Università degli Studi di Padova

DIPARTIMENTO DI DIRITTO PUBBLICO, INTERNAZIONALE E COMUNITARIO

CORSO DI LAUREA IN DIRITTO E TECNOLOGIA

ANNO ACCADEMICO 2022-2023

Titolo tesi:

Diritto all'Oblio e Large Language Models: la necessità di un equilibrio tra privacy e innovazione

Relatore: Prof. Andrea Pin

Laureando: Luca Bona

INDICE

Introduzione

Questo elaborato offre un'analisi delle complesse intersezioni tra il diritto all'oblio, i Large Language Models (LLM) e la normativa sull'intelligenza artificiale (AI) in ambito europeo, anche in vista dell'approvazione dell'Artificial Intelligence Act. Nella prima parte, sono state esaminate le definizioni e le linee guida europee relative al diritto all'oblio, delineando il contesto normativo che ha ispirato le successive indagini. Successivamente, l'attenzione si sposta ai Large Language Models, investigando la loro natura, il funzionamento e le sfide intrinseche legate al diritto all'oblio. Particolare attenzione è stata riservata alla distinzione cruciale tra i LLM e i motori di ricerca, con un'analisi dettagliata delle implicazioni del General Data Protection Regulation (GDPR) nell'applicazione di normative sulla privacy ai Large Language Models. In ultima battuta, la tesi affronta l'ambito più ampio dell'intelligenza artificiale, esaminando l'Artificial Intelligence Act come strumento normativo chiave nell'Unione europea. Sono state identificate e analizzate le sfide specifiche legate all'applicazione di questo atto normativo ai sistemi di intelligenza artificiale, considerando le complessità etiche e giuridiche connesse alla regolamentazione di tecnologie avanzate. In conclusione, questo studio fornisce una prospettiva critica e approfondita sulle questioni giuridiche ed etiche riguardanti il diritto all'oblio nei Large Language Models nel contesto dell'Artificial Intelligence Act.

Capitolo 1

Il Concetto di "Diritto all'Oblio"

Questo capitolo introduce il tema del Diritto all'Oblio e le rispettive Linee guida europee

1.1 Definizione

Il "diritto all'oblio" è una parte del panorama legale in evoluzione in relazione alla privacy online. È un concetto giuridico che ha acquisito notorietà negli ultimi anni, in risposta all'esponenziale evoluzione delle tecnologie e alla sempre più ampia condivisione di informazioni personali online. Il "diritto all'oblio" rappresenta il diritto inalienabile di ciascun cittadino di richiedere la rimozione, ma anche l'aggiornamento o la modifica di informazioni che lo riguardano direttamente, specialmente quando tali informazioni incidono sulla sua reputazione. Essendo quest'ultima uno dei diritti inviolabili dell'essere umano, è riconosciuta e tutelata dalla nostra Costituzione (art.2). È legittimato per tanto il diritto di ogni individuo di essere dimenticato e di non essere più ricordato per eventi del passato che ormai appartengono alla sua sfera privata[1].

Nel 2018 l'Unione europea (UE) ha adottato il Regolamento Generale sulla Protezione dei Dati (GDPR); l'articolo 17 stabilisce un "diritto all'oblio" simile al diritto riconosciuto dalla Corte di giustizia dell'Unione europea (CGUE) sotto la legge precedente che il GDPR ha sostituito. Questo diritto è stato riconosciuto in varie giurisdizioni in tutto il mondo come una forma di bilanciamento tra la libertà di espressione e l'interesse pubblico[2].

Tuttavia, il "diritto all'oblio" non è assoluto e spesso coinvolge complessi equilibri giuridici. Nel 2014, con il caso Google Spain SL e Inc. contro AEPD, Mario Costeja

[1] La Costituzione, Principi fondamentali, articolo 2; https://www.ildirittoalloblio.it/#discover
[2] https://support.google.com/legal/answer/10769224?hl=en

González[3], cercando il proprio nome sul motore di ricerca, il soggetto scopre che un annuncio riguardante aste fallimentari e debiti propri passati era stato indicizzato da Google. González chiede di rimuovere i collegamenti a questa informazione dai risultati di ricerca, temendo conseguenze per la sua reputazione. Google respinge la richiesta sostenendo che l'azienda non aveva alcun controllo sull'informazione in questione e che aveva il diritto di mostrare i risultati in modo oggettivo. La Corte di giustizia dell'Unione europea ha stabilito che i motori di ricerca sono responsabili del trattamento dei dati personali e devono rispettare la legislazione europea sulla protezione dei dati. La CGUE ha stabilito che Google deve rimuovere o adeguare i collegamenti ai dati personali se tali sono ritenuti obsoleti o non rilevanti, purché non vi siano ragioni legittime per la loro inclusione. La Corte afferma che questo diritto di cancellazione o modifica dei dati deve essere bilanciato con il diritto alla libertà di espressione e di informazione. Inoltre, il "diritto all'oblio" può prevalere sull'interesse pubblico in determinate circostanze, ma ogni caso deve essere valutato individualmente e spetta ai motori di ricerca fare questa valutazione[4].

Dal maggio 2014 ad oggi le richieste di rimozione degli URL da Ricerca Google per privacy sono aumentate e con esse però anche il numero di quelle approvate. Si può constatare dunque che da un lato il tema del diritto all'oblio è sempre più diffuso, dall'altro Google si è impegnata in termini di accessibilità alla privacy e precisione di intervento nei singoli casi[5].

1.2 Linee guida europee sul "Diritto all'Oblio"

Secondo quanto enunciato nella sentenza della Corte di giustizia dell'Unione europea, in considerazione della seria minaccia che l'elaborazione di dati personali rappresenta per i diritti fondamentali alla privacy e alla protezione dei dati, i diritti della persona interessata devono, in generale, avere la precedenza rispetto all'interesse economico dei motori di ricerca e alla volontà degli utenti di Internet di accedere a dati personali tramite essi. Tuttavia, è necessario effettuare una ponderazione tra i diritti e gli interessi in gioco, il cui esito può variare in base alla natura e alla sensibilità dei dati coinvolti, nonché

[3] https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A62012CJ0131
[4] https://www.osservatoriosullefonti.it/archivio-rubriche-2014/fonti-dellunione-europea-e-internazionali/986-ue-corte-di-giustizia-causa-c-13112-google-spain
[5] https://transparencyreport.google.com/eu-privacy/overview

all'interesse del pubblico a ottenere determinate informazioni. In particolare, l'interesse del pubblico potrebbe essere considerevolmente più rilevante quando la persona interessata ricopre un ruolo di rilievo nella sfera pubblica[6].

La sentenza stabilisce che il diritto alla cancellazione si applica solo ai risultati delle ricerche basate sul nome di una persona e non richiede la rimozione completa del link dagli indici del motore di ricerca. Quindi, se sono soddisfatte le condizioni di cui agli Articoli 12 e 14 della Direttiva 95/46/CE, gli utenti possono ottenere la deindicizzazione di link a pagine web pubblicate da terzi contenenti informazioni che li riguardano dall'elenco dei risultati che compaiono digitando appunto il nome di una persona. In altre parole, l'informazione originale rimarrà accessibile tramite altre chiavi di ricerca o accedendo direttamente alla fonte originale dell'editore[7].

