

8 novembre 2023 13:23

'L'intelligenza artificiale va resa più trasparente e affidabile'di [Redazione](#)**Intervista a Melanie Mitchell, che da quarant'anni studia la storia, i limiti e le potenzialità dell'intelligenza artificiale**

[Melanie Mitchell](#) si occupa di intelligenza artificiale da quarant'anni. Da quando, fulminata dalla lettura di "Gödel, Escher, Bach: un'eterna ghirlanda brillante" dello scienziato e divulgatore Douglas Hofstadter, lasciò il lavoro di insegnante di matematica a scuola per seguire i corsi introduttivi di computer science al MIT, propedeutici al dottorato in informatica. Al MIT incontrò casualmente proprio Hofstadter, che si trovava lì nel corso di un anno sabbatico e cominciò la carriera accademica.

Oggi Melanie Mitchell insegna Computer Science alla Portland State University ed è external professor al Santa Fe Institute, il maggiore centro di ricerca al mondo sui fenomeni complessi. Divulgatrice d'eccezione, nel 2010 con "Complexity: a guided tour" (Oxford University Press, 2009) ha vinto il Phi Beta Kappa Science Book Award. Con ["L'intelligenza artificiale"](#) (Einaudi, 2022), è in Italia, a Reggio Calabria, finalista del premio Cosmos. La guida è stata scritta nel 2018, ben prima che l'IA diventasse un fenomeno mediatico e socio-culturale, e spiega in modo semplice, appassionato e a volte ironico sia la storia della ricerca di settore sia i limiti e le potenzialità dei meccanismi su cui si basano le diverse tecniche chiamate intelligenza artificiale.

Il sottotitolo del volume è "Una guida per esseri umani pensanti", la dedica è ai suoi genitori, che, scrive, «mi hanno insegnato a essere una persona pensante, e molto di più»... cosa significa pensare per lei?

Per me, pensare è avere una sorta di rappresentazione del mondo nella propria mente e usarla per agire, fare, prendere decisioni... e prevedere ciò che accadrà. Possiamo usarla per capire una nuova situazione, una nuova persona, per dare un senso a qualsiasi cosa incontriamo nel mondo.

Questa definizione vale sia per gli esseri umani che per le macchine?

Riguarda gli esseri umani, ma forse ci sono altri modi attraverso cui le macchine possono capire il mondo. O forse esistono altri modi di pensare che non sono quelli umani. Le parole che usiamo cambiano, man mano che conosciamo meglio noi stessi e cerchiamo di costruire macchine. Ad esempio: prima che fossero realizzate le macchine che giocano a scacchi a livello professionistico, le persone erano convinte che queste macchine per funzionare avrebbero dovuto pensare come gli esseri umani. Oggi sappiamo che non è così, eppure riescono a giocare estremamente bene. Quindi forse c'è un livello differente di pensiero che non è uguale a quello umano.

Nel libro riporta la preoccupazione del suo mentore Hofstadter sul fatto che l'intelligenza artificiale possa rivelarci la facilità di meccanizzare capacità umane per noi preziose. Ad oggi qual è la differenza tra pensiero umano e funzionamento della macchina?

Una delle differenze più grandi riguarda proprio la rappresentazione mentale che noi umani abbiamo del mondo. In questo momento, ho un modello nella mia mente di come funzionano gli oggetti: se prendo questa sedia e la faccio cadere, ho idea di cosa potrebbe succedere. Ho anche dei modelli rispetto al comportamento delle altre persone:

so che tieni al tuo smartphone e non lo lancerai in mezzo alla strada, come so che non lo mangerai. E tutto questo fa parte della fisica e della psicologia intuitive, che appartengono già ai bambini molto piccoli ma non alle macchine.

Non sto dicendo che le macchine non possano avere il nostro modello di mondo, ma ad oggi quello che possiedono, dopo essere state addestrate, è un gran numero di associazioni statistiche tra le parole e i significati. Se chiedessi a una macchina: «Josephine sta per lanciare il suo smartphone in strada?» probabilmente risponderebbe di no, ma il metodo di decisione sarebbe basato su come le parole della frase sono associate le une alle altre, quindi solo sul linguaggio. Un processo molto diverso da quello umano.

Come scrive riportando il paradosso di uno dei pionieri dell'IA, Marvin Minsky, «le cose facili sono difficili»...

Sì, le macchine riescono a fare cose che sono molto difficili per le persone, come giocare a scacchi, tradurre lingue, risolvere problemi matematici. Ma a loro risulta molto difficile fare cose per noi semplici, come, nel caso dei robot, passare tra due sedie, oppure, nel caso di una macchina a guida autonoma, capire e reagire al movimento delle persone in strada. Queste cose riguardano percezione, pianificazione, azione. Noi umani sappiamo farle quasi dalla nascita.

