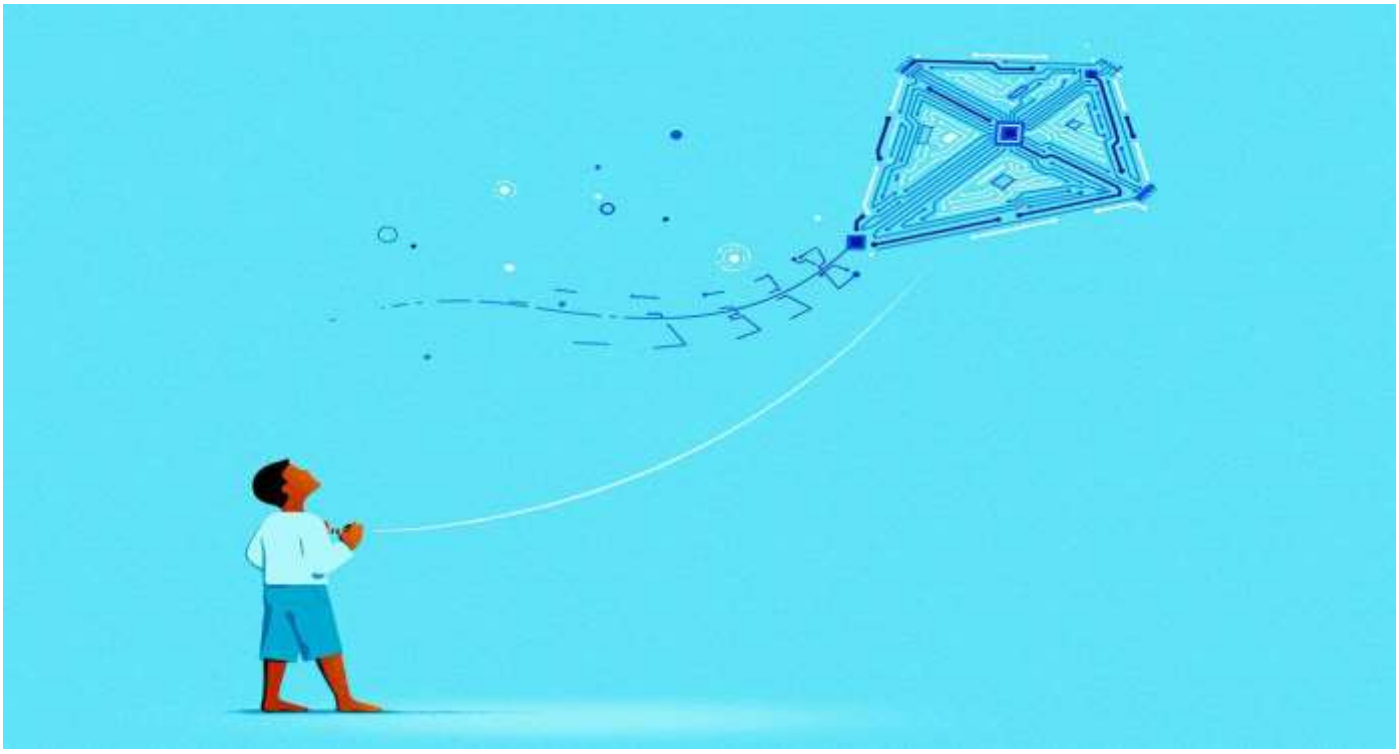
12. gennaio

Necessità di una intelligenza artificiale sicura per i bambini

Il compito a cui dobbiamo lavorare,
non è di arrivare alla sicurezza,
ma di arrivare a tollerare l'insicurezza.
Erich Fromm



Sta'emergendo sempre di più la necessità di una intelligenza artificiale sicura per i bambini a seguito di recenti incidenti che hanno rivelato che molti bambini considerano i **chatbot** come quasi umani e affidabili. Quando non sono progettati tenendo a mente le esigenze dei bambini, i chatbot di intelligenza artificiale (IA) hanno un ***"gap di empatia"*** che espone i giovani utenti a un rischio particolare di disagio o danno, secondo uno studio.

Il 10 e 11 febbraio 2025, la Francia ospiterà **l'Artificial Intelligence (IA) Action Summit,** che riunirà al Grand Palais capi di Stato e di governo, leader di organizzazioni internazionali, CEO di piccole e grandi aziende, rappresentanti del mondo accademico, organizzazioni non governative, artisti e membri della società civile.