Per quanto riguarda le modalità di esercizio del diritto all'oblio, gli utenti non sono tenuti a contattare il sito web originale. Possono rivolgersi infatti ad uno o più motori di ricerca presentando un'istanza. Così facendo, è il soggetto stesso a decidere le misure sufficienti a ridurre o eliminare gli effetti dell'indicizzazione delle informazioni in questione. La Direttiva 95/46/CE, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, non tratta di procedure specifiche per l'esercizio dei diritti in essa riconosciuti, ma suggerisce che vi sono diversi metodi di presentazione dell'istanza e offre una certa flessibilità in merito. Di conseguenza, i motori di ricerca devono seguire le normative nazionali in merito alla protezione dei dati riguardo alle istanze presentate e fornire risposte adeguate. Se una richiesta viene respinta, l'utente ha il diritto di ricevere spiegazioni e di rivolgersi all'autorità di protezione dei dati (ADP) o all'autorità giudiziaria[8].

Un altro punto molto importane è quello della "portata" dell'intervento. La sentenza del "caso Google Spain SL e Inc. contro AEPD, Mario Costeja González" (*cap. 1.1*) riguarda principalmente i motori di ricerca generalisti, ma può essere applicata ad altri intermediari. I motori di ricerca interni ai siti web sono differenti da quelli esterni, in

---

[6] Linee-guida sull'attuazione della sentenza della Corte di Giustizia Dell'Unione europea nel caso c-131/12 "Google Spain e Inc. Contro Agencia Española de Protección de Datos (AEPD) e Mario Costeja González" Parte I, punto A, paragrafo 5.

[7] Linee-guida sull'attuazione della sentenza della Corte di Giustizia Dell'Unione europea nel caso c-131/12 "Google Spain e Inc. Contro Agencia Española de Protección de Datos (AEPD) e Mario Costeja González" Parte I, punto A, paragrafi 6-9.

[8] Linee-guida sull'attuazione della sentenza della Corte di Giustizia Dell'Unione europea nel caso c-131/12 "Google Spain e Inc. Contro Agencia Española de Protección de Datos (AEPD) e Mario Costeja González" Parte I, punto B;
Direttiva 95/46/CE https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX%3A31995L0046

quanto reperiscono solo informazioni contenute in specifiche pagine web e non creano un profilo completo delle persone cercate. Di conseguenza, il diritto alla deindicizzazione di solito non si applica a motori di ricerca con una portata limitata, come quelli interni ai siti di notizie online. Quindi, affinché i diritti dell'individuo, conformemente a quanto stabilito dalla Corte, siano pienamente garantiti, è essenziale attuare le decisioni di deindicizzazione in modo tale da assicurare una tutela completa ed efficace dei diritti della persona interessata, evitando qualsiasi tentativo di eludere le normative dell'Unione europea. In questo contesto, limitare la deindicizzazione ai domini dell'UE sulla base dell'usanza degli utenti di accedere ai motori di ricerca dai rispettivi domini nazionali, non è ritenuto sufficiente per garantire adeguatamente i diritti degli interessati, come specificato dalla sentenza. La deindicizzazione, quindi, deve essere applicata anche su tutti i domini internazionali rilevanti, ad esempio il dominio ".com" [9].

Capitolo 2

Large Language Models

Questo capitolo approfondisce i Large Language Models, le problematiche generali e le sfide derivanti dall'applicazione del GDPR

2.1 Cosa sono?

I Large Language Models (abbreviati con LLM) rappresentano una categoria di modelli di intelligenza artificiale noti come "foundation models" o modelli di base. Questi modelli di base sono sottoposti a una fase di pre-addestramento, in cui vengono esposti a grandi quantità di dati non etichettati e ad auto-supervisione. Questo processo consente ai

[9] Linee-guida sull'attuazione della sentenza della Corte di Giustizia Dell'Unione europea nel caso c-131/12 "Google Spain e Inc. Contro Agencia Española de Protección de Datos (AEPD) e Mario Costeja González" Parte I, punto C.

modelli di apprendere da schemi intrinseci presenti nei dati, al fine di generare un output generalizzabile e adattabile a una vasta gamma di applicazioni. Gli LLM sono particolarmente mirati all'elaborazione di testo scritto o dati simili, come il codice informatico. Per il loro addestramento, attingono a enormi dataset di testo, quali libri, articoli e conversazioni. L'aggettivo "Large" si riferisce alle dimensioni impressionanti di questi modelli, che possono occupare decine di gigabyte in termini di spazio di archiviazione e sono addestrati su quantità di dati potenzialmente ragguardevoli, arrivando a livelli dell'ordine dei petabyte (10^{15} byte). Per avere un'idea delle proporzioni, un file di testo di un gigabyte (10^{9} byte) può contenere circa 178 milioni di parole, mentre un petabyte equivale a circa un milione di gigabyte. Oltre alle dimensioni impressionanti, gli LLM si distinguono anche per il loro elevato numero di parametri. Un parametro rappresenta un valore che il modello può regolare in modo indipendente durante il processo di apprendimento. In generale, più parametri ha un modello, maggiore è la sua complessità. Ad esempio, GPT-3.5, uno dei LLM più noti, è pre-addestrato su un vasto corpus di dati che raggiunge i 45 terabyte (10^{12} byte) e utilizza l'incredibile numero di 175 miliardi di parametri di *machine learning.*[10]

2.2 Come funzionano?

Nel contesto dei Large Language Models, tre elementi principali svolgono un ruolo chiave: i dati, l'architettura e l'addestramento. Il ruolo che gli enormi volumi di dati testuali svolgono in questo contesto è stato esaminato al *paragrafo 2.1*. Per quanto riguarda l'aspetto architetturale, è importante notare che i LLM si basano su una specifica architettura di rete neurale, nota come "transformer" o trasformatore. Questa architettura è fondamentale per consentire al modello di gestire sequenze di dati, come frasi o porzioni di codice. All'interno di un trasformatore si distinguono due componenti fondamentali: il codificatore e il decodificatore. Il compito del codificatore è elaborare la sequenza di input, mentre il decodificatore si concentra sulla generazione della sequenza di output desiderata. L'operatività dei trasformatori è basata sull'apprendimento sequenza-sequenza, in cui il modello riceve una sequenza di *token*, come le parole in una frase, e

[10] https://youtu.be/5sLYAQS9sWQ
https://www.nvidia.com/en-us/glossary/data-science/large-language-models/#:~:text=Large%20language%20models%20(LLMs)%20are,content%20using%20very%20large%20datasets

predice la parola successiva nella sequenza di output. Questo processo comporta iterazioni attraverso vari livelli del codificatore, consentendogli di generare rappresentazioni che indicano quali parti delle sequenze di input sono strettamente correlate. Queste rappresentazioni vengono poi trasmesse al livello successivo. Successivamente, il decodificatore sfrutta queste rappresentazioni per generare la sequenza di output desiderata.[11]

