FONDAZIONE BRUNO KESSLER
MAGAZINE

Regolamentare e conoscere l'Intelligenza Artificiale attraverso le sfide dei deepfake

8 Luglio 2024

La regolamentazione dell'IA è una questione complessa e sfaccettata, che richiede un approccio equilibrato.

Ne hanno discusso il 10 maggio scorso **Silvio Ranise e Massimo Leone**, direttori rispettivamente del Centro per la Cybersecurity e del Centro per le Scienze Religiose della Fondazione Bruno Kessler. Potete visionare ed ascoltare l'incontro integralmente qui.

Sebbene l'obiettivo principale della recente legislazione in Europa sia stato il finanziamento della ricerca e dello sviluppo dell'IA, l'aspetto altrettanto importante è **come controllare e regolare l'implementazione delle tecnologie di IA**.
Un principio chiave alla base di questo regolamento è **l'approccio basato sul rischio**, che riconosce il potenziale impatto dell'IA sui diritti e sulle libertà fondamentali delle persone. La portata degli strumenti basati sull'IA fa sì che anche una minaccia con una bassa probabilità di materializzarsi possa avere un impatto negativo enorme a causa della scala (potenzialmente decine o centinaia di milioni di utenti) alla quale operano tali algoritmi.

Le **autorità di regolamentazione** devono valutare attentamente la probabilità e la gravità degli esiti negativi per le diverse parti interessate, con l'obiettivo dichiarato di trovare un equilibrio ottimale che tenga conto di queste diverse prospettive. Tecnicamente, questo tipo di problemi di **ottimizzazione** si dicono **multiobiettivo** ed è noto (grazie alla teoria elaborata dall economista, sociologo ed ingegnere italiano Vilfredo Pareto) che non ammettono una singola soluzione ottimale ma molte che hanno la seguente caratteristica: non si può migliorare la condizione di una parte interessata senza peggiorare quella di una altra. Per risolvere il problema di ottimizzazione, si deve quindi scegliere tra le molteplici soluzioni ottimali quella che meglio bilancia gli svantaggi delle parti interessate. Tale scelta richiede considerazioni di varia natura, incluse quelle di tipo etico, giuridico, economico e politico.

A complicare ulteriormente la situazione è la **percezione nelle interazioni tra umano ed IA** che tende a creare la percezione di **onniscienza algoritmica.** Per evitare questo effetto, bisognerebbe creare maggiore consapevolezza sui vantaggi e sui limiti degli algoritmi di IA.
A tale scopo, "la trasparenza è un elemento cruciale", nota **Ranise**, "poiché gli utenti devono essere informati quando interagiscono con i contenuti generati dall'IA".

Per facilitare un **rapporto più consapevole con l'IA** , sono a disposizioni tecniche, come il watermarking (un meccanismo che serve a trarre informazioni sull'origine di contenuti digitali, come immagini, video e documenti, inserendo informazioni nascoste al loro interno), che permettono d'identificare più facilmente la provenienza di contenuti generati dall'IA. Tuttavia, si tratta di una sfida complessa, poiché gli attaccanti sanno adattarsi rapidamente alle misure tecniche adottate per ridurre i rischi. La vera soluzione potrebbe risiedere nel coltivare una comprensione culturale e un consenso più ampi sugli usi appropriati dell'IA generativa.
"Proprio come il linguaggio può essere utilizzato sia per rappresentare la realtà sia per creare finzioni," aggiunge **Leone**, "gli strumenti di IA hanno il potere di aumentare la comunicazione umana in modi sia benefici che dannosi." I **sandbox normativi** (ambienti controllati dove intermediari vigilati e operatori del settore tecnologico possono testare, per un periodo di tempo limitato, prodotti e servizi innovativi) offrono già un approccio promettente, consentendo la sperimentazione controllata e la messa a punto delle regole prima di un'implementazione su più ampia scala, aiutando a gestire i rischi che l'adozione di tali tecniche possono portare. In definitiva, il progresso dipenderà dallo sviluppo non solo di garanzie tecniche, ma anche dalla saggezza necessaria per utilizzare queste potenti tecnologie in modo responsabile.

