L'incontro di Bletchley

L'intelligenza artificiale cambia ed evolve La sfida mondiale è scrivere regole al buio

ILARIA SOLAINI


PRIMO PIANO

L'intelligenza artificiale cambia ed evolve
La sfida mondiale è scrivere regole al buio

AI SAFETY SUMMIT
HOSTED BY THE UK
1-2 NOVEMBER 2023

L'IMPEGNO
L'Italia pronta a mettere il tema al centro dell'agenda del G7

Quali usi dell'intelligenza artificiale sono così pericolosi da dover essere vietati o classificaticome ad alto rischio? Questa è la domanda cruciale, posta al centro del primo summit su sicurezza eintelligenza artificiale che si è chiuso, ieri, in un luogo iconico a 75 chilometri da Londra, comeBletchley Park, dove furono decifrati i codici segreti dei nazisti durante la seconda guerra mondiale.Il risultato è stato un impegno condiviso, racchiuso nella "Dichiarazione di Bletchley" che coinvolgei 25 Paesi presenti che si sono detti pronti a lavorare insieme e stabilire un approccio comune sullasupervisione degli sviluppi dell'intelligenza artificiale. Anche Wu Zhaohui, vice ministro cinesedella Scienza e della Tecnologia, ha sottolineato che Pechino aumenterà la collaborazione sullasicurezza dell'intelligenza artificiale per aiutare a costruire un "quadro di governance"internazionale.

Quest'impegno sulla carta si va a integrare agli sforzi compiuti, finora singolarmente, perregolamentare i possibili rischi e conflitti etici legati all'utilizzo dell'intelligenza artificiale.Ad esempio il riconoscimento facciale in ogni luogo pubblico, o la giustizia predittiva, che è lapossibilità di calcolare in anticipo l'esito di una sentenza oppure il cosiddetto social scoring,ossia un punteggio sociale che misura l'affidabilità delle persone attraverso gli algoritmi cheproducono valutazioni numeriche e classificatorie a partire da comportamenti, preferenze e e datianche personali raccolti sul web.

Dalla dichiarazione di Bletchley all'AI Act Lunedì il presidente degli Stati Uniti, Joe Biden hafirmato un ordine esecutivo sull'intelligenza artificiale, richiedendo che le aziende riferiscano algoverno federale sui rischi che i loro sistemi possano aiutare altri Paesi o gruppi terroristici aprodurre armi di distruzione di massa. L'ordinanza mira anche a ridurre i pericoli dei "deep fake" chepotrebbero influenzare il voto o truffare i consumatori. « I deep fake utilizzano audio e videogenerati dall'intelligenza artificiale per diffamare la reputazione, diffondere notizie false ecommettere frodi», ha affermato Biden alla firma dell'ordine alla Casa Bianca, ponendo sul tavoloanche il rischio della disinformazione. Nello stesso giorno, il 30 ottobre, i Paesi del Gruppo deiSette hanno concordato un Codice di condotta in 11 punti per le aziende che sviluppano sistemiavanzati di IA, che «mira a promuovere un'IA sicura, protetta e affidabile in tutto il mondo».

Dal canto suo, la Cina già aveva presentato la " Global AI Governance Initiative", preoccupando nonpoco il resto del mondo e in particolare gli Usa: Pechino aveva pubblicato una proposta di requisitidi sicurezza per le aziende che offrono servizi basati sull'intelligenza artificiale, inclusa unalista nera di fonti che non possono essere utilizzate per addestrare modelli di intelligenza

artificiale. E ha imposto nel contempo ai fornitori di servizi di presentare valutazioni di sicurezzae ricevere l'autorizzazione prima di rilasciare prodotti di intelligenza artificiale sul mercato dimassa. In questa direzione di divieti e richieste di autorizzazioni si muove anche l'Europa, dove inegoziati sono appena entrati in una fase cruciale per finalizzare l'ambiziosa legge sull'IAdell'Unione Europea entro la fine dell'anno.
Gli incentivi ad agire, e a farlo insieme, dunque, sono molti.
In primo luogo, perché l'intelligenza artificiale è davvero una tecnologia globale che offre grandipromesse come diagnosticare malattie, prevedere inondazioni e altri effetti del cambiamento climatico,ma comporta nel contempo rischi che includono potenziali violazioni della privacy e l'eliminazione diposti di lavoro. L'IA si basa su modelli linguistici di grandi dimensioni (Large language model, LLM)che vengono allenati usando una grande quantità di testi provenienti da diverse fonti: possono essereeseguiti su un computer qualsiasi, in qualunque parte del mondo. Dunque, sarebbe di scarsa utilitàregolamentare in modo stringente l'IA in alcuni Paesi, se poi in altri rimanessero dei vuotinormativo.
La volontà di agire c'è, ma manca «il consenso su quali siano i problemi che dobbiamo governare, pernon parlare di come dovremmo governarli», ha sintetizzato Henry Farrell, docente di Affariinternazionali alla Johns Hopkins University di Baltimora. Nello specifico vanno distinte tre domandeche aprono, di fatto, tre aree di discussione. Di cosa dovrebbe preoccuparsi il mondo con lo sviluppodell'IA? A cosa dovrebbero mirare le regole? E come dovrebbero essere applicate? Se prima siopponevano alla regolamentazione, oggi, anche giganti come Alphabet e Microsoft fanno pressioni perchénuove regole vengano negoziate, concordate, applicate.
Per il timore che una concorrenza sfrenata possa spingere altri ad agire in modo sconsiderato,rilasciando modelli, di cui si potrebbe abusare.

