EDPS

2004 2024 20 EDPS

EUROPEAN DATA PROTECTION SUPERVISOR

The EU's independent data protection authority

03 giugno 2024

IA generativa e l'EUDPR.

Primi orientamenti del GEPD per garantire la conformità alla protezione dei dati durante l'utilizzo Sistemi di IA generativa.

I presenti orientamenti del GEPD sull'intelligenza artificiale generativa (IA generativa) e sulla protezione dei dati personali intendono fornire consigli pratici e istruzioni alle istituzioni, agli organi, agli uffici e alle agenzie (IUE) dell'UE sul trattamento dei dati personali quando si utilizzano sistemi di IA generativa, per facilitarne la conformità con i loro obblighi in materia di protezione dei dati stabiliti, in particolare, nel regolamento (UE) 2018/1725. Questi orientamenti sono stati elaborati per coprire il maggior numero possibile di scenari e applicazioni e non prescrivono misure tecniche specifiche. Mettono invece l'accento sui principi generali della protezione dei dati che dovrebbero aiutare gli IUE a rispettare i requisiti di protezione dei dati secondo il Regolamento (UE) 2018/1725.

Questi orientamenti rappresentano un primo passo verso orientamenti più dettagliati che terranno conto dell'evoluzione dei sistemi e delle tecnologie di intelligenza artificiale generativa, del loro utilizzo da parte delle IUE e dei risultati delle attività di monitoraggio e supervisione del GEPD.

Il GEPD emana questi orientamenti nel suo ruolo di autorità di controllo della protezione dei dati e non nel suo nuovo ruolo di autorità di controllo sull'IA ai sensi della legge sull'AI.

Questi orientamenti lasciano impregiudicata la legge sull'intelligenza artificiale.

Introduzione e ambito

1. Questi orientamenti intendono fornire alcuni consigli pratici alle istituzioni, agli organi, agli uffici e alle agenzie dell'UE (IUE) 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 come stabilito nel Regolamento (UE) 2018/1725 ("il Regolamento", o EUDPR). Anche se il Regolamento non menziona esplicitamente il concetto di Intelligenza Artificiale (AI), la corretta interpretazione e applicazione dei principi di protezione dei dati è essenziale per ottenere un uso vantaggioso di questi sistemi che non pregiudichi i diritti e le libertà fondamentali delle persone.

2. Il GEPD emana questi orientamenti nel suo ruolo di autorità di controllo della protezione dei dati e non nel suo nuovo ruolo di autorità di controllo sull'IA ai sensi della legge sull'AI.

3. Questi orientamenti non mirano a coprire in modo completo tutte le questioni rilevanti relative al trattamento dei dati personali nell'uso di sistemi di IA generativa che sono soggetti ad analisi da parte delle autorità di protezione dei dati. Alcune di queste domande 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 un modo che consente una migliore comprensione di come funziona l'intelligenza artificiale generativa.

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

5. Questi orientamenti sono strutturati come segue: domande chiave, seguite da risposte iniziali insieme ad alcune conclusioni preliminari e ulteriori chiarimenti o esempi.

6. Questi orientamenti iniziali costituiscono un passo preliminare verso lo sviluppo di orientamenti più completi. Nel corso del tempo, questi orientamenti saranno aggiornati, perfezionati e ampliati per affrontare ulteriori elementi necessari per supportare le IUE nello sviluppo e nell'implementazione di questi sistemi. Tale aggiornamento dovrebbe avvenire entro e non oltre dodici mesi dalla pubblicazione del presente documento.

1. Cos’è l’IA generativa?

L’intelligenza artificiale generativa è un sottoinsieme dell’intelligenza artificiale che utilizza modelli specializzati di apprendimento automatico progettati per produrre un’ampia e generale varietà di output, in grado di svolgere una gamma di attività e applicazioni, come la generazione di testo, immagini o audio. Concretamente, si basa sull'uso dei cosiddetti modelli di base, che fungono da modelli di base per altri sistemi di intelligenza artificiale generativa che saranno "perfezionati" da essi.

Un modello di fondazione funge da architettura centrale o base su cui vengono costruiti altri modelli più specializzati. Questi modelli vengono addestrati sulla base di set di dati diversi ed estesi, compresi quelli contenenti informazioni disponibili al pubblico. Possono rappresentare strutture complesse come immagini, audio, video o linguaggio e possono essere ottimizzati per compiti o applicazioni specifici.

Grandi modelli linguistici sono un tipo specifico di modello di base addestrato su enormi quantità di dati di testo (da milioni a miliardi di parole) che possono generare risposte in linguaggio naturale a un'ampia gamma di input basati su modelli e relazioni tra parole e frasi. Questa grande quantità di testo utilizzata per addestrare il modello può essere presa da Internet, libri e altre fonti disponibili. Alcune applicazioni già in uso sono sistemi di generazione di codice, assistenti virtuali, strumenti di creazione di contenuti, motori di traduzione linguistica, riconoscimento vocale automatizzato, sistemi di diagnosi medica, strumenti di ricerca scientifica, ecc.

La relazione tra questi concetti è gerarchica. L’intelligenza artificiale generativa è l’ampia categoria che comprende 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, basati sul modello di base, soddisfano compiti o applicazioni specifici, utilizzando la conoscenza e le capacità dell'architettura di base.

Il ciclo di vita di un modello di IA generativa copre 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 fondazione adeguato con cui iniziare, in altri casi potrebbe essere costruito da zero un nuovo modello. La fase successiva prevede l'addestramento del modello con set di dati rilevanti per lo scopo del sistema futuro, inclusa la messa a punto del sistema con set di dati specifici e personalizzati richiesti per soddisfare il caso d'uso del modello. Per finalizzare la formazione, vengono utilizzate tecniche specifiche che richiedono l'intervento umano per garantire informazioni più accurate e comportamenti controllati. La fase successiva mira a valutare il modello e a stabilire metriche per valutare regolarmente fattori come l'accuratezza e l'allineamento del modello con il caso d'uso. Infine, vengono implementati e implementati i modelli, compreso il monitoraggio continuo e la valutazione regolare utilizzando le metriche stabilite nelle fasi precedenti.

I casi d'uso rilevanti nell'intelligenza artificiale generativa sono applicazioni generali orientate al consumatore (come ChatGPT e sistemi simili che possono già essere trovati in diverse versioni e dimensioni[1], compresi quelli che possono essere eseguiti su un telefono cellulare). Esistono anche applicazioni aziendali in aree specifiche, modelli preaddestrati, applicazioni basate su modelli preaddestrati ottimizzati per l'uso specifico in un'area

[1] La dimensione di un modello linguistico di grandi dimensioni viene solitamente misurata come il numero di parametri (token che contiene). La dimensione di un modello LLM è importante poiché alcune funzionalità appaiono solo quando il modello cresce oltre determinati limiti.

di attività e, infine, modelli in cui l'intero sviluppo, compreso il processo di formazione, è svolto dall'entità responsabile.