Crede che questa differenza sia legata al fatto che le macchine non hanno (ancora) un corpo?

In parte. Noi abbiamo un corpo che si è evoluto per adattarsi a lavorare in un mondo fisico, le macchine no. Il nostro mondo fisico non è configurato per il modo in cui oggi funzionano le macchine, che infatti possono calcolare, riconoscere pattern, ma non hanno un corpo che possa agire nel mondo.

Se mettessimo un set di algoritmi in un robot, le cose cambierebbero?

Potrebbero. La robotica è difficile, ancora non esistono gli algoritmi e i meccanismi di apprendimento corretti per fare questo. Se succedesse, le macchine sarebbero più capaci di pensare.

Nel testo sottolinea che poniamo «troppa attenzione ai rischi di una IA super intelligente e troppo poca alla scarsa affidabilità e trasparenza del deep learning, oltretutto alla sua vulnerabilità agli attacchi». Cosa intende dire?

Le persone che pensano ai pericoli dell'IA si dividono principalmente in due gruppi: chi è preoccupato da una IA super intelligente che sviluppi propri desideri e minacci il genere umano, una preoccupazione quindi di lungo termine; e chi invece si focalizza nell'immediato, sui bias e le vulnerabilità agli attacchi che stanno minacciando le persone adesso. Molte persone pensano che il pericolo sia un'IA fuori controllo, infatti parlano di rischio esistenziale per il genere umano, di estinzione. Penso che, invece, dobbiamo focalizzarci sui pericoli immediati.

E cosa suggerisce di fare per mitigare questi rischi?

Possiamo fare delle regolamentazioni per rendere le IA più trasparenti e affidabili, come si sta facendo nell'Unione europea. Stare attenti alla privacy, essere sicuri che le macchine possano spiegare le loro decisioni e che i dati con cui sono allenate siano trasparenti. Sono questioni importanti, difficili perché intaccano il modello di business delle aziende del settore. Ad oggi, il modo in cui le macchine sono allenate e prendono decisioni è spesso una black box, una scatola nera, ma sono fiduciosa, anche nell'Explainable AI, il campo di ricerca che lavora per rendere comprensibili agli esseri umani i processi con cui le macchine producono gli output. Non è un compito impossibile.

Un altro tema è la sostenibilità ambientale. Qualche giorno fa papa Francesco ha pubblicato un'esortazione apostolica sulla crisi climatica e sulle sue cause umane, soffermandosi su un paradigma tecnocratico che non tiene conto delle risorse fisiche limitate. L'intelligenza artificiale potrebbe aiutare il contrasto ai cambiamenti climatici o accelerare il disastro?

Entrambe le cose. Indubbiamente oggi le macchine con algoritmi IA consumano enormi quantità di energia, acqua, elettricità e sono molto inefficienti rispetto all'apprendimento umano. Penso che ci siano modi per renderle più efficienti e più sostenibili, facendole lavorare ugualmente bene ma in una dimensione ridotta. Molte persone sono interessate a lavorare su questo. Oggi lo spreco ambientale esiste. Allo stesso tempo, penso che l'IA possa essere usata nella scienza per produrre energia sostenibile. Ogni tecnologia può essere usata in entrambi i modi.

Nella sua guida riprende il paragone tra elettricità e IA per sottolineare che, nel primo caso, la tecnologia era molto ben conosciuta prima di essere commercializzata. Ora che l'IA è sul mercato, pensa che

aumenterà le possibilità delle persone di conoscere?

Si pensa che l'IA renderà la conoscenza più accessibile e la democratizzerà. In effetti, questo è sempre successo nella storia della tecnologia, dal linguaggio ai libri, alla stampa, al computer, a Wikipedia, e l'IA non farà eccezione. Può lavorare su una quantità enorme di dati che possono essere analizzati in modo utile, ed è uno step enorme dell'evoluzione umana. Abbiamo sempre esternalizzato alcune nostre capacità: dalle calcolatrici al computer, adesso l'intelligenza artificiale. Possiamo usarle, così non dovremo "pensare" più a nulla.

Allo stesso tempo, le persone capaci di programmare queste macchine non sono molte, non trova?

Sono in aumento ma non è la programmazione il problema. La questione è più centrata sulla disponibilità dei dati e sulla potenza computazionale. Per eseguire un programma come ChatGPT oggi devo passare per forza dall'azienda OpenAI, perché il mio computer non ha la potenza computazionale per farlo girare. Questo cambierà. Non so quando, ma penso che già tra qualche anno questi sistemi potranno girare direttamente sui nostri dispositivi, siano computer, smartphone, occhiali smart.

