

Alla c.a. del Presidente Alberto Luigi Gusmeroli
Alla c.a. della Commissione Attività produttive della Camera dei deputati

Roma, 3 ottobre 2023

Illustre Presidente Gusmeroli,
Illustre Commissione,

EGAIR - *European Guild for Artificial Intelligence Regulation* - associazione senza scopo di lucro promossa dall'associazione italiana MeFu - Mestieri del Fumetto - Vi ringrazia per l'invito a partecipare al ciclo di audizioni nell'ambito dell'indagine conoscitiva sulle opportunità e i rischi dell'intelligenza artificiale per il sistema produttivo.

L'Associazione - che riunisce oggi oltre 100.000 creativi, artisti visuali, fumettisti, illustratori, doppiatori insieme a case editrici e sindacati professionali di diversi Paesi europei - è nata per proporre alle Istituzioni europee una regolamentazione sullo sfruttamento dei dati e delle opere creative da parte delle società AI.

Fin dalla sua costituzione, l'Associazione opera a livello nazionale ed europeo per sensibilizzare le Istituzioni sui danni prodotti e i rischi futuri di un utilizzo non regolamentato dell'intelligenza artificiale.

A questo proposito, EGAIR illustra di seguito le proprie considerazioni e proposte di modifica del testo della proposta di Regolamento UE (COM (2021) 2026) (cosiddetto AI Act).

**L'AI ACT e l'importanza della tutela delle opere,
delle professioni e dell'identità culturale europea**

*Considerazioni tecniche sul testo approvato
dal Parlamento europeo*

Ottobre 2023

Premessa

L'Intelligenza Artificiale applicata alla creatività

Le applicazioni di intelligenza artificiale consentono di progettare sistemi *hardware* e *software* atti a fornire all'elaboratore elettronico prestazioni, come ad esempio l'apprendimento continuo, per riprodurre o emulare alcune capacità umane. Queste applicazioni comportano significative implicazioni su tutti gli aspetti della vita culturale, lavorativa e sociale degli esseri umani, con conseguenze etiche e pratiche su cui è tempo di interrogarsi.

I settori delle arti e della creatività sono direttamente interessati da questo fenomeno. Le AI generative si propongono, infatti, sul mercato dell'arte come prodotti in grado di offrire un servizio sostitutivo del lavoro di artisti e creativi, producendo *output* comparabili alle opere di un essere umano in termini di qualità. L'esempio più famoso di questa tipologia di intelligenza artificiale sono le "AI Text to Image" (AI TTI). Il "risultato artistico" viene ottenuto con l'inserimento, da parte dell'utente, di un semplice *brief* testuale dal quale vengono elaborate delle immagini/disegni. Le AI TTI sono dunque in grado di interpretare un linguaggio di testo e utilizzarlo per generare immagini. Il risultato finale si basa esclusivamente su un processo di elaborazione di un insieme di dati/opere precedentemente raccolte. La qualità del prodotto finale (immagine o un disegno) dipende essenzialmente dai dati/opere raccolti nel tempo. Più è alta la qualità dei dati/opere, migliore è l'immagine elaborata e la sua conseguente competitività sul mercato. Se viceversa vengono inseriti dati scadenti il prodotto elaborato sarà di bassa qualità. Di conseguenza, sono le opere (immagini, disegni, bozzetti ecc) a determinare il valore di un'applicazione AI TTI. Ad esempio, un'applicazione che lavora sul *dataset* di un singolo stile visivo, sarà esclusivamente in grado di riprodurre immagini in quello specifico stile. È dunque grazie al contenuto dei *dataset*, le opere degli artisti in questo caso, che l'AI riesce a esprimersi producendo risultati anche di notevole qualità e varietà.

