AI E PRIVACY: ANALISI DEL DOCUMENTO EDPB "AI PRIVACY RISKS & MITIGATIONS – LLMS"

14 aprile 2025

AUTORI

Aurora Agostini
Partner

Giulietta Minucci
Counsel

Giovanni Lombardi
Associate

Alessandro Carlini
Associate

Il documento dell'EDPB analizza i rischi privacy associati ai LLM e propone una metodologia operativa per la loro mitigazione. Il presente contributo offre una lettura critica del testo, alla luce delle normative vigenti e in corso di adozione, e propone linee guida concrete per imprese e sviluppatori impegnati nell'adozione responsabile dell'AI generativa.

INTRODUZIONE

L'avvento dei modelli linguistici di grandi dimensioni – comunemente noti come *Large Language Models* (LLMs) – rappresenta una delle trasformazioni più dirompenti nell'evoluzione dell'intelligenza artificiale. Grazie alla loro capacità di comprendere, generare e manipolare il linguaggio naturale, tali sistemi sono oggi impiegati in una molteplicità di contesti: dall'assistenza virtuale alla redazione automatica di testi, dalla programmazione al supporto decisionale in ambito medico-legale, fino all'integrazione in agenti intelligenti dotati di autonomia operativa.

A fronte delle straordinarie opportunità offerte da questi strumenti, si pongono tuttavia interrogativi in merito alla tutela dei diritti fondamentali degli interessati, con particolare riguardo alla protezione dei dati personali. I LLM, per loro stessa natura, si basano su processi di addestramento che prevedono l'elaborazione di **enormi volumi di dati**, spesso tratti da fonti eterogenee e talvolta contenenti informazioni riconducibili a persone fisiche identificate o identificabili. Inoltre, le modalità di funzionamento di questi modelli – basate su approcci statistici e probabilistici – rendono difficile, se non impossibile, controllare con certezza il contenuto degli output generati o garantire la non rievocazione di dati personali appresi in fase di training.

In questo scenario, il documento elaborato dal Support Pool of Experts per conto dell'European Data Protection Board (EDPB), intitolato "AI Privacy Risks & Mitigations – Large Language Models (LLMs)" (marzo 2025), costituisce un contributo di fondamentale importanza per giudare l'attività di sviluppatori, fornitori e utilizzatori di LLM nel rispetto delle previsioni del Regolamento (UE) 2016/679 (GDPR) e dell'AI Act.

Il documento non si limita a delineare in via teorica le principali categorie di rischio in materia di *privacy*, ma propone una metodologia operativa per l'identificazione,

la valutazione e la mitigazione dei rischi connessi ai LLM, adottando un approccio ispirato ai principi della ***data protection by design and by default***, della **minimizzazione** dei dati e della responsabilizzazione (***accountability***). Esso include, inoltre, una rassegna sistematica delle **misure tecniche e organizzative** suggerite per ridurre l'esposizione al rischio, e individua le implicazioni derivanti dalla distribuzione dei ruoli e delle responsabilità tra i diversi **attori** coinvolti nel ciclo di vita di un sistema AI.

1. RICONOSCIMENTO E CLASSIFICAZIONE DEI RISCHI PRIVACY NEGLI LLM

l documento dell'EDPB affronta in modo analitico la questione dei rischi *privacy* associati ai modelli linguistici di grandi dimensioni lungo l'intero ciclo di vita del sistema, articolando l'analisi su una pluralità di scenari d'uso e configurazioni tecniche. La prima operazione di rilievo consiste proprio nella tipizzazione sistematica delle principali categorie di rischio, che si presentano sia durante la fase di addestramento del modello (***training phase***), sia nel momento in cui il modello è impiegato per generare risposte a *input* ricevuti dagli utenti (***inference phase***), sia nei processi di aggiornamento iterativo basati sui *feedback* degli utenti (***fine-tuning / reinforcement learning***).

