Copyright, Intelligenza Artificiale Generativa e legislazioni

FIMI (Federazione Industria Musicale Italiana) tutela e promuove le attività connesse all'industria discografica: nata nel 1992, è socia fondatrice di Confindustria Cultura Italia e membro IFPI (Federazione Industria Fonografica Internazionale), rappresenta le maggiori imprese produttrici e distributrici del settore discografico per un totale di oltre 2.500 marchi tra i più famosi del mondo.

Per generazioni, varie tecnologie sono state utilizzate con successo per supportare la creatività umana. Prendiamo la musica, per esempio... dal pianoforte all'amplificazione dei pedali per chitarra, dai sintetizzatori alle drum machine, dalle workstation di digital audio digitali alle librerie di suoni e oltre, i creatori di musica hanno sfruttato a lungo la tecnologia per esprimere le loro visioni attraverso voci, strumenti e dispositivi diversi. L'intelligenza artificiale svolge già - e lo farà sempre più nel futuro - quel ruolo di strumento di assistenza al processo creativo, consentendo a una gamma più ampia di persone di esprimersi in modo creativo.

Inoltre, l'intelligenza artificiale ha molti usi preziosi al di fuori del processo creativo stesso, inclusi quelli che amplificano le connessioni dei fan, perfezionano i consigli personalizzati, identificano i contenuti in modo rapido e accurato, aiutano con la pianificazione, automatizzano e migliorano i sistemi di pagamento e altro ancora.

I modelli di intelligenza artificiale generativa sono una forma distinta di algoritmo di apprendimento automatico programmato per uno scopo specifico: produrre output (quando richiesto) sulla base di lavori preesistenti che vengono acquisiti dal sistema. L'intelligenza artificiale generativa è diversa da quella tradizionale (a volte chiamata analitica) L'intelligenza artificiale generativa assorbe lavori creativi per produrre nuovo materiale, mentre l'intelligenza artificiale tradizionale viene utilizzata per svolgere compiti e calcoli più meccanici o automatici come rilevare modelli, affinare l'analisi o classificare i dati.

L'output dei modelli generativi potrebbe essere testo, immagini, materiale audio o audiovisivo, che il modello AI in genere produce solo dopo aver importato simili opere protette da copyright.

Queste attività della "fase di importazione" includono la raccolta (che include lo scraping) e la curatela di opere protette da copyright a fini di formazione, indipendentemente dal fatto che l'entità impegnata in tali atti sia la stessa entità che possiede o gestisce il sistema di intelligenza artificiale generativa che importa queste opere.

Gli artisti che hanno problemi fisici, mentali o di altro tipo potrebbero aver avuto difficoltà o non essere stati in grado di eseguire alcuni aspetti del processo creativo, ma grazie all'assistenza dell'intelligenza artificiale ora potrebbero essere in grado di fare ciò che prima non potevano. Allo stesso modo, i gruppi emarginati i cui sforzi creativi potrebbero essere stati limitati in passato a causa della mancanza di risorse o di accesso a strumenti creativi potrebbero scoprire che la democratizzazione dell'intelligenza artificiale generativa "livella il campo di gioco" e si traduce in nuovi sforzi creativi.

Gli autori di narrativa che lottano con il “blocco dello scrittore” utilizzano l’intelligenza artificiale generativa per assistere nella loro ideazione, nonché per fare cose come sviluppare personaggi, nomi di luoghi, trame, ecc.

L’intelligenza artificiale consente agli artisti di intraprendere progetti più ambiziosi che in precedenza richiedevano manodopera e outsourcing. Ad esempio, gli artisti hanno riferito di utilizzare generatori di immagini AI per creare elementi di sfondo di grandi dimensioni che poi combinano con opere d'arte originali create fuori piattaforma.

Qui di seguito si cerca di fornire risposte alle domande chiave relative all'obbligo proposto di registrazione e divulgazione. Queste risposte sono fornite per convalidare e giustificare l'importanza di un obbligo efficace ed equilibrato di registrazione e divulgazione, mentre sfatano alcuni dei miti sulla difficoltà di conformarsi a esso.

1) **Quali sono le preoccupazioni relative all'IA generativa per i detentori dei diritti e altre parti interessate?**
2) **Perché la registrazione e la divulgazione da parte delle entità chiave nella catena dell'IA generativa (incluso lo sviluppo e l'operatività dei modelli di base) sono una soluzione necessaria a tali preoccupazioni?**
3) **Perché è necessaria ora una legge per imporre tale obbligo di registrazione e divulgazione?**
4) **Quali entità dovrebbero essere soggette all'obbligo di registrazione e divulgazione?**
5) **Quali documenti dovrebbero essere tenuti a conservare gli sviluppatori e gli operatori dell'IA ai fini della divulgazione?**
6) **Perché è tecnicamente fattibile conformarsi all'obbligo di registrazione e divulgazione?**

1) Quali sono le preoccupazioni relative all'IA generativa per i titolari dei diritti e altre parti interessate?

I modelli di IA (compresi i modelli di base) che alimentano i sistemi di IA generativa vengono "addestrati" su input costituiti da vaste quantità di dati che comprendono testo, immagini, audio e video e possono produrre output costituiti da testo, immagini, audio e video. Quando tali dati contengono contenuti protetti da diritto d’autore o altri dati protetti, i titolari dei diritti hanno (e devono continuare a avere) il potere discrezionale di scegliere se concedere o negare il permesso per l'uso delle opere, determinare un adeguato livello di compensazione e avere voce nel modello economico applicabile.

Infatti, lo sviluppo di sistemi IA sicuri e significativi, compresi i modelli di base, dipende dalla disponibilità continua di input di alta qualità, la creazione dei quali si basa sul mantenimento di un solido quadro normativo sul diritto d'autore che continua a incentivare la creazione e gli investimenti

in nuove opere. Alcuni potrebbero cercare di caratterizzare le autorizzazioni di proprietà intellettuale come un ostacolo all'accesso o un freno all'innovazione dell'IA, ma nulla potrebbe essere più lontano dalla verità. Quando l'innovazione creativa e l'innovazione tecnologica vanno di pari passo, vi è una crescita.

Con l'avvento dell'intelligenza artificiale, l'industria musicale sta ancora una volta cogliendo le opportunità offerte dalle nuove tecnologie. L'intelligenza artificiale sta già migliorando il modo in cui gli artisti creano musica, il modo in cui le piattaforme organizzano la musica in playlist e raccomandazioni, il modo in cui gli artisti si connettono con i fan e il modo in cui i fan accedono alla musica.

Sta inoltre migliorando e accelerando molte delle attività dietro le quinte, dalla produzione, al mixaggio, al mastering, alla programmazione, alla pianificazione, all'amministrazione, alla comprensione delle tendenze dei consumatori e alla lotta alla pirateria. In effetti, vale la pena notare che sono già in corso discussioni tra le principali case discografiche e le società di intelligenza artificiale per creare partnership per l'uso autorizzato della musica e delle voci degli artisti in strumenti di intelligenza artificiale generativa attraverso licenze e autorizzazioni commerciali.

Il quadro normativo dell'UE sul diritto d'autore richiede che i sistemi di IA siano addestrati sui contenuti a cui si accede legalmente, nel rispetto degli "opt-out" da parte dei titolari dei diritti.