L'escamotage dell'uso per ricerca e il furto di opere sul web

Nel corso degli ultimi anni, attraverso l'escamotage dell'eccezione per uso di ricerca ("*fair use*" statunitense) molte start-up che operano nell'intelligenza artificiale hanno potuto rastrellare sul web miliardi di dati/opere, la gran parte delle quali sono state ottenute senza il consenso dei titolari aventi diritto. Titolari che, nella stragrande maggioranza dei casi, erano del tutto ignari che le opere presenti sul web potessero essere utilizzate e manipolate liberamente per essere strumento di *training* per i software AI. Nel lasso di pochi anni, si è assistito allo *spin-off* di queste società per creare aziende che invece hanno un manifesto e pubblico intento commerciale. Sono aziende che hanno potuto offrire dei servizi agli utenti esclusivamente grazie all'utilizzo e allo sfruttamento delle opere di migliaia di disegnatori, fumettisti, illustratori di tutto il mondo. Queste società, alcune quotate in borsa o partecipate da Fondi di investimento - si stanno arricchendo sfruttando il lavoro, il talento e la creatività di professionisti senza aver mai corrisposto loro alcun compenso.

Il testo approvato dal Parlamento europeo

Il Parlamento europeo ha approvato con modifiche la Proposta di Regolamento (COM (2021) 2026) che si pone l'obiettivo di armonizzare in ambito europeo le normative nazionali in materia di AI, lasciando agli Stati membri alcuni compiti sussidiari e relativi alla vigilanza e alle sanzioni (*governance* a livello nazionale). Lo spirito di fondo del Regolamento è di incentivare l'utilizzo e lo sviluppo di sistemi di AI in alcuni ambiti, fermo restando il rispetto di alcuni diritti fondamentali dell'Unione Europea: dignità umana, vita privata, non discriminazione ecc.

Nel corso dell'esame della Proposta di Regolamento, il Parlamento europeo ha approvato un'importante modifica al testo, proposta da EG AIR, che introduce nuovi obblighi in capo ai fornitori di modelli di base utilizzati nei sistemi di AI generativa.

Il nuovo articolo 28 b stabilisce, infatti, che essi devono:

- **rispettare** gli obblighi di trasparenza;
- **assicurare** adeguate garanzie contro la generazione di contenuti in violazione del diritto dell'Unione;
- **documentare e rendere disponibile al pubblico** un riepilogo sufficientemente dettagliato dell'uso dei dati protetti dal diritto d'autore utilizzati per allenare i sistemi di AI generativa.

Per quanto si tratti di un passo in avanti, EG AIR ritiene indispensabile **rafforzare il principio della trasparenza sui dati utilizzati** fino ad ora, ma anche **regolamentare per il futuro il fenomeno del "web scraping"** per evitare il perdurare di questa pratica illegittima in cui i dati degli artisti vengono sottratti e utilizzati per fornire servizi alternativi e in diretta concorrenza con quelli degli effettivi titolari dei dati.

Le soluzioni oggi prospettate, da una parte il "riepilogo" di cui all'articolo 28 b della proposta dell'AI Act e, dall'altra, il meccanismo dell'*opt-out*, rischiano di continuare ad avvantaggiare le società proprietarie di AI generative.

L'*opt-out*: un meccanismo inadeguato e tardivo

Ad oggi l'unico strumento previsto dalla normativa per tutelare gli artisti dal furto delle loro opere da parte delle AI è il cosiddetto "*opt-out*": un **meccanismo del tutto inadeguato e inefficace**, oltre che a completo ed esclusivo vantaggio delle aziende di AI.

Tale strumento è stato introdotto dall'articolo 4 della Direttiva Copyright (n. 2019/790) che prevede un'eccezione al *copyright* per operazioni di "*text and data mining*". Tale eccezione si applica "a condizione che l'utilizzo delle opere e di altri materiali non sia stato espressamente riservato dai titolari dei diritti in modo appropriato, ad esempio attraverso strumenti che consentano lettura automatizzata ("*machine readable*" in inglese) in caso di contenuti resi pubblicamente disponibili *online*." L'*opt-out* è questa espressione di utilizzo riservato.