Tra i rischi più ricorrenti e significativi si segnalano:

- la memorizzazione non intenzionale di dati personali nei pesi del modello, che può condurre alla loro rievocazione in fase di inferenza (fenomeno noto come *data leakage* o *regurgitation*): questo rischio si manifesta con particolare evidenza nei modelli di grandi dimensioni addestrati su *dataset* generalisti non controllati, come nel caso delle fonti *web crawled* (es. *common crawl*), che possono contenere dati identificabili, comprese categorie particolari ai sensi dell'art. 9 GDPR;
- l'impossibilità di prevedere in modo deterministico gli *output* generati dal modello: trattandosi di sistemi basati su architetture neurali e meccanismi di attenzione distribuita (*transformer*), il comportamento del modello non è pienamente verificabile a priori, e ciò complica sia l'analisi di impatto (DPIA) sia l'individuazione preventiva dei rischi;
- il trattamento indiretto di dati personali tramite *prompt* inviati dall'utente: anche in assenza di una funzione esplicita di raccolta dati, l'interazione utente-modello può comportare il trattamento di informazioni personali fornite dall'interessato stesso o da terzi, configurando una responsabilità del deployer ai sensi dell'art. 4, par. 2 GDPR;
- la conservazione eccessiva dei dati di *input/output* da parte dei provider del modello, spesso per finalità di miglioramento del servizio, senza base giuridica o in assenza di una corretta informativa (violazione degli artt. 5,

13 e 14 GDPR): questo punto è rilevante nei modelli erogati in modalità LLM-as-a-Service, dove l'utente finale non ha visibilità né controllo sul trattamento dei dati all'interno dell'infrastruttura del *provider*;

- la mancata compatibilità tra le finalità del trattamento e l'uso secondario dei dati per il *fine-tuning*: in assenza di una base giuridica solida – ad esempio, un consenso valido o un interesse legittimo debitamente bilanciato – l'utilizzo dei dati raccolti durante le interazioni per addestrare nuovamente il modello può configurare una violazione del principio di limitazione della finalità (art. 5, par. 1, lett. b), GDPR);
- la difficoltà di garantire l'esercizio dei diritti da parte degli interessati, in particolare il diritto alla cancellazione, alla limitazione del trattamento e all'opposizione: una volta che i dati personali sono stati inglobati nel modello, anche attraverso processi di *embedding* e *tokenizzazione*, il loro isolamento risulta tecnicamente complesso; questo genera un'asimmetria tra l'obbligo giuridico (art. 17 GDPR) e le reali possibilità tecniche di adempimento, con conseguenze rilevanti anche sotto il profilo del principio di responsabilizzazione (art. 5, par. 2 GDPR);
- l'assenza di trasparenza sugli algoritmi e sulle logiche inferenziali impiegate: il documento evidenzia come gli LLM – per natura – tendano a funzionare come "scatole nere" (*black-box systems*), rendendo arduo per gli utenti e gli interessati comprendere il funzionamento del sistema, le fonti utilizzate, le ragioni di determinate risposte. Questo elemento può generare un conflitto diretto con l'art. 22 GDPR in caso di processi decisionali automatizzati, soprattutto laddove il modello sia utilizzato in ambito HR, assicurativo o creditizio.

Il merito del documento dell'EDPB risiede nel fatto che non si limita a descrivere i rischi, ma li contestualizza rispetto alle diverse configurazioni architetturali e contrattuali dei sistemi LLM, tenendo conto della crescente diffusione di modelli multimodali, di architetture agentiche e di processi di *orchestration* tra LLM e *Small Language Models* (SLM). Ogni evoluzione tecnologica, infatti, determina nuove superfici di attacco (*attack surfaces*) per la protezione dei dati e richiede una ricalibrazione del modello di rischio, tanto sotto il profilo probabilistico, quanto rispetto alla gravità del potenziale impatto.