A proposito di oggetti smart, nel testo riporta una delle previsioni di Ray Kurzweil, ingegnere capo di Google e profeta della Singolarità, ovvero il momento in cui l'intelligenza artificiale supererà l'intelligenza umana. Una di queste previsioni recita appunto che entro gli anni Trenta del Duemila «trasmettitori di esperienze [...] invieranno sul web tutto il flusso delle loro esperienze sensoriali e i correlati neurologici delle loro reazioni emotive». Considerato lo sviluppo dell'IoT, delle tecnologie indossabili, della realtà virtuale e della realtà mista, pensa che questa previsione si stia realizzando?

In una certa misura sì. Ma questa previsione si basa su uno scenario in cui la realtà virtuale sia indistinguibile dalla realtà fisica. E non sono così sicura che questo possa succedere.

E se pensa alla realtà mista? Ovvero al mix tra realtà aumentata e IoT?

Vorresti avere sensori che tracciano tutto di te e dati che possono essere disponibili a chiunque?

A proposito di questo. Nelle pagine che spiegano la differenza tra visione umana ed artificiale, sottolinea come noi esseri umani «sappiamo cosa ignorare», ad esempio in una fotografia, rispetto a quello che riteniamo irrilevante. Ma vale ancora oggi, in cui le nostre cornici di significato sono co-costruite dagli algoritmi delle piattaforme che utilizziamo?

Un'enorme quantità di lavoro è stata impiegata per capire a cosa reagiscono gli esseri umani e cosa ignorano: i pubblicitari lo sanno molto bene quando progettano contenuti per la propaganda o i social media specificatamente per coinvolgere le persone. Sappiamo cosa ignorare in ambienti naturali ma siamo anche manipolabili.

Nel libro sottolinea l'importanza del buon senso (common sense) per la conoscenza umana. Sarà possibile aggiungerlo alle macchine?

Non lo so. Sono stata sorpresa da come i Large Language Model riescano a ricavare conoscenza di senso comune solo dal linguaggio, ma penso che ci siano altre cose non codificabili in questo modo. Non so se avremo bisogno di un corpo o di simulare scenari in realtà virtuale, quindi senza corpo, ma penso che essere presenti nel mondo, interagire nel mondo, anche in un mondo virtuale, sia essenziale per generare common sense. È possibile che lo faremo con le macchine.

Nella guida riporta come il primo etichettamento di immagini per l'apprendimento di algoritmi di intelligenza artificiale nel progetto ImageNet sia stato realizzato con Amazon Mechanical Turk, il servizio di crowdsourcing che esternalizza compiti facili e ripetitivi a una forza lavoro distribuita e pagata con piccole (se non piccolissime) somme di denaro. Una dinamica simile si è ripetuta per l'addestramento di ChatGPT. Siamo destinati a utilizzare (o sfruttare) persone come macchine per allenare le macchine?

Le macchine, per essere usabili, devono essere allenate a riconoscere testi e immagini da internet, e devono riconoscere quali siano appropriate, quali tossiche, quali rilevanti. Queste informazioni arrivano dagli esseri umani. Si arriva a una situazione per cui le persone vengono pagate davvero poco per un lavoro che le traumatizza, perché spesso le immagini sono disturbanti. È ancora una questione importante nell'IA perché gli algoritmi non imparano completamente da soli ma hanno ancora bisogno del feedback umano. Si stanno cercando altre soluzioni, come dare regole generali alla macchina (Constitutional AI). Ma le regole sono ambigue, vanno calate nei contesti e le macchine hanno problemi a farlo.

Nel libro sottolinea come l'astronave di Star Trek abbia ispirato sia la progettazione della versione avanzata di Google, sia l'assistente vocale di Amazon Alexa, sia il sistema Watson di IBM. Abbiamo bisogno di altri immaginari per collaborare con l'IA?

Sì, forse abbiamo bisogno di scrittori di fantascienza. Spesso hanno scritto distopie, cose orribili causate dall'intelligenza artificiale... forse abbiamo bisogno che immaginino un'IA che lavori insieme agli umani. E, così facendo, riconfigurino il futuro.

(Josephine Condemi su Valori.it del 06/11/2023)

CHI PAGA ADUC

l'associazione non **percepisce ed è contraria ai finanziamenti pubblici** (anche il 5 per mille)

La sua forza economica sono iscrizioni e contributi donati da chi la ritiene utile

DONA ORA (<http://www.aduc.it/info/sostienici.php>)