A una prima lettura, potrebbe dunque sembrare che qualunque dato che non sia stato appropriatamente salvato e caricato *online* con le dovute modalità possa essere liberamente scaricato, riprodotto e utilizzato da aziende per il *training* di AI. A tal proposito, però, si osservi che:

- la Direttiva e le eccezioni che essa prevede sono state formulate in un momento storico precedente all'ascesa delle aziende che offrono servizi commerciali di intelligenza artificiale (Open AI, Stability AI, Midjourney, etc.);
- oggi sappiamo che servizi di AI generative possono essere usati e vengono usati per generare contenuti su una scala inimmaginabile nel 2019 e per scopi commerciali diversi da quelli previsti dal testo della Direttiva;
- i contenuti in questione vengono ottenuti addestrando le AI su materiale coperto da *copyright*. I servizi di queste aziende AI vengono poi offerti negli stessi mercati in cui lavorano i titolari dei dati originali a prezzi insostenibili per un artista.

Diversi trattati internazionali, tra i quali la Convenzione di Berna, stabiliscono che le eccezioni al *copyright* devono aderire a specifici passaggi. La Direttiva UE adotta questo strumento di protezione, prevedendo che le eccezioni siano applicabili in casi in cui non vadano a **pregiudicare in maniera irragionevole il legittimo interesse** dei diritti dei titolari. Tuttavia, al momento, questo non avviene e materiale coperto da *copyright*, di proprietà di cittadini europei, viene usato per competere con gli stessi sul mercato del lavoro, spesso con lo scopo di generare contenuti che siano simili agli stili degli autori.

Inoltre, anche qualora i creativi volessero esercitare il proprio diritto di *opt-out*, essi si trovano **spesso impossibilitati nel farlo** o nel verificare che la loro volontà sia stata rispettata. La maggior parte dei *dataset* delle aziende di AI sono infatti inaccessibili per la consultazione da parte del cittadino europeo, mentre i siti che permettono una consultazione del *dataset*¹ non sono facilmente navigabili e la ricerca di immagini richiede molto tempo². È inoltre da precisare che l'utilizzo che si fa di queste immagini non è mai specificato, ciò rende impossibile determinare quali immagini siano state usate per il *training* dei diversi modelli di AI.

È inoltre poco chiaro quale possa essere un meccanismo di *opt-out* efficace. La rapidità con cui vengono sviluppate nuove tecniche di *scraping* e nuovi *bot* rende infatti l'utilizzo di appositi metadati inutile. Tra l'altro non è neanche chiaro come i proprietari dei dati possano tutelarsi nel caso in cui siano soggetti terzi a caricare online materiale coperto da *copyright*. Quando si parla di immagini, ad esempio, tra le principali fonti per operazioni di *data mining* ci sono servizi di e-

¹ L'azienda Stability AI offre uno dei pochi casi di *dataset* aperti e consultabili in LAION 5b, una raccolta di 5 miliardi di immagini raccolte dalla no-profit tedesca LAION per scopi di ricerca e usato da Stability per sviluppare modelli commerciali.

² Di recente Open AI è stata coinvolta in diverse cause legali per aver addestrato il suo Chat-GPT usando anche copie pirata di decine di migliaia di romanzi, il cui accesso da parte dell'azienda non è chiaramente legale. Allo stesso modo è impossibile verificare se queste aziende abbiano preso materiale anche da siti i cui *terms of use* vieterebbero operazioni di *text and data mining* (come la popolare piattaforma Pinterest).

commerce come Amazon ed Ebay: l'autore della copertina di un libro non ha alcun controllo sui metadati che vengono usati da un soggetto terzo, sia esso un editore straniero o un privato, nel caricare la sua illustrazione nello store di Amazon.

Si osservi peraltro che alcune di queste aziende offrono la possibilità di esercitare *opt-out* "postumi", ovvero successivi alla raccolta dei dati e all'addestramento dei loro modelli³. Tuttavia, è ormai noto che il *machine un-learning* è, allo stato dell'arte, tecnicamente impossibile: un modello di AI non può "dimenticare" singoli dati senza azzerare tutto il modello.

Come più volte descritto al suo interno, la Direttiva n. 2019/790 auspica un mercato europeo in cui creativi e aziende trovino accordi e stipulino licenze economiche per lo sfruttamento di testi e dati. L'attuale situazione di **scarsa trasparenza e di opacità**, unita alla rapidità con cui si sono mosse le aziende di AI e allo squilibrio di potere tra esse e gli artisti europei, rende tuttavia impossibile una corretta applicazione dei principi della Direttiva e la nascita di un mercato europeo dei dati che sia sano e profittevole per tutti.