A differenza dei modelli basati su Reti Neurali Ricorrenti (RNN), i trasformatori si avvalgono di un meccanismo chiamato "self-attention" o "attenzione". Questo meccanismo fornisce un contesto che evidenzia le relazioni tra gli elementi nella sequenza di input, contribuendo a identificare il contesto che attribuisce significato a ciascuna parola all'interno della sequenza. I trasformatori sono in grado di elaborare più sequenze in parallelo, il che accelera notevolmente i tempi di addestramento. Questo approccio consente di costruire una comprensione dettagliata della struttura delle frasi e del significato delle parole al loro interno.[12]

Infine, l'architettura del LLM viene sottoposta a un intenso processo di addestramento su una vasta quantità di dati. Durante questa fase di addestramento il modello si allena a prevedere la parola successiva in una frase. Inizialmente, effettua previsioni casuali, ad esempio, iniziando con una frase come "la mela è una macchina". Tuttavia, attraverso iterazioni successive, il modello regola i suoi parametri interni per ridurre la discrepanza tra le sue previsioni e i risultati reali. Questo continuo processo di affinamento consente al modello di migliorare progressivamente le sue capacità di generare frasi coerenti e significative, arrivando quindi alla frase corretta "la mela è un frutto". È inoltre possibile eseguire un ulteriore affinamento del modello su dataset più piccoli e specifici, al fine di perfezionare ulteriormente la sua capacità di eseguire determinati compiti con maggiore precisione. Questa fase di affinamento è ciò che consente a un modello di linguaggio generale di diventare esperto in un dominio particolare e viene definita "fine tuning".[13]

[11] https://www.elastic.co/what-is/large-language-models
[12] https://cloud.google.com/ai/llms
[13] https://youtu.be/ZXiruGOCn9s

2.3 Problemi e sfide dei Large Language Models

I Large Language Models, pur dando l'illusione di comprendere il significato e di poter rispondere con precisione, conservano la loro natura di strumento tecnologico e affrontano una serie di sfide significative.

Allucinazione: si verifica il fenomeno chiamato "allucinazione" quando un LLM produce una risposta errata o che non è in sintonia con l'intenzione dell'utente. Ad esempio, potrebbe erroneamente sostenere di possedere sentimenti umani o di essere innamorato dell'utente. Questo accade perché i LLM predicono la prossima parola o frase sulla base della grammatica e della sintassi, ma non sono in grado di comprenderne il significato umano.

Sicurezza: i LLM presentano gravi rischi per la sicurezza quando non vengono adeguatamente gestiti o monitorati. Possono divulgare informazioni private delle persone, partecipare a truffe di *phishing* e generare spam. Alcuni utenti malintenzionati riescono a riconfigurare l'intelligenza artificiale secondo le proprie convinzioni o pregiudizi e contribuire alla diffusione di disinformazione. Le conseguenze di ciò possono avere un impatto devastante a livello globale.

Bias: le uscite prodotte da un grande modello di linguaggio sono influenzate dai dati utilizzati per il suo addestramento. Se questi dati rappresentano principalmente un gruppo demografico o mancano di varietà, le risposte del modello saranno influenzate da tale limitazione, mostrando a loro volta una mancanza di diversità.[14]

Deployment: l'impiego di LLM richiede competenze tecniche avanzate, poiché coinvolge il *deep learning*, l'impiego di "transformer models", software e hardware distribuiti.

Consenso: i LLM vengono addestrati su vasti e molteplici dataset, alcuni dei quali potrebbero essere stati raccolti e creati senza il consenso dei soggetti coinvolti. Quando vengono estratti dati da Internet, è noto che questi modelli possano ignorare le restrizioni del copyright, plagino contenuti scritti e riutilizzino contenuti protetti senza ottenere il permesso dai legittimi proprietari o dagli artisti originali. Inoltre, non vi è alcun tracciamento dell'origine dei dati quando il modello produce risultati, spesso senza riconoscere o dare credito ai creatori, esponendo gli utenti a potenziali problemi di violazione del copyright.

[14] https://www.elastic.co/what-is/large-language-models#limitations-and-challenges-of-large-language-models

Privacy: i LLM possono anche estrarre dati personali, come nomi di soggetti o fotografi dalle descrizioni di immagini, mettendo a rischio la privacy delle persone. In alcuni casi, ciò ha portato a controversie legali.
Memorizzazione dei Dati di Addestramento: è stato notato che i LLM sono in grado di memorizzare informazioni personali, le quali possono successivamente emergere nelle loro risposte. Anche se le aziende si adoperano per eliminare dati personali dai dataset utilizzati nell'addestramento, esiste ancora la possibilità che tali dati siano presenti all'interno del materiale di addestramento.[15]

2.4 Large Language Models e motori di ricerca

I Large Language Models (LLM) e i motori di ricerca sono entrambi strumenti che permettono di accedere a informazioni online, ma hanno scopi e funzionalità diverse. Mentre i LLM si concentrano sul comprendere e generare testo, i motori di ricerca si focalizzano sull'indicizzazione, il recupero di pagine web e risorse online in base alle richieste degli utenti.

Differenze:

Scopo principale:

- LLM: i Large Language Models, come GPT-3.5, sono modelli di linguaggio addestrati a comprendere e generare testo in linguaggio naturale. Sono spesso utilizzati per rispondere a domande, completare testi, tradurre lingue e altre attività linguistiche.
- Motori di ricerca: i motori di ricerca, come Google, Bing e Yahoo, sono progettati per trovare pagine web, documenti o risorse online in base a parole chiave o frasi di ricerca. Il loro scopo principale è aiutare gli utenti a trovare informazioni specifiche su Internet.[16]

[15] https://www.elastic.co/what-is/large-language-models#limitations-and-challenges-of-large-language-models
[16] https://www.elastic.co/what-is/large-language-models#large-language-models-use-cases

Funzionalità:

- LLM: con essi si può ottenere testo generico in risposta a domande o input specifici, ma non hanno la capacità di navigare o indicizzare pagine web o risorse online.
- Motori di ricerca: esaminano e indicizzano le pagine web, in modo da poter restituire elenchi di risultati pertinenti quando gli utenti cercano informazioni specifiche.

Tipo di contenuto:

- LLM: generano principalmente testo basato su input testuali e possono rispondere a domande o fornire spiegazioni in linguaggio naturale.
- Motori di ricerca: forniscono principalmente collegamenti ipertestuali a pagine web, immagini, video e altri tipi di contenuto online.

Similitudini:

Natural Language Processing (NLP):

- Il NPL rappresenta una branca dell'intelligenza artificiale che si dedica all'analisi delle modalità con cui computer e individui possono interagire attraverso l'uso del linguaggio umano. I computer quindi analizzano, comprendono e rispondono attraverso i nostri mezzi di comunicazione, quali il linguaggio orale e il testo scritto. Entrambi i LLM e i motori di ricerca utilizzano il NPL per comprendere e generare testo in linguaggio naturale. I motori di ricerca utilizzano il NLP per interpretare le richiesta degli utenti e restituire risultati pertinenti, mentre i LLM possono generare testo in risposta a domande o input testuali.[17]

Accesso a informazioni online:

- Sia i LLM che i motori di ricerca consentono agli utenti di accedere a informazioni online, anche se lo fanno in modi diversi. I LLM forniscono risposte testuali, mentre i motori di ricerca forniscono collegamenti a pagine web e altre risorse.