L'intelligenza artificiale generativa, come altre nuove tecnologie, offre soluzioni in diversi campi intese a supportare e migliorare le capacità umane. Tuttavia, crea anche sfide potenziale impatto sui diritti e sulle libertà fondamentali che rischia di passare inosservato, trascurato, non adeguatamente considerato e valutato.

-L'addestramento di un Large Language Model (LLM) (e in generale di qualsiasi modello di machine learning) è un processo iterativo, complesso e ad alta intensità di risorse che coinvolge diverse fasi e tecniche volte a creare un modello in grado di generare testo simile a quello umano in reazione a comandi (o prompt) forniti dagli utenti. Il processo inizia con l'addestramento del modello su enormi set di dati, la maggior parte dei quali normalmente non etichettati e ottenuti da fonti pubbliche utilizzando tecnologie di webscraping (- le autorità di 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 accessibili al pubblico dati personali). Successivamente, gli LLM vengono, non in tutti i casi, perfezionati utilizzando l'apprendimento supervisionato o attraverso tecniche che coinvolgono l'intervento umano (come l'apprendimento per rinforzo con feedback umano (RLHF) o il test contraddittorio tramite esperti di dominio) per aiutare il sistema a riconoscere ed elaborare meglio informazioni e contesto, nonché per determinare le risposte preferite, se limitare la produzione in risposta a domande sensibili e allinearla ai valori degli sviluppatori (ad esempio evitare di produrre risultati 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 affinare il modello.

---

2. Le IUE possono utilizzare l'IA generativa?

In quanto IUE, in linea di principio non vi sono ostacoli allo sviluppo, alla diffusione e all'uso di sistemi di IA generativa nella fornitura di servizi pubblici, a condizione che le norme dell'IUE lo consentano e che tutti i requisiti legali applicabili siano soddisfatti, soprattutto considerando la responsabilità speciale dell'IUE. settore pubblico per garantire il pieno rispetto dei diritti e delle libertà fondamentali degli individui nell'uso delle nuove tecnologie.

In ogni caso, qualora l'utilizzo di sistemi di IA generativa comporti il trattamento di dati personali, il Regolamento si applica integralmente. Il regolamento è tecnologicamente neutro e si applica a tutte le attività di trattamento dei dati personali, indipendentemente dalle tecnologie utilizzate e fatti salvi altri quadri giuridici, in particolare la legge sull'AI. Il principio di accountability prevede che le responsabilità siano chiaramente identificate e rispettate tra i diversi attori coinvolti nella filiera del modello di IA generativa.

Le IUE possono sviluppare e implementare le proprie soluzioni di intelligenza artificiale generativa o, in alternativa, utilizzare per proprio uso le soluzioni disponibili sul mercato. In entrambi i casi, le IUE possono utilizzare fornitori per ottenere tutti o alcuni degli elementi che fanno parte del sistema di IA generativa. In questo contesto, gli IUE devono chiaramente determinare i ruoli specifici - titolare, responsabile, contitolare - per lo specifico trattamento operazioni effettuate e le loro implicazioni in termini di obblighi e responsabilità previste dal Regolamento.

Poiché le tecnologie dell'intelligenza artificiale avanzano rapidamente, gli IUE devono considerare attentamente quando e come utilizzare l'intelligenza artificiale generativa in modo responsabile e vantaggioso per il bene pubblico. Tutte le fasi di un'intelligenza artificiale generativa Il ciclo di vita della soluzione dovrebbe operare in conformità con i quadri giuridici applicabili, compreso il regolamento, quando il sistema prevede il trattamento di dati personali.

-I termini IA affidabile o responsabile si riferiscono alla necessità di garantire che i sistemi di IA siano sviluppati in modo etico e legale. Ciò implica considerare le conseguenze indesiderate dell'uso della tecnologia IA e la necessità di seguire un approccio basato sul rischio che copra tutte le fasi del ciclo di vita del sistema. Implica anche trasparenza riguardo all'uso dei dati di formazione e alle loro fonti, su come gli algoritmi sono progettati e implementati, che tipo di pregiudizi potrebbero essere presenti nel sistema e come vengono affrontati i possibili impatti sui diritti e sulle 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.

3. Come sapere se l'utilizzo di un sistema di IA generativa comporta il trattamento di dati personali?

Il trattamento dei dati personali in un sistema di IA generativa può avvenire a vari livelli e fasi del suo ciclo di vita, senza necessariamente essere ovvio a prima vista. Ciò include la creazione dei set di dati di addestramento, nella fase di addestramento stessa, deducendo 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 esecuzione.

Quando uno sviluppatore o un fornitore di un sistema di IA generativa dichiara che il proprio sistema non tratta dati personali (per motivi quali il presunto utilizzo di set di dati anonimizzati o dati sintetici durante la progettazione, lo sviluppo e il test), è fondamentale chiedere informazioni specifiche controlli messi in atto per garantire ciò. In sostanza, gli IUE potrebbero voler sapere quali passaggi o procedure utilizza il fornitore per garantire che i dati personali non vengano elaborati dal modello.

Il GEPD ha già messo in guardia[2] contro l'uso di tecniche di web scraping per raccogliere dati personali, attraverso le quali gli individui potrebbero perdere il controllo delle proprie informazioni personali quando queste vengono raccolte a loro insaputa, contro le loro aspettative e per scopi diversi da quelli della raccolta originaria. Il GEPD ha inoltre sottolineato che il trattamento dei dati personali accessibili al pubblico resta soggetto alla legislazione dell'UE sulla protezione dei dati. A tale riguardo, l'uso di tecniche di web scraping per raccogliere dati da siti web e il loro utilizzo a fini di formazione potrebbe non rispettare i principi pertinenti in materia di protezione dei dati, compresi la minimizzazione dei dati e il principio di accuratezza, nella misura in cui non vi è alcuna valutazione sull'affidabilità dei dati. fonti.

Il monitoraggio regolare e l'attuazione 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 per tale trattamento.

-EUI-X, un'istituzione fittizia dell'UE, sta valutando l'acquisizione di un prodotto per il riconoscimento e la trascrizione automatica del parlato. Dopo aver studiato le opzioni disponibili, si è concentrato 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 vocale e la traduzione. Poiché tale modello sarà utilizzato per la trascrizione delle riunioni mediante file vocali registrati, si è stabilito che l'utilizzo di tale modello richiede il trattamento di dati personali e pertanto deve garantire il rispetto del Regolamento.

