

LEZIONE V: i lati oscuri dell'IA

Il primo e più fondamentale lato oscuro dell'IA è la **nozione di intelligenza**, che è tutt'altro che trasparente. Le IA non sembrano imitazioni dell'intelligenza umana se consideriamo, ad es., la 37^a mossa dell'incontro a GO.

Alcune sono progettate per risolvere problemi insuperabili per l'intelligenza umana. Perciò sembrano intelligenze «aliene». Queste IA sono dunque inquietanti per definizione, e se ci sfuggissero potrebbero rappresentare un pericolo.

Per il filosofo Floridi l'IA è fondamentalmente baconiana, sintattica, mentre la nostra intelligenza è platonica, semantica.

Come apprendono le reti neurali secondo Dehaene

Gli algoritmi di deep learning convenzionali si limitano ad imitare le prime tappe dell'elaborazione dei dati sensoriali, i famosi 200 o 300 ms durante i quali il nostro cervello funziona in una modalità non cosciente. L'elaborazione rimane strettamente ascendente, bottom-up, ossia priva di una vera capacità di riflessione.

Le reti neurali oggi disponibili tendono ad apprendere le regolarità statistiche superficiali di dati, piuttosto che concetti astratti di alto livello. A differenza della macchina, il cervello domanda, rianalizza, concentra la propria attenzione su questo o quell'aspetto dell'immagine che non corrisponde alla sua prima impressione. Questa seconda analisi, cosciente e intelligente, fa appello alle nostre capacità di ragionamento e astrazione. Le reti neurali trascurano il fatto che imparare significa costruirsi un modello astratto del mondo, non solo un filtro di riconoscimento delle forme.

Le reti neurali imparano troppo lentamente: hanno bisogno di migliaia, milioni, persino miliardi di dati per sviluppare «l'intuizione» in un dominio specifico.

Come apprende il nostro cervello secondo Dehaene

Il cervello umano è capace di integrare velocemente una singola informazione, ad esempio memorizzare un nuovo verbo e poi coniugarlo e inserirlo in altre frasi, un compito che le macchine non svolgono ancora bene. Le regole grammaticali sono solo un esempio del particolare talento del nostro cervello che ci consente di scoprire le regole generali che si nascondono dietro i casi particolari. Per esempio molti bambini scoprono che ogni numero n ha un successore $n+1$, e dunque che la sequenza di numeri interi è infinita. Nessuna rete neurale è attualmente in grado di rappresentare una conoscenza così sistematica per cui “ogni numero ha un successore”.

I valori di verità assoluti, vero o falso, non sono la loro specialità. La sistematicità, la capacità di generalizzare sulla base di una regola simbolica anziché sulla base di una somiglianza superficiale, sfugge ancora ai modelli attuali: l'apprendimento che chiamiamo profondo manca ancora di una comprensione profonda.

La singolarità del nostro apprendimento secondo Dehaene

L'apprendimento è una sorta di programmazione: imparare significa selezionare, tra tutte le formule disponibili nel linguaggio del pensiero, la più semplice tra quelle che si adattano ai dati. Le reti neurali non sono oggi in grado di rappresentare il ventaglio di frasi, formule, regole e teorie astratte con cui il nostro cervello modella il mondo. Non è una coincidenza: c'è qualcosa di profondamente umano esclusivo del nostro cervello, e che le neuroscienze non sono ancora in grado di spiegare, una vera propria singolarità della nostra specie. Il nostro cervello sembra essere l'unico a disporre di una serie di simboli combinabili secondo una sintassi complessa e dalla forma ad albero. Il cervello umano sembra disporre di una specie di computer interno, una macchina di Turing, in grado di concatenare operazioni in un ordine arbitrario, con la massima flessibilità. Le attuali reti neurali non hanno ancora questa flessibilità. Ciò che apprendono rimane confinato in connessioni nascoste, inaccessibili, disperse e molto difficili da riutilizzare per altri compiti, anche se simili. Manca loro la capacità di comporre le conoscenze che hanno appreso, vale a dire ricombinarle per risolvere nuovi problemi. [Dehaene]

Stupida ma sempre più ubiqua e decisiva

Eppure ogni giorno l'IA realizza nuove conquiste, che fino al giorno precedente sembravano prerogative dell'intelligenza umana. Forse qualcosa di veramente intelligente dell'IA sfugge a quegli studiosi che la giudicano in fondo stupida.

Se l'IA è stupida e noi intelligenti perché ci affidiamo sempre più ad essa? Non siamo ancora più stupidi a fidarci di essa?

Se è stupida dobbiamo chiederci come mai le affidiamo la gestione dei mercati finanziari, che sono il tempio della nostra società.

La sicurezza degli Stati significa ogni giorno di più IA.

Forse non ci viene raccontato tutto dell'IA.

La nozione stessa di IA è oscura?

Alcuni studiosi denunciano che **la definizione stessa d'intelligenza artificiale è opaca**. Oggi questa espressione abbraccia una gamma vastissima di tecnologie, dalle più banali alle più inquietanti.