Intelligenza artificiale:

- I LLM sono alimentati da modelli di intelligenza artificiale avanzati, mentre i motori di ricerca utilizzano algoritmi complessi e tecnologie avanzate per indicizzare e classificare i contenuti online. Nell'ultimo

[17] https://www.elastic.co/what-is/natural-language-processing

periodo però, molti browser tra cui *Opera* stanno implementando le intelligenze artificiali come supporto nativo alla normale navigazione web, senza quindi che l'utente debba ricorrere all'uso di estensioni di terze parti.

La comprensione delle interconnessioni tra questi due strumenti agevola l'analisi delle potenziali problematiche, in particolare dal punto di vista legale, riguardo alla privacy dei dati personali. Entrambi i Large Language Models e i motori di ricerca hanno la capacità di raccogliere e sfruttare dati personali degli utenti. Nel caso dei motori di ricerca, ciò si manifesta attraverso la memorizzazione e l'utilizzo delle ricerche, migliorando la pertinenza dei risultati. D'altra parte, i LLM, quando sottoposti ad addestramento su testi contenenti dati sensibili, potrebbero assimilare e mantenere queste informazioni. Inoltre, questi dati possono essere impiegati dai motori di ricerca per scopi di profilazione, ovvero l'analisi del comportamento online delle persone al fine di proporre loro annunci mirati. Questo solleva legittime preoccupazioni in merito alla privacy, poiché gli utenti potrebbero non essere adeguatamente informati riguardo al processo di profilazione e all'uso dei loro dati personali. Oltretutto le informazioni personali possono essere conservate all'interno delle pagine web, dei documenti e dei dati di input su cui operano sia i motori di ricerca che i LLM, e ciò potrebbe comportare il rischio di esposizione o accesso da parte di terzi non autorizzati, a meno che non siano adottate adeguate misure di sicurezza.[18]

Pertanto, apprendere il funzionamento di LLM e motori di ricerca riveste un'importanza fondamentale nell'ambito dell'educazione degli utenti sulle migliori pratiche per salvaguardare la loro privacy online. Inoltre, la comprensione di questi meccanismi di raccolta e utilizzo dei dati da parte di LLM e motori di ricerca rappresenta una tappa cruciale nella formulazione di regolamentazioni sulla privacy, finalizzate a garantire che le aziende rispettino le leggi sulla protezione dei dati e preservino la privacy degli utenti.[19]

[18] https://www.csiro.au/en/news/all/articles/2023/september/your-right-to-be-forgotten-ai

[19] https://montrealethics.ai/right-to-be-forgotten-in-the-era-of-large-language-models-implications-challenges-and-solutions/#:~:text=RTBF%20may%20apply%20in%20two,even%20extend%20this%20difficulty%20further.

2.5 Il GDPR nei Large Language Models

Come visto nei paragrafi precedenti, i dati posso essere forniti dagli utenti stessi o possono esistere all'interno dei modelli. In ognuno di questi casi, secondo il GDPR, gli utenti hanno il diritto di richiedere l'accesso, la rettifica e la cancellazione di questi dati (articoli 15,16 e 17).[20]

In più, il GDPR stabilisce che l'utente debba concedere il proprio consenso al trattamento dei dati in modo informato ed esplicito. Tuttavia, il processo per ottenere il consenso durante le interazioni con gli LLM può risultare ambiguo, dal momento che gli utenti potrebbero non essere completamente consapevoli di come i propri dati vengano impiegati. L'articolo 9 del GDPR sancisce la protezione di tutti i dati ritenuti sensibili, imponendo restrizioni per quanto concerne il loro trattamento.[21]

In particolare, casi abbastanza frequenti riguardano l'ambito sanitario: i LLM sono risultati molto efficaci agli oftalmologi nel prendere appunti durante un intervento chirurgico agli occhi, ma molto spesso questi vengono mischiati alla storia medica e ai sintomi dei pazienti, scontrandosi quindi con l'articolo 9 del GDPR.[22]

Un altro punto in cui il Regolamento va ad intervenire è quello del trasferimento internazionale dei dati. Questa problematica diventa particolarmente critica quando i dati vengono trasferiti al di fuori dell'Unione europea, poiché molte società che sviluppano modelli come i LLM hanno sede all'estero. Per essere conformi al GDPR (articoli 44-50), tali organizzazioni devono rispettare le norme internazionali sui trasferimenti di dati. Questo potrebbe comportare l'adozione di clausole contrattuali standard o la verifica che il paese di destinazione abbia leggi adeguate sulla protezione dei dati. Nonostante la "decisione di adeguatezza" del GDPR offra una forma di protezione per i soggetti dei dati che intendono condividere le proprie informazioni al di fuori dell'UE, persistono comunque restrizioni procedurali. Ciò include la necessità di verificare le garanzie nei paesi terzi che non hanno ottenuto una simile "Decisione di adeguatezza" e che quindi non fanno parte dell'EEA (European Economic Area).[23]

[20] https://gdpr-info.eu/art-15-gdpr/, https://gdpr-info.eu/art-16-gdpr/, https://gdpr-info.eu/art-17-gdpr/
[21] https://gdpr-info.eu/art-9-gdpr/
[22] https://digi-con.org/large-language-models-and-eu-data-protection-mapping-some-of-the-problems/#:~:text=Under%20the%20General%20Data%20Protection,which%20the%20LLM%20is%20trained
[23] https://gdpr-info.eu/chapter-5/

Altro scenario in cui il GDPR opera è quello del rischio di profilazione e di presa di decisioni automatizzate. Sorgono implicazioni legali quando un LLM raccoglie i dati degli utenti per alimentare algoritmi di apprendimento automatico, generando previsioni sulle loro preferenze personali, le quali vengono a loro volta utilizzate per la pubblicità mirata. Questa pratica si scontra con l'articolo 22 del GDPR, il quale sancisce il diritto di essere informati su attività di profilazione e di contestare le decisioni dei sistemi automatizzati.[24]

Capitolo 3

Artificial Intelligence Act: risponde alla sfida?