Tuttavia, esistono prove evidenti che noti sviluppatori e operatori di IA si sono appropriati di contenuti su larga scala, compresi contenuti protetti da diritto d'autore, provenienti da siti web e servizi senza rispettare i loro obblighi legali. Comprendiamo che, in questa fase iniziale dell'IA generativa, molti sviluppatori hanno effettivamente ignorato la legge e hanno estratto illegalmente contenuti da internet per scopi di addestramento, compresi, ma non limitati, a contenuti caricati dagli utenti sulle piattaforme (User Uploaded Content - UUC). Questi sviluppatori hanno quindi reso disponibili al pubblico e, in alcuni casi commercializzato, questi sistemi, il che significa che la comunità creativa ha di fatto sovvenzionato lo sviluppo dell'IA da parte delle aziende tecnologiche che acquisiscono contenuti protetti per generare "nuovi" contenuti che, al tempo stesso, competono direttamente con il materiale utilizzato per addestrare i modelli.

Inoltre, ciò sta accadendo non solo con i modelli di base e altri sistemi di IA generativa che hanno assimilato opere creative per generare "nuovi" contenuti che competono direttamente con i contenuti coperti dal diritto d'autore, ma anche con i modelli di clonazione vocale che consentono ad altri di creare discorsi o brani musicali non autorizzati che imitano le voci degli artisti senza il loro consenso o autorizzazione. Ciò ha portato gli artisti e le etichette musicali a dover affrontare una proliferazione di voci clonate dall'IA non autorizzate che violano, non solo i diritti degli artisti, il cui timbro viene clonato, ma anche i diritti di coloro che possiedono le composizioni musicali e le registrazioni audio in ciascuna traccia musicale sottostante. A titolo illustrativo, su un solo server di Discord, sono stati pubblicati più di 30 modelli vocali IA di persone in un solo periodo di 24 ore e, negli ultimi mesi, un solo account su un altro servizio di rilievo ha pubblicato oltre 800 modelli di clonazione vocale IA. Allo stesso modo, il numero di servizi ampiamente disponibili che offrono la possibilità di creare deepfake sta aumentando. Mentre ci sono centinaia di modelli vocali IA non autorizzati di cantanti e attori famosi, e modelli di

immagini non autorizzate di varie celebrità, è altrettanto facile trovare modelli vocali o di immagini IA di figure politiche o aziendali liberamente disponibili, come il presidente Biden, Barack Obama, Benjamin Netanyahu, Yoon Suk Yeol e Elon Musk. Inoltre, sono stati recentemente resi disponibili diversi servizi e app in cui è possibile creare un clone vocale di qualsiasi voce senza la necessità di alcuna conoscenza tecnica. Questi strumenti possono essere utilizzati da chiunque per creare modelli di cloni vocali di altre figure politiche o aziendali, o di cittadini comuni, e tali modelli che creano output che suonano o appaiono inquietantemente realistici e possono essere utilizzati per creare e diffondere disinformazione.

L'incapacità di determinati sviluppatori di IA di interagire in modo appropriato con i titolari dei diritti e di utilizzare contenuti protetti senza autorizzazione, oltre a violare i diritti, provoca danni agli artisti e ai consumatori. Tale pratica è anche miope, poiché il rispetto del diritto d'autore e delle altre leggi sulla proprietà intellettuale fornisce una base sicura e stabile su cui possono essere trovate soluzioni commerciali - i colloqui tra i titolari dei diritti e sviluppatori di IA sono già in corso.

2) **Perché la registrazione e la divulgazione da parte di tutti gli attori della catena dell'IA generativa (sviluppatori e operatori dei modelli di base) sono una soluzione necessaria a queste preoccupazioni?**

In base alle norme sul diritto d'autore dell'UE, i sistemi di intelligenza artificiale devono essere addestrati su opere accessibili legalmente, con le appropriate autorizzazioni preventive laddove il titolare dei diritti le abbia riservate, come previsto dall'articolo 4 della Direttiva DSM. Sebbene il mancato rispetto delle norme sul diritto d'autore non sia un problema nuovo, la grande capacità dei modelli di intelligenza artificiale generativa di ingerire, copiare e appropriarsi di contenuti protetti crea nuove sfide per i titolari dei diritti nell'applicare le norme sul diritto d'autore esistenti. In particolare, potrebbe essere impossibile stabilire se e come i contenuti siano stati utilizzati per addestrare tali modelli e se vi siano state violazioni di opere o registrazioni audio preesistenti. Anche se sono in corso sperimentazioni a livello tecnologico per determinare se gli input utilizzati per addestrare i modelli di IA possano essere identificati, determinare quali contenuti siano stati copiati o utilizzati durante il processo di IA potrebbe richiedere risorse significative o rivelarsi impossibile per i titolari dei diritti. Inoltre, lo sviluppo o l'attuazione di tecniche investigative per determinare l'uso del proprio contenuto protetto da diritto d'autore non dovrebbe, anche se fosse possibile, ricadere sui titolari dei diritti come un onere. Questo sarebbe non solo irragionevole, ma significativamente meno efficiente rispetto al tenere semplicemente traccia del materiale utilizzato.

Questo problema fondamentale può essere risolto se le entità chiave nella catena dell'IA generativa (ossia gli sviluppatori e gli operatori di sistemi e modelli di IA, compresi i modelli di base) tengano registri delle opere di parti terze o di altri soggetti protetti. Ciò consentirà ai detentori dei diritti di prevenire utilizzi non autorizzati delle loro opere e di facilitare eventuali azioni legali per far valere i loro diritti, poiché sarà possibile verificare che l'accesso ai loro contenuti sia stato legittimo, sia attraverso le licenze e le autorizzazioni appropriate ottenute dai pertinenti titolari dei diritti, sia tramite una relativa eccezione. In effetti, la registrazione e la divulgazione dei registri da parte di coloro che

intendono fare affidamento sugli articoli 3 o 4 della Direttiva DSM rappresenta l'unico vero modo in cui l'eccezione può funzionare; senza di essa, i titolari dei diritti non saranno in grado di stabilire se i requisiti dell'eccezione siano stati soddisfatti. Questo garantirà, pertanto, che l'esistente quadro normativo sul diritto d'autore, così come le altre norme applicabili, rimangano efficaci nella pratica. Contribuirà anche a garantire che i sistemi di intelligenza artificiale siano progettati e diffusi in modo responsabile, trasparente e affidabile, tutelando i diritti e la sicurezza delle persone, compresi i loro diritti di proprietà intellettuale, i diritti di personalità, i diritti alla privacy e altri diritti umani. Un obbligo di conservare i dati usati per l'addestramento dei sistemi AI sarebbe altresì importante per gli operatori dei sistemi di intelligenza artificiale che desiderano avere la certezza che gli strumenti di IA che stanno utilizzando siano sicuri, etici e non comportino rischi per loro, compresa la violazione di contenuti di terze parti. Infatti, secondo il National Institute of Standards and Technology, l'adempimento di tali obblighi di registrazione e divulgazione può "migliorare la trasparenza, migliorare i processi di revisione umana e rafforzare la responsabilità" nei sistemi di IA[1].

1) Perché è ora necessaria una legislazione per garantire la conservazione dei registri e l'obbligo di divulgazione?

Le aziende di IA e i titolari dei diritti comprendono il valore dei contenuti protetti dal diritto d'autore e dai diritti correlati nel contesto dell'IA. Tuttavia, mentre le aziende di IA parlano di trasparenza, equità e della necessità di regolamentazione, questa conversazione non si estende sempre a una vera trasparenza sull'uso di contenuti protetti nell'addestramento dei loro sistemi e modelli di IA (compresi i modelli di base). Tutto ciò è centralmente importante per coloro i cui contenuti protetti vengono ampiamente utilizzati.