---

[2] *Parere 41/2023, del 25 settembre 2023, sulla proposta di regolamento relativo alle statistiche del mercato del lavoro dell'Unione europea relative alle imprese*

4. Qual è il ruolo dei DPO nel processo di sviluppo o implementazione di sistemi di IA generativa?

L'articolo 45 del Regolamento stabilisce i compiti del responsabile della protezione dei dati. I DPO informano e consigliano sui pertinenti obblighi in materia di protezione dei dati, assistono i titolari del trattamento nel monitorare la conformità interna, forniscono consulenza ove richiesto in merito alle DPIA e fungono da punto di contatto per gli interessati e il GEPD.

Nel contesto dell'implementazione da parte delle IUE di sistemi di IA generativa che trattano dati personali, è importante garantire che i DPO, nell'ambito del loro ruolo, forniscano consulenza e assistenza in modo indipendente sull'applicazione del regolamento, abbiano una corretta comprensione del ciclo di vita dei sistemi di intelligenza artificiale generativa che trattano dati personali. il sistema di IA generativa che l'IUE intende acquisire, progettare o implementare e come funziona. Ciò significa ottenere informazioni su quando e come questi sistemi trattano i dati personali e su come funzionano i meccanismi di input e output, nonché i processi decisionali implementati attraverso il modello. È importante, come sottolinea il regolamento[3], per fornire consulenza ai titolari del trattamento durante lo svolgimento di valutazioni d'impatto sulla protezione dei dati. I titolari del trattamento devono garantire che tutti i processi siano adeguatamente documentati e che la trasparenza sia garantita, compreso l'aggiornamento dei registri dei trattamenti e, come migliore pratica, la realizzazione di un inventario specifico sui sistemi e sulle applicazioni generative basati sull'intelligenza artificiale. Infine, il DPO dovrebbe essere coinvolto nella revisione delle questioni di conformità nel contesto degli accordi di condivisione dei dati firmati con i fornitori di modelli.

Dal punto di vista organizzativo, l'attuazione di sistemi di IA generativa conformi al regolamento non dovrebbe essere l'impegno di una sola persona. Dovrebbe esserci un dialogo continuo tra tutte le parti interessate coinvolte nel ciclo di vita del prodotto. Pertanto, i controllori dovrebbero collaborare con tutte le funzioni rilevanti all'interno dell'organizzazione, in particolare il DPO, il servizio legale, il servizio IT e il responsabile locale della sicurezza informatica (LISO), al fine di garantire che l'IUE operi entro i parametri di un'affidabilità generativa. AI, buona governance dei dati e conformità al Regolamento. La creazione di una task force sull'intelligenza artificiale, compreso il DPO, e la preparazione di un piano d'azione, comprese azioni di sensibilizzazione a tutti i livelli dell'organizzazione e la preparazione di orientamenti interni possono contribuire al raggiungimento di questi obiettivi.

-Come esempio di clausole contrattuali, la Commissione Europea, attraverso l'iniziativa "Procurement of AI Community", ha riunito le parti interessate nel procurare soluzioni AI per sviluppare un'ampia gamma diclausole contrattuali tipo per l'acquisizione di Intelligenza Artificiale da parte di organizzazioni pubbliche. È inoltre rilevante considerare lclausole contrattuali standard tra titolari e responsabili del trattamento ai sensi dell'artRegolamento[1].

[3] *Articolo 39, paragrafo 2, del regolamento*

5. AnEUI vuole sviluppare o implementare sistemi di IA generativa. Quando dovrebbe essere effettuata una DPIA?

I principi della protezione dei dati fin dalla progettazione e per impostazione predefinita[4] mirano a proteggere i dati personali durante l'intero ciclo di vita del trattamento dei dati, a partire dalla fase iniziale. Rispettando questo principio del Regolamento, sulla base di un approccio orientato al rischio, le minacce e i rischi che l'IA generativa può comportare possono essere considerati e mitigati con sufficiente anticipo. Gli sviluppatori e gli operatori potrebbero dover effettuare le proprie valutazioni dei rischi e documentare qualsiasi azione di mitigazione intrapresa.

Il Regolamento richiede che una DPIA[5] deve essere effettuato prima di qualsiasi operazione di trattamento che possa comportare un rischio elevato[6] ai diritti e alle libertà fondamentali delle persone. Il regolamento sottolinea l'importanza di effettuare tale valutazione laddove debbano essere utilizzate nuove tecnologie o siano di un nuovo tipo in relazione alle quali non è stata effettuata in precedenza alcuna valutazione da parte del titolare del trattamento, ad esempio nel caso dei sistemi di IA generativa.

Il titolare del trattamento è obbligato a chiedere il parere del responsabile della protezione dei dati (DPO) quando effettua una DPIA. A seguito della valutazione, devono essere adottate misure tecniche e organizzative adeguate per mitigare i rischi identificati, tenendo conto delle responsabilità, del contesto e delle misure all'avanguardia disponibili.

Potrebbe essere opportuno, nel contesto dell'uso dell'intelligenza artificiale generativa, chiedere il parere delle persone interessate dal sistema, ovvero gli interessati stessi o i loro rappresentanti nell'ambito del trattamento previsto. Oltre alle revisioni per valutare se la DPIA è correttamente attuata, è necessario effettuare un monitoraggio regolare e revisioni delle valutazioni del rischio, dal momento che il funzionamento della DPIA modello può esacerbare i rischi identificati o crearne di nuovi. Tali rischi sono legati alla protezione dei dati personali, ma sono legati anche ad altri diritti e libertà fondamentali.

Tutti gli attori coinvolti nella DPIA devono garantire che qualsiasi 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'IUE gestire adeguatamente i rischi connessi all'utilizzo di 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 evolve, se i rischi già identificati stanno peggiorando o se stanno emergendo nuovi rischi. La comprensione dei rischi legati all'uso dell'intelligenza artificiale generativa è ancora in corso, quindi è necessario mantenere un approccio vigile nei confronti

---

[4] Articolo 27 del Regolamento

[5] Articoli 39 e 89 del Regolamento.

[6] La classificazione di un sistema IA come "ad alto rischio" a causa del suo impatto sui diritti fondamentali ai sensi della legge sull'AI fa scattare una presunzione di "alto rischio" ai sensi del GDPR, dell'EUDPR e della LED nella misura in cui i dati personali i dati vengono elaborati.

rischi emergenti non identificati. Se vengono identificati rischi che non possono essere mitigati con mezzi ragionevoli, è tempo di consultare il GEPD.

-Il GEPD ha stabilito un modello che consente ai titolari del trattamento di valutare se devono effettuare una DPIA [allegato sei alla parte I dello strumentario sulla responsabilità]. Inoltre, il GEPD ha istituito unelenco aperto delle operazioni di trattamento soggette all'obbligo di una DPIA. Ove necessario, il titolare del trattamento effettua una revisione per valutare se il trattamento dei dati viene effettuato in conformità con la valutazione d'impatto sulla protezione dei dati, almeno quando si verifica una modifica dei rischi rappresentati dalle operazioni di trattamento. Se seguendo la DPIA, i titolari del trattamento non sono sicuri che i rischi siano adeguatamente mitigati, dovrebbero procedere a una consultazione preventiva con il GEPD.

