

Incoscienza artificiale. Intervista a Massimo Chiriatti

di Giacomo Bottos

19-03-2022

Qual è la natura dell'intelligenza artificiale? Quali sono le potenzialità e i rischi di delegare alcune funzioni finora appannaggio dell'essere umano a sistemi autonomi che operano in maniera profondamente diversa? A partire da queste domande abbiamo intervistato Massimo Chiriatti, autore di *Incoscienza artificiale*. Come fanno le macchine a prevedere per noi - edito da Luiss University Press con prefazione di Luciano Floridi e postfazione di Vincenzo Paglia.

Massimo Chiriatti, tecnologo e dirigente informatico, collabora con università e consorzi per eventi di formazione sull'economia digitale, è esperto di innovazione e futuro del lavoro ed è autore anche di *#Humanless*. L'*algoritmo egoista* (Hoepli 2019).

Fin dal titolo del libro si sottolinea il carattere di "incoscienza" legato alle tecnologie di intelligenza artificiale. Cosa si intende qui per coscienza? Questo carattere di incoscienza è temporaneo - legato allo sviluppo attuale della tecnologia - o strutturale?

Massimo Chiriatti: Senza pretendere di definirla in modo esaustivo e definitivo, del resto è un tema ancora aperto, qui intendiamo la coscienza come la funzione della mente umana che percepisce i pensieri su noi stessi e sul mondo che ci circonda. È un sentire in senso figurato, è "quello che sentiamo mentre stiamo pensando". Caratteristiche importanti della coscienza includono la soggettività (la sensazione che l'evento mentale, come il piacere o il dolore, ci appartenga), la continuità (sembra ininterrotta) e l'intenzionalità (è diretta verso un oggetto). Resta da vedere se tutte queste straordinarie funzioni possano essere realizzate un giorno all'esterno del nostro corpo. Nel frattempo, la macchina non decide quali attività eseguire e soprattutto non sa spiegare perché ha scelto di farlo in un certo modo: siamo noi che decidiamo di far svolgere alla macchina un lavoro. Per questo le macchine ci costringono a dare segni e atti concreti, non ambigui. L'ambiguità, l'incertezza e l'intenzionalità sono quello che ci differenzia e che ci preserva dal non diventare artificiali. Queste attività sono così complesse che non si vede ancora come le macchine possano autonomamente svilupparle e sono prive - strutturalmente - anche di uno strato biologico. Se pensiamo al classico modello input-elaborazione-output, qui si tratta di tradurre le percezioni (input), le decisioni (elaborazione) e i giudizi umani (output) in algoritmi. Ma l'elaborazione consapevole ci differenzia dalle macchine e la consapevolezza è proprio ciò che dà significato alla vita. Quindi, nel caso delle macchine, non resta che far apprendere loro modelli di comportamento, se ne abbiamo di buoni da insegnare. E noi che siamo in grado di percepire cosa sente l'altro abbiamo un vantaggio competitivo.

Nel libro viene usato il nome "lasima" per definire l'intelligenza artificiale. Quali le ragioni di questa scelta?

Massimo Chiriatti: L'intelligenza artificiale (IA), prima di essere qualcosa o addirittura qualcuno, esiste solo e sempre in relazione agli esseri umani. Noi le diamo in mano il nostro universo probabilistico, ma senza accettare a scatola chiusa il suo rigido calcolo e i suoi risultati, perché siamo in possesso della coscienza. Pensiamo grazie agli altri, ora anche grazie all'IA, e dipenderemo sempre tutti gli uni dagli altri, nel bene e nel male. Per queste ragioni non prevarranno né gli scenari utopici né quelli distopici: il futuro non dipenderà solo dalla selezione naturale e dai

limiti ambientali, ma soprattutto dalle scelte e omissioni che faremo nel campo dell'IA. A proposito, non dovremmo più chiamarla macchina, perché si tratta di un soggetto agente: *lasima*. Così le diamo una connotazione attiva, di agente artificiale guidato dai numeri, ma non da una coscienza. *lasima* può fare molto più di noi, non solo perché è più veloce e precisa, ma anche perché non sa cosa siano l'errore e la paura. E per l'appunto, chi non ha paura di sbagliare è incosciente.