- Il CEO di OpenAI, Sam Altman, ha dichiarato che "i creatori meritano il controllo su come vengono utilizzate le loro creazioni e su ciò che accade dopo che le hanno rese disponibili nel mondo", e ha inoltre affermato che "la cosa giusta è assicurarsi che [i creatori] ottengano un significativo vantaggio" dalle tecnologie IA e "i proprietari di contenuti, somiglianze e le persone meritano completamente il controllo su come vengono utilizzati e di trarne beneficio"[2]. Tuttavia, ciò è impossibile finché OpenAI non pubblica alcuna informazione[3] riguardo ai dati utilizzati nell'addestramento.

[1] Come osservato dal National Institute of Standards and Technology (NIST), "il mantenimento della provenienza dei dati di addestramento e il supporto dell'attribuzione delle decisioni del sistema di intelligenza artificiale a sottoinsiemi di dati di addestramento possono aiutare sia in termini di trasparenza che di responsabilità". National Institute of Standard and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)", NIST AI 100-1, gennaio 2023, pag.16, disponibile su https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

[2] Video dell'udienza del 16 maggio 2023 della sottocommissione della magistratura del Senato per la privacy, la tecnologia e la legge intitolata Oversight of AI: Rules for Artificial Intelligence, disponibile all'indirizzo https://www.c-span.org/video/?528117-1/openai-ceo-testifies-artificial-intelligence a partire da 1:07:44.

[3] Documentazione di OpenAI https://cdn.openai.com/papers/gpt-4.pdf

OpenAI sembra offrire un opt-out solo per coloro che hanno un account[4] OpenAI e solo "in futuro"[5]. È troppo tardi per la vasta quantità di contenuti protetti già appropriati dai loro sistemi.

- Microsoft ha dichiarato che "è fondamentale che gli autori mantengano il controllo dei loro diritti ai sensi del diritto d'autore e ottengano un ritorno economico adeguato dalle loro creazioni"[6]. Tuttavia, le varie misure nell'annuncio di Microsoft riguardano i controlli per evitare che i contenuti protetti da diritto d'autore 'fuoriescano' nei risultati, nonché le risposte nel caso ciò accada. Questo elude completamente la questione della trasparenza sui contenuti protetti da diritto d'autore negli input per l'addestramento, che sarebbe il primo passo verso il vero mantenimento del controllo da parte degli autori.
- Meta ha dichiarato di aver utilizzato il dataset Books3 nell'addestramento della sua acclamata famiglia di grandi modelli linguistici (LLMa)[7]. È emerso che Books3 potrebbe contenere oltre 170.000 copie che violano i diritti d'autore di libri protetti, estratte tramite ElutherAI da Bibliotik[8], presumibilmente tramite la rete di condivisione di file BitTorrent[9]. Bibliotik è elencata da addictivetips[10] tra i "migliori" siti torrent del 2023. Books3 è stato citato in un'azione legale contro Meta, così come nel contesto dell'addestramento di vari modelli LLM di altre aziende. Il gruppo danese Rights Alliance ha coordinato l'attività di rimozione di questo dataset, che ha portato a una rimozione globale. Rights Alliance ha specificamente menzionato la necessità di trasparenza in relazione a questa azione di successo[11]. Tuttavia, al rilascio dei modelli LLaMA 2, Meta non ha fornito alcuna informazione significativa sui testi dei libri utilizzati per l'addestramento[12,13], muovendosi nella direzione sbagliata.

[4] Politica OpenAI all'indirizzo https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model performance e vedere il modulo collegato che dice "* Assicurati che l'e-mail fornita sia associato al tuo account e che l'ID organizzazione abbia il formato "org-eXam3pleOr9giD", altrimenti non saremo in grado di elaborare la tua richiesta."

[5]Politica OpenAI su https://help.openai.com/en/articles/7039943-data-usage-for-consumer-services-faq

[6]Impegno di Microsoft sul copyright per i clienti all'indirizzo: https://blogs.microsoft.com/on-theissue/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/

[7]Meta documentazione tecnica: https://arxiv.org/pdf/2302.13971.pdf

[8]Tweet in cui la persona che si dice essere l'autore di Books3 dichiara che la fonte è Bibliotik.

[9] https://www.theatlantic.com/technology/archive/2023/08/books3-ai-meta-llama-pirated-books/675063/

[10]Recensione Addictivetips dei siti di torrenting https://www.addictivetips.com/vpn/best-private-torrenting-sites

[11]Rights Alliance pubblicazione su Books3 su https://rettighedsalliancen.com/the-books3-case-highlights-the-need-for-transparency-in-the-training-of-artificial-intelligence/

[12]Meta documentazione tecnica su LLaMa2: https://arxiv.org/pdf/2307.09288.pdf

[13] Meta documentazione tecnica su LLaMa2: https://ai.meta.com/resources/models-and-libraries/llama/

- I modelli GPT di OpenAI sono stati addestrati, tra le altre cose, su dati di testi musicali protetti da diritto d'autore, come dimostrato fino a metà del 2023, poiché i modelli GPT di OpenAI producevano l'intero testo completo quando sollecitati a farlo. Nella descrizione di GPT-4[14], OpenAI non ha pubblicato informazioni su dove tali dati testuali fossero stati ottenuti, né ha fornito informazioni sull'entità dei corpora inclusi nell'addestramento[15]. Per quanto riguarda i dati di addestramento complessivi, le informazioni precedenti su GPT-3 di OpenAI forniscono una categoria molto limitata di informazioni sui dati di addestramento. Con il lancio del suo modello GPT-4 più recente, OpenAI ha citato "il panorama competitivo e le implicazioni per la sicurezza"[16] come motivazioni vaghe per cui non fornirà "ulteriori dettagli" sui dati di addestramento utilizzati. Tuttavia, il vantaggio competitivo non può giustificare l'uso di opere protette da diritto d'autore senza il permesso dei titolari dei diritti e, allo stesso tempo, è difficile immaginare reali implicazioni per la sicurezza nel divulgare la presenza di testi di canzoni nel dataset di addestramento.
- Microsoft ha lanciato il modello di generazione rap neurale DeepRapper come parte del suo progetto di ricerca Muzic e della suite di tool. I documenti e le pubblicazioni di Microsoft affermano che, come passo nell'addestramento del modello DeepRapper, "sviluppiamo un processo di data mining per raccogliere un dataset di rap su larga scala che include un gran numero di canzoni rap"[17]. Tuttavia, attualmente il progetto su Github di Microsoft non fornisce accesso pubblico alle informazioni sui dati di addestramento, mentre l'archivio restante di campioni di testi delle canzoni del progetto include solo i testi di una canzone di un artista indipendente cinese, Kung Fu Pen[18]. Questo sembra non riflettere il "dataset rap su larga scala" che Microsoft afferma di aver raccolto; nel frattempo, i titolari dei diritti non hanno modo di scoprire se i loro testi siano stati utilizzati.
- Google MusicLM è un modello musicale generativo all'avanguardia che è stato addestrato in parte " relying on pretrained and frozen MuLan"[19]. MuLan è un modello di base multimodale testo/audio sviluppato da Google. La documentazione di MuLan afferma che è stato addestrato, in parte, utilizzando "un set di addestramento di oltre 44 milioni di video musicali su internet"[20]. Non è noto di cosa fossero composti questi 44 milioni di video musicali, poiché i dettagli non sono stati divulgati da Google, ma i set di addestramento di dimensioni molto più ridotte assemblati da Google, MusicCaps e AudioSet, sono derivati da video di YouTube, compresi video protetti da diritto d'autore.