Proposte

EGAIR ritiene indispensabile e urgente intervenire, tramite la proposta di Regolamento, sui seguenti temi:

1. rafforzare la trasparenza sui dati e le opere già "rubati" dalle società di AI, senza chiedere il consenso né corrispondere un compenso per il loro utilizzo: il fatto che sia già avvenuto non li legittima;
2. garantire che questa situazione non si ripeta negli anni a venire, assicurando che dati e opere personali siano utilizzate per allenare i modelli di base delle AI previo consenso esplicito.

Sebbene il tema della protezione futura del diritto d'autore sembri trovare una più corretta collocazione nella disciplina del *copyright* o della tutela dei dati personali, riteniamo che sia fondamentale intervenire sull'AI Act, in quanto le tempistiche della legislazione europea, a cui si aggiunge la tornata elettorale del prossimo anno, non sono compatibili con la dinamicità del settore. Se la discussione si chiudesse con il testo attuale dell'AI Act, rimarrebbero ampie zone grigie nella normativa ed è evidente che, almeno per il prossimo triennio, non ci sarà alcuna

³ L'esempio più recente di queste pratiche è Dall-e3, un programma di AI generativo di immagini di Open AI. Il modello è già stato addestrato su un set di immagini il cui contenuto non è dato sapere. Eppure, Open AI offre agli artisti una sua interpretazione dell'*opt-out*: secondo le dichiarazioni dell'azienda, ogni artista può compilare un form online per ognuna delle sue immagini presenti su internet e inviarlo a Open AI. Future versioni di Dall-e (non si specifica quali) terranno in considerazione questi *opt-out* bloccando la generazione di immagini simili ai disegni inviati e agli stili dei loro lavori.

possibilità per i creativi europei di veder tutelati i propri diritti. L'Unione europea non può abdicare al proprio ruolo di fronte all'urgenza del tema.

Effettiva trasparenza

La previsione normativa di cui all'articolo 28 b dispone che i fornitori dei modelli di base utilizzati nei sistemi di AI generativa mettano a disposizione un **riepilogo** dell'uso dei dati protetti dal diritto d'autore utilizzati per allenare i sistemi di AI generativa.

È evidente che un riepilogo non possa essere efficace, trattandosi di uno strumento sommario e discrezionale. Le società che offrono servizi di AI dovrebbero mettere a disposizione **registri esaustivi** - contenenti cioè tutto il materiale coperto da *copyright* da loro utilizzato - aperti e facilmente consultabili da parte di chiunque (Istituzioni, operatori, associazioni, singoli cittadini).

Le caratteristiche essenziali che EGAIR ritiene indispensabile che tali registri abbiano sono:

- che siano consultabili per tipologia di file (testo, audio, video, immagini);
- che siano consultabili usando i nomi degli artisti;
- a ogni file dovrebbe essere associato l'url di provenienza e il nome del file;
- dovrebbe essere possibile effettuare ricerche testuali e per url;
- nel caso di immagini dovrebbe essere possibile effettuare ricerche per immagini;
- per ogni file dovrebbe essere incluso il tag usato per descrivere il file per scopi di *machine learning*;
- per ogni file dovrebbe essere disponibile l'url da cui il file proviene;
- nel caso di file audio dovrebbero essere disponibili tre secondi di file in previsione di futuri programmi di riconoscimento audio;
- per ogni file si dovrebbe specificare il numero di versioni di quel file presenti nel *dataset* e i rispettivi url di provenienza;
- se il file fosse stato ceduto sulla base di una licenza, dovrebbe essere specificato chi ha fornito il file.