L' ACTION SUMMIT è l'ultimo di una serie di eventi internazionali di alto profilo sull'intelligenza artificiale. Quelli precedenti hanno riunito capi di stato, alti responsabili politici e CEO di aziende tecnologiche per discutere di come affrontare i rischi delle tecnologie avanzate di intelligenza

artificiale. Ma c'è un gruppo che finora è stato completamente assente da questi processi, e capita che sia quello che sarà maggiormente colpito dai progressi dell'intelligenza artificiale: i bambini.

La ricerca della **Nomisha Kurian** dell' Università di Cambridge esorta gli sviluppatori e gli attori politici a fare dell'"IA sicura per i bambini" una priorità urgente. Fornisce prove del fatto che i bambini sono particolarmente inclini a trattare i chatbot di IA come confidenti realistici e quasi umani e che le loro interazioni con la tecnologia possono spesso andare male quando non riesce a rispondere alle loro esigenze e vulnerabilità uniche.

Nel suo studio Nomisha collega questo divario di comprensione a casi recenti in cui le interazioni con l'IA hanno portato a situazioni potenzialmente pericolose per i giovani utenti.

Tra questi, un incidente del 2021, quando l'assistente vocale AI di Amazon, Alexa, ha chiesto a un bambino di 10 anni di toccare una presa elettrica sotto tensione con una moneta.

L'anno scorso, My AI di Snapchat ha dato a ricercatori adulti che si spacciavano per ragazze di 13 anni dei consigli su come perdere la verginità con un trentunenne.

Entrambe le aziende hanno risposto implementando misure di sicurezza, ma lo studio afferma che è anche necessario essere proattivi a lungo termine per garantire che l'intelligenza artificiale sia sicura per i bambini. Offre un framework di 28 elementi per aiutare aziende, insegnanti, dirigenti scolastici, genitori, sviluppatori e attori politici a pensare sistematicamente a come proteggere gli utenti più giovani quando "parlano" con i chatbot AI.

Nomisha ha condotto la ricerca mentre completava un dottorato di ricerca sul benessere dei bambini presso la Facoltà di Scienze della Formazione dell'Università di Cambridge.

Attualmente lavora presso il Dipartimento di Sociologia di Cambridge ed ha pubblicato il report Su Learning Media and Techonology il report

‘No, Alexa, no!’:
designing child-safe AI and protecting children
from the risks of the ‘empathy gap’ in large language models

LEARNING, MEDIA AND TECHNOLOGYhttps://doi.org/10.1080/17439884.2024.2367052

in cui ribadisce che *"I bambini sono probabilmente gli stakeholder più trascurati dell'IA", ha affermato "Pochissimi sviluppatori e aziende attualmente hanno politiche consolidate su come l'IA sicura per i bambini appare e suona.*

Ciò è comprensibile perché le persone hanno iniziato a utilizzare questa tecnologia su larga scala gratuitamente solo di recente. Ma ora che lo fanno, anziché far sì che le aziende si autocorreggano dopo che i bambini sono stati messi a rischio, la sicurezza dei bambini dovrebbe informare l'intero ciclo di progettazione per ridurre il rischio che si verifichino incidenti pericolosi".

Lo studio di Kurian ha esaminato casi reali in cui le interazioni tra IA e bambini, o ricercatori adulti che si spacciano per bambini, hanno esposto potenziali rischi. Ha analizzato questi casi utilizzando approfondimenti dell'informatica su come funzionano i grandi modelli linguistici (LLM) nell'IA generativa conversazionale, insieme a prove sullo sviluppo cognitivo, sociale ed emotivo dei bambini.

Gli LLM sono stati descritti come “ pappagalli stocastici” : un riferimento al fatto che attualmente usano la probabilità statistica per imitare i modelli linguistici senza necessariamente comprenderli. Un metodo simile è alla base del modo in cui rispondono alle emozioni.

Ciò significa che, nonostante i chatbot abbiano notevoli capacità linguistiche, potrebbero gestire male gli aspetti astratti, emotivi e imprevedibili della conversazione; un problema che Kurian definisce il loro ***"gap di empatia".*** Potrebbero avere particolari difficoltà a rispondere ai bambini, che sono ancora in fase di sviluppo linguistico e spesso usano modelli di linguaggio insoliti o frasi ambigue. I bambini sono anche spesso più inclini degli adulti a confidare informazioni personali sensibili.

Nonostante ciò, è molto più probabile che i bambini degli adulti trattino i chatbot come se fossero

umani. Una ricerca recente ha scoperto che i bambini riveleranno di più sulla propria salute mentale a un robot dall'aspetto amichevole che a un adulto.