Come dimostrato negli esempi precedenti (che non sono un elenco esaustivo), le aziende di intelligenza artificiale non forniscono volontariamente informazioni sufficienti sui dati che hanno

[14] Documentazione tecnica OpenAI GPT4: https://arxiv.org/pdf/2303.08774.pdf

[15] https://www.vice.com/en/article/ak3w5a/openais-gpt-4-is-closed-source-and-shrouded-in-secrecy

[16] Documentazione tecnica OpenAI GPT4: https://arxiv.org/pdf/2303.08774.pdf

[17] Documentazione di Microsoft DeepRapper https://arxiv.org/pdf/2107.01875.pdf

[18] Deeprapper Github su https://github.com/microsoft/muzic/tree/main/deeprapper/data/lyrics/lyrics_samples/raw/rap_功夫胖

[19] Documentazione tecnica di Google MusicLM: https://arxiv.org/pdf/2301.11325.pdf

[20] Documentazione tecnica di Google MuLan all'indirizzo https://arxiv.org/pdf/2208.12415.pdf

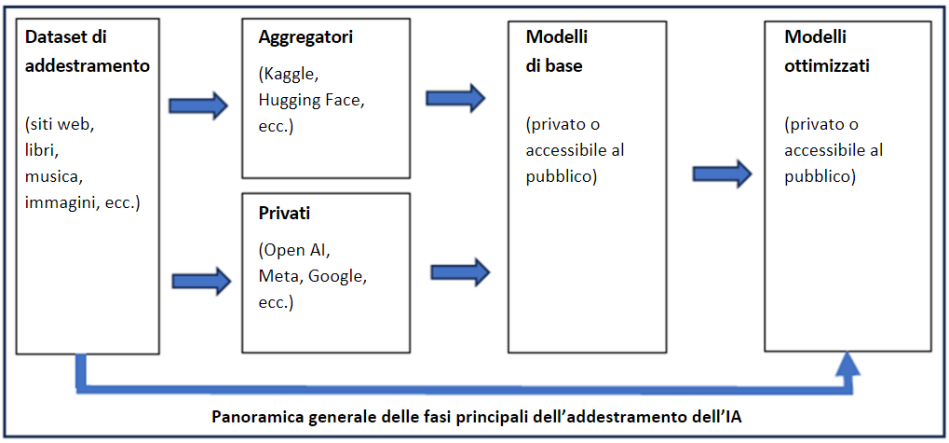
utilizzato nei dataset di addestramento. Ciò rende ancora più evidente la necessità di un obbligo legale di conservazione dei registri e di divulgazione: le misure volontarie non funzioneranno.

È quindi necessaria una legislazione per invertire questa tendenza e proteggere gli interessi legittimi di tutte le parti interessate, non solo i titolari dei diritti ma anche gli utenti dei servizi di IA che devono sapere che non saranno esposti a rischi etici o legali derivanti dall'uso di sistemi di IA.

2) Quali entità dovrebbero essere soggette all'obbligo di registrazione e divulgazione?

***Risposta breve**: lo sviluppo e la diffusione di sistemi e modelli di IA (compresi i modelli di base) coinvolge in genere molti attori diversi in vari punti della catena, spesso operanti in giurisdizioni diverse. La legislazione deve essere ampia per evitare di creare opportunità di eludere le regole.*

Ci sono diverse fasi e diversi attori in ciascuna fase dello sviluppo dei set di dati di addestramento e nell'addestramento dei modelli di intelligenza artificiale. La Figura 1 mostra una panoramica generale:



Panoramica generale delle fasi principali dell'addestramento dell'IA

La Figura 2, di seguito, mostra lo sviluppo e la diffusione di più ampi sistemi e modelli di IA (compresi i modelli di base). Le attività descritte nella figura 1 sono mostrate in un contesto più ampio di sviluppo e diffusione.

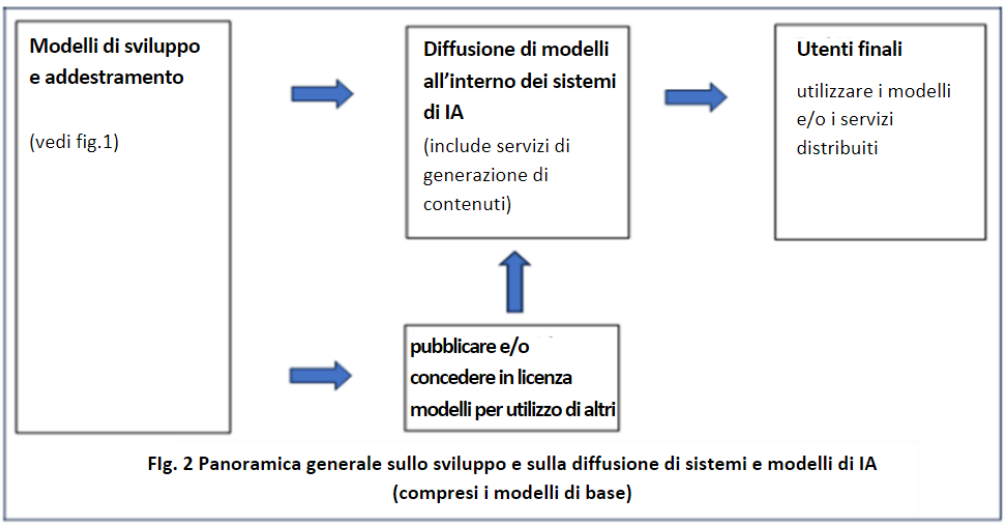


FIg. 2 Panoramica generale sullo sviluppo e sulla diffusione di sistemi e modelli di IA (compresi i modelli di base)

L'intelligenza artificiale è spesso sviluppata da diversi attori, tra cui entità che sviluppano e pre-addestrano modelli di base, entità che ottimizzano i modelli, entità che sviluppano o addestrano modelli generativi e entità che diffondono questi modelli in sistemi di intelligenza artificiale, rendendoli disponibili agli utenti finali. Specialmente nel caso dell'IA generativa, i sistemi di IA possono essere offerti come strumenti o servizi per generare contenuti. Inoltre, molti tipi di modelli di IA: (a) includono rappresentazioni di dati di addestramento; e (b) sono concessi in licenza o resi open source per l'uso da parte di altri. Le entità coinvolte in questi processi o nei vari passaggi di un processo possono trovarsi in giurisdizioni diverse. Per essere efficaci, è necessario creare, conservare e trasmettere i registri dei dati di addestramento in questo ecosistema.

Di conseguenza, le norme devono essere redatte in modo ampio per essere significative e non essere facilmente aggirate. L'obbligo di conservare registri accurati dovrebbe risalire all'inizio dello sviluppo del modello e continuare a essere applicato alle entità successive che forniscono un modello di base (indipendentemente dal fatto che abbiano sviluppato il fondamento da sole o che lo abbiano ottenuto tramite assegnazione/licenza) per fornire una catena completa di utilizzo e garantire che l'obbligo non possa essere facilmente eluso.