È inoltre importante che i file non siano semplicemente dei link ai siti fonte dell'immagine, come avviene per il dataset LAION 5b. Questo per diversi motivi:

- il "*link rot*": 1 link su 200 statisticamente "muore" entro un mese dalla sua creazione per rimozione del suo contenuto dal sito. Diversi studi sui *dataset* ci confermano che il 20% dei link all'interno dei *dataset* più popolari sono morti dalla costruzione di detti *dataset*. Questo non solo impedisce di verificare se un file sia stato usato nel *training* di un AI, ma è anche dannoso per la scienza e la ricerca, dato che inficia la riproducibilità del *training* di un dato modello;
- la pratica di ospitare, in un *dataset* aperto, esclusivamente url è particolarmente vulnerabile ad attacchi di *data-poisoning*, come, ad esempio, soggetti che comprino url "morti" e rimpiazzino i dati originali con dati corrotti o contenuti illegali.

L'opportunità di creare un *Marketplace*

Tali caratteristiche renderebbero possibile, in futuro, la trasformazione dei registri in ***Marketplace* altamente controllati e monitorati**, in cui possano essere i titolari stessi dei dati a caricare i loro lavori in alta qualità e definizione a seguito della stipulazione di accordi con le aziende o gratuitamente. Questo spazio ideale dovrebbe:

- garantire la tracciabilità della proprietà di ogni file e dato caricato, in modo da rendere legalmente responsabile chi sceglie di farlo e da scoraggiare *upload* illegittimi;
- offrire la possibilità di scegliere per quali scopi di *machine learning* caricare i propri dati;
- garantire il *tracking* delle chiavi API usate per il download dei dati;
- avere un numero casuale di dati con *watermark* invisibili che possano essere usati per tracciare la fonte originale, nel caso qualcuno decidesse di diffondere il *dataset online* sotto una licenza differente.

Così facendo si avrebbe la garanzia che i dati usati per il *training* siano stati effettivamente autorizzati dai legittimi titolari. Un simile spazio sarebbe anche **in linea con gli obiettivi perseguiti dalla Direttiva 2019/790 e dal Data Act** nello stabilire delle possibili licenze per *text and data mining* e per stimolare la nascita di un mercato europeo delle AI.

L'AI Act si è trasformato in un documento molto più ambizioso di ciò che era originariamente, il cui scopo è quello di inquadrare il tema delle Intelligenze Artificiali nelle regole e nei principi già vigenti all'interno dell'Unione. Non garantire livelli di trasparenza elevatissimi porterà a un cortocircuito tra l'AI Act e questi principi.

Consenso preventivo

È fondamentale, per il futuro, evitare che le società di AI continuino a sfruttare indebitamente dati e materiale coperto da *copyright*. Pertanto, EGAIR chiede di:

- prevedere che i fornitori di modelli di base utilizzati nei sistemi di AI generativa debbano richiedere preventivamente il consenso esplicito e informato dei titolari dei dati relativi a persone e opere, in qualsivoglia forma, utilizzati per il training di modelli di AI. Sarà quindi necessario assicurarsi che ogni singolo dato/opera di terzi che viene inserito nei modelli di training sia autorizzato dai titolari aventi diritto;
- introdurre il "*training right*", che prevede che i contenuti di terzi siano utilizzati secondo termini e condizioni contrattuali precedentemente stabiliti con i titolari aventi diritto e che tale utilizzo, salvo diverso accordo tra le parti, sia soggetto ad una remunerazione adeguata e proporzionata ai ricavi derivanti dallo sfruttamento delle opere, come avviene per tutto il materiale coperto dal diritto d'autore. Ciò inoltre potrebbe essere attuato tramite accordi di licenza con il titolare avente o con accordi collettivi tramite le associazioni di categoria maggiormente rappresentative dei titolari aventi diritto.