Va infine sottolineato come il rischio *privacy* negli LLM non sia soltanto diretto, ma anche potenziale e futuro: molti dei rischi si manifestano a posteriori, nel tempo, man mano che i modelli vengono integrati in sistemi più complessi (es. agenti AI) o combinati con fonti di dati aziendali, CRM o *knowledge base* interne: in tale ottica, la gestione del rischio deve essere vista non come un'attività *una tantum*, ma come un processo dinamico e iterativo, che accompagna lo sviluppo e l'evoluzione dell'LLM nel tempo.

2. LA VALUTAZIONE DEL RISCHIO PRIVACY NEI LLM

Uno dei contributi più rilevanti del documento EDPB è l'elaborazione di una metodologia strutturata per la valutazione del rischio *privacy* in relazione ai modelli linguistici di grandi dimensioni. Tale metodologia si innesta pienamente nei principi generali di cui agli articoli 5, 24 e 32 del GDPR, e trova il proprio perno operativo nell'art. 35, relativo alla **valutazione d'impatto sulla protezione dei dati (DPIA)**.

Il documento evidenzia che, data la natura potenzialmente sistemica dei trattamenti effettuati tramite LLM, la DPIA non solo è raccomandabile, ma nella maggior parte dei casi è da ritenersi obbligatoria ai sensi della **"*blacklist*" del WP29 (Opinion 248/2017)**, fatta propria dall'EDPB, che include tra i trattamenti ad alto rischio quelli che comportano:

- l'uso innovativo o applicazione di nuove soluzioni tecnologiche;
- trattamenti su larga scala;
- trattamenti automatizzati con effetti significativi sull'interessato.

Tutti elementi che, a ben vedere, sono presenti in modo strutturale nei sistemi LLM, soprattutto quando adottati in contesti di interazione utente o di supporto a decisioni aziendali critiche.

1.1 Dalla identificazione alla stima del rischio: una matrice multidimensionale

La metodologia proposta si articola in tre livelli principali:

1) identificazione del rischio, mediante la mappatura dei flussi di dati lungo il ciclo di vita dell'LLM (raccolta, addestramento, inferenza, *logging*, *fine-tuning*, aggiornamento e dismissione): questo esercizio deve comprendere sia dati "a monte" (*training set*) sia dati "a valle" (*input/output* e *feedback* degli utenti). È fondamentale in questa fase definire chiaramente le categorie di dati, i contesti di utilizzo e i soggetti coinvolti (titolari, responsabili, fornitori terzi, utenti finali, ecc.);
2) stima della probabilità di accadimento del rischio, secondo criteri empirici che includono: la qualità delle fonti dati, la presenza di filtri o meccanismi di disidentificazione, il tipo di architettura usata (es. modelli *decoder-only* o *encoder-decoder*), la modalità di accesso al modello (*on-premise, SaaS, API*), la documentazione disponibile, l'eventuale certificazione o *audit* del fornitore;
3) valutazione della severità dell'impatto, che deve considerare non solo la natura dei dati potenzialmente coinvolti (es. dati sanitari, giudiziari, biometrici), ma anche il contesto di utilizzo del modello, il grado di autonomia decisionale del sistema e la possibilità concreta di re-identificare gli interessati.

Il risultato atteso di questa valutazione è una classificazione del rischio, che il documento propone di rappresentare in forma di **matrice** (basso/medio/alto, combinando probabilità e impatto), utile per orientare le scelte successive in materia di misure di mitigazione e accettazione del rischio residuo.

1.2 DPIA, accountability e ciclo di vita del modello

Una delle osservazioni più incisive del documento è che la DPIA, nel contesto degli LLM, non può essere considerata un adempimento puntuale e statico, ma deve evolvere come strumento di ***privacy governance*** **dinamica**. Ogni modifica del modello – sia essa architetturale (ad es. l'integrazione di moduli agentici o di SLM), sia funzionale (nuovi use case o sorgenti dati), sia organizzativa (cambio del fornitore, migrazione da *on-premise* a *SaaS*) – impone un aggiornamento della valutazione d'impatto.