---

6. Quando è lecito il trattamento dei dati personali nelle fasi di progettazione, sviluppo e validazione di sistemi di IA generativa?

Il trattamento dei dati personali nei sistemi di IA generativa può coprire l'intero ciclo di vita del sistema, comprendendo tutte le attività di trattamento legate alla raccolta dei dati, alla formazione, all'interazione con il sistema e alla generazione di contenuti dei sistemi. Le attività di raccolta e di trattamento legate alla formazione comprendono l'ottenimento di dati da fonti accessibili al pubblico su Internet, direttamente, da terzi o dagli archivi degli IUE. I dati personali possono anche essere ottenuti dal modello di intelligenza artificiale generativa direttamente dagli utenti, tramite gli input al sistema o attraverso l'inferenza di nuove informazioni. Nel contesto dei sistemi di IA generativa, la formazione e l'utilizzo dei sistemi si basano normalmente sul trattamento sistematico e su larga scala di dati personali, in molti casi senza che le persone i cui dati vengono trattati ne siano consapevoli.

Il trattamento di qualsiasi dato personale da parte degli IUE è lecito se sussiste almeno uno dei motivi di liceità[7] elencati nel regolamento sono applicabili. Inoltre, affinché il trattamento di categorie particolari di dati personali sia lecito, esiste una delle eccezioni[8]elencati nel regolamento devono applicarsi. Quando il trattamento viene effettuato per l'esecuzione di compiti svolti nell'interesse pubblico[9]o è necessario per l'adempimento di un obbligo legale[10]cui è soggetto il titolare del trattamento, la base giuridica del trattamento deve essere individuata nel diritto dell'UE. Inoltre, il diritto comunitario in questione dovrebbe essere chiaro e preciso e la sua applicazione dovrebbe essere prevedibile per i soggetti ad esso soggetti, conformemente alla normativa 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, laddove una base giuridica dà luogo a una grave interferenza con i diritti fondamentali alla protezione dei dati e alla 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 robuste e dettagliate dovrebbero essere le norme e le garanzie. Facendo affidamento su norme interne, tali norme interne dovrebbero definire con precisione la portata dell'ingerenza nel diritto alla protezione dei dati personali, attraverso l'identificazione della finalità del trattamento, delle categorie di interessati, delle categorie di dati personali che verrebbero trattati, del titolare del trattamento e dei responsabili del trattamento, nonché i periodi di conservazione, unitamente alla descrizione delle concrete garanzie e misure minime per la tutela dei diritti delle persone fisiche.

L'uso del consenso[11]come base giuridica può applicarsi in alcune circostanze nel contesto dell'uso di sistemi di IA generativa. Ottenere il consenso[12]ai sensi del regolamento, e affinché tale consenso sia valido, è necessario che soddisfi tutti i requisiti legali, compresa la necessità di una chiara azione affermativa da parte dell'individuo, sia dato liberamente, specifico, informato e inequivocabile. Considerato il modo in cui vengono addestrati i sistemi di intelligenza artificiale generativa e le fonti dei dati di addestramento, comprese le informazioni disponibili al pubblico, l'uso del consenso in quanto tale deve essere attentamente considerato, anche nel contesto del suo utilizzo da parte del pubblico.

---

[7]Articolo 5 del Regolamento.

[8]Articolo 10, paragrafo 2, del regolamento.

[9]Articolo 5, paragrafo 1, lettera a), del regolamento.

[10]Articolo 5, paragrafo 1, lettera b), del regolamento

[11]Articoli 5, comma 1, lettera d), e 7 del Regolamento.

[12]Linee guida EDPB 05/2020 sul consenso ai sensi del Regolamento 2016/679, disponibili all'indirizzo https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en .pdf

organismi, come gli IUE. Inoltre, in caso di revoca del consenso, tutte le operazioni di trattamento dei dati basate su tale consenso ed effettuate prima della revoca – e in conformità al Regolamento – rimangono lecite. Tuttavia, in questo caso, il titolare del trattamento deve interrompere le operazioni di trattamento in questione. Se non esiste altro fondamento giuridico che giustifichi il trattamento dei dati, i dati rilevanti devono essere cancellati dal titolare del trattamento.

I fornitori di servizi di modelli di IA generativa possono avvalersi dell'interesse legittimo ai sensi del Regolamento generale sulla protezione dei dati dell'UE[13](GDPR) come base giuridica per il trattamento dei dati, in particolare per quanto riguarda la raccolta dei dati utilizzati per sviluppare il sistema, compresi i processi di formazione e validazione. Lo ha affermato la Corte di Giustizia dell'Unione Europea (CGUE).[14]che l'utilizzo dell'interesse legittimo prevede tre condizioni cumulative affinché il trattamento dei dati personali coperto da tale base giuridica sia lecito. In primo luogo, il perseguimento di un legittimo interesse da parte del titolare del trattamento o di terzi; in secondo luogo, la necessità di trattare i dati personali per finalità legate ai legittimi interessi perseguiti; e in terzo luogo, che gli interessi o le libertà e i diritti fondamentali della persona interessata dalla protezione dei dati non prevalgono sull'interesse legittimo del titolare del trattamento o di terzi. Nel caso del trattamento dei dati mediante sistemi di IA generativa, molte circostanze possono influenzare il processo di bilanciamento insito nella disposizione, che porta a effetti quali l'imprevedibilità per gli interessati, nonché l'incertezza giuridica per i titolari del trattamento. A tale riguardo, le IUE hanno la responsabilità specifica di verificare che i fornitori di sistemi di IA generativa rispettino le condizioni
dell'applicazione di tale base giuridica, tenendo conto delle condizioni specifiche del trattamento effettuato da tali sistemi.

In qualità di titolari del trattamento dei dati personali, gli IUE sono responsabili dei trasferimenti di dati personali da loro avviati e di quelli effettuati per loro conto all'interno e all'esterno dello Spazio economico europeo. Questi trasferimenti possono avvenire solo se l'IUE in questione li ha istruiti o consentiti, o se tali trasferimenti sono richiesti dal diritto dell'UE o dal diritto degli Stati membri dell'UE. I trasferimenti possono avvenire a diversi livelli nel contesto dello sviluppo o dell'uso di sistemi di IA generativa, anche quando le IUE fanno uso di sistemi basati su servizi cloud o quando devono fornire, in determinati casi, dati personali da utilizzare per formare, testare o convalidare un modello. In entrambi i casi, questi trasferimenti di dati devono rispettare le disposizioni di cui al capo V[15]del Regolamento, fatte salve anche le altre disposizioni del Regolamento, ed essere coerenti con la finalità originaria del trattamento dei dati.