Inoltre, per evitare il "riciclaggio dell'IA", è fondamentale che tale obbligo si estenda a tutti i sistemi e modelli di IA resi disponibili nell'UE o che generano output utilizzati nell'UE, indipendentemente dalla giurisdizione in cui lo sviluppo (compreso il pre-addestramento, l'addestramento, l'ottimizzazione, la convalida, il collaudo o l'adattamento di un sistema di IA pre-addestrato e la generazione di contenuti) potrebbe aver avuto luogo. Pertanto, questa disposizione deve essere mantenuta e rafforzata per raggiungere questo obiettivo, anche per garantire che i contenuti illegali generati dall'IA utilizzando un modello di base non vengano resi disponibili nell'UE.

Infine, oltre a regolamentare i sistemi di IA generativa che coinvolgono modelli di base, i quadri normativi dovrebbero applicarsi anche ai sistemi di IA che imitano la voce, l'immagine o l'identità di un individuo o che forniscono la capacità di clonare la voce, l'immagine o l'identità di un individuo senza autorizzazione (i cosiddetti "deep fakes").

3) Quali registri devono essere tenuti dagli sviluppatori e dagli operatori di IA a fini di divulgazione?

Risposta breve: *Proponiamo un "sistema a due livelli". Agli sviluppatori e operatori dei sistemi e modelli di IA dovrebbe essere richiesto di:*

1. *rendere pubblicamente disponibili un primo livello di informazioni sufficientemente dettagliate sull'uso dei dati di addestramento e di altri materiali o contenuti protetti dal diritto d'autore, al fine di consentire a parti con legittimo interesse, come i titolari del diritto d'autore, di effettuare una determinazione iniziale (prima facie) su se e come i loro contenuti siano stati coinvolti nel processo di addestramento e, successivamente, di intraprendere azioni appropriate.*

2. *su richiesta di parti con legittimo interesse, compresi i titolari di diritti, fornire accesso completo ai registri dei dati di addestramento.*

Questo sistema a due livelli garantirebbe che le parti con interessi legittimi possano far valere i propri diritti, proteggendo al contempo gli sviluppatori e gli operatori di sistemi e modelli di intelligenza artificiale (compresi i modelli di base) da richieste futili di soggetti senza interessi legittimi o da parte di concorrenti.

Allo scopo di presentare pubblicamente il primo livello di informazioni sui dati di addestramento utilizzati, le "Model Cards", se preparate con sufficiente dettaglio, forniscono una base esistente per fare ciò. Nel sottolineare l'importanza delle "Model Cards", l'International Association of Privacy Professionals (IAPP) afferma:

> *È chiaro il motivo per cui l'intelligenza artificiale responsabile è importante. La paura della discriminazione, l'incapacità di distinguere le informazioni errate e persino il futuro della TV dipendono dall'uso sicuro e trasparente della tecnologia. Le organizzazioni che sviluppano e distribuiscono l'IA, in particolare l'IA generativa, hanno adottato le "Model Cards" come uno dei modi per promuovere la decifrabilità e raggiungere quella trasparenza*[21].

Le "Model Cards"[22] sono state introdotte da Google con l'obiettivo specifico di aumentare la trasparenza,[23] inclusa quella relativa ai dati di addestramento[24].

Per la seconda categoria di registrazione e divulgazione, le informazioni di primo livello, compresa una "sufficientemente dettagliata sintesi", non sono sufficienti. I record dei dati deve indicare, in modo dettagliato, i dati di addestramento utilizzati. Senza questo, i titolari dei diritti sarebbero incapaci di determinare se i loro specifici elementi di contenuto sono stati utilizzati, cioè a meno che i registri di addestramento includano dettagli chiari sul contenuto utilizzato nell'addestramento.

Ad esempio, nella musica registrata, come minimo, sarebbero necessari identificatori standard (come l'ISRC se disponibile), l'artista e il titolo della traccia. Per i testi, sarebbe necessario un titolo dell'opera e alcune informazioni sull'artista (esecuzione o composizione).

[21] IAPP '5 cose da sapere sulle Model Cards: https://iapp.org/news/a/5-things-to-know-about-ai-model-cards/

[22] Model cards sono descritte in https://modelcards.withgoogle.com/about

[23] Il documento tecnico "Model Cards for Model Reporting": https://arxiv.org/abs/1810.03993 spiega come aumentare la trasparenza

[24] Un mezzo per creare un model card utilizzando il «Model Card Toolkit» è definito in https://cloud.google.com/blog/products/ai-machine-learning/create-a-model-card-with-scikit-learn. Questa risorsa spiega che "Una model card fornisce anche informazioni sulla costruzione del modello, compresa la sua architettura e i dati di addestramento utilizzati".

Il dataset stesso è un record a sé. Ciò è generalmente evidente o almeno può essere dedotto nei dataset aperti, offerti, ad esempio, attraverso Hugging Face (una piattaforma di apprendimento automatico e scienza dei dati in cui gli utenti possono condividere set di dati di addestramento, condividere modelli addestrati e diffondere applicazioni di intelligenza artificiale).

Ad esempio, un dataset DISCO-10M[25] include informazioni sull'artista e sul nome della traccia; un altro dataset MusicCaps[26] include identificatori che possono essere collegati facilmente al contenuto sottostante. Un ulteriore dataset AudioSet[27] contiene collegamenti al contenuto sottostante. Questi dataset mostrano come i dati di addestramento possano (e già lo fanno) includere, identificare e collegarsi al contenuto in modo efficacemente auto-documentato e al livello di dettaglio proposto in questo documento.

Questi dataset di esempio vengono forniti in modo utile anche insieme a uno strumento per sfogliare ed eseguire facilmente ricerche tra i dati. Anche se alcuni dataset accessibili pubblicamente potrebbero non contenere uno strumento di visualizzazione del dataset, una volta scaricato il dataset, i dati al suo interno diventano accessibili e possono essere registrati.

Questo è vero anche per i dataset privati, sebbene sarebbe necessaria una base per fornire accesso a questi dati. Tale base non deve essere onerosa come avviare cause legali e procedure di scoperta. Dovrebbero essere disponibili informazioni sufficientemente dettagliate al pubblico e le parti con interessi legittimi dovrebbero poter richiedere l'accesso ai registri completi dei dati di addestramento.

Nel caso di risorse internet che comprendono contenuti protetti da diritto d'autore o potenzialmente protetti da diritto d'autore, come immagini, testi, registrazioni audio, ecc., è pratica comune che i collegamenti al contenuto siano compresi o inclusi nel dataset di addestramento[28]. Questi collegamenti sarebbero essenziali per il registro. Nel caso di dataset privati, sia il contenuto stesso che i collegamenti al contenuto sarebbero parte integrante dei dati di addestramento e costituirebbero i registri.

Quando i modelli addestrati sono resi disponibili in base ad accordi aperti o privati, devono essere forniti dati sui dataset di addestramento utilizzati nel pre-addestramento o nell'ulteriore addestramento del modello insieme al modello. Questi dati potrebbero, ad esempio, essere inclusi come parte della "Model Card" (spiegata ulteriormente di seguito) o accanto ad essa.