In tal senso, l'EDPB invita a integrare la DPIA con logiche e strumenti propri del *risk management*, tra cui:

- revisione periodica dei controlli di sicurezza (art. 32 GDPR);
- monitoraggio dei *log* di sistema con finalità di *audit* e *detection* di anomalie;
- procedure documentate per la gestione del rischio residuo (accettazione, trasferimento, ulteriore mitigazione);
- coinvolgimento sistematico del DPO in fase di sviluppo, *testing* e deployment del sistema.

È utile ricordare, sul punto, la posizione costante del Garante per la protezione dei dati personali italiano, secondo cui "*laddove il sistema sia in grado di produrre effetti significativi sui diritti e le libertà delle persone fisiche, la DPIA assume carattere non solo obbligatorio, ma anche fondativo dell'intero impianto di accountability*" (cfr. Provv. 21.12.2023, n. 1012, su sistemi AI in ambito HR).

1.3 Integrazione tra DPIA e AI Risk Assessment

Il documento EDPB, infine, anticipa uno degli snodi più rilevanti dell'AI Act, ovvero l'obbligo – per i sistemi classificati ad alto rischio – di effettuare un *AI Risk Assessment ex* art. 9 e ss., distinto ma parzialmente sovrapponibile alla DPIA.

L'invito alle imprese è chiaro: dotarsi fin da ora di una metodologia unificata di valutazione del rischio, che soddisfi i requisiti di entrambi i *framework* normativi. In particolare, ciò implica:

- mappare il sistema AI non solo per finalità di protezione dati, ma anche in termini di sicurezza, robustezza, accuratezza e trasparenza;
- collegare ogni rischio privacy a un technical control verificabile (es. output filtering, policy di retention, metadati anonimizzati);
- predisporre evidenze documentali riutilizzabili in sede ispettiva, anche in chiave multi-normativa.

3. TITOLARI, RESPONSABILI E ALTRI RUOLI TRA GDPR E AI ACT

Una delle questioni più complesse nell'applicazione del GDPR ai sistemi basati su modelli linguistici di grandi dimensioni è l'identificazione dei soggetti che, a vario titolo, partecipano alla progettazione, sviluppo, erogazione, personalizzazione o utilizzo del modello. Il documento dell'EDPB affronta tale problematica con un approccio casistico, distinguendo i ruoli in base ai diversi modelli di erogazione del servizio e alla configurazione architetturale dei sistemi.

In particolare, si pongono almeno tre livelli di attori che possono assumere ruoli distinti (o talvolta sovrapposti) nella filiera del trattamento:

- il **fornitore del modello** (*model provider*): sviluppa l'LLM, lo addestra e lo mette a disposizione – tipicamente tramite API – per usi da parte di soggetti terzi;
- il ***deployer*** (utilizzatore integratore): configura e integra il modello nei propri sistemi (es. *chatbot* aziendali, *virtual assistant*), ne definisce lo scopo e ne controlla i parametri;
- **l'utente finale**: interagisce con il sistema attraverso *prompt*, eventualmente fornendo dati personali propri o altrui.

1.4 LLM-as-a-Service: co-titolarità o responsabilità delegata?

Nel caso più frequente oggi sul mercato – l'utilizzo di un modello in modalità LLM-as-a-Service (es. GPT-4 o Claude integrati via API in un'applicazione web o mobile) – la domanda principale è se il *provider* sia da considerarsi **titolare** autonomo, **responsabile** del trattamento oppure **co-titolare** insieme al *deployer*.

