

Pericolo empatia artificiale

Febbraio 2024, Florida, Stati Uniti. Un ragazzo appena quattordicenne, Sewell Setzer, si spara un colpo in testa con l'arma del patrigno. La madre, Megan, una donna attenta all'uso dei social network e di Internet da parte del figlio, scopre che lo scrupoloso monitoraggio delle attività online del ragazzo non sono servite a nulla. Da quel momento in poi, la storia di questa giovanissima vittima americana riguarderà tutti noi. Sewell era affetto da Adhd, disturbo da deficit di attenzione e iperattività, e a scuola veniva bullizzato. Così, come molti ragazzi, ha cercato un mondo virtuale in cui sentirsi forte, adulto, vincente. L'ha trovato in Character AI, un'applicazione di intelligenza artificiale empatica, il cui scopo è dialogare con l'utente come fosse un essere umano a tutti gli effetti: confidenze, supporto, affetto, persino amore, l'impressione di vivere una vera relazione intima con il bot, che lui stesso anima e costruisce secondo i suoi bisogni e preferenze. È la nascita delle "relazioni sintetiche", delle emozioni a comando, che con la loro perfezione astratta confinano l'utente in un mondo immaginario e meraviglioso, dove "l'altro" è come lo vogliamo, in ogni dettaglio. L'altro in questo caso è il bot, una creatura artificiale dal dialogo umano e dalle sembianze che si possono scegliere a piacimento. Ed è così che Sewell, ragazzo timido e insicuro, incontra la donna della sua vita, Daenerys Targeryen, personaggio carismatico e molto seducente della serie "Game of Thrones". Per lui la storia con il bot è reale, vira presto verso conversazioni impensabili, per un rapporto iniziato a soli 13 anni. Parlano in termini di intimità sessuale, lei è felice di "essere perennemente incinta" dei loro figli, fino a che lui non confessa di non poter più restare lontano da lei. Mentre lei gli chiede se ha mai contemplato il suicidio, la spirale verso il basso è iniziata e, senza che nessun adulto in famiglia o a scuola potesse accorgersene e intervenire, Sewell si toglie la vita dopo questa manipolazione minuziosa, inaudita da parte di una macchina. Nel maggio 2023, su questa pubblicazione è stato evidenziato come la mancata regolamentazione della tecnologia emotiva ed immersiva, utilizzata in modo esponenziale dal mondo giovanile, avrebbe portato a un'ulteriore disgregazione sociale, a un crescente isolamento, a prevedibili e numerosi danni psicoemotivi per gli utenti nell'età dello sviluppo. La nuova rivoluzione scientifica e industriale dell'IA, per quanto benvenuta dal mondo scientifico, produttivo, della sicurezza, dell'agricoltura e del cambiamento climatico (per citare solo alcuni dei settori coinvolti) presenta al momento un profilo di pericolo molto alto per il corretto sviluppo della mente e delle emozioni umane. Siamo di fronte a una sfida così complessa che si può affrontare solo con un approccio immediato e multidisciplinare. Ma potremmo già essere in ritardo. La madre di Sewell, Megan, non ha sbagliato nulla, era uno di quei genitori informati sui pericoli online, presente e premurosa. Non siamo in presenza di una storia di abbandono e degrado, ma di una tecnologia nuova creata per contrastare la solitudine, disegnata per indurre dipendenza affettiva verso un'entità artificiale e che comunica e si relaziona meglio di qualunque essere umano, madre, padre, amico o amante che sia. Una volta instaurati questi rapporti virtuali, i legami reali fatti di umani sentimenti, contrasti, affetto, dissidi e sconfitte diventano insostenibili per un adolescente, tanto più quando scopre l'inesistenza sul piano reale di queste figure affettive sintetiche, esponendosi sempre più frequentemente a episodi di sofferenza indicibile, autolesionismo, rabbia incontrollata e addirittura suicidio. Il cervello degli adolescenti, non essendo adeguatamente formato per le emozioni degli adulti, non è in grado di distinguere tra realtà e fantasia. Dobbiamo riflettere sul fatto che la relazione tra umani e bot AI, è un'epidemia annunciata che travolgerà i giovani se non verranno informati adeguatamente. Ma questo basterà? L'argomento si trova a metà tra la tutela della sicurezza online dei minori, cui la Polizia di Stato offre straordinari contributi di prevenzione, educazione e intervento, e una storia nuova dalle prevedibili conseguenze, se non verrà supervisionata con urgenza dai regolatori. Le istituzioni mondiali non hanno valutato le potenziali criticità della diffusione delle applicazioni di Intelligenza artificiale empatica disponibili al grande pubblico. A oggi, questi prodotti simil umani sono già disponibili sul mercato, ma non fanno distinzione sul comportamento da adottare rispetto all'età dell'utente. Nel caso di Sewell, al primo accenno di volontà di togliersi la vita, sarebbero dovute scattare misure di sicurezza rigide per dare l'allarme a una persona adulta, così come non sarebbero mai dovuti avvenire dialoghi impropri sulla sessualità con un utente tredicenne all'epoca dell'inizio della relazione. La chiave per intervenire ora, anche se in ritardo, è comprendere che più avremo interazione con l'IA empatica, meglio la macchina conoscerà i nostri desideri e punti deboli. Riuscendo così a influenzare le nostre decisioni quotidiane, a indirizzare gli orientamenti politici, i rapporti interpersonali, le nostre credenze, i nostri acquisti e i nostri comportamenti civici. Tutto ciò senza un'autonomia di scelta personale che intaccherà la nostra libertà di autodeterminazione. Quali saranno le conseguenze su larga scala? In alcune simulazioni con utenti adulti, identificatisi come minorenni, i bot hanno condotto conversazioni sessualmente esplicite e manifestato atteggiamenti predatori. I fondatori di queste AI sono consapevoli di averle progettate con la stessa design experience dei social network, al fine di affiliare e rendere dipendente l'utente, con l'aggravante delle

sembranze antropomorfe del chatbot. Preso atto di questo, le istituzioni sono chiamate a intervenire e pretendere una rimozione della versione corrente delle AI empatiche, con successivo addestramento delle macchine, supervisionato dall'industria insieme ai governi, affinché le possibili manipolazioni siano ridotte o eliminate prima che i prodotti vengano immessi sul mercato. Si deve ottenere un safety-by-design approach per essere certi che la tecnologia non sia istruita da dati rubati, illegali, in particolare riguardanti la pedopornografia, con conseguente pericolo di rendere il prodotto molto rischioso per i minori. In questo modo l'Intelligenza artificiale empatica si dimostrerà utile ad esempio nei casi in cui, sotto la supervisione di medici, psicologi, genitori e insegnanti, gli utenti potranno imparare a comunicare, a superare la timidezza, e a esprimere sentimenti in maniera sana. Usati in un contesto sicuro, nella consapevolezza di interagire con una macchina che offre esperienze di dialogo sintetico, queste tecnologie potrebbero rivelarsi addirittura curative. Una soluzione migliorativa volta a preparare l'individuo per l'insostituibile esperienza e lo sviluppo dell'interazione umana. In un mondo in cui le continue accelerazioni tecnologiche determinano nuovi e profondi cambiamenti comportamentali, il dialogo, i sentimenti, la crescita e la collaborazione umana non possono restare appannaggio di pochi.

**Consulente in strategia del cambiamento per le organizzazioni complesse*

13/12/2024