Questo capitolo mira a delineare gli aspetti chiave dell'Artificial Intelligence Act, analizzandone la struttura, l'evoluzione nel tempo e gli obiettivi principali, soffermandosi poi sul riscontro pragmatico nei LLM e nella società

3.1 Introduzione all'Intelligenza Artificiale

Con il termine "intelligenza artificiale" (IA) si denota una serie di tecnologie in rapida evoluzione, capaci di generare un ampio spettro di vantaggi economici e sociali in vari contesti, sia industriali che sociali. L'implementazione dell'intelligenza artificiale, attraverso il miglioramento delle previsioni, l'ottimizzazione delle operazioni, la gestione efficiente delle risorse e la personalizzazione dei servizi, può contribuire significativamente al raggiungimento di risultati positivi dal punto di vista sociale e ambientale. Inoltre, offre vantaggi competitivi cruciali per le imprese e l'economia. Questa tecnologia risulta particolarmente necessaria in settori come l'ambiente, la sanità, il settore pubblico, la finanza, la mobilità, gli affari interni e l'agricoltura. Tuttavia, è

[24] https://gdpr-info.eu/art-22-gdpr/

importante riconoscere che gli stessi elementi e le tecniche che contribuiscono ai benefici socio-economici dell'IA possono anche presentare nuovi rischi o generare conseguenze negative per le persone fisiche e la società. L'intelligenza artificiale dovrebbe essere concepita come uno strumento al servizio delle persone e come un elemento positivo per la società, finalizzato al miglioramento del benessere umano. Purtroppo però il rapido avanzamento dell'IA solleva la necessità di una regolamentazione efficace, poiché non tutte le tecnologie vengono impiegate a scopi benigni o con modalità neutre, rendendo cruciale definire linee guida etiche e legali per mitigare potenziali rischi. Ecco che entra in gioco l'Artificial Intelligence Act (AIA).[25]

La legge sull'intelligenza artificiale (AIA) dell'Unione europea è costituita su due pilastri fondamentali: garantire che lo sviluppo e l'impiego dell'IA rispettino i valori e le regole europee e sfruttare appieno il potenziale dell'IA a fini industriali. Una caratteristica distintiva dell'AIA è il suo sistema di classificazione, il quale valuta il livello di rischio che una tecnologia di intelligenza artificiale potrebbe comportare alla salute, alla sicurezza o ai diritti fondamentali delle persone. Tale quadro prevede quattro livelli di rischio: inaccettabile, alto, limitato e minimo. Tra i sistemi di intelligenza artificiale con rischio limitato e minimo troviamo ad esempio i filtri antispam o i videogiochi, che godono di un grado maggiore di flessibilità. Dall'altro lato abbiamo i sistemi a rischio inaccettabile, come quelli governativi di punteggio sociale, i dispositivi di identificazione biometrica in tempo reale negli spazi pubblici, i veicoli a guida autonoma (Self driving cars). Il progetto normativo introduce inoltre regolamentazioni riguardanti ciò che è comunemente definita "intelligenza artificiale a scopo generale", ossia sistemi di IA in grado di essere impiegati per svariate finalità con differenti livelli di rischio. Tra le tecnologie comprese in questa categoria figurano ad esempio i grandi sistemi di IA generativi basati su Large Language Models, come nel caso di ChatGPT.[26]

L'Artificial Intelligence Act è suddiviso in quattro parti: la prima parte riguarda le disposizioni generali, l'ambito di applicazione e le definizioni; la seconda contiene l'elenco delle pratiche di IA vietate, quindi tutti i sistemi in cui l'uso di IA raggiunge il grado di "inaccettabile"; la terza parte scende più nello specifico, fornendo le linee guida per l'impiego della tecnologia di riconoscimento facciale combinata con l'AI; l'ultima parte definisce gli obblighi di trasparenza.[27]

---

[25] https://eur-lex.europa.eu/legal-content/IT-BG-DE/TXT/?from=EN&uri=CELEX%3A52021PC0206 paragrafo 1.1

[26] https://www.weforum.org/agenda/2023/06/european-union-ai-act-explained/

[27] https://www.agendadigitale.eu/cultura-digitale/lai-act-approvato-dal-parlamento-ue-luci-e-ombre-di-un-regolamento-di-portata-storica/

3.2 Base giuridica ed evoluzione dell'AIA

Nel 2017 il Consiglio europeo ha sollecitato l'attenzione sulla necessità di affrontare con urgenza le tendenze emergenti, compresa l'intelligenza artificiale, mentre garantiva una protezione elevata dei dati, dei diritti digitali e delle norme etiche. Nel 2019, attraverso le sue conclusioni sul piano "Made in Europe" per lo sviluppo dell'intelligenza artificiale, il Consiglio ha enfatizzato l'importanza di rispettare pienamente i diritti dei cittadini europei e ha invitato a una revisione della normativa esistente alla luce delle nuove sfide dell'IA. Ulteriori conclusioni del Consiglio nel 2020 hanno evidenziato la necessità di affrontare l'opacità, la complessità e la faziosità nei sistemi di intelligenza artificiale, assicurandone la compatibilità con i diritti fondamentali dell'uomo. Anche il Parlamento europeo ha giocato un ruolo attivo, adottando risoluzioni nel 2020 e 2021 su varie questioni legate all'IA, dalla responsabilità all'etica, all'uso nell'ambito penale e nell'istruzione. La risoluzione del Parlamento delinea una proposta legislativa che rispetta i principi di proporzionalità e sussidiarietà. Nel febbraio 2020 la Commissione europea si stava muovendo sul tema dell'IA, pubblicando il "Libro bianco sull'intelligenza artificiale". Il Libro Bianco delinea le strategie per raggiungere il duplice obiettivo di promuovere l'adozione dell'IA e mitigare i rischi associati a specifici impieghi di tale tecnologia. Questa proposta si concentra sull'attuazione del secondo obiettivo, mirando a sviluppare un ecosistema di fiducia mediante l'introduzione di un quadro giuridico per un utilizzo affidabile dell'IA. Fondata sui valori e sui diritti fondamentali dell'Unione europea, l'iniziativa si propone di ispirare fiducia nelle soluzioni basate sull'IA tra individui e altri utilizzatori, incentivando al contempo le imprese a contribuire al loro sviluppo. Questa proposta è una risposta alle chiare richieste del Parlamento europeo e del Consiglio, i quali hanno sollecitato ripetutamente un intervento legislativo per garantire il corretto funzionamento del mercato interno per i sistemi di intelligenza artificiale.[28] L'AIA sostanzialmente è l'unione delle bozze riguardanti la regolamentazione dell'IA da parte del Parlamento europeo e del Consiglio, sulla base della proposta della Commissione europea. La base giuridica della proposta è costituita innanzitutto dall'articolo 114 del Trattato sul funzionamento dell'Unione europea (TFUE), che prevede l'adozione di misure destinate ad assicurare l'instaurazione ed il

[28] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206 paragrafo 1: Contesto della proposta

funzionamento del mercato interno. La presente proposta costituisce un elemento chiave della strategia dell'Unione per il mercato unico digitale, mirando principalmente a garantire un corretto funzionamento del mercato interno attraverso regole armonizzate. L'attenzione si focalizza sullo sviluppo, l'immissione sul mercato e l'utilizzo di prodotti e servizi basati su tecnologie di intelligenza artificiale o forniti come sistemi di IA indipendenti ("stand-alone"). Alcuni Stati membri stanno valutando regole nazionali per la sicurezza dell'IA e il rispetto dei diritti fondamentali. Ciò potrebbe generare frammentazione e incertezza giuridica, specialmente su questioni chiave come i requisiti, la commercializzazione e l'uso dei prodotti e servizi di IA. Al fine di mitigare questi problemi, la proposta indica requisiti comuni obbligatori per la progettazione e lo sviluppo dei sistemi di IA prima dell'immissione sul mercato, con norme tecniche armonizzate per la loro implementazione. Inoltre, affronta il periodo successivo all'immissione sul mercato, stabilendo procedure armonizzate per i controlli ex post. Particolare attenzione è dedicata alla protezione delle persone fisiche nel trattamento dei dati personali, con restrizioni sull'uso di sistemi di IA per l'identificazione biometrica remota in spazi pubblici. Tali disposizioni specifiche trovano base giuridica nell'articolo 16 del Trattato sul funzionamento dell'Unione europea.[29]

3.3 Obiettivi e Disposizioni principali del AIA

I vari titoli di cui è composto l'AIA illustrano dettagliatamente le disposizioni contenute nella proposta.