Il documento EDPB riconosce che la risposta dipende da una valutazione caso per caso, tenendo conto dei seguenti fattori:

- chi determina le **finalità** del trattamento? Se è il *deployer* a definire lo scopo dell'uso del modello (es. assistenza clienti, profilazione, assistenza medica), sarà normalmente titolare;
- chi decide le **modalità** del trattamento? Se il *provider* esercita un controllo autonomo sulle modalità operative (ad es. conservazione dei *prompt*, uso dei dati per il *fine-tuning*, logica inferenziale), egli sarà da qualificare come titolare autonomo, anche se "a valle";
- esiste una **clausola contrattuale** che definisca i ruoli? L'EDPB precisa che le clausole contrattuali sono rilevanti ma non determinanti: occorre guardare alla sostanza dei rapporti.

Nella prassi, emerge sempre più spesso una configurazione mista, in cui:

- il *deployer* è titolare per il trattamento dei dati degli utenti tramite il proprio servizio;

- il *provider* è titolare per l'uso autonomo dei dati per addestramento, *logging*, miglioramento dei modelli;
- entrambi assumono obblighi informativi e contrattuali (ex art. 26 GDPR) se il trattamento è congiunto o interdipendente.

In tale prospettiva, la qualificazione dei ruoli non è solo una questione di diritto, ma anche di *data governance*: occorre stabilire chi raccoglie i dati, chi li conserva, chi ne regola l'uso secondario, chi risponde agli interessati.

1.5 Responsabilità nei modelli agentici e nei sistemi orchestrati

Un'area particolarmente innovativa – e ancora in parte inesplorata dal legislatore – è quella degli agenti AI: sistemi autonomi capaci di interagire con ambienti esterni, prendere decisioni, pianificare e agire secondo obiettivi prefissati, in cui il rischio di decentramento del controllo è massimo.

Nel documento EDPB si sottolinea che, nei sistemi agentici basati su LLM, la responsabilità deve essere ricondotta all'organizzazione che orchestra il comportamento dell'agente, configurandone il perimetro d'azione, le regole decisionali e le interfacce esterne. Tale soggetto dovrà (i) predisporre misure di sicurezza adeguate (art. 32 GDPR); (ii) valutare il rischio sistemico derivante da decisioni automatizzate (art. 22), e (iii) definire *ex ante* i canali di *escalation* e supervisione umana.

L'AI Act, all'art. 28 e ss., introduce una nuova categoria soggettiva – il "deployer" del sistema AI – che può coincidere con il titolare *ex* GDPR, ma anche assumere obblighi propri (audit, monitoraggio, registrazione eventi). È evidente quindi la necessità di armonizzare le due discipline in un sistema coerente di allocazione delle responsabilità.

1.6 Soluzioni (contrattuali e operative)

Alla luce di quanto sopra, il documento EDPB raccomanda di adottare strumenti contrattuali e organizzativi volti a:

- definire chiaramente i ruoli *privacy* nelle relazioni *provider–deployer*, anche con referimento ai modelli standard EDPB (es. Data Processing Agreement, Joint Controller Agreement);
- mappare i trattamenti nei rigistri;
- introdurre obblighi reciproci di notifica in caso di *data breach* o richiesta da parte dell'interessato;
- stabilire regole per la conservazione, l'anonimizzazione o la pseudonimizzazione dei *prompt*;
- prevedere un registro condiviso delle interazioni AI (*ex* art. 29 AI Act) e un sistema di tracciabilità delle decisioni automatizzate.

In sintesi, nel mondo dell'AI generativa non è più sufficiente identificare "chi è il titolare", ma occorre strutturare una **rete di *accountability* condivisa**, con obblighi espliciti, verificabili e proporzionati al rischio.

2. LE MISURE DI MITIGAZIONE DEL RISCHIO

Nel documento "AI Privacy Risks & Mitigations – LLMs", il Support Pool of Experts dell'EDPB dedica ampio spazio all'individuazione di misure concrete e proporzionate per la mitigazione dei rischi *privacy*, calibrate sulle specificità dei modelli linguistici di grandi dimensioni. L'approccio è pienamente in linea con il principio di *privacy by design and by default* (art. 25 GDPR), ma si estende anche alle logiche di *AI governance* previste dall'AI Act, con particolare riferimento agli articoli 9, 17, 28 e 29 del Regolamento europeo.