Si dice spesso che certi reti neurali, per via della loro complessità, sono oscure in quanto **non sappiamo come prendono le loro decisioni**. Si definisce «scatola nera» “qualsiasi sistema d'intelligenza artificiale i cui meccanismi e la cui logica interna siano opachi o inaccessibili alla comprensione umana”.

Ma per alcuni critici sono le aziende che le producono che ci impediscono di sapere come le IA funzionano per proteggere i loro segreti industriali, in particolare quello di non svelare le scelte discutibili dei proprietari sugli scopi del sistema.

IA non umana?

Infatti non bisogna fare l'errore di descrivere l'intelligenza artificiale solo in termini di imitazione dell'intelligenza umana, sostengono questi critici. L'obiettivo potrebbe essere creare una macchina in grado di risolvere problemi che per un essere umano sarebbero insormontabili. In questo caso avremmo a che fare con un'intelligenza non umana. La prima domanda che dovremmo fare di fronte a un sistema d'intelligenza artificiale è quali tecniche sta impiegando. L'evidente mancanza di trasparenza contraddice il concetto di "intelligenza artificiale fidata" che l'Unione europea sembra così ansiosa di promuovere. La *Ico-At guidance* mostra che si può fare molto per spiegare il funzionamento interno di un sistema d'intelligenza artificiale se i soggetti coinvolti sono disposti a fare uno sforzo in questo senso. Ma le imprese impegnate nell'IA sono estremamente riservate sulle loro ricerche.

Il problema del pregiudizio nei dati

In assenza di controllo, pregiudizi ed errori presenti nei dati possono amplificarsi a dismisura, di pari passo con la crescita del sistema e il diffondersi di soluzioni AI sempre più complesse. Ad es. un'AI addestrata con dati sbagliati potrebbe applicare pregiudizi razziali, di genere o di religione (o di altro tipo), quando seleziona i candidati per un lavoro o decide a chi assegnare un prestito.

Una delle strategie più comuni, ad oggi, consiste nella selezione accurata dei dati. Se si vuole evitare che l'AI si trasformi in un mostro "alla Skynet« [un'IA che ha prodotto Terminator] allora bisogna alimentare il sistema di Machine Learning con dati privi di rischi.

Una cosa che è molto più facile a dirsi che a farsi.

Per i filosofi eliminare i pregiudizi è utopistico.

Per alcuni una soluzione sta nelle XAI secondo cui il sistema AI deve essere sempre in grado di spiegare dettagliatamente perché ha preso una certa decisione. Ma la loro efficienza è inversamente proporzionale alla loro efficienza.

L'IA ci sorveglia continuamente e prevede i nostri comportamenti

L'IA è il mezzo essenziale del «capitalismo della sorveglianza». Google, Facebook, ecc. ci tracciano per conoscere tutto quello che facciamo. I numerosissimi dati sulle nostre vite sono raccolti e processati da IA che elaborano i nostri profili psicologici per cui sanno quello che vogliamo, desideriamo, pensiamo e scegliamo, e soprattutto pretendono di prevedere esattamente cosa faremo. In Cina, all'avanguardia del riconoscimento facciale, oltre 200 milioni di telecamere associate all'IA sorvegliano in ogni istante i cittadini per prevenire ogni reato e stroncare il dissenso. Tutti i cittadini cinesi sono costantemente sorvegliati. Alcuni esperti affermano che queste IA, oltreché ficcare il naso nelle nostre vite molto più di quanto facesse la STASI nella DDR, ci conoscono molto meglio di quanto noi conosciamo noi stessi.

La manina nascosta dei proletari digitali

C'è un altro lato oscuro, ancora più sfuggente, per alcuni, dal momento che le grandi aziende di software si adopererebbero per rafforzare l'illusione che l'IA sia davvero autonoma.

Alcuni affermano che i macchinari, soprattutto i migliori e i più nuovi, hanno bisogno di un continuo intervento umano. Molti dei progressi più recenti dell'intelligenza artificiale nell'ambito del *machine learning* si devono non solo alle innovazioni tecniche di hardware e software, ma anche alle nuove opportunità di mobilitare rapidamente e a basso costo risorse umane da mettere al servizio delle macchine. L'intelligenza artificiale ha bisogno di scienziati ed esperti, ma anche di schiere di operai digitali, che raccolgono il materiale didattico per i sistemi e ne sorvegliano i progressi nell'apprendimento.

Poco rispettosa dell'ambiente

Un ultimo aspetto assai poco noto interessa l'ecologia. Uno studio dell'università di Amherst nel Massachusetts ha evidenziato che un progetto standard di apprendimento automatico emette nel suo ciclo di sviluppo 284 tonnellate di CO₂ equivalente.

Le ricerche di IA non considerano neppure l'efficienza energetica anche perché ci si è resi conto che le grandi reti neurali sono più funzionalmente efficienti anche se sotto il profilo energetico lasciano a desiderare. Insomma una rete neurale per dare risultati rilevanti deve consumare enormi quantità di energia.

L'IA purtroppo è poco compatibile con l'ambiente.