Il trattamento dei dati personali nel contesto dei sistemi di IA 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 pubblici poteri, tale base giuridica deve essere chiaramente e precisamente individuata nel diritto dell'Unione. L'utilizzo del consenso come base giuridica richiede un'attenta considerazione per garantire che soddisfi i requisiti del Regolamento, per essere valido.

---

[13]Regolamento (UE) 2016/679 del Parlamento Europeo e del Consiglio del 27 aprile 2016 sulla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché sulla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (Regolamento generale sulla protezione dei dati)

[14]Sentenza del 4 luglio 2023, Meta Platforms e altri (Condizioni generali d'uso di una rete sociale), C-252/21, EU:C:2023:537, punto 106 e giurisprudenza ivi citata

[15]Articoli da 46 a 51 del Regolamento

-Ad esempio, ilRisoluzione GPA sui sistemi di intelligenza artificiale generativa afferma che, ove richiesto dalla legislazione pertinente, gli sviluppatori, i fornitori e gli utilizzatori di sistemi di IA generativa devono identificare fin dall'inizio la base giuridica per il trattamento dei dati personali relativi a: a) raccolta di dati utilizzati per sviluppare sistemi di IA generativa; b) formazione, convalida e test di set di dati utilizzati per sviluppare o migliorare i sistemi di IA generativa; c) interazioni degli individui con sistemi di IA generativa; d) contenuti generati da sistemi di IA generativa.

7. Come si può garantire il principio della minimizzazione dei dati quando si utilizzano sistemi di IA generativa?

Il principio di minimizzazione dei dati implica che i titolari del trattamento garantiscono che i dati personali oggetto del trattamento siano adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati. C'è un malinteso sul principio della minimizzazione dei dati[16]non ha posto nel contesto dell'intelligenza artificiale. Tuttavia, i titolari del trattamento hanno l'obbligo di limitare la raccolta e comunque il trattamento dei dati personali a quanto necessario per le finalità del trattamento, evitando trattamenti indiscriminati dei dati personali. Tale obbligo copre l'intero ciclo di vita del sistema, compresi il collaudo, l'accettazione e l'immissione nelle fasi di produzione. I dati personali non dovrebbero essere raccolti e trattati indiscriminatamente. Le IUE devono garantire che il personale coinvolto nello sviluppo di modelli di IA generativa 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.

Le IUE dovrebbero sviluppare e utilizzare modelli addestrati con set di dati di alta qualità limitati ai dati personali necessari per raggiungere lo scopo del trattamento. In questo modo, questi set di dati dovrebbero essere ben etichettati e curati, nel quadro di adeguate procedure di governance dei dati, inclusa la revisione periodica e sistematica del contenuto. I set di dati e i modelli devono essere accompagnati dalla documentazione sulla loro struttura, manutenzione e destinazione d'uso. Quando utilizzano sistemi progettati o gestiti da fornitori di servizi terzi, le IUE dovrebbero includere nelle loro valutazioni considerazioni relative al principio di minimizzazione dei dati.

L'utilizzo di grandi quantità di dati per addestrare un sistema di IA generativa non implica necessariamente una maggiore efficacia o risultati migliori. L'attenta progettazione di dataset ben strutturati, da utilizzare in sistemi che privilegiano la qualità rispetto alla quantità, a seguito di un processo di formazione adeguatamente supervisionato e soggetto a monitoraggio regolare, è essenziale per raggiungere i risultati attesi, non solo in termini di minimizzazione dei dati, ma anche quando si tratta di qualità dell'output e di sicurezza dei dati.

-EUI-X intende addestrare un sistema di intelligenza artificiale per essere in grado di assistere con attività legate allo sviluppo e alla programmazione del software. Per questo vorrebbero 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 che non elaboreranno dati personali che non sarebbero utili per lo scopo previsto. Ad esempio, possono 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 particolari di dati personali. Inoltre, dovranno esaminare la tipologia dei dati (vale a dire sintetizzati, anonimizzati o pseudonimizzati). Infine, dovranno verificare tutti gli elementi tecnici e giuridici rilevanti delle fonti di dati utilizzate, compresa la loro liceità, trasparenza e accuratezza.

---

[16]Ai sensi dell'articolo 4, comma 1, lettera c), del Regolamento, i dati personali oggetto di trattamento devono essere adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati.

8. I sistemi di IA rigenerativi rispettano il principio di accuratezza dei dati?

I sistemi di IA generativa possono utilizzare in tutte le fasi del loro ciclo di vita, in particolare durante la fase di formazione, enormi quantità di informazioni, compresi i dati personali.

Il principio di accuratezza dei dati[17] richiede che i dati siano esatti, aggiornati, mentre il titolare del trattamento è tenuto ad aggiornare o cancellare i dati inesatti. I titolari del trattamento dei dati devono garantire l'accuratezza dei dati in tutte le fasi dello sviluppo e dell'utilizzo di un sistema di IA generativa. Devono infatti attuare le misure necessarie per integrare la protezione dei dati fin dalla progettazione che contribuiranno ad aumentare l'accuratezza dei dati in tutte le fasi.

Ciò implica verificare la struttura e il contenuto dei set di dati utilizzati per i modelli di addestramento, compresi quelli provenienti o ottenuti da terze parti. È altrettanto importante avere il controllo sui dati di output, comprese le inferenze fatte dal modello, che richiede un monitoraggio regolare di tali informazioni, compresa la supervisione umana. Gli sviluppatori dovrebbero utilizzare set di convalida[18] durante la formazione e set di test separati per la valutazione finale per ottenere una stima delle prestazioni del sistema. Sebbene generalmente non siano orientati alla protezione dei dati, i parametri sull'accuratezza statistica (la capacità dei modelli di produrre risultati o previsioni corretti in base ai dati su cui sono stati addestrati), quando disponibili, possono offrire un indicatore dell'accuratezza dei dati che il modello utilizza come nonché sulle prestazioni attese.

Quando le IUE utilizzano un sistema di IA generativa o set di dati di formazione, test o convalida forniti da terzi, è necessario ottenere garanzie contrattuali e documentazione sulle procedure utilizzate per garantire l'accuratezza dei dati utilizzati per lo sviluppo del sistema. Ciò include procedure di raccolta dei dati, procedure di preparazione, come annotazione, etichettatura, pulizia, arricchimento e aggregazione, nonché l'identificazione di possibili lacune e problemi che possono influire sull'accuratezza. La documentazione tecnica e per l'utente del sistema, comprese le schede modello, dovrebbe consentire al responsabile del controllo del sistema di effettuare regolarmente controlli e azioni adeguati per garantire il principio di accuratezza. Ciò è ancora più importante poiché i modelli, anche se addestrati con dati rappresentativi di alta qualità, possono generare output contenenti informazioni inaccurate o false, inclusi dati personali, le cosiddette "allucinazioni".