Il Titolo I del regolamento definisce il campo di applicazione delle nuove norme relative all'immissione sul mercato, alla messa in servizio e all'utilizzo dei sistemi di intelligenza artificiale, fornendo inoltre definizioni chiare utilizzate in tutto l'atto. La definizione di sistema di IA è formulata in modo tecnologicamente neutro, progettata per adattarsi alle future esigenze e tenere conto delle veloci evoluzioni in questo campo. Per garantire una certezza giuridica adeguata, il Titolo I è accompagnato dall'Allegato I, che presenta un elenco dettagliato di approcci e tecniche per lo sviluppo dell'IA, adattabile dalla Commissione in risposta ai progressi tecnologici. Il ruolo dei partecipanti chiave lungo

[29] https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:12012E/TXT:it:PDF
https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206

l'intera catena del valore dell'IA, compresi fornitori e utenti, sia pubblici che privati, è definito in modo chiaro, assicurando parità di condizioni.[30]

Il Titolo II stabilisce un elenco di pratiche di intelligenza artificiale vietate, adottando un approccio basato sul rischio che classifica gli utilizzi dell'IA in base a: i) rischio inaccettabile; ii) rischio alto; iii) rischio basso o minimo. L'elenco di pratiche vietate include tutti i sistemi di IA il cui utilizzo è considerato inaccettabile in quanto contravviene ai valori imprescindibili dell'Unione, violando ad esempio i diritti fondamentali. Questi divieti riguardano pratiche con elevato potenziale di manipolazione subliminale delle persone o sfruttamento di vulnerabilità specifiche, come minori o persone con disabilità, causando danni psicologici o fisici. Per quanto riguarda pratiche manipolative o di sfruttamento sugli adulti, esse possono essere soggette alle normative esistenti sulla protezione dei dati, tutela dei consumatori e servizi digitali. La proposta proibisce inoltre l'assegnazione di un punteggio sociale basato sull'IA da parte di autorità pubbliche e vieta l'utilizzo di sistemi di identificazione biometrica remota "in tempo reale" in spazi pubblici per attività di contrasto, salvo alcune eccezioni limitate.[31]

Il Titolo III contiene disposizioni specifiche per i sistemi di intelligenza artificiale che comportano un rischio elevato per la salute, la sicurezza o i diritti fondamentali delle persone fisiche. Adottando un approccio basato sul rischio, tali sistemi "ad alto rischio" sono ammessi sul mercato europeo solo a condizione che soddisfino requisiti obbligatori e siano soggetti a una valutazione della conformità preventiva. La classificazione di un sistema di IA come "ad alto rischio" dipende dalla sua finalità prevista, seguendo la normativa dell'UE sulla sicurezza dei prodotti. Il Capitolo 1 del Titolo III delinea le regole di classificazione, identificando due categorie principali di sistemi di IA "ad alto rischio": quelli destinati a essere utilizzati come componenti di sicurezza di prodotti e altri sistemi indipendenti che presentano implicazioni significative per i diritti fondamentali. Il Capitolo 2 stabilisce requisiti giuridici per i sistemi di IA "ad alto rischio" relativi a dati e governance dei dati, documentazione, trasparenza, sorveglianza umana, robustezza, accuratezza e sicurezza. Questi requisiti, già avanzati e basati su orientamenti etici, rispecchiano standard internazionali e possono essere adattati dai fornitori in conformità con le conoscenze ingegneristiche e scientifiche. Il Capitolo 3 impone obblighi orizzontali ai fornitori, agli utenti e ad altri partecipanti nella catena del valore dell'IA, garantendo un approccio proporzionato. Il Capitolo 4 stabilisce il quadro per gli

[30] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, I.
[31] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, II.

organismi notificati coinvolti nelle procedure di valutazione della conformità, mentre il Capitolo 5 dettaglia le procedure di valutazione per i vari tipi di sistemi di IA "ad alto rischio", con l'obiettivo di ridurre l'onere per gli operatori economici e aumentare progressivamente la capacità degli organismi notificati nel tempo. I sistemi di IA destinati a essere componenti di sicurezza di prodotti saranno soggetti agli stessi meccanismi di conformità e applicazione dei prodotti ai quali sono collegati.[32]

Il Titolo IV si concentra su specifici sistemi di intelligenza artificiale al fine di affrontare i rischi particolari di manipolazione associati ad essi. Gli obblighi di trasparenza saranno applicati ai sistemi che: interagiscono con esseri umani; sono utilizzati per rilevare emozioni o stabilire associazioni (sociali) basate su dati biometrici; generano o manipolano contenuti ("deepfake"). Quando le persone interagiscono con un sistema di IA oppure le loro emozioni o caratteristiche vengono riconosciute attraverso mezzi automatizzati, è obbligatorio informarle. Se un sistema di IA è impiegato per creare o manipolare immagini, audio o video che assomigliano notevolmente a contenuti autentici, è richiesto che venga chiaramente indicato che tali contenuti sono generati tramite mezzi automatizzati, salvo eccezioni per scopi legittimi come attività di contrasto o libertà di espressione. Ciò garantisce alle persone la possibilità di prendere decisioni informate o di adottare una posizione consapevole in determinate situazioni.[33]

Il Titolo V contribuisce all'obiettivo di creare un contesto giuridico che favorisca l'innovazione, sia adatto alle sfide future e sia resiliente alle perturbazioni. Di conseguenza, incoraggia le autorità nazionali competenti a istituire spazi di sperimentazione normativa e stabilisce un quadro di base per governance, controllo e responsabilità. Gli spazi di sperimentazione normativa per l'intelligenza artificiale forniscono un ambiente controllato per testare tecnologie innovative per un periodo limitato, basato su un piano di prova concordato con le autorità competenti. Il Titolo V include anche misure volte a ridurre gli oneri normativi per le PMI e le start-up.[34]

Il Titolo VI istituisce i sistemi di governance a livello dell'Unione e nazionale. A livello dell'Unione, la proposta crea un Comitato europeo per l'intelligenza artificiale ("Comitato"), composto da rappresentanti degli Stati membri e della Commissione. Questo Comitato offre un'implementazione agevole, efficace e armonizzata del regolamento, contribuendo all'efficacia della collaborazione tra le autorità nazionali di controllo e la Commissione, fornendo consulenza e competenze. Inoltre, il Comitato

[32] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, III.
[33] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, IV.
[34] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, V.

raccoglie e condivide le migliori pratiche tra gli Stati membri. A livello nazionale, gli Stati membri dovranno designare una o più autorità competenti, tra cui l'Autorità Nazionale di Controllo, per monitorare l'applicazione e l'attuazione del regolamento. Il Garante europeo della Protezione dei Dati agirà come autorità competente per la supervisione delle istituzioni, agenzie e organismi dell'Unione quando rientrano nell'ambito di applicazione del regolamento.[35]

