AI: la “spada di Damocle” dei deepfake sulle prossime elezioni

Alessandro Foti, Federico Aluigi, Jacopo Piemonte
26.1.2024

Tempo di lettura: 3'

La regolamentazione dell'intelligenza artificiale (AI), con focus sulle linee guida emanate da OpenAI in vista delle elezioni che si susseguiranno nel 2024 in vari paesi del mondo (Usa, inclusa). Come scongiurare il pericolo dei deepfake audio/video

Il **30 novembre 2022**, con il lancio di **ChatGpt**, si è aperto un capitolo fondamentale nell'utilizzo dell'intelligenza artificiale (AI), catalizzando il dibattito sulle implicazioni di questa tecnologia nelle nostre vite. Particolarmente rilevante è l'emergere dell'I**A generativa,** incarnata, per sentire diffuso, da **ChatGpt**, che rappresenta una frontiera complessa e critica nell'ampio panorama dell'IA.

A differenza dei modelli convenzionali, limitati alla realizzazione di compiti specifici, i sistemi generativi come ChatGpt sono progettati per creare autonomamente nuovi contenuti, che spaziano da testi a immagini, persino opere d'arte e video.

Leggi anche

ChatGpt: i vantaggi (attesi) del suo uso per consulenti e investitori

Il funzionamento delle intelligenze artificiali (AI)

Il funzionamento delle intelligenze artificiali generative è basato sulle **reti neurali complesse**, intese come modelli computazionali addestrati su enormi quantità di dati. Questi modelli, alimentati da un vasto corpus di informazioni, sono in grado di apprendere schemi, stili e strutture linguistiche e visive, consentendo loro di generare contenuti in modo realistico. Questa capacità di creare da sé nuovi contenuti offre un potenziale immenso in **svariati settori**, dalla creatività artistica alla scrittura automatica di testi complessi.

I deepfake audio/video

Tuttavia, detta innovazione solleva in parallelo una serie di **questioni etiche e rischi connessi**, soprattutto in relazione alla possibile generazione di contenuti falsi o manipolati. Esempi eclatanti come i deepfake utilizzati nelle recenti elezioni in Taiwan, evidenziano infatti il potenziale impatto negativo di queste tecnologie. Nel caso di specie, era stato diffuso via social un falso video del **deputato statunitense Rob Wittman** che si impegnava a sostenere militarmente Taiwan in caso di affermazione del Democratic progressive party. Tutto ciò è ovviamente suscettibile di influenzare la percezione della realtà.

A ciò, si aggiungano fenomeni come la **potenziale perdita di controllo sulla produzione di contenuti dannosi o illegali**, nonché la perpetuazione di pregiudizi e discriminazioni potenzialmente presenti nei dati di addestramento degli algoritmi.

È fondamentale, di conseguenza, affrontare questi rischi sviluppando meccanismi di regolamentazione e controllo che garantiscano un utilizzo etico e responsabile di tali tecnologie.

I meccanismi di regolamentazione e controllo: AI Act

L'Unione europea ha risposto a questa esigenza con il **Regolamento sull'intelligenza artificiale (AI Act)**, il quale, nella sua poliedricità, si occupa anche delle tecnologie generative. Durante la lunga stagione dei triloghi, il **19 novembre 2023**, Italia, Francia e Germania avevano a tal proposito avanzato un documento informale proponendo un approccio regolatorio soft che si basasse sull'autoregolamentazione obbligatoria attraverso codici di condotta.

Tuttavia, l'accordo finale sull'**AI Act** tra Parlamento europeo e Consiglio dell'**8 dicembre 2023** ha infine regolamentato la materia introducendo requisiti di trasparenza per tutta la categoria dei cosiddetti **Gpai** (General purpose AI), compresi al suo interno i modelli generativi. Gli sviluppatori dovranno dunque, sulla base di questo nuovo strumento legislativo, fornire documentazione tecnica, rispettare

le leggi sul diritto d'autore e fornire dettagliati resoconti sui dati di addestramento, con ulteriori obblighi più rigorosi per i soli modelli ad alto impatto.

Le linee guida di Open AI

Nonostante queste premesse, l'AI Act inizierà a dispiegare i suoi **effetti solo fra qualche mese**, rendendo di conseguenza cruciali le eventuali iniziative di autoregolamentazione in capo alle imprese. In tal senso, Open AI ha giocato un ruolo proattivo, pubblicando delle linee guida informali in data 16 gennaio 2024, al fine di prevenire abusi dell'IA nelle campagne di disinformazione, in particolare **in vista delle elezioni politiche** che si susseguiranno in svariati Stati nel mondo durante il 2024 (fra tutti, si pensi alle elezioni Europee, Stati Uniti, India e Brasile).

In primo luogo, viene posta un'**enfasi particolare sulla prevenzione di possibili abusi**. Open AI ha assunto l'impegno di anticipare, nella maggiore misura possibile, minacce come i deepfake ingannevoli, le operazioni di influenza su larga scala e l'impiego di chatbot per impersonare candidati nel corso delle elezioni. Prima del lancio di nuovi sistemi, Open AI dichiara dunque di condurre rigorosi test, coinvolgendo utenti e collaboratori esterni per raccogliere feedback, e di implementare misure di sicurezza finalizzate a ridurre il potenziale rischio.

Nelle stesse linee guida di Open AI viene peraltro sottolineato come la stessa società sia ancora nel processo di apprendimento delle modalità con cui gli individui possano abusare della tecnologia in esame.

