Intervista. Taddeo: «Così l'intelligenza artificiale inquina il dibattito politico»

La studiosa di Oxford: «Serve un'etica digitale che faccia emergere i rischi di sistemi come ChatGPT. Gli esseri umani si differenziano dalle macchine per la capacità di fare domande sul loro mondo»

«La capacità di fare domande e la spinta naturale a comprendere il mondo che ci circonda» sono quello che caratterizza l'essere umano di fronte all'intelligenza artificiale (IA) che, oltre alle sue grandi potenzialità positive, rischia di diventare strumento per inquinare e falsare il dibattito pubblico. Risulta quindi sempre più necessaria un'etica digitale. Ed è esattamente Digital Ethics e Defence Technologies che insegna all'Internet Institute dell'Università di Oxford **Mariarosaria Taddeo, studi a Bari e Padova e poi brillante carriera accademica in Inghilterra. L'esperta è a Milano** per la quarta edizione di "Il Verde e il Blu Festival – Buone idee per il futuro del pianeta", il 19 e 20 ottobre presso gli spazi di The Mall (Piazza Lina Bo Bardi, Piazza Alvar Alto). La manifestazione – ricca di eventi e ospiti, promossa dalla multinazionale di consulenza BIP e prodotta da Beulcke&Partners - intreccia due temi cruciali di questi anni: il verde della sostenibilità e il blu dell'innovazione digitale, dove a farla da padrone è l'IA con i suoi rapidissimi sviluppi.

Professoressa Taddeo, perché l'intelligenza artificiale, che da anni è presente nelle nostre vite, solo recentemente sembra essere balzata all'interesse generale? È solo l'effetto ChatGPT e quindi, passata la sorpresa e abituati a dialogare con strumenti di quel tipo, l'IA tornerà sotto traccia?
Nell'ultimo anno c'è stata sicuramente un'attenzione maggiore dovuta alla diffusione dei *large language models* (LLMs) – ChatGPT appunto — che hanno richiamato l'attenzione per varie ragioni: gli LLMs svolgono molto bene compiti tipicamente attribuiti agli esseri umani, per esempio comporre un testo, creare una nuova immagine; le interazioni con questi modelli sono molto semplici, si danno comandi ai modelli attraverso il linguaggio naturale, quindi tutti quelli che sanno scrivere e leggere possono usarli direttamente; perciò vi è stata un'adozione larga e repentina. Questi modelli hanno un grande potenziale ma pongono anche rischi concreti (si pensi alla questione del copyright del materiale che producono o all'uso di questi modelli per diffondere disinformazione, per esempio). Servono quindi uno sforzo condiviso da diversi attori nelle nostre società (accademici, tech providers, legislatori) e un dibattito pubblico informato. Per questo credo che parleremo d'Intelligenza Artificiale (IA) e di etica dell'IA ancora per un po'. Del resto, non abbiamo iniziato a parlare di IA con ChatGPT.
Va ricordato che le questioni etiche e di governance dell'IA non nascono con gli LLMs né sono diventate rilevanti con questi modelli. L'etica dell'IA inizia con il lavoro di Wiener, negli anni 50 del Novecento. Anche senza voler andare così indietro, l'analisi delle implicazioni etiche, legali e sociali (ELSI) dell'IA ha preso nuovo vigore nel primo decennio degli anni 2000 (in particolare nel 2012 quando abbiamo capito bene il potenziale delle reti neurali). Sono almeno 11 anni che parliamo di *bias*, di fiducia nelle macchine, di protezione dell'autodeterminazione degli esseri umani e di *responsability gap*. Questo dibattito non è stato

solo accademico, anzi è andato quasi di pari passo con quello sulla governance. Si pensi per esempio alle linee guida dell'Unione Europea sull'IA affidabile (2019) e al framework regolamentativo per il digital che l'EU sta definendo e che culmina proprio con l'AI Act.

In quanto organismi frutto di una lenta evoluzione naturale, ci stiamo adattando sufficientemente bene al mondo digitale o i disturbi psicologici che sembrano in aumento segnalano un disagio e una difficoltà rispetto a questa transizione così rapida?
Non credo sia una questione di adattamento biologico (o non solo), credo che sia una questione di strategia e 'leadership'. La rivoluzione digitale è una rivoluzione tanto operativa (come facciamo le cose) quanto concettuale (come capiamo la realtà intorno a noi). A volte questi cambiamenti sono molto radicali e richiedono una riflessione filosofica, etica, culturale per poterli comprendere e anche guidare nella direzione che ci sembra più opportuna. I disagi, i danni e i rischi legati alla rivoluzione digitale sono frutto di una mancata comprensione delle implicazioni dell'innovazione digitale magari a seguito di un'adozione frettolosa della stessa e soprattutto di una mancata visione di come vogliamo sfruttare il digitale.
Dovremmo ricordarci che il digitale è una tecnologia trasformativa, e quindi chiederci come vogliamo che il digitale trasformi la nostra realtà, le nostre interazioni con gli altri e con l'ambiente, le nostre società. E cercare di regolare le tecnologie digitali in modo che ci portino dove vogliamo. L'alternativa è un adattarsi passivo e pericoloso, che non lascia presumere niente di entusiasmante.