Il Titolo VII mira ad agevolare il monitoraggio da parte della Commissione e delle autorità nazionali mediante la creazione di una banca dati a livello dell'Unione per i sistemi di intelligenza artificiale "ad alto rischio" indipendenti, che presentano principalmente implicazioni relative ai diritti fondamentali. Questa banca dati, gestita dalla Commissione, sarà alimentata con i dati forniti dai produttori dei sistemi di IA, i quali saranno tenuti a registrare i propri sistemi prima di immetterli sul mercato o impiegarli in altro modo.[36]

Il Titolo VIII stabilisce gli obblighi di monitoraggio e segnalazione per i fornitori di sistemi di intelligenza artificiale riguardanti il post-immissione sul mercato, la notifica di incidenti e malfunzionamenti correlati all'IA e le relative indagini. L'applicazione post-immissione garantirà che, una volta sul mercato, i sistemi di IA siano soggetti a un monitoraggio efficace da parte delle autorità pubbliche, con la capacità di intervenire in caso di rischi imprevisti che richiedono una rapida azione. Inoltre, le autorità monitoreranno il rispetto degli obblighi da parte degli operatori in conformità con il regolamento. Se necessario per il loro mandato, le autorità di controllo e contrasto esistenti avranno il potere di richiedere tutta la documentazione conservata in conformità con il regolamento, nonché di accedervi e, ove necessario, di richiedere alle autorità di vigilanza del mercato di condurre prove sui sistemi di IA "ad alto rischio" mediante mezzi tecnici.[37]

Il Titolo IX istituisce un quadro per la creazione di codici di condotta, incentivando i fornitori di sistemi di intelligenza artificiale non "ad alto rischio" a conformarsi volontariamente ai requisiti obbligatori delineati per i sistemi di IA "ad alto rischio" (come specificato nel Titolo III). I fornitori di sistemi di IA non "ad alto rischio" sono liberi di sviluppare e attuare autonomamente tali codici di condotta. Questi codici possono anche includere impegni volontari riguardanti, ad esempio, la sostenibilità ambientale, l'accessibilità per le persone con disabilità, la partecipazione degli interessati alla

[35] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, VI.
[36] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, VII.
[37] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, VIII.

progettazione e allo sviluppo dei sistemi di IA, nonché la promozione della diversità nei gruppi coinvolti nello sviluppo.[38]

Il Titolo X sottolinea l'obbligo per tutte le parti di rispettare la riservatezza delle informazioni e dei dati, stabilendo regole per lo scambio delle informazioni acquisite durante l'attuazione del regolamento. Inoltre, il Titolo X include misure volte a garantire l'efficace attuazione del regolamento attraverso sanzioni proporzionate e dissuasive in caso di violazione delle disposizioni.[39]

Il Titolo XI delinea le regole per l'esercizio delle deleghe e delle competenze di esecuzione. La proposta concede alla Commissione il potere di adottare, se necessario, atti di esecuzione volti a garantire un'applicazione uniforme del regolamento.[40]

Il Titolo XII impone alla Commissione l'obbligo di valutare periodicamente la necessità di aggiornare l'Allegato III (il quale elenca i sistemi definiti "ad alto rischio") e prevede la preparazione di relazioni periodiche di valutazione e riesame del regolamento. Inoltre, contiene disposizioni finali, compreso un periodo transitorio differenziato per la data iniziale di applicabilità del regolamento, al fine di agevolare una corretta attuazione da parte di tutte le parti interessate.[41]

3.4 AIA e LLM: regolamentazione e prospettive

Queste disposizioni sono valide per tutti quei sistemi di intelligenza artificiale che hanno uno scopo predefinito e un uso specifico come: l'identificazione biometrica e la categorizzazione di persone naturali; la gestione e operazione di infrastrutture critiche; l'istruzione e la formazione professionale, l'occupazione, la gestione dei lavori e accesso all'autoimpiego; l'accesso e fruizione di servizi essenziali privati e servizi pubblici benefici; le forze dell'ordine; la gestione delle migrazioni, asilo e controllo delle frontiere; l'amministrazione della giustizia e processi democratici. Di conseguenza, questi sistemi posso essere meglio regolamentati e monitorati. Sfida ben più complicata è quella che riguarda invece tutti i sistemi che vengono definiti "ad uso generico". Questi sistemi non sono altro che i cosiddetti "foundation models" dai quali derivano i Large Language Models, trattati nel *Capitolo 2*. Anche questi modelli sono stati trattati nonostante risulti

[38] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, IX.
[39] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, X.
[40] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, XI.
[41] https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206, XII.

complicato, in quanto gli usi che possono derivare da questo tipo di IA sono molteplici e non tutti prevedibili, rendendo quindi impossibile una regolamentazione a priori. In generale, sono state introdotte delle regole di trasparenza più avanzate riguardanti: la natura dei dataset utilizzati per l'addestramento, andando a ridurre i problemi legati alla privacy e al copyright; il contrasto ai deepfakes con l'obbligo di esplicitare la natura veritiera o meno del contenuto che si sta riproducendo; la creazione di materiali ritenuti illegali. Tutte queste misure di sicurezza si agganciano a delle previsioni fatte per i social network attraverso il Digital Services Act (DSA). Di conseguenza, attraverso l'AIA si potranno applicare molte delle disposizioni contenute nel DSA anche ai sistemi di intelligenza artificiale.[42]

Nella bozza della proposta, il Consiglio dell'Unione europea in particolare, ha ridefinito questi sistemi di intelligenza artificiale come: "intesi dal fornitore per svolgere funzioni generalmente applicabili, come riconoscimento di immagini e suoni, generazione di audio e video, rilevamento di modelli, risposta a domande, traduzione e altri; un sistema di IA a uso generale può essere utilizzato in una pluralità di contesti ed essere integrato in una pluralità di altri sistemi di IA". Il Consiglio, dunque, stabilisce che le regole sui sistemi di IA "ad alto rischio" si applicano ai sistemi di IA ad uso generale che possono essere utilizzati in tali contesti, a meno che questi utilizzi siano esplicitamente esclusi. Inoltre, conferisce alla Commissione l'autorità di "specificare e adattare" i requisiti per questo tipo di sistemi. A differenza dei fornitori di IA "ad alto rischio", i quali sottostanno agli obblighi definiti dagli articoli 16-25 del regolamento, i fornitori di IA ad uso generale sono vincolati ad un sistema di obblighi inferiore: fornire il proprio nome e marchio (art.16aa); effettuare una valutazione di conformità (art.16e); obblighi di registrazione (art.16f); azioni correttive per cause di non conformità (art.16g); obbligo di marcature CE (art.16i); dimostrazione della conformità (art.16j); nomina di un rappresentate autorizzato (art.25); monitoraggio post-vendita (art.61); condivisione di informazioni con i concorrenti entranti (art.4b(5), proposta del Consiglio); Il Consiglio, quindi, introduce un nuovo ruolo all'interno della Commissione, con l'intento di limitare gli obblighi dei fornitori di sistemi di intelligenza artificiale a uso generale. Questo implica l'instaurazione di un nuovo obbligo che richiede ai fornitori di condividere direttamente la loro conoscenza con i concorrenti. Fondamentalmente, ciò rappresenta un cambiamento nell'approccio della Commissione, che adotta una variante per determinati