2.1 Misure tecniche: dal filtering all'*unlearning*

Le misure tecniche rappresentano la prima linea di difesa contro i rischi derivanti da trattamento illecito, *leak* o esposizione involontaria di dati personali. Tra le più efficaci, e già applicate da diversi fornitori nel mercato, si segnalano:

- ***data filtering* e *curation* nei *dataset* di addestramento**: consiste nella selezione preventiva di fonti di dati "*low risk*" (es. documentazione tecnica, fonti *open-access* con licenze appropriate), e nella rimozione di contenuti che possano includere informazioni personali, come indirizzi email, codici fiscali, testi provenienti da *social network* o archivi giudiziari non anonimizzati;
- ***prompt and output filtering***: si tratta dell'implementazione di filtri automatici che rilevano (e bloccano) l'inserimento nei *prompt* di dati potenzialmente personali o sensibili, così come la generazione di *output* che includano *pattern* identificabili; i filtri possono operare a livello lessicale, semantico o sintattico, ed essere integrati con tecnologie NLP di classificazione o Named Entity Recognition (NER);
- ***machine unlearning***: l'EDPB valorizza l'impiego di tecniche che consentano al modello di "dimenticare" selettivamente dati personali già appresi, su richiesta dell'interessato. Sebbene l'efficacia di tali soluzioni sia ancora limitata, si stanno rapidamente evolvendo approcci come SISA (*Sharded, Isolated, Sliced and Aggregated Training*) e metodi basati su distillazione *o low-impact re-training*;
- ***retrospective logging minimization***: consiste nel ridurre al minimo i dati registrati nei *log* di sistema, privilegiando identificatori aggregati o pseudonimi, e configurando periodi di conservazione brevi. Questo risponde sia all'art. 5, par. 1, lett. e) GDPR (limitazione della

conservazione), sia all'art. 29 AI Act (registrazione degli eventi e tracciabilità).

- *Retrieval-Augmented Generation* (RAG): questa architettura separa la conoscenza aggiornata del sistema dalla "memoria" statica dell'LLM, impiegando una base dati esterna interrogabile in tempo reale; in questo modo, si riduce il rischio che il modello memorizzi dati personali strutturalmente, consentendo una più agile gestione dei diritti ex artt. 15-22 GDPR.

2.2 Misure organizzative: accountability, processi e controlli interni

Oltre alle misure tecniche, il documento sottolinea l'importanza di interventi di natura organizzativa, volti a garantire una governance del rischio coerente e documentabile. Tra le più rilevanti:

- definizione di *policy* interne per l'uso degli LLM: ogni organizzazione che adotti o integri LLM dovrebbe dotarsi di una policy dedicata che disciplini modalità d'uso, autorizzazioni, limitazioni operative e casi d'uso vietati;
- formazione del personale: è cruciale che chi interagisce con LLM (es. *customer care, marketing, HR*) riceva una formazione specifica sui rischi *privacy* connessi, sulle buone pratiche per la gestione dei *prompt*, e sull'obbligo di evitare l'inserimento di dati personali senza autorizzazione;
- audit periodici sul comportamento del modello: è raccomandabile l'adozione di programmi di *red-teaming* o *adversarial testing*, che consentano di verificare se l'LLM sia suscettibile di rivelare informazioni apprese durante il *training*. Tali audit devono essere documentati, ripetibili e integrati nei processi di *risk management* aziendale;
- registro delle interazioni: sebbene non obbligatorio ai sensi del GDPR, il mantenimento dei *log* delle interazioni (eventualmente anonimizzati) può rivelarsi utile per individuare abusi, migliorare la qualità del sistema e tracciare eventuali violazioni;
- designazione del DPO e coinvolgimento nei processi AI: il DPO deve essere coinvolto sin dalla fase di scelta e integrazione dell'LLM, con particolare attenzione alla DPIA, alle informative rese agli utenti e alla risposta alle richieste di esercizio dei diritti.