In considerazione di ciò, le **misure provvisorie** per evitare abusi che Open AI avrebbe implementato consistono in:

1. un divieto di sviluppo di applicazioni da utilizzarsi in campagne politiche e attività di lobbying;
2. un divieto di sviluppo di chatbot che fingano di essere persone reali (ad esempio, candidati) o istituzioni;
3. un divieto di sviluppo di applicazioni che scoraggino la partecipazione ai processi democratici, ad esempio, rappresentando in modo distorto i processi di voto e le qualifiche (quando, dove o chi è idoneo a votare) o che scoraggino il voto sostenendone la futilità;
4. un sistema di report delle potenziali violazioni a disposizione degli utenti.

Il modello text-to-image "Dall-E 3" di Open AI: in cosa consiste

In secondo luogo, le linee guida prospettano una maggiore trasparenza sui contenuti generati dall'intelligenza artificiale. In questo ambito l'attenzione è posta su **un modello text-to-image** sviluppato da Open AI, denominato **Dall-E 3**. Tale applicazione utilizza metodologie di deep learning per generare immagini digitali partendo da descrizioni in linguaggio naturale, anche dette "pro**m**pt".

A questo riguardo, le linee guida anticipano l'imminente implementazione delle credenziali digitali elaborate dalla **Coalition for content provenance and authenticity**, che dovrebbero codificare e "restituire" all'utente i dettagli sull'origine delle immagini generate da Dall-E 3.

Per quanto riguarda, d'altra parte, la piattaforma ChatGpt, si evidenzia, nelle stesse linee guida, come quest'applicazione si stia integrando sempre di più con fonti esistenti di informazioni. Ad esempio, utilizzando lo strumento in parola, gli utenti inizieranno ad avere accesso a notizie in tempo reale a livello globale, con l'inclusione dei relativi link riferiti alle fonti.

La trasparenza sulla provenienza delle informazioni e l'equilibrio nelle fonti delle notizie potranno così aiutare gli elettori a valutare meglio le informazioni cosicché gli stessi possano decidere autonomamente i contenuti a cui dare fiducia.

Infine, viene riportata la notizia di una **collaborazione in corso d'opera, negli Stati Uniti, tra Open AI e la National association of secretaries of state (Nass)**, l'organizzazione professionale non-partisan più antica del paese per funzionari pubblici. Nel contesto di questa partnership, ChatGpt è stato progettato per indirizzare gli utenti verso CanIVote.org, l'autorevole sito che fornisce informazioni sul voto negli Stati Uniti quando sorgono domande procedurali legate alle elezioni, ad esempio circa la localizzazione dei seggi elettorali o dei requisiti per l'ammissione al voto.

Conclusioni

L'imponente rivoluzione portata dall'intelligenza artificiale generativa richiede un intervento regolamentare celere e ponderato. Mentre l'**Unione europea** si appresta a introdurre l'AI Act, è importante riconoscere che l'effettiva attuazione di questa legislazione potrebbe richiedere un periodo prolungato, aprendo la strada a una fase di transizione in cui gli attori del settore possono optare per l'autoregolamentazione.

In questo contesto, emerge chiaramente la **necessità di instaurare un dialogo continuo tra gli stakeholder,** poiché solo attraverso una collaborazione attiva sarà possibile affrontare con successo le sfide emergenti e garantire che le normative siano allineate all'incessante evoluzione della tecnologia generativa.

La raffinatezza delle soluzioni richiederà un approccio sinergico, un equilibrio sottile tra la tutela dei diritti e lo stimolo all'innovazione, per plasmare un quadro regolatorio che sia al passo con la dinamicità di questo ambito rivoluzionario.

(Articolo scritto in collaborazione con Federico Aluigi e Jacopo Piemonte)

Leggi anche

Intelligenza artificiale: la prospettiva globale della regolamentazione

Alessandro Foti, Federico Aluigi, Jacopo Piemonte

Opinione personale dell'autore

Avvocato tributarista senior presso lo studio De Berti-Jacchia in Milano, si occupa della materia sia in ambito nazionale sia internazionale con particolare attenzione a Hnwi e multinazionali altamente digitalizzate, quali quelle operanti nei settori big data, Ai, cloud, cybersecurity, IoT, blockchain.

La redazione vi consiglia altri articoli

SU FINTECH

- Meta sta cercando di liquidare la sua stablecoin mai nata, Diem
- Digitalizzare la banca fra smart working e multicanale

- Se le crypto non sono più decorrelate, sono inutili per diversificare

SU DIGITAL TRANSFORMATION & TECH

- Helvetia: ai clienti risponde Clara, l'assistente basata su ChatGpt
- Intelligenza artificiale: come cambierà la vita dei professionisti?
- Open banking, all'interno degli istituti non c'è allineamento

Cosa vorresti fare?

Ascoltare	*Leggere*	*Guardare*
Apprendere	*Cercare un consulente*	*Scoprire i Talents*
Seguire i Brands	*Pleasure Assets*	*Millennials*

We Wealth
Chi siamo
Contatti
FAQ

Top Pagina
Home
news
voices
podcasts
Cerca un consulente
Scopri i Talents
Segui i Brands
Live broadcast
Weekly Bell
Chiedilo ai Talents
We Wealth Must

Categorie
Investimenti
Consulenza patrimoniale
Filantropia
SRI-impact investing
Pleasure asset
Fintech
Aziende & protagonisti
Secret places
Agorà
Trova il tuo Advisor
Iscriviti alla newsletter
Abbonati al mensile
Privacy investitori
Privacy professionisti
T&C investitori
T&C professionisti
Cookie policy

Partner di:

TORNA SU