[42] https://www.youtube.com/watch?v=Qjvtt5YdWZs&t=580s

sistemi, indicando una via per adattare i requisiti della Legge sull'intelligenza artificiale per sistemi considerati "ad alto rischio".[43]

Per quanto concerne invece la proposta del Parlamento europeo, esso definisce i LLM come "modelli di intelligenza artificiale addestrati su una vasta gamma di dati su larga scala, i quali sono progettati per la generalità dell'output e possono essere adattati a una vasta gamma di compiti specifici" (art.3(1c), proposta del Parlamento). Questa definizione sposta il focus sui possibili utilizzi e la capacità di adattamento di questi modelli a compiti specifici. All'articolo 28b(1) della proposta del Parlamento troviamo i criteri da adottare da parte dei sistemi di IA ad uso generico: obbligo di istituire una governance del rischio; obbligo di istituire una governance dei dati; requisiti per avere livelli appropriati di prestazioni, prevedibilità, interpretabilità, correzione, sicurezza e cyber sicurezza; obbligo di ridurre l'uso di energia, di risorse e gli sprechi, nonché di aumentare l'efficienza energetica e complessiva del modello; redigere documentazione tecnica dettagliata e istruzioni comprendibili per l'uso; istituire un sistema di gestione della qualità; registrare quel foundation model nel database dell'Unione europea. Inoltre, nell'articolo 28(b)(4) della proposta del Parlamento si trovano tre criteri da affiancare agli obblighi più generali dei LLM quali: conformarsi agli obblighi di trasparenza delineati nell'articolo 52(1); addestrare e, nel caso, progettare e sviluppare il modello di base in modo da garantire adeguate salvaguardie contro la generazione di contenuti in violazione del diritto dell'Unione, in linea con lo stato dell'arte generalmente riconosciuto e senza pregiudizio dei diritti fondamentali, compresa la libertà di espressione; senza pregiudizio per la legislazione dell'Unione o nazionale sull'Unione in materia di copyright, documentare e rendere pubblicamente disponibile un riepilogo sufficientemente dettagliato dell'uso dei dati di formazione protetti dalla legge sul copyright.[44]

L'analisi degli strumenti normativi sull'intelligenza artificiale si concentra sulla sfida di bilanciare l'impatto negativo sull'innovazione con la promozione di un uso responsabile dell'IA. Le posizioni del Consiglio e del Parlamento, pur mantenendo il quadro generale della Commissione, presentano emendamenti che alleviano gli oneri per i fornitori di sistemi di Intelligenza Artificiale Generale. Gli emendamenti del Parlamento sembrano favorire l'innovazione, ma possono complicare l'approccio basato sul rischio. La

[43] https://artificialintelligenceact.eu/wp-content/uploads/2022/12/AIA-%E2%80%93-CZ-%E2%80%93-General-Approach-25-Nov-22.pdf
https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206

[44] https://artificialintelligenceact.eu/wp-content/uploads/2023/06/AIA-%E2%80%93-IMCO-LIBE-Draft-Compromise-Amendments-14-June-2023.pdf,
https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206

descrizione generale dei requisiti per i modelli di base richiede una revisione per evitare ambiguità. Il Consiglio si distingue per il focus sulla responsabilità, adattando gli obblighi dei fornitori in base al potenziale di influenza sui sistemi. Questo approccio personalizzato si adatta meglio alla natura dinamica dell'IA rispetto alla categorizzazione rigida proposta dal Parlamento.[10]

3.5 Sfide operative e normative tra Diritto all'Oblio, LLM e AIA

Le sfide e gli ostacoli derivanti da un intervento normativo di tale portata nei confronti di queste tecnologie sono molteplici e di diversa natura. Dal punto di vista economico, l'Europa necessita di un maggiore impegno in termini di risorse finanziarie e di progetti transnazionali. L'unità politica europea è fondamentale per garantire un'efficace competitività tecnologica a livello globale. Ciò implica la necessità di condividere impegni, strutture, personale, sforzi economici e attività di ricerca. Sebbene il primato dell'Europa nella creazione di framework normativi, successivamente adattati e utilizzati in tutto il mondo, rappresenti un valore aggiunto, esso da solo non è sufficiente per competere con potenze come la Cina o gli Stati Uniti. Scendendo nello specifico delle PMI, è fondamentale garantire l'accesso a competenze e risorse adeguate affinché le imprese siano consapevoli delle sfide legali che possono sorgere e possano quindi sfruttare appieno le potenzialità dei sistemi di intelligenza artificiale.[45]

Dal punto di vista tecnico, riaddestrare intere parti o porzioni di LLM, rimuovendo determinate informazioni a causa dell'applicazione del diritto all'oblio e l'AIA, ha diverse conseguenze: perdita di informazioni cruciali e rappresentative che potrebbero compromettere la performance del modello; creazione o espansione di bias nel modello; perdita della capacità di generalizzare su nuovi dati (overfitting), compromettendo l'utilità del modello di adattarsi alla vita reale; difficoltà nel garantire la qualità e l'affidabilità del modello, essendo intervenuti sul dataset di addestramento. Presupposto che il risultato ottenuto tramite un avvenuto riaddestramento di questi modelli abbia pienamente successo, resta irrisolto il problema delle tempistiche d'intervento: il cosiddetto "undue delay" o ritardo ingiustificato. Questo concetto, sebbene non rigidamente definito dal GDPR, viene interpretato dal legislatore come un mese, ma con

[45] https://www.iterinformatica.it/intelligenza-artificiale-cose-ai-act-e-quale-impatto-avra/ https://www.youtube.com/watch?v=Qjvtt5YdWZs&t=583s

le attuali capacità tecniche questo intervallo di tempo potrebbe essere impossibile da rispettare.[46]

[46] https://arxiv.org/pdf/2307.03941.pdf

Conclusioni

In conclusione, emerge chiaramente la necessità di un approccio bilanciato e attento nella regolamentazione delle tecnologie emergenti come i Large Language Models. Il "diritto all'oblio", sebbene fondamentale per la tutela della privacy individuale, richiede una riconsiderazione e un adattamento per affrontare le sfide uniche presentate da queste nuove forme di intelligenza artificiale. L'Artificial Intelligence Act rappresenta un passo significativo verso una regolamentazione più ampia, ma la sua efficacia nei confronti dei Large Language Models richiederà un continuo dialogo e adattamento normativo.
Ecco, quindi, che nasce la necessità di cercare e trovare un efficace punto di equilibrio tra ordinamento giuridico e tecnologia. Uno sforzo per riuscire ad assottigliare il più possibile quel divario che vi è tra privacy e innovazione e che accresca la sinergia tra le parti, andando a plasmare un futuro in cui il potenziale dell'IA venga sfruttato nel pieno rispetto della sicurezza dell'uomo. Il rapido ed esponenziale innovamento tecnologico prescinde da un intervento normativo efficace, tempestivo e bilanciato, che non rallenti il progresso tecnologico ed economico e mantenga come priorità assoluta la protezione dei diritti fondamentali dell'uomo.