoltre, EU-X si avvarrà dei servizi di un revisore esterno per verificare la corretta implementazione del sistema, anche per quanto riguarda la sicurezza.

Per saperne di più

EDPS work on AI

45th Closed Session of the Global Privacy Assembly - Resolution on Generative Artificial Intelligence Systems - 20 October 2023

EDPS TechDispatch #2/2023 - Explainable Artificial Intelligence

EDPS at work: data protection and AI (includes links to several documents published by the EDPS alone or in cooperation with other authorities)

EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)

EDPS Opinion 44/2023 on the Proposal for Artificial Intelligence Act in the light of legislative developments Large Language Models (EDPS website, part of the EDPS "TechSonar" report 2023-2024)

Other relevant documents

Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01)

CNIL: AI how-to-sheets

Spanish Data Protection Authority: Artificial Intelligence: accuracy principle in the processing activity

Italian Data Protection Authority: Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale – September 2023 (Italian)

The Hamburg Commissioner for Data Protection and Freedom of Information - Checklist for the use of LLM-based chatbots - 15/11/2023

AI Security Concerns in a nutshell (DE Federal Office for Information Security, March 2023) o Multilayer Framework for Good Cybersecurity Practices for AI (ENISA, June 2023)

Ethics Guidelines for Trustworthy AI (EC High-Level Expert Group on AI, 2019)

Living Guidelines on the responsible use of Generative AI in research (ERA Forum Stakeholders' document, March 2024)

OECD AI Incidents Monitor (AIM)

OECD Catalogue or tools and metrics for trustworthy AI