La proposta di emendamento

Articolo 28 b AI Act	Proposta di emendamento di EGAIR
<p>4. I fornitori di modelli di base utilizzati in sistemi di IA specificamente destinati a generare, con vari livelli di autonomia, contenuti quali testi complessi, immagini, audio o video ("sistema di IA generativa") e i fornitori che specializzano un modello di base in un sistema di IA generativa, devono inoltre:</p>	<p>4. I fornitori di modelli di base utilizzati in sistemi di IA specificamente destinati a generare, con vari livelli di autonomia, contenuti quali testi complessi, immagini, audio o video ("sistema di IA generativa") e i fornitori che specializzano un modello di base in un sistema di IA generativa, devono inoltre:</p>
	<p>(a) richiedere preventivamente il consenso esplicito e informato dei titolari dei dati relativi a persone o opere, in qualsivoglia forma, utilizzati per il training di modelli di AI,</p>
<p>(a) rispettare gli obblighi di trasparenza di cui all'articolo 52, paragrafo 1,</p>	<p>(b) rispettare gli obblighi di trasparenza di cui all'articolo 52, paragrafo 1, e assicurarsi che ogni singolo dato/opera di terzi che viene inserito nei modelli di training sia autorizzato dai titolari aventi diritto,</p>
<p>(b) formano e, se del caso, progettano e sviluppano il modello di base in modo da garantire adeguate garanzie contro la generazione di contenuti in violazione del diritto dell'Unione, in linea con lo stato dell'arte generalmente riconosciuto e senza pregiudicare i diritti fondamentali, compresa la libertà di espressione,</p>	<p>(c) formano e, se del caso, progettano e sviluppano il modello di base in modo da garantire adeguate garanzie contro la generazione di contenuti in violazione del diritto dell'Unione, in linea con lo stato dell'arte generalmente riconosciuto e senza pregiudicare i diritti fondamentali, compresa la libertà di espressione,</p>
<p>(c) fatta salva la legislazione nazionale o dell'Unione in materia di diritto d'autore, documentare e rendere disponibile al pubblico un riepilogo sufficientemente dettagliato dell'uso dei dati di formazione protetti dal diritto d'autore.</p>	<p>(d) fatta salva la legislazione nazionale o dell'Unione in materia di diritto d'autore, documentare e rendere disponibile al pubblico un registro contenente tutti i dati di formazione protetti dal diritto d'autore,</p>
	<p>(e) utilizzare contenuti di terzi secondo termini e condizioni contrattuali precedentemente stabiliti con i titolari aventi diritto,</p>
	<p>(f) salvo diverso accordo fra le Parti, l'utilizzo di opere altrui da parte di società di AI è soggetto ad una remunerazione adeguata e proporzionata ai ricavi derivanti dallo sfruttamento delle opere, anche tramite accordi di licenza con il titolare avente diritto o con accordi collettivi tramite le associazioni di categoria maggiormente rappresentative dei titolari aventi diritto.</p>
	<p>4-bis. I registri di cui al comma 3 lettera d) devono presentare le seguenti caratteristiche:</p>

	(a) essere aperti e facilmente consultabili da parte di chiunque,
	(b) essere consultabili per tipologia di file (testo, audio, video, immagini),
	(c) essere consultabili usando i nomi degli artisti,
	(d) a ogni file dovrebbe essere associato l'url di provenienza e il nome del file,
	(e) deve essere possibile effettuare ricerche testuali e per url,
	(f) nel caso di immagini, deve essere possibile effettuare ricerche per immagini,
	(g) per ogni file deve essere incluso il tag usato per descrivere il file per scopi di machine learning,
	(h) per ogni file deve essere disponibile l'url da cui il file proviene,
	(i) nel caso di file audio, devono essere disponibili tre secondi di file in previsione di futuri programmi di riconoscimento audio,
	(l) per ogni file si deve specificare il numero di versioni di quel file presenti nel dataset e i rispettivi url di provenienza,
	(m) se il file viene sulla base di una licenza, deve essere specificato chi fornisce il file.

Chi Siamo

La EUROPEAN GUILD FOR AI REGULATION (www.egair.eu) è un'associazione senza scopo di lucro che riunisce oltre 100.000 creativi, artisti visuali, fumettisti, illustratori, doppiatori insieme a case editrici e sindacati professionali di diversi Paesi europei per proporre alle istituzioni europee una regolamentazione sullo sfruttamento dei dati e delle opere creative da parte delle società AI.