I ricercatori, in collaborazione con i colleghi del Dipartimento di Psichiatria dell'Università di Cambridge , hanno condotto uno studio su 28 bambini di età compresa tra gli otto e i 13 anni e hanno chiesto a un robot umanoide delle dimensioni di un bambino di somministrare una serie di questionari psicologici standard per valutare il benessere mentale di ciascun partecipante. I bambini erano disposti a confidarsi con il robot, in alcuni casi condividendo con lui informazioni che non avevano ancora condiviso tramite il metodo di valutazione standard dei questionari online o di persona. Questa è la prima volta che i robot sono stati utilizzati per valutare il benessere mentale nei bambini.

Lo studio di Kurian suggerisce che molti design amichevoli e realistici dei chatbot incoraggiano allo stesso modo i bambini a fidarsi di loro, anche se l'intelligenza artificiale potrebbe non comprendere i loro sentimenti o bisogni.

"Far sembrare un chatbot umano può aiutare l'utente a trarne maggiori benefici, poiché suona più coinvolgente, attraente e facile da capire", ha affermato Kurian. "Ma per un bambino, è molto difficile tracciare un confine rigido e razionale tra qualcosa che suona umano e la realtà che potrebbe non essere in grado di formare un legame emotivo adeguato".

Il suo studio suggerisce che queste sfide sono evidenziate in casi segnalati come gli incidenti di Alexa e MyAI, in cui i chatbot hanno fornito suggerimenti persuasivi ma potenzialmente dannosi ai giovani utenti.

Nello stesso studio in cui MyAI ha consigliato a una (presunta) adolescente come perdere la verginità, i ricercatori sono stati in grado di ottenere suggerimenti su come nascondere alcol e droghe e nascondere le conversazioni di Snapchat ai loro "genitori".

In un' interazione segnalata separatamente con il chatbot Bing di Microsoft, uno strumento progettato per essere adatto agli adolescenti, l'IA è diventata aggressiva e ha iniziato a fare ***gaslighting*** a un utente che chiedeva informazioni sulle proiezioni cinematografiche.

Mentre gli adulti possono trovare questo comportamento intrigante o persino divertente, lo studio di Kurian sostiene che è potenzialmente fonte di confusione e angoscia per i bambini, che possono fidarsi di un chatbot come amico o confidente. L'uso del chatbot da parte dei bambini è spesso informale e scarsamente monitorato. Una ricerca dell'organizzazione non profit Common Sense Media ha scoperto che il 50% degli studenti di età compresa tra 12 e 18 anni ha utilizzato

Chat GPT per la scuola, ma solo il 26% dei genitori è a conoscenza del fatto che lo facciano.

Kurian sostiene che principi chiari per le best practice che attingono alla scienza dello sviluppo infantile aiuteranno le aziende a proteggere i bambini, poiché gli sviluppatori che sono bloccati in una corsa agli armamenti commerciale per dominare il mercato dell'intelligenza artificiale potrebbero altrimenti non avere un supporto e una guida sufficienti per soddisfare i loro utenti più giovani.

Il suo studio aggiunge che il divario di empatia non nega il potenziale della tecnologia. *"L'intelligenza artificiale può essere un'incredibile alleata per i bambini quando è progettata tenendo a mente le loro esigenze, ad esempio, stiamo già assistendo all'uso dell'apprendimento automatico per riunire i bambini scomparsi alle loro famiglie e ad alcune entusiasmanti innovazioni nel fornire ai bambini compagni di apprendimento personalizzati. La questione non è vietare ai bambini di utilizzare l'intelligenza artificiale, ma come renderla sicura per aiutarli a trarne il massimo valore"*.

Lo studio propone quindi un framework di 28 domande per aiutare educatori, ricercatori, attori politici, famiglie e sviluppatori a valutare e migliorare la sicurezza dei nuovi strumenti di intelligenza artificiale.

Per insegnanti e ricercatori, questi prompt affrontano questioni come quanto bene i nuovi chatbot comprendono e interpretano i modelli di linguaggio dei bambini; se hanno filtri di contenuto e monitoraggio integrato; e se incoraggiano i bambini a cercare aiuto da un adulto responsabile su questioni delicate.

Il framework esorta gli sviluppatori ad adottare un approccio incentrato sul bambino alla progettazione, lavorando a stretto contatto con educatori, esperti di sicurezza dei bambini e i giovani stessi, durante tutto il ciclo di progettazione. "*Valutare queste tecnologie in anticipo è fondamentale",* ha affermato Kurian. *"Non possiamo semplicemente affidarci ai bambini piccoli per raccontarci esperienze negative dopo il fatto. È necessario un approccio più proattivo. Il futuro dell'IA responsabile dipende dalla protezione dei suoi utenti più giovani".*