2.3 Misure contrattuali: la gestione del rischio nel rapporto provider-deployer

Il documento dell'EDPB sottolinea come, nel rapporto tra *provider* e *deployer* di LLM, le misure contrattuali rivestano un ruolo cruciale per garantire una gestione responsabile dei rischi *privacy*. È fondamentale, anzitutto, formalizzare in modo chiaro i ruoli delle parti: se il *provider* agisce come responsabile del trattamento, è necessario un *Data Processing Agreement* ai sensi dell'art. 28 GDPR; in caso di

contitolarità, occorre un accordo ex art. 26 che disciplini finalità, responsabilità e modalità operative.

Particolare attenzione va riservata all'inserimento di clausole che regolino la conservazione dei *prompt*, l'uso dei dati per il *fine-tuning* e l'eventuale riutilizzo dei *feedback* dell'utente, prevedendo limiti chiari e verificabili. È inoltre raccomandabile inserire clausole di *audit* a favore del *deployer*, e richiedere al provider attestazioni di conformità o certificazioni, come la ISO/IEC 42001 o i report SOC 2.

Queste previsioni contrattuali, se ben strutturate, contribuiscono a rendere effettivo il principio di *accountability* e ad assicurare che l'utilizzo degli LLM si inserisca in un contesto di governance del rischio solido e documentabile.

3. VERSO UN'AI GENERATIVA SOSTENIBILE, TRASPARENTE E CONFORME

Il documento dell'EDPB fornisce un importante riferimento europeo per valutare e mitigare i rischi *privacy* associati ai LLM, ma non può essere letto isolatamente. Anche al di fuori dell'Unione, infatti, Stati Uniti, Regno Unito e Canada stanno costruendo approcci distinti – e in parte complementari – per regolamentare l'uso responsabile dell'intelligenza artificiale generativa.

Negli **Stati Uniti**, l'Executive Order sull'AI ha avviato un percorso normativo più strutturato, rafforzato dal ruolo attivo della FTC e dal framework volontario del NIST; l'attenzione si concentra soprattutto su sicurezza, trasparenza e supervisione umana. Nel **Regno Unito**, invece, si è scelto un approccio graduale e pro-innovazione, fondato su cinque principi generali e sulla guida delle autorità di settore, con l'ICO in prima linea nel fornire indicazioni pratiche. Il **Canada**, infine, si prepara all'introduzione dell'AIDA, mentre già oggi le autorità di vigilanza pubblicano raccomandazioni e conducono indagini sull'uso degli LLM.

In questo contesto in evoluzione, anche gli operatori europei devono attrezzarsi, adottando da subito pratiche solide di responsabilità e prevenzione. Ai **fornitori** di LLM si raccomanda di documentare accuratamente il ciclo di vita del modello, applicare misure efficaci di filtering, valutare tecniche di unlearning e rafforzare i meccanismi di audit. Le **imprese** che li adottano dovrebbero dotarsi di policy interne chiare, valutare l'opportunità di una DPIA e formare il personale, limitando l'inserimento di dati personali nei prompt. Quanto ai **DPO** e ai **consulenti**, è essenziale guidare queste scelte con visione strategica, integrando *governance privacy* e gestione del rischio AI.

L'esperienza internazionale dimostra che il valore dei LLM può realizzarsi appieno solo se sostenuto da un impianto giuridico e organizzativo robusto. In definitiva, la sfida per gli operatori oggi non è se conformarsi alle norme, ma come farlo bene, anticipando i rischi e valorizzando la compliance come leva di fiducia e

competitività; perché solo un'intelligenza artificiale rispettosa dei diritti sarà davvero utile, affidabile e pronta per il futuro.

AUTORI

Aurora Agostini

Partner

Giulietta Minucci

Counsel

Giovanni Lombardi

Associate

Alessandro Carlini

Associate