L'iniziativa è stata promossa dagli artisti italiani Lorenzo Ceccotti LRNZ e Francesco Archidiacono di MeFu, l'associazione Mestieri del Fumetto che ha svolto, nei mesi scorsi, un intenso lavoro di raccordo con le reti dei professionisti del settore di tutta Europa proponendo la sottoscrizione di un Manifesto in difesa dei diritti e della creatività dell'essere umano.

Il Manifesto (in appendice) è stato sottoscritto, fino ad oggi, da centinaia di artisti di fama mondiale come Milo Manara, Claire Wendling, AleKsi Briclot, Greg Rutkowksi, dalle associazioni Central Vapoeur (BEL), Ligue des Auteurs Professionnels (FR), SNAC (FR), Illustratoren Organisation (D) AÉ ARTEsÉTICA (SP), Associazione Nazionale Attori Doppiatori (IT), European Illustrators Forum, CAA Concept Art Association (USA), dalle case editrici Bao, Coconino e da festival del fumetto come il ComiCon e ARFI.

Per qualsiasi ulteriore informazione:

Francesco.schlitzer@verastudio.it

[Gaia Frascella@verastudio.it](mailto:Gaia.Frascella@verastudio.it)



EUROPEAN GUILD
FOR ARTIFICIAL
INTELLIGENCE
REGULATION

Siamo un Gruppo di artisti, creative, editori e associazioni europei uniti nel portare all'attenzione del pubblico come i nostri dati e le nostre proprietà intellettuali stiano venendo sfruttati senza il nostro consenso e su una scala mai vista prima. Questa situazione senza precedenti ci ha portati a unire le nostre forze per raggiungere le istituzioni europee. Se anche tu credi che i tuoi dati personali e il frutto del tuo lavoro creative non debbano essere impunemente sfruttati per profitto da una manciata di società e multinazionali, unisciti a noi in questa battaglia.

L'estate 2022 ha visto l'emergere di una nuova e incredibile tecnologia: le AI text-to-image (TTI) o image-to image. Ogni volta che viene sviluppata una tecnologia così rivoluzionaria, la nostra società deve monitorare il suo sviluppo, al fine di evitare qualunque tipo di danno o l'infrazione dei diritti umani. Questo non è ancora accaduto con la tecnologia AI. Queste tecnologie sono in grado di generare immagini o altri contenuti a partire da prompt (comandi) testuali o altri media forniti dall'utilizzatore. Per fare ciò, un'AI deve essere addestrata su un dataset di media. La qualità di un'AI generative è definita dalla qualità del suo dataset – per esempio, riguardo alle immagini, più foto e illustrazioni vengono usate nell'addestramento di un'AI, più stili quell'AI potrà replicare e più cose potrà fare. Ne consegue che i prodotti venduti dalle società di AI sono il risultato di operazioni su dataset, i quali contengono ogni tipo di dato, inclusi milioni di immagini coperte da copyright, foto private e altro materiale sensibile. Questi file sono stati raccolti attraverso un'operazione di scraping indiscriminato di tutto il web, senza alcun consenso da parte dei proprietari e delle persone rappresentate e stanno venendo utilizzati dalle società di AI per trarne profitto. Alcune di queste società stanno inoltre utilizzando i nomi degli artisti i cui lavori sono stati estratti dal web per pubblicizzare le loro AI e gli stili che esse possono replicare. Questo sfruttamento del nostro lavoro e dei nostri dati non solo non rispetta i requisiti minimi imposti dai diritti umani che regolano la nostra società: stanno anche danneggiando severamente il mercato creativo, segnandolo potenzialmente per sempre. Vediamo tutto ciò come solo l'inizio di una crisi che colpirà ogni tipologia di lavoro, creative o no. Il mercato delle arti visive è stato il primo a venir colpito unicamente per le sue vulnerabilità, che lo rendono una preda facile. È tempo di cambiare tutto ciò.