[25] Il dataset DISCO-10M è disponibile all'indirizzo https://huggingface.co/datasets/DISCOX/DISCO-10M

[26] MusicCaps è su https://www.kaggle.com/datasets/googleai/musiccaps e ci sono alcune ulteriori copie di questo set di dati anche su https://huggingface.co/datasets/google/MusicCaps

[27] È possibile accedere al set di dati AudioSet all'indirizzo http://research.google.com/audioset/dataset/index.html

[28] V., ad esempio, la FAQ LAION https://laion.ai/faq in cui si afferma che i set di dati LAION sono "elenchi di URL" alle immagini originali insieme agli ALT Text trovati collegati a quelle immagini"

I registri completi dovrebbero essere in formato leggibile dalla macchina (in gergo “machine-readable”) e conservati per 10 anni. Dati sul formato del dataset e informazioni necessarie per comprendere il dataset dovrebbero anche essere archiviati.

Per quanto riguarda come questi registri di dati potrebbero essere espressi, il concetto di "Model Cards" introdotto da Google offre una possibile formulazione per riassumere i dati di addestramento.

Le "Model Cards" mirano ad aumentare la trasparenza e forniscono regolarmente dettagli sui dati di convalida e sono state progettate anche per fornire informazioni sui dati di addestramento. I registri dettagliati potrebbero essere conservati in un database o in un sistema di gestione o monitoraggio dei dati di apprendimento automatico.

4) Perché è tecnicamente fattibile rispettare l'obbligo di conservazione dei registri e di divulgazione?

***Risposta breve**: le entità che conducono lo sviluppo e la diffusione di sistemi e modelli di intelligenza artificiale (compresi i modelli di base) gestiscono processi estremamente sofisticati, specialmente per quanto riguarda l'accesso, la preparazione e l'uso dei dati di addestramento. La registrazione e la divulgazione dei dataset di addestramento da parte sia dei grandi che dei piccoli sviluppatori e operatori già avvengono ed è un passo molto semplice nel processo complessivo. Gli sviluppi dell'industria e la ricerca accademica forniscono già strumenti, stabiliscono le best practice e dimostrano i benefici pratici che derivano dal monitoraggio attento e dalla cura dei dati di addestramento.*

Una quantità significativa di registrazione dei contenuti utilizzati nei processi di intelligenza artificiale avviene già, anche al momento della generazione del dataset (come indicato sopra).

Il lavoro di assemblaggio di dataset per l’apprendimento automatico e l’intelligenza artificiale riguarda la costruzione, la pulizia e, ove necessario, l’etichettatura degli elementi del dataset. Un lavoro significativo viene svolto anche per isolare elementi di dati distorti, tossici o patologici che avrebbero un impatto negativo sui risultati dell’addestramento. I registri dei dati di addestramento fanno già parte dei processi intrapresi dai team responsabili della creazione o della diffusione dei dataset, poiché la registrazione dei dati è parte integrante della preparazione e dell'utilizzo dei dati di addestramento.

È quindi del tutto fattibile per gli sviluppatori e gli operatori di intelligenza artificiale che si addestrano sui contenuti e sui dati altrui generare, conservare e divulgare registrazioni di tali contenuti e dati. Ogni suggestione che ciò non sia fattibile è semplicemente non credibile.

Quali best practice sono già stabilite?

Esistono significativi vantaggi per le aziende di machine learning e intelligenza artificiale che derivano dalla gestione, monitoraggio e cura dei dataset di addestramento a livello dettagliato. Ecco perché esiste una vasta gamma di strumenti e framework per svolgere questo lavoro.

Poiché i moderni processi di apprendimento automatico dipendono da dataset molto ampi e da un'accurata messa a punto, è pratica comune per gli operatori di intelligenza artificiale utilizzare sistemi di gestione dati su larga scala. Gli strumenti di gestione dei dati per l'apprendimento automatico vengono utilizzati principalmente perché forniscono dati di addestramento migliorati che portano a risultati e modelli migliori. Il beneficio delle soluzioni di gestione dei dati per l'apprendimento automatico è quello di "aiutare a comprendere, visualizzare e curare i dati per l'addestramento, scoprire dati corrotti come esempi con etichette sbagliate e individuare casi limite difficili", mentre "gli strumenti di gestione dei dati per l'apprendimento automatico sono spesso utilizzati quotidianamente, dimostrando il loro valore per individui e team. Nel parlare con gli acquirenti, il team di apprendimento automatico di Waymo (un'azienda di tecnologia per la guida autonoma) sembra aver valutato il ritorno sugli investimenti in diversi modi: 1) realizzando il valore dei dataset esistenti; 2) riducendo la spesa per l'annotazione; 3) debugging del dataset; e 4) messa a punto dell'apprendimento automatico e miglioramenti delle prestazioni."[29]

Esempi:

- Le soluzioni di gestione dati e registrazione per l'apprendimento automatico sono già ampiamente utilizzate. Whylabs[30] è uno dei fornitori di soluzioni di registrazione dati per l'apprendimento automatico, offrendo il prodotto whylogs[31] in forma di software open-source. È importante notare che l'efficienza e i miglioramenti delle prestazioni nell'addestramento dell'apprendimento automatico sono citati come risultato della registrazione dei dati di addestramento[32]. WhyLabs afferma che "una scarsa qualità dei dati... può causare costosi fallimenti" e che "mantenere sano un modello di IA richiede che gli sviluppatori e i data scientist siano consapevoli dei cambiamenti nella qualità e coerenza dei loro dati", cosa che può essere affrontata "con la libreria open-source di monitoraggio dei dati whylogs". L'azienda Scale[33] fornisce Nucleus, una soluzione di gestione dati per l'apprendimento automatico. Nucleus consente "migliori modelli di apprendimento automatico attraverso l'esplorazione, la cura e l'assicurazione della qualità dei dati".[34]
- SuperAnnotate offre "soluzioni avanzate di cura e gestione dei dati per creare dataset sani per modelli ad alte prestazioni".[35] Tra le sue ampie capacità di gestione dei dati, SuperAnnotate può gestire ed esportare i metadati del set di addestramento a livello di progetto e di elemento.[36]

[29] https://medium.com/memory-leak/ml-data-management-a-primer-a635a5eac858

[30] Informazioni su Whylabs sono disponibili all'indirizzo https://whylabs.ai/about

[31] Le informazioni sui prodotti Whylogs sono disponibili su https://whylabs.ai/whylogs

[32] Miglioramenti delle prestazioni derivanti dalla gestione dei dati di machine learning citati in https://medium.com/whylabs/whylogs-embrace-data-logging-a9449cd121d

[33] Scale può essere trovato su https://scale.com/

[34] Informazioni sulla gestione dei dati di Machine Learning di Nucleus in https://scale.com/nucleus

[35] Superannotate https://www.superannotate.com/data-curation

[36] Le informazioni sull'API dei metadati SuperAnnotate sono disponibili su https://superannotate.readthedocs.io/en/stable/api_reference/api_metadata.html#item-metadata

- Apache Airflow[37] ha funzionalità di registrazione integrate che utilizzano un framework[38] open-source per registrare i log. Airflow è altamente scalabile e adatto per elaborare dataset molto grandi.
- Alcuni ambienti di apprendimento automatico gestiscono dati in streaming, e Apache Kafka[39] è uno strumento progettato per agire come intermediario tra grandi fonti di dati e l'ambiente di creazione del modello. Può gestire e registrare i dati. Un progetto realizzato con Kafka traccia i dati dai sistemi di gestione del motore delle auto per l'analisi predittiva dei guasti. In questa applicazione, Kafka acquisisce e memorizza flussi di dati nello scenario dell'Internet of Things (IoT), elaborando dati derivati da 100.000 auto in tempo reale.[40]
  Splunk[41] è uno strumento di ingegneria dei dati orientato all'elaborazione e alla gestione di dataset molto grandi con l'obiettivo di produrre dataset di qualità per l'utilizzo in applicazioni basate sui dati e di potenziare prestazioni e risultati attraverso l'osservabilità dei dati. In particolare, una funzionalità di Splunk permette il tracciamento della derivazione dei dati che "consente agli ingegneri di: Identificare la fonte dei record di dati; Tracciare la cronologia delle trasformazioni dei record attraverso un flusso di lavoro dei dati."