La necessità di un approccio all'intelligenza artificiale basato sul rischio

Al netto delle osservazioni di cui Ranise e Leone hanno ampiamente discusso, l'attenzione dell' **Unione Europea sulla regolamentazione dello sviluppo e della diffusione dell'IA** è un passo fondamentale nella gestione del profondo impatto che questa tecnologia può avere sulla nostra società. "Sebbene le vaste possibilità dell'IA promettano immensi benefici, dobbiamo anche bilanciare attentamente questi progressi tecnologici con i diritti e le libertà fondamentali dei cittadini," avverte **Leone**. "L'assoluta complessità e portata dell'influenza dell'IA richiedono un approccio sfumato e basato sul rischio per la sua governance".

Questo quadro, basato sul rischio, riconosce perciò che l'IA non è intrinsecamente "buona o cattiva", ma i suoi impatti **dipendono fortemente dal contesto applicativo**. Comprendere a fondo i molteplici rischi legati al caso d'uso, consente l'adozione controllata di nuove e innovative soluzioni tecnologiche pur mantenendo adeguate garanzie contro usi impropri o conseguenze indesiderate.

Il potere e il pericolo dei progressi tecnologici

Il potere del linguaggio nel creare "falsi" è un fatto peraltro storico, che precede di molto l'ascesa delle tecnologie digitali. Gli esseri umani hanno sempre posseduto la straordinaria capacità di usare le parole per costruire narrazioni, fondate o meno sulla realtà, essendo il linguaggio "la tecnologia più potente che abbiamo per generare dei falsi", avverte **Leone**. Questa capacità ha consentito una ricca espressione culturale attraverso la narrativa, la poesia e la filosofia, ma ha anche **facilitato la diffusione della disinformazione e della propaganda**. L'attuale gioco del gatto e del topo tra "metodi di rilevamento" e "tecniche in evoluzione" per la generazione di media cosiddetti sintetici è semplicemente l'ultima iterazione di questa dinamica secolare. Poiché gli strumenti per la manipolazione di audio, video e testo diventano sempre più sofisticati, le sfide legate alla distinzione tra contenuto autentico e contenuto fittizio diventano più complesse.

Tuttavia, Ranise e Leone ricordano come ciò non debba significare che siamo condannati a un futuro d'inganno onnipresente. Il ruolo della comunicazione e della costruzione del consenso sarà

cruciale nel modellare le percezioni e le norme sociali attorno a queste nuove tecnologie. Proprio come abbiamo sviluppato collettivamente pratiche culturali e quadri giuridici per gestire i rischi dei “falsi” basati sul linguaggio, dobbiamo ora fare lo stesso per le loro controparti digitali.

Ciò richiederà un delicato equilibrio: **abbracciare il potenziale creativo dell’IA generativa e allo stesso tempo stabilire barriere per prevenirne l’uso improprio.**

Gli scenari normativi, gli standard di trasparenza e gli sforzi educativi possono tutti svolgere un ruolo nel coltivare questo nuovo consenso sociale. La chiave è evitare reazioni istintive di **tecno-utopismo o di paura apocalittica**, distinzione quasi “manichea” che peraltro si è presentata spesso pure nei decenni precedenti a oggi. Dobbiamo invece affrontare attentamente le sfumature, sfruttando il potere della tecnologia, pur rimanendo radicati nei nostri valori umani condivisi.

Sarà fondamentale combinare una sana consapevolezza tecnologica ed una comprensione condivisa delle capacità e dei limiti dell’IA, sia tra gli sviluppatori che tra gli utenti finali. Solo attraverso questo sforzo combinato di rigore tecnico e consenso culturale, concludono i relatori, possiamo affrontare le sfide poste dall’intelligenza artificiale in modo ponderato ed equilibrato. Abbracciare la complessità, piuttosto che arrendersi a divieti semplicistici o a un’adozione sfrenata dell’IA, è il percorso verso la realizzazione dell’intero potenziale di questa tecnologia trasformativa, che preservi al tempo stesso i nostri valori umani fondamentali.

LINK

https://magazine.fbk.eu/it/news/regolamentare-e-conoscere-lintelligenza-artificiale-attraverso-le-sfide-dei-deepfake/

TAG

- #cybersicurezza
- #deepfake
- #disinformazione
- #IA
- #intelligenzaartificiale
- #scienzereligiose

VIDEO COLLEGATI

- https://www.youtube.com/watch?v=C2quaJiT5As

AUTORI

- Andrea Franceschini