Lei mutua i contatti di Sistema 1 e Sistema 2 da Daniel Kahneman, innestandovi una riflessione sull'intelligenza artificiale. Cosa si intende con questi termini e perché sono concetti rilevanti per innestarvi una riflessione sull'intelligenza artificiale?

Massimo Chiriatti: Per descrivere i nostri processi cognitivi, il premio Nobel Daniel Kahneman usa i concetti di Sistema 1 e Sistema 2. Il Sistema 1 funziona automaticamente e rapidamente, con poco o nessun sforzo e nessun senso di controllo volontario. È responsabile di tutte le cose a cui non dobbiamo pensare molto, come lavare i piatti, lanciare una palla o acquistare beni indifferenziati. Il Sistema 2, invece, svolge attività mentali più impegnative, che richiedono un approfondimento cognitivo, come per esempio l'esecuzione di calcoli complessi. Il Sistema 1 è veloce, intuitivo ed emotivo, mentre il Sistema 2 è lento e logico. Un esempio di pensiero del Sistema 1 è rilevare che un oggetto è più distante di un altro, mentre un esempio di pensiero del Sistema 2 è trovare il modo di parcheggiare in uno spazio ristretto. Usando i due Sistemi, Kahneman descrive come si formano i processi decisionali, con tutti i pregiudizi connessi. Il processo di arricchimento della nostra conoscenza, con il ragionamento (Sistema 2) e con l'intuito (Sistema 1), è la strada che ci ha fatto avanzare nel progresso umano; adesso una nuova macchina (il Sistema 0, chiamato *lasima*) ci aiuta nel prendere decisioni. *lasima* è esterna al nostro corpo e osserva tutti i nostri atti: li cattura sotto forma di dati che elabora e filtra. Il punto fondamentale è con quali criteri faccia tutto questo. Poiché non impiega regole intenzionalmente precostituite dalle persone, pertanto siamo nelle peggiori condizioni possibili: *lasima* "apprende" dai dati con modalità tutte sue. Se usiamo il termine "apprendimento" rischiamo di confonderci - e meravigliarci - quando essa simula la nostra percezione, pensando erroneamente che un mucchio di silicio e bit abbia le nostre stesse caratteristiche. *lasima* ci confonde perché è un sistema adattivo, che cambia imparando. La proposta di aggiungere il Sistema 0 può suonare come una trovata letteraria, un semplice strato che si aggiunge a quelli di Kahneman, ma in realtà è fondamentale per capire come ci stiamo evolvendo. Ricordiamoci però che il Sistema 0 è solo una simulazione del ragionamento umano: le sue non sono "decisioni" vere e proprie.

In che modo, a suo avviso, la prospettiva di John Searle fornisce un contributo utile alla discussione sull'intelligenza artificiale?

Massimo Chiriatti: La coscienza è l'essenza stessa della nostra esistenza. Per spiegare come *lasima* ne sia sprovvista, ricorriamo al filosofo statunitense John Searle, riprendendo in particolare il suo esperimento della "stanza cinese", una critica fondamentale all'idea di una vera IA. Perché se l'IA non è cosciente, allora non può essere neanche intelligente. Searle ritiene, seppur criticato in letteratura, che non possa esserci coscienza senza un supporto biologico adeguato, mentre i tecno-ottimisti credono che basti aumentare a dismisura la potenza elaborativa affinché essa emerga. Per lui l'esecuzione di un programma, ossia la semplice manipolazione dei simboli, non è una condizione sufficiente per spiegare la comprensione. Il cuore della sua argomentazione è un esperimento mentale in cui immagina un soggetto solo, chiuso dentro una stanza, che possiede tutte le istruzioni necessarie per rispondere in modo appropriato ad alcune domande scritte in cinese (lingua che il soggetto non conosce). A chi è all'esterno della stanza e analizza le risposte, sembra che "la stanza" conosca il cinese, quando in realtà si tratta di una manipolazione di simboli. Perciò non si può dire che un computer comprenda una lingua per il solo fatto che sia in grado di