Regolamentare sì, ma in che modo?

Veniamo agli obiettivi della regolamentazione. Questi sono difficili da impostare perchél'intelligenza artificiale si sta evolvendo rapidamente: non passa giorno senza che una start-upinventi qualcosa di nuovo. E perfino gli sviluppatori non sono in grado dire con certezza qualifunzionalità avranno in futuro i sistemi basati sull'IA. Durante un'audizione al Senato di Washingtonnel luglio scorso, Dario Amodei, amministratore delegato di Anthropic, aveva avvertito che i modellidi intelligenza artificiale saranno in grado di fornire in pochi anni tutte le informazioni necessarieper costruire armi biologiche, consentendo «a molti più attori di effettuare attacchi biologici sularga scala». Simili previsioni disastrose sono state fatte anche riguardo alle armi informatiche.Altri rischi riguardano l'indebolimento del processo democratico.
Per tutte queste ragioni il dibattito su che cosa regolamentare non sembra essere facile da risolvere.Le aziende tecnologiche suggeriscono per lo più di limitare l'esame ai modelli più potenti e piùevoluti. Microsoft ha chiesto un regime di licenze che imponga alle aziende di registrare quei modelliche superano determinate soglie di prestazione. Ma la maggior parte delle aziende ritiene che siano leapplicazioni dei modelli, e non i modelli stessi, a dover essere regolamentate.

Mentre gli sviluppatori di intelligenza artificiale avvertono che norme più invasive rallenterebbero l'innovazione.
Chi controllerà gli sviluppi dell'IA? Fino allo scorso anno Stati Uniti, Gran Bretagna e Unioneeuropea sembravano concordare su questo approccio basato sul rischio. L'ascesa mozzafiato del LLM dallancio di Chat gpt un anno fa li sta facendo ripensare. L'Ue ora si chiede se, dopo tutto, i modellistessi debbano essere supervisionati. Il Parlamento europeo vuole che i creatori di modelli testinogli LLM per il potenziale impatto su tutto, dalla salute umana ai diritti umani. E insiste perottenere informazioni sui dati su cui vengono addestrati i modelli. Queste regolamentazioni più severerappresenterebbero un cambiamento di rotta rispetto ai codici di condotta non vincolanti, che finorahanno rappresentato l'approccio preferito. L'estate scorsa la Casa Bianca aveva già negoziato unaserie di "impegni volontari" con alcune aziende per far testare i loro modelli internamente edesternamente prima del rilascio, chiedendo anche di condividere le informazioni su come gestiscono irischi dell'IA. Poi c'è la questione di chi dovrebbe fare la regolamentazione. Stati Uniti e la GranBretagna pensano che le agenzie governative esistenti possano svolgere la maggior parte del lavoro.L'Ue vuole creare un nuovo organismo di regolamentazione. A livello internazionale, alcuni dirigentitecnologici hanno chiesto la creazione di qualcosa di simile al Gruppo intergovernativo suicambiamenti climatici (Ipcc), a cui le Nazioni Unite hanno il compito di tenersi al passo con laricerca sul riscaldamento globale e di sviluppare modi per misurarne l'impatto. Considerate tuttequeste questioni aperte è difficile prevedere cosa possa aggiungere l'odierna "Dichiarazione diBletchley". Per dirla con Keegan McBride, docente di Intelligenza artificiale e politica all'InternetInstitute dell'Università di Oxford, ha senso capire come poter aiutare le aziende a competere in unmomento di rapidi cambiamenti e di enormi investimenti nell'ambito dell'AI, senza ignorare i rischipiù immediati nel mondo reale della tecnologia. RIPRODUZIONE RISERVATA La dichiarazione di Bletchley:25 Paesi pronti a cooperare sulla sicurezza e contro la disinformazione Si pone la questione di chicontrollerà gli sviluppi dell'IA e le sue potenziali applicazioni Ursula Von der Leyen, presidentedella Commissione europea, all'arrivo al vertice di Bletchley.