Esistono anche sistemi di monitoraggio per l'apprendimento automatico che forniscono ampie capacità di monitoraggio e registrazione in tutto il processo di apprendimento automatico, compreso il monitoraggio del processo e delle macchine utilizzate, e questi strumenti forniscono anche la possibilità di tenere traccia dei dati e della registrazione. Alcuni esempi includono:

- NetflixMetaflow[42]. Questo framework di apprendimento automatico, che include il monitoraggio tramite un'interfaccia utente grafica open source, "consente ai data scientist di monitorare i propri flussi di lavoro in tempo reale, tenere traccia degli esperimenti e visualizzare registri e risultati dettagliati per ogni attività eseguita" e può "gestire il nostro archivio esistente composto da milioni di esecuzioni, alcune delle quali contengono decine di migliaia di attività senza intoppi". Prevede inoltre il "versioning integrato dei dati"[43] all'interno degli ambienti, inclusa l'addestramento dei modelli di base.
- Uber Michaelangelo[44], che copre il "flusso di lavoro ML end-to-end" ed è progettato per "gestire dati, addestrare, valutare e diffondere modelli, effettuare previsioni e monitorare le previsioni".

[37] Le informazioni sul flusso di Apache Airflow sono disponibili all'indirizzo https://airflow.apache.org

[38] La documentazione della funzione di registrazione open source Python è disponibile su https://docs.python.org/3/library/logging.html

[39] Apache Kafka è su https://kafka.apache.org

[40] Registrazione dei dati nel case study di gestione di AI engine: https://github.com/kaiwaehner/hivemq-mqtttensorflow-kafka-realtime-iot-machine-learning-training-inference (I dati sono memorizzati in MongoDB).

[41] Strumento di osservabilità dei dati Splunk: https://www.splunk.com/en_us/blog/learn/data-observability.html

[42] Netflix Metaflow: https://netflixtechblog.com/open-sourcing-a-monitoring-gui-for-metaflow-75ff465f0d60

[43] Controllo delle versioni dei dati in Dolly con Metaflow: https://outerbounds.com/blog/train-dolly-metaflow/

[44] Uber Michaelangelo: https://www.uber.com/en-GB/blog/michelangelo-machine-learningplatform/

Uber afferma che "abbiamo scoperto che la creazione e la gestione delle pipeline dei dati sono tipicamente una delle parti più costose di una soluzione completa di apprendimento automatico" e Michelangelo è progettato specificamente "per costruire pipeline affidabili, uniformi e riproducibili per la creazione e la gestione di dati di addestramento e previsione su larga scala".

Cosa dice la letteratura accademica?

Esistono numerose pubblicazioni accademiche nei settori dell'IA e dell'apprendimento automatico che propongono metodi di registrazione dei dati di addestramento e citano i benefici derivanti da tale pratica. Molte di queste pubblicazioni affrontano in modo specifico la gestione e la cura dei dati come chiavi per aumentare l'efficienza e l'accuratezza nell'addestramento su dataset molto ampi, spesso utilizzando approcci ottimizzati per ridurre o pulire i dati mediante strumenti di gestione dei dati. Nessuna di queste pubblicazioni cita il tracciamento di dati come un problema; al contrario, citano notevoli vantaggi dalla gestione dei dati. Per citare solo alcune di queste pubblicazioni:

- "A Survey on Data Collection for Machine Learning"[45] riconosce la raccolta e la cura dei dati come un ostacolo e propone strumenti di gestione dei dati (come quelli descritti sopra) come parte fondamentale della soluzione.
- "Coresets for Data-efficient Training of Machine Learning Models"[46] mostra che un'attenta selezione di un sottoinsieme di dati di addestramento può aumentare l'efficienza dell'addestramento senza compromettere i risultati dell'addestramento. Questo articolo mostra che non solo è possibile curare enormi dataset di machine learning, ma che è possibile eseguire l'ottimizzazione matematica per selezionare un sottoinsieme ottimale.

Naturalmente, quel sottoinsieme deve essere indicato e selezionato per l'addestramento che può essere facilmente registrato.

- 'Data Quality Toolkit: Automatic assessment of data quality and remediation for machine learning datasets'[47], in cui gli autori "creano uno strumento in grado di rilevare, spiegare e risolvere problemi nei dati e catturare in modo sistematico e automatico tutte le modifiche apportate ai dati", dimostrando che questo può migliorare significativamente i risultati dell'addestramento sui dati.
- 'Dataset Pruning: Reducing Training Data by Examining Generalization Influence'[48] riporta un lavoro in cui un'analisi di ciascun elemento di dati "riduce del 50% il tempo di convergenza con una diminuzione dell'accuratezza del test dell'1,3%". Il tracciamento degli elementi di dati è essenziale per questo processo che produce miglioramenti delle prestazioni.

[45] Un'indagine sulla raccolta di dati per l'apprendimento automatico: https://arxiv.org/abs/1811.03402

[46] Coreset per l'addestramento efficiente dei dati dei modelli di Machine Learning: https://arxiv.org/abs/1906.01827

[47] Data Quality Toolkit: valutazione automatica della qualità dei dati e correzione per i set di dati di machine learning: https://arxiv.org/abs/2108.05935

[48] Dataset pruning: riduzione dei dati di addestramento esaminando l'influenza della generalizzazione: https://arxiv.org/abs/2205.09329

- 'Management of Machine Learning Lifecycle Artifacts: A Survey'[49] identifica dataset e registri come artefatti del processo generale di apprendimento automatico ed esamina i metodi per gestire in modo sistematico questi dati. Gli autori valutano 60 di tali sistemi.
- 'Quantifying Transparency of Machine Learning Systems through Analysis of Contributions'[50] si concentra sui benefici della trasparenza e della fiducia derivanti dalla chiarezza sull'addestramento dell'apprendimento automatico, compreso il tracciamento dei dati di addestramento sottostanti. Guardando all'apprendimento automatico in settori come la sanità o il business, gli autori individuano la necessità di "continua comprensione dell'adeguatezza delle parti che hanno creato il modello o dei dati utilizzati per addestrarlo". Quindi "presentano un metodo per ottenere una misura quantificabile in grado di classificare la trasparenza delle pipeline di processo utilizzate per generare modelli di apprendimento automatico e altri asset di dati, in modo che gli utenti, i revisori e altre parti interessate possano acquisire fiducia nella possibilità di convalidare e fidarsi delle fonti di dati e dei contributori umani nei sistemi su cui si basano".
- Gli autori di 'Software Logging for Machine Learning'[51] partono dal presupposto che "i log di sistema svolgano una funzione critica nei sistemi intensivi di software" e quindi esaminano la complessità della gestione dei dati di log in un'azienda di grandi dimensioni. Poi gli autori "presentano un approccio sistematico e strutturato per generare dati di log che non soffrono delle sfide e dei problemi identificati e che viene poi ottimizzato per l'uso nell'apprendimento automatico". Ciò fornisce conferma "che questo approccio affronta le sfide e i problemi identificati".
- Gli autori di 'Training Data Distribution Search with Ensemble Active Learning'[52] sostengono che alcuni elementi di dati in grandi dataset di addestramento possano influire negativamente sui risultati e presentano un modo per ispezionare e filtrare gli elementi dei dati di addestramento. I "risultati forniscono forti evidenze empiriche che ottimizzare la diffusione dei dati di addestramento può comportare significativi benefici". Naturalmente, il tracciamento dei dati da utilizzare è essenziale per il processo proposto.