Nonostante gli sforzi per garantire l'accuratezza dei dati, i sistemi di intelligenza artificiale generativa sono ancora inclini a risultati imprecisi che possono avere un impatto sui diritti e sulle libertà fondamentali degli individui.

Mentre i fornitori stanno implementando sistemi di formazione avanzati per garantire che i modelli utilizzino e generino dati accurati, gli IUE dovrebbero valutare attentamente l'accuratezza dei dati durante l'intero ciclo di vita dei sistemi di intelligenza artificiale generativa e considerare l'uso di tali sistemi se l'accuratezza non può essere mantenuta.

[17] Articolo 4, paragrafo 1, lettera d), del regolamento.

[18] I set di validazione vengono utilizzati per mettere a punto i parametri di un modello e per valutarne le prestazioni.

-L'EUI-X, seguendo il parere del DPO, ha deciso che i risultati del modello ASR, quando utilizzati per la trascrizione di riunioni e udienze ufficiali, saranno soggetti a validazione da parte di personale qualificato dell'IUE. Nei casi in cui il modello venga utilizzato per altri incontri meno sensibili, la trascrizione sarà sempre accompagnata da una chiara indicazione che si tratta di un documento generato da un sistema di AI. EUI-X ha predisposto e approvato a livello di alta direzione una policy per l'utilizzo del modello nonché informative sulla privacy conformi al Regolamento che richiedono il consenso dei soggetti, sia per la registrazione della loro voce durante le riunioni sia per il suo trattamento da parte il sistema di trascrizione. È stata inoltre effettuata una DPIA prima dell'implementazione del sistema di IA da parte dell'IUE.

9. Come informare gli individui sul trattamento dei dati personali quando le IUE utilizzano sistemi di IA generativa?

Adeguate politiche di informazione e trasparenza possono contribuire a mitigare i rischi per le persone e garantire la conformità ai requisiti del regolamento, in particolare fornendo informazioni dettagliate su come, quando e perché le IUE trattano i dati personali nei sistemi di IA generativa. Ciò implica disporre di informazioni complete - che devono essere fornite dagli sviluppatori o dai fornitori a seconda dei casi - sulle attività di trattamento svolte nelle diverse fasi di sviluppo, inclusa l'origine dei set di dati, la procedura di curation/tagging, nonché qualsiasi altra attività associata in lavorazione. In particolare, le IUE dovrebbero garantire di ottenere informazioni adeguate e pertinenti sui set di dati utilizzati dai loro fornitori e che tali informazioni siano affidabili e regolarmente aggiornate. Alcuni sistemi (ad esempio i chatbot) possono richiedere specifici requisiti di trasparenza, tra cui informare le persone che stanno interagendo con un sistema di intelligenza artificiale senza intervento umano.

Come il diritto all'informazione[19] comprende l'obbligo di fornire alle persone, nei casi di profilazione e decisioni automatizzate, informazioni significative sulla logica di tali decisioni, nonché sul loro significato e sulle possibili conseguenze sulle persone, è importante che l'IUE mantenga informazioni aggiornate, non solo sulle sul funzionamento degli algoritmi utilizzati, ma anche sui dataset di elaborazione. Tale obbligo dovrebbe in generale essere esteso ai casi in cui, pur non essendo il procedimento decisionale interamente automatizzato, esso comporta atti preparatori basati su un trattamento automatizzato.

Le IUE devono fornire alle persone tutte le informazioni richieste dal regolamento quando utilizzano sistemi di IA generativa che trattano dati personali. Le informazioni fornite agli interessati devono essere aggiornate quando necessario per mantenerli adeguatamente informati e in controllo dei propri dati.

-EU-X sta preparando un chatbot che assisterà le persone durante l'accesso a determinate aree del suo sito web. I titolari del trattamento interessati, con la consulenza del DPO, hanno predisposto un'informativa sulla protezione dei dati, disponibile sul sito web EU-X. L'informativa contiene informazioni sulle finalità del trattamento, sulla base giuridica, sugli estremi identificativi del titolare e sui dati di contatto del DPO, sui destinatari dei dati, sulle categorie di dati personali raccolti, sulla conservazione dei dati nonché sulle modalità esercitare i diritti individuali. L'informativa contiene anche informazioni sul funzionamento del sistema e sull'eventuale utilizzo degli input dell'utente per affinare la funzionalità di 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'utilizzo del chatbot non è consentito ai minorenni. Prima di utilizzare il chatbot dell'IUE, le persone possono fornire il consenso dopo aver letto l'informativa sulla protezione dei dati.

---

[19] Articolo 14 del Regolamento.

10. Che dire delle decisioni automatizzate ai sensi dell'articolo 24 del Regolamento?

L'uso di un sistema di IA generativa non implica necessariamente un processo decisionale automatizzato[20] ai sensi del regolamento. Esistono tuttavia sistemi di IA generativa che forniscono informazioni decisionali ottenute con mezzi automatizzati che comportano profilazione e/o valutazioni individuali. A seconda dell'uso di tali informazioni nel prendere la decisione finale da parte di un servizio pubblico, gli IUE possono rientrare nell'ambito di applicazione dell'articolo 24 del regolamento, quindi devono garantire che siano garantite salvaguardie individuali, compreso almeno il diritto di ottenere intervento umano da parte del responsabile del trattamento, per esprimere il proprio punto di vista e contestare la decisione.

Nella gestione degli strumenti decisionali dell'IA, gli IUE devono considerare attentamente come garantire che il diritto di ottenere l'intervento umano sia adeguatamente attuato. Ciò è di fondamentale importanza nel caso in cui le IUE utilizzino agenti IA autonomi in grado di eseguire compiti e prendere decisioni senza intervento o guida umana.

Gli IUE devono prestare molta attenzione al peso che le informazioni fornite dal sistema hanno nelle fasi finali del processo decisionale e se hanno un'influenza decisiva sulla decisione finale presa dal responsabile del trattamento. È importante riconoscere i rischi specifici e i potenziali danni dei sistemi di IA generativa nel contesto del processo decisionale automatizzato, in particolare sulle popolazioni vulnerabili e sui bambini[21].

Laddove siano previsti sistemi di IA generativa per supportare le procedure decisionali, le IUE devono valutare attentamente se metterli in funzione se il loro utilizzo solleva dubbi sulla loro liceità o sul loro potenziale di essere decisioni ingiuste, non etiche o discriminatorie.

[20] Articolo 24 del Regolamento.
[21] Assemblea globale sulla privacy (GPA) (2023). Risoluzione sui sistemi di intelligenza artificiale generativa.

-EUI-X sta valutando l'utilizzo di un sistema di intelligenza artificiale per lo screening iniziale e il filtraggio delle domande di lavoro. Il fornitore di servizi C ha offerto un sistema di intelligenza artificiale generativa che effettua un'analisi dei requisiti formali e una valutazione automatizzata 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'accuratezza statistica (misure sulla precisione e 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 vi saranno chiare indicazioni che il rischio di bias sia stato eliminato e le misure sulla precisione non siano migliorate, all'analisi dei requisiti formali.