Questi sono i punti chiave che vogliamo introdurre:

- 1) Tutti i dati relativi a persone o opere, in ogni forma essi siano, siano essi già digitalizzati come file di testo, audio, video o di immagine o catturati dalla realtà attraverso videocamere, microfoni o qualsiasi mezzo di registrazione, non devono poter essere utilizzati per il training di modelli di AI senza l'esplicito consenso informato dei loro legittimi proprietari.
- 2) L'utilizzo di nomi di persone, nomi d'arte o opere che non sia coperto da una licenza di sfruttamento per il training di AI deve essere proibito dai software che consentono di inserire un "prompt" testuale o vocale per richiedere la generazione di un'immagine, video, testo o suono.
- 3) L'utilizzo di video, immagini, suoni e testi che non sia coperto da una licenza di sfruttamento per il training di AI deve essere proibito dai software che consentono di inserire media per richiedere la generazione di un'immagine, video, testo o suono.
- 4) Deve essere istituito un sistema di indicizzazione e certificazione standardizzato, "human e machine readable" delle attività delle AI e dei dataset di immagini, testi o suoni interamente o parzialmente prodotte con AI. Diciture come "interamente generato con AI", "realizzato utilizzando materiale generato con AI" dovrebbero diventare lo standard.
- 5) Non appare sufficiente il discrimine tra materiale "copyrighted" e "di pubblico dominio". Nei dataset possono infatti trovarsi anche dati personali sensibili, quindi protetti dalle leggi sulla privacy, ma non da copyright. Al tempo stesso, esistono casi di materiale pubblicato in un momento in cui non si poteva prevederne l'utilizzo nel contesto di dataset per il training dei modelli di Intelligenza Artificiale. Ogni singolo dato che viene inserito nei modelli di training deve essere curato e autorizzato dai legittimi proprietari e quindi inserito in maniera volontaria e informata dai singoli autori. Le aziende di AI devono produrre internamente materiali originali per il training o licenziare contenuti esterni secondo termini e condizioni contrattuali precedentemente stabilite con gli autori interessati.

Scopri di più su www.egair.eu

EGAIR™

EUROPEAN GUILD
FOR ARTIFICIAL
INTELLIGENCE
REGULATION

www.egair.eu

EGAIR

European Guild for AI Regulation

Un network di creativi, associazioni e aziende europei, nato in Italia su iniziativa di MeFu - Mestieri del Fumetto.

Il nostro obiettivo è la tutela dei dati personali e delle opere degli artisti europei.

Egair rappresenta più di 100.000 creativi da tutta Europa.

www.egair.eu



Lo scenario

- Le AI generative dipendono dai dati e dalle opere dei creativi. Senza di essi non possono essere addestrate.
- Le principali aziende (tutte extra-europee) che offrono servizi e modelli di AI generativa hanno illecitamente raccolto e usato i dati degli artisti europei per trarne profitto.
- I modelli di AI addestrati con questi dati vengono usati per fare una concorrenza diretta e sleale ai proprietari dei dati originali.
- Lo scenario attuale, privo di regole chiare, si presta ad abusi e pone enormi rischi per tutti.

Ragazze «spogliate» con l'Intelligenza artificiale, l'allarme nelle scuole. «Girano le foto di me nuda»

Le immagini sono create da un software generativo, ma appaiono incredibilmente realistiche

Regole

- Gli artisti e i creativi non sono contrari alle AI, ma contro lo sfruttamento non autorizzato delle loro opere a fini economici.
- Regole chiare sono il presupposto per un mercato sano e concorrenziale.
- I principi cardine per il legislatore dovrebbero essere il consenso e la trasparenza.
- Un mercato legale europeo delle AI deve coinvolgere anche i titolari dei diritti, a cui va data la possibilità di cedere le proprie opere con licenze apposite.
- L'AI Act è centrale. È essenziale che questo Regolamento stabilisca regole chiare per le AI generative e la tutela dei creativi.
- Importante anche il ruolo delle Autorità per la Protezione dei Dati Personali, per intervenire sugli illeciti e vigilare sul rispetto delle regole.

L'intervento del Parlamento

- Presidiare e stimolare le Istituzioni europee per rafforzare l'AI Act, introducendo strumenti concreti a garanzia della trasparenza e della manifestazione del consenso esplicito.
- Valutare di introdurre a livello nazionale norme per modificare nella stessa direzione la normativa nazionale in materia di copyright.