La proposta di Regolamento per l'Intelligenza Artificiale (AI Act)

Con la proposta di regolamento sull'Intelligenza Artificiale (Artificial Intelligence Act) ora in corso di approvazione finale, l'Unione europea ha l'opportunità unica di assumere un ruolo guida a livello globale nella creazione di un quadro trasparente ed efficace per l'intelligenza artificiale (IA). L'obiettivo del regolamento è stimolare l'innovazione e creare nuove opportunità commerciali, garantendo al tempo stesso che l'intelligenza artificiale si sviluppi in modo responsabile, comprensibile e sostenibile.

[49] Management of Machine Learning Lifecycle Artifacts: A Survey, è disponibile all'indirizzo: https://arxiv.org/abs/2210.11831

[50] Quantificare la trasparenza dei sistemi di apprendimento automatico attraverso l'analisi dei contributi: https://arxiv.org/abs/1907.03483

[51] Registrazione software per l'apprendimento automatico: https://arxiv.org/abs/2001.10794

[52] Ricerca della diffusione dei dati di addestramento con Ensemble Active Learning: https://arxiv.org/abs/1905.12737v2

Per garantire ciò e assicurare che i diritti fondamentali esistenti siano rispettati nella pratica, è essenziale che gli sviluppatori e gli operatori di sistemi e dei modelli di intelligenza artificiale (compresi i modelli di base) conservino registri dettagliati sull'uso dei dati utilizzati per l'addestramento di tali sistemi e modelli, comprese informazioni sull'utilizzo e la provenienza di materiali o contenuti di parti terze, e successivamente rendano queste informazioni disponibili alle parti che hanno interessi legittimi. Ciò includerebbe coloro che detengono diritti di proprietà intellettuale, compreso il diritto d'autore, un diritto fondamentale ai sensi dell'articolo 17, paragrafo 2, della Carta dei diritti fondamentali dell'UE[53].

La proposta del Parlamento Europeo all'articolo 28b(4) di obbligare i fornitori di modelli di base a registrare i contenuti utilizzati per addestrare i modelli è un passo nella giusta direzione, ma per essere veramente significativa tale proposta deve essere ulteriormente sviluppata, o migliorando il testo del Parlamento o integrandola nel testo del Consiglio.

Questo documento fornisce una guida tecnica per spiegare la necessità di un obbligo di mantenimento di registri dettagliati sui dati utilizzati per l'addestramento dei sistemi AI e include una proposta di emendamento su come tale obbligo possa essere implementato in modo efficace, ragionevole ed equilibrato.

SINTESI: SPIEGAZIONE DELLA RACCOMANDAZIONE IN BREVE

Il problema: sebbene la mancata conformità al diritto d'autore non sia un problema nuovo, la capacità dei modelli di intelligenza artificiale generativa di assimilare, copiare e appropriarsi di contenuti protetti solleva nuove questioni per i titolari dei diritti. La sfida principale è essere in grado di applicare le leggi esistenti quando è quasi impossibile accertare se e come tali contenuti siano stati utilizzati per addestrare i modelli di intelligenza artificiale.

La soluzione: questo problema fondamentale può essere risolto se le entità chiave nella catena dell'intelligenza artificiale generativa (ossia gli sviluppatori e gli operatori dei sistemi e dei modelli di intelligenza artificiale, compresi i modelli di base) sono tenute a rispettare i principi fondamentali di governance dei dati. In particolare, in prima battuta, dovrebbero conservare e rendere pubblicamente disponibili informazioni sufficientemente dettagliate sull'uso dei dati di addestramento e di altri materiali o contenuti protetti dal diritto d'autore, al fine di consentire alle parti con un interesse legittimo, come i titolari dei diritti d'autore, di determinare in via preliminare (prima facie) se e come i loro diritti siano stati lesi e, in seconda istanza, coloro che hanno interessi legittimi dovrebbero essere in grado di richiedere e ricevere registrazioni complete dei dati di addestramento, comprese opere o altri materiali protetti. Questo sistema a due livelli garantirebbe che le parti legittimate possano far valere i propri diritti, proteggendo al tempo stesso gli sviluppatori e gli operatori di sistemi e modelli di AI (compresi i modelli di base) da richieste futili da parte di soggetti senza interessi legittimi da parte dei concorrenti.

[53] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

In ambito di diritto d'autore, questo è fondamentale per far funzionare nella pratica le eccezioni previste dagli articoli 3 e 4 della Direttiva DSM, laddove applicabili. Senza una registrazione dei dati, i titolari dei diritti non hanno alcun modo di garantire che l'accesso al loro contenuto utilizzato per l'addestramento dei modelli IA sia avvenuto legalmente, che siano stati rispettati eventuali opt-out e che siano state ottenute le licenze e le autorizzazioni necessarie. Tuttavia, è altrettanto fondamentale per tutti i cittadini, ma anche per le entità operanti nel campo dell'IA, di essere in grado di garantire la responsabilità in relazione alle violazioni dei loro diritti fondamentali.

Ambito di applicazione: l'obbligo di conservare registrazioni accurate dei dati dovrebbe essere posto in capo allo sviluppatore del modello di base e continuare a essere applicato alle entità a valle che forniscono tale modello (sia che lo abbiano sviluppato in proprio o che gli sia stato assegnato/licenziato in altro modo), ciò per garantire che l'obbligo non possa essere facilmente eluso. Inoltre, per evitare il "riciclaggio di intelligenza artificiale", è fondamentale che tale obbligo si estenda a tutti i sistemi resi disponibili nell'UE o che generano output utilizzati nell'UE, indipendentemente dalla giurisdizione in cui potrebbe essere avvenuto lo sviluppo (incluso il pre-addestramento, l'addestramento, il raffinamento, la convalida, i test, l'adattamento di un sistema di intelligenza artificiale pre-addestrato o la generazione degli output). Ciò anche per garantire che contenuti illeciti, generati da modelli di base addestrati fuori dall'UE, non possano essere resi disponibili all'interno dell'UE.

Fattibilità: le entità che conducono lo sviluppo e la diffusione di sistemi e modelli di intelligenza artificiale (compresi i modelli di base) gestiscono processi estremamente sofisticati, specialmente per quanto riguarda l'accesso, la preparazione e l'uso dei dati di addestramento. La conservazione dei registri viene già effettuata da sviluppatori e fornitori di sistemi e modelli di intelligenza artificiale generativa, sia grandi che piccoli ed è una pratica molto semplice all'interno del processo complessivo. Le innovazioni industriali e la ricerca accademica forniscono già strumenti, stabiliscono le best practice e dimostrano i benefici pratici derivanti dal monitoraggio e dalla cura dei dati di addestramento. Pertanto, è del tutto fattibile per gli sviluppatori e gli operatori di sistemi di intelligenza artificiale che addestrano i loro modelli utilizzando i contenuti e i dati di altri generare, conservare e dare accesso alle registrazioni di tali contenuti e dati.