In ogni caso, se tale sistema è considerato "idoneo allo scopo" (ovvero lo screening dei candidati) e conforme a tutte le norme applicabili all'IUE, l'IUE dovrebbe essere in grado di dimostrare di poter validamente avvalersi di una delle eccezioni di cui all'art. 24, comma 2, del Regolamento; che l'IUE ha attuato misure adeguate per salvaguardare i diritti degli individui, compreso il diritto di ottenere l'intervento umano da parte dell'IUE, di esprimere il proprio punto di vista e di contestare la decisione (ad esempio, non ammissibilità).

L'IUE deve fornire informazioni, ai sensi dell'articolo 15, paragrafo 2, lettera f), del Regolamento, se i dati sono raccolti presso l'individuo, sulla logica utilizzata dal sistema di IA, nonché sulle conseguenze previste di tale elaborazione per l'individuo. Una DPIA deve essere effettuata anche prima dell'implementazione del sistema di IA da parte dell'IUE.

L'IUE-X potrebbe decidere di utilizzare, invece di un sistema di intelligenza artificiale generativa, uno strumento automatizzato online "più semplice" per lo screening delle domande di lavoro (ad esempio, uno strumento informatico che controlla automaticamente il numero di anni di esperienza professionale o di istruzione).

11. Come si può garantire un trattamento equo ed evitare distorsioni quando si utilizzano sistemi di IA generativa?

In generale, le soluzioni di intelligenza artificiale tendono ad amplificare i pregiudizi umani esistenti e possibilmente a incorporarne di nuovi, il che può creare nuove sfide etiche e rischi di conformità legale. Possono sorgere pregiudizi in qualsiasi fase dello sviluppo di un sistema di intelligenza artificiale generativa attraverso l'addestramento di set di dati, algoritmi o attraverso le persone che sviluppano o utilizzano il sistema. I pregiudizi nei sistemi di IA generativa possono portare a conseguenze negative significative per i diritti e le libertà fondamentali degli individui, compresi trattamenti iniqui e discriminazioni, in particolare in settori quali la gestione delle risorse umane, l'assistenza sanitaria pubblica e la fornitura di servizi sociali, scientifici e pratiche ingegneristiche, processi politici e culturali, settore finanziario, ambiente ed ecosistemi, nonché pubblica amministrazione.

Le principali fonti di distorsione possono provenire, tra le altre, da modelli esistenti nei dati di addestramento, mancanza di informazioni (totale o parziale) sulla popolazione interessata, inclusione o omissione di variabili e dati che non dovrebbero o dovrebbero far parte dei set di dati, metodi metodologici errori o addirittura pregiudizi introdotti attraverso il monitoraggio.

È essenziale che i set 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 nei set di dati di addestramento, implementando allo stesso tempo meccanismi di responsabilità e supervisione. che consentono un monitoraggio continuo per prevenire il verificarsi di pregiudizi che si ripercuotono sugli individui, nonché per correggere tali comportamenti. Ciò include garantire che le attività di trattamento siano tracciabili e verificabili[22]e che gli IUE conservino la documentazione di supporto. A tale riguardo, è importante che le IUE adottino e attuino modelli di documentazione tecnica, che possono essere di particolare importanza quando i modelli utilizzano diversi set di dati e/o combinano diverse fonti di dati.

I fornitori di sistemi di intelligenza artificiale generativa cercano di rilevare e mitigare i pregiudizi nei loro sistemi. Tuttavia, gli IUE conoscono meglio il proprio business case e dovrebbero testare e monitorare regolarmente se l'output del sistema è distorto utilizzando dati di input adattati alle loro esigenze aziendali.

Le IUE, in quanto autorità pubbliche, dovrebbero mettere in atto misure di salvaguardia per evitare un eccessivo affidamento sui risultati forniti dai sistemi che può portare all'automazione e a errori di conferma.

L'applicazione di procedure e migliori pratiche per la minimizzazione e l'attenuazione dei bias 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. Per questo è necessaria la supervisione e la comprensione di come funzionano gli algoritmi e dei dati utilizzati per addestrare il modello.

[22] L'audit dei dati di addestramento può aiutare a rilevare bias e altre questioni problematiche studiando il modo in cui i dati di addestramento vengono raccolti, etichettati, curati e annotati. La qualità dell'audit e i suoi risultati dipendono dall'accesso alle informazioni rilevanti, inclusi i set di dati di formazione, la documentazione e i dettagli di implementazione.

-EU-X sta valutando l'esistenza di errori di campionamento nel sistema di riconoscimento vocale automatizzato. I servizi di traduzione hanno segnalato tassi di errore di parole significativamente più elevati per alcuni parlanti rispetto ad altri. Sembra che il sistema abbia difficoltà a gestire alcuni accenti inglesi. Dopo essersi consultato con lo sviluppatore, è giunto alla conclusione che esiste un deficit nei dati di addestramento per alcuni accenti, in particolare quando i parlanti non sono nativi. Poiché è sistematico, EU-X sta valutando la possibilità di perfezionare il modello utilizzando i propri set di dati generati.

12. Che dire dell'esercizio dei diritti individuali?

Le particolari caratteristiche dei sistemi di IA generativa rendono impossibile l'esercizio dei diritti individuali [23]possono 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 nell'identificazione e nell'accesso ai dati personali archiviati dal sistema. Nei modelli linguistici di grandi dimensioni, ad esempio, le singole parole come "gatto" o "cane" non vengono memorizzate come stringhe di testo. Sono invece rappresentati come vettori numerici attraverso un processo chiamato word embedding. Questi vettori derivano dall'addestramento del modello su grandi quantità di dati di testo. La conseguenza è che accedere, aggiornare o cancellare i dati memorizzati in questi modelli, se possibile, è molto difficile. In questo senso, una corretta gestione dei set di dati può facilitare l'accesso alle informazioni, cosa difficile nel caso di una formazione non supervisionata basata su fonti accessibili al pubblico che incorporano dati personali. Altrettanto complesso è gestire la produzione di dati personali ottenuti tramite inferenza. Infine, l'esercizio di alcuni diritti, come quello alla cancellazione, può avere un impatto sull'efficacia del modello.

Mantenere una registrazione tracciabile del trattamento dei dati personali, nonché gestire i set di dati in modo da consentire la tracciabilità del loro utilizzo, può supportare 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.

Gli IUE, in quanto titolari del trattamento, sono responsabili dell'attuazione di misure tecniche, organizzative e procedurali adeguate per garantire l'effettivo esercizio dei diritti individuali. Tali misure dovrebbero essere progettate e attuate sin dalle prime fasi del ciclo di vita del sistema, consentendo una registrazione e una valutazione dettagliate.

tracciabilità delle attività di trattamento.