Dalla sua prospettiva, ci sono reali motivi di preoccupazione per l'espansione dell'intelligenza artificiale, al di là della perdita di posti di lavoro in molti settori economici?
Ci sono rischi concreti da considerare. Chiariamolo subito, questi rischi non sono quelli relativi allo sviluppo di macchine realmente intelligenti, senzienti. Questa è fantascienza. Chiunque abbia studiato un po' i meccanismi degli LLMs capisce che questi modelli non fanno che applicare regole di statistica a grandissimi volumi di dati, senza avere alcuna comprensione né delle regole né dei dati.
I rischi sono diversi e in parte sono gli stessi che riguardano l'IA predittiva, per esempio, la possibilità per questi modelli di diffondere su larga scala errori e pregiudizi, la limitata predicibilità dei comportamenti dei modelli IA e la loro limitata trasparenza, due fattori che ne limitano il controllo; i problemi nell'attribuire la responsabilità agli esseri umani per le azioni di queste macchine. Ci sono tre aree che, a mio avviso, sono forse più rilevanti quando si pensa agli LLMs.

La prima è l'integrazione di questi modelli nelle nostre attività quotidiane. Andiamo verso scenari in cui l'IA diventa un membro dei nostri team professionali, per esempio. Qui la questione non è tanto quanti posti di lavoro si perderanno o genereranno (trovo queste stime sempre problematiche, perché i numeri dei posti di lavoro generati, richiesti o persi in un settore dipendono da tantissime variabili che hanno numerose dipendenze ed *externalities*, questo rende difficile fare delle stime accurate) ma come integriamo con l'IA nei processi professionali in modo da delegare i compiti senza erodere le capacità degli esseri umani di eseguire gli stessi compiti. Per esempio, possiamo delegare all'IA la lettura di immagini diagnostiche, ma vogliamo che i medici continuino a saper leggere le stesse per capire quando l'IA sbaglia. Dobbiamo essere in grado di tutelare gli esseri umani proteggendo la loro autonomia decisionale, che succede se un medico prende una decisione diversa da quella suggerita dall'IA? Chi è responsabile degli errori di un sistema di intelligenza artificiale?

La seconda area afferisce al dibattito pubblico e ai processi democratici. Abbiamo visto già con Cambridge Analytica, che la combinazione IA e social network può porre rischi molto seri per il dibattito pubblico e per lo svolgimento di processi democratici, per esempio elezioni e referendum, attraverso la creazione e diffusione capillare e personalizzata di disinformazione (fake news). Credo che con gli LLMs questi rischi aumentino esponenzialmente nella misura in cui questi modelli possono creare a basso costo testi personalizzati su certi tipi di utenti e indistinguibili da quelli creati da un essere umano. Come controlliamo le fake news, che responsabilità hanno gli sviluppatori degli LLMs in questo senso e che ruolo dovrebbero avere i gestori dei social networks?

La terza è la cybersecurity. L'IA è una tecnologia vulnerabile dal punto di vista della cybersicurezza, gli LLMs non fanno eccezione. I dati ci dicono che è più facile attaccare un modello LLMs di altri tipi di IA. Questo perché è possibile manipolare questi modelli usando il linguaggio naturale. Stiamo integrando questi sistemi in maniera quasi capillare, si pensi a Google Bard che è stato reso disponibile per un miliardo di utenti. C'è da chiedersi se gli LLMs non siano un cavallo di Troia per le società digitali.

In questo senso, che cosa è l'etica digitale e quali sono i temi più urgenti da affrontare?
L'etica del digitale è un'area di ricerca che guarda alle sfide (sia i rischi sia le opportunità da cogliere) legati ai dati (aspetti relativi alla generazione, registrazione, raccolta, cura, elaborazione, diffusione, condivisione e utilizzo), agli algoritmi (inclusi l'intelligenza artificiale, gli agenti artificiali, l'apprendimento automatico e i robot) e alle deontologie professionali (inclusa l'innovazione responsabile, la programmazione, l'hacking e i codici professionali). L'obbiettivo è formulare e sostenere soluzioni eticamente valide (ad esempio, condotte

corrette o valori giusti) che permettano di bilanciare interessi legittimi ma contrastanti alla luce dei valori e diritti fondamentali delle nostre società.
Per fare questo l'etica del digitale non può avere approcci 'verticali', un'etica che si occupi per esempio solo degli algoritmi senza guardare ai dati e alle deontologie professionali fallisce in partenza, perché non è in grado di cogliere l'aspetto sistemico delle sfide etiche che il digitale pone.

Che cosa resterà tipico dell'intelligenza umana davanti al continuo miglioramento dell'intelligenza artificiale in tutti gli ambiti conosciuti (comprese l'arte e la religione)?
Le due cose non sono comparabili, quando pensiamo all'IA dovremmo concentrarci su 'artificiale' più che su 'intelligenza' (la cui definizione è di per sé problematica) che rimane appannaggio degli umani, con le intuizioni, i dubbi, la creatività, l'empatia, le emozioni, la socialità e tutto ciò che ci rende senzienti. Tra queste, l'aspetto che credo sia più importante è la capacità di fare domande e la spinta naturale a comprendere il mondo che ci circonda.

Avvenire, Andrea Lavazza, mercoledì 18 ottobre 2023
https://www.avvenire.it/agora/pagine/intelligenza-artificiale-taddeo-ecco-i-rischi-per-la-politica