-EU-X ha incluso nell'informativa sulla protezione dei dati per il chatbot un riferimento all'esercizio dei diritti individuali, inclusi l'accesso, la rettifica, la cancellazione, l'opposizione e la limitazione del trattamento in conformità con l'EUDPR. L'avviso include i dati di contatto del titolare del trattamento e dell'RPD EU-X, nonché un riferimento alla possibilità di presentare un reclamo al GEPD. A seguito di una richiesta di accesso da parte di un soggetto in merito al contenuto delle sue conversazioni con il chatbot, EU-X ha risposto, dopo aver effettuato le relative verifiche, che nessun contenuto viene conservato di dette conversazioni oltre il periodo di conservazione stabilito, 30 giorni. Le conversazioni, come indicato all'individuo, non sono state utilizzate per addestrare il modello di chatbot.

---

[23]Capo III del Regolamento.

13. E la sicurezza dei dati?

L'uso di sistemi di IA generativa può amplificare i rischi per la sicurezza esistenti o crearne di nuovi, inclusa la creazione di 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 da dati di addestramento inaffidabili, complessità dei sistemi, opacità, problemi nell'effettuare test adeguati, vulnerabilità nelle salvaguardie del sistema, ecc. L'offerta limitata di modelli in settori critici per la fornitura dei servizi pubblici come la sanità possono amplificare l'impatto delle vulnerabilità di questi sistemi. Il Regolamento impone alle IUE di attuare misure tecniche e organizzative adeguate per garantire un livello di sicurezza[24]adeguato al rischio per i diritti e le libertà delle persone fisiche.

I titolari del trattamento dovrebbero, oltre ai tradizionali controlli di sicurezza per i sistemi IT, integrare controlli specifici adattati alle vulnerabilità già note di questi sistemi - modello di attacchi di inversione[25], iniezione rapida[26], jailbreak[27]- in modo da agevolare il monitoraggio continuo e la valutazione della loro efficacia. Si consiglia ai titolari del trattamento di utilizzare solo set di dati forniti da fonti attendibili e di eseguire regolarmente procedure di verifica e convalida, anche per i set di dati interni. Le IUE dovrebbero formare il proprio personale su come identificare e gestire i rischi per la sicurezza legati all'uso di sistemi di IA generativa. Poiché i rischi evolvono rapidamente, sono necessari un monitoraggio regolare e aggiornamenti della valutazione del rischio. Allo stesso modo, poiché le modalità degli attacchi possono cambiare, è necessario garantire un accesso adeguato alle conoscenze e alle competenze avanzate. Un possibile modo per affrontare i rischi sconosciuti è utilizzare il "red teaming".[28]"tecniche per cercare di trovare ed esporre le vulnerabilità.

Quando si utilizza la generazione aumentata di recupero[29]con i sistemi di IA generativa, è necessario verificare che il sistema di IA generativa non perda dati personali che potrebbero essere presenti nella base di conoscenza del sistema.

La mancanza di informazioni sui rischi per la sicurezza legati all'uso dei sistemi di intelligenza artificiale generativa e su come potrebbero evolversi richiede agli IUE di esercitare estrema cautela e di effettuare una pianificazione dettagliata di tutti gli aspetti relativi alla sicurezza informatica, compreso il monitoraggio continuo e il supporto tecnico specializzato . Gli IUE devono essere consapevoli dei rischi derivanti dagli attacchi di terzi dannosi e degli strumenti disponibili per mitigarli.

[24]Articolo 33 del Regolamento.
[25]Un attacco di inversione del modello avviene quando un utente malintenzionato ne estrae informazioni tramite il reverse engineering.
[26]Gli autori malintenzionati utilizzano attacchi di tipo prompt injection per introdurre istruzioni dannose come se fossero innocue.
[27]Gli attori malintenzionati utilizzano tecniche di jailbreak per ignorare le garanzie del modello.
[28]Una squadra rossa utilizza tecniche di attacco mirate a trovare vulnerabilità nel sistema.
[29]Sistemi di IA in cui un Large Language Model basa le sue risposte su una base di conoscenza preparata dal proprietario del sistema di IA generativa (ad esempio un IUE) con fonti interne e non sulla conoscenza archiviata dal LLM stesso.

-EU-X, a seguito di una valutazione della sicurezza, ha deciso di implementare il sistema ASR on premise, invece di utilizzare i servizi API forniti allo sviluppatore del modello. EU-X formerà il proprio personale IT sull'uso e sull'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 in termini di sicurezza.

14. Vuoi saperne di più?

- **Il GEPD lavora sull'IA**

  - 45a sessione chiusa dell'Assemblea globale sulla privacy -Risoluzione sui sistemi di intelligenza artificiale generativa - 20 ottobre 2023

  oTechDispatch n. 2/2023 del GEPD -Intelligenza artificiale spiegabile

  - Il GEPD al lavoro:protezione dei dati e intelligenza artificiale (include collegamenti a diversi documenti pubblicati dal GEPD da solo o in collaborazione con altre autorità)

  - EDPB-EDPSParere congiunto 5/2021 sulla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce norme armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale)

  - GEPDParere 44/2023 sulla proposta di legge sull'intelligenza artificiale alla luce degli sviluppi legislativi

    Modelli linguistici di grandi dimensioni (sito web del GEPD, parte delRelazione "TechSonar" del GEPD 2023-2024)

- **Altri documenti rilevanti**

  - Linee guida sul processo decisionale individuale automatizzato e sulla profilazione ai fini del Regolamento 2016/679 (wp251rev.01)

  oCNIL:Fogli didattici AI

  - Autorità spagnola per la protezione dei dati:Intelligenza Artificiale: principio di accuratezza nell'attività di trattamento

  - Garante italiano per la protezione dei dati personali:Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale – Settembre 2023 (italiano)

  - Il Commissario di Amburgo per la protezione dei dati e la libertà d'informazione -Lista di controllo per l'utilizzo di chatbot basati su LLM - 15/11/2023

  - Preoccupazioni per la sicurezza dell'intelligenza artificiale in poche parole (Ufficio federale tedesco per la sicurezza informatica, marzo 2023)

  - Quadro multistrato per buone pratiche di sicurezza informatica per l'intelligenza artificiale (ENISA, giugno 2023)

  - Linee guida etiche per un'intelligenza artificiale affidabile (Gruppo di esperti di alto livello della CE sull'intelligenza artificiale, 2019)

  - Living Linee guida sull'uso responsabile dell'IA generativa nella ricerca (Documento delle parti interessate del Forum ERA, marzo 2024)

  - Monitoraggio degli incidenti di intelligenza artificiale dell'OCSE (AIM)

  - Catalogo o strumenti e parametri dell'OCSE per un'intelligenza artificiale affidabile