"usarla". Nel ragionamento di Searle, la comprensione non può essere manipolazione di simboli formali. Il computer non può comprendere il linguaggio umano, ma può solo essere in grado di analizzare formalmente le strutture sintattiche. IASIMA ha la competenza sintattica, sa leggere e scrivere una frase grammaticalmente corretta, ma non è in grado di estrapolare il significato dalle frasi. Quindi, per un modello computazionale che manipola simboli, non è possibile fare un'esperienza: è incapace di riprodurre fenomeni intenzionali umani o esperienze soggettive in prima persona. Perché il pensare, come lo intendiamo noi esseri umani, richiede un contenuto semantico, mentre il calcolo ha soltanto una sintassi, cioè dei processi formali e simbolici, ed è insufficiente per realizzare intenzionalità, sentimenti, emozioni, dolore, e così via.

Cosa intende per IA debole, forte e generale? Ritiene effettivamente possibile l'emergere di un'IA generale?

Massimo Chiriatti: L'IA può essere qui tipizzata su tre livelli: 1) IA debole, che corrisponde alla percezione aumentata ed è al servizio dell'essere umano; 2) IA forte, che corrisponde alla cognizione aumentata ed è appena paragonabile all'essere umano; 3) IA generale, che corrisponde a un essere umano aumentato ed è, come vedremo, un obiettivo utopistico, al momento. In quest'ultima visione utopistica, si tratta di colmare in poco tempo una distanza che l'evoluzione naturale ha coperto in milioni di anni, replicando artificialmente la complessità evolutiva via hardware e software.

Da dove nasce l'opacità dei meccanismi di deep learning e quali problematiche introduce?

Massimo Chiriatti: In alcune implementazioni è oggettivamente complesso rendere trasparente il funzionamento: sia perché i dati mentono, sia per gli errori nelle loro elaborazioni. Ciò che accade dentro la black box è un paradosso, perché le previsioni sono verificabili ma non è possibile comprendere come siano state ottenute. Come si è arrivati a determinare quale sia l'oggetto rappresentato in milioni di immagini con miliardi di pixel analizzati da milioni di neuroni con miliardi di connessioni è per noi molto arduo da comprendere. Quando qualcosa non va per il verso giusto, noi ci assumiamo la colpa, oppure la diamo a qualche fenomeno naturale (il meteo), oppure a qualche fenomeno sociale (la mancata collaborazione altrui). Ma se sbaglia l'IA, e sappiamo che è previsto che sbaglia, a chi diamo la colpa? Non sono chiare le quote di responsabilità di chi l'ha prodotta, chi l'ha allenata e chi l'ha impiegata. Occorre un nostro intervento, consapevoli però che anche il nostro cervello non pensa in modo lineare e dunque anche per noi risulta difficile spiegare le vere ragioni per le quali abbiamo preso una decisione. Anche se lo strumento lavora in modo oscuro, noi abbiamo il dovere di scegliere il fine, che deve essere sempre trasparente.

Quali problemi introduce il concetto di delega a un sistema autonomo?

Massimo Chiriatti: Potremmo classificare la tecnologia in tre forme: 1) semplice, che richiede il fare, estendendo e rinforzando le nostre capacità manuali. Per esempio, il martello e la forbice; 2) automatica, che automatizza le regole che noi abbiamo immesso. Per esempio, il robot e la calcolatrice; 3) autonoma, che calcola con le regole che ha "imparato" autonomamente, con nostra delega, dalla realtà. L'atto di automatizzare è una forma di delega e corrisponde alla diminuzione della nostra attenzione, con annesso un certo livello di rischio. Il fatto che da parte nostra non vi sia né una previsione del risultato né un intervento durante il processo, è un bene o un male? Questa domanda apre a un paradosso: progettiamo l'architettura delle macchine per fare previsioni, ma non è possibile progettare ex ante le risposte che ci forniranno, perché nel contesto del deep learning, per esempio, non sarà possibile spiegare come si è giunti alle conclusioni. Ma la domanda più importante è: stiamo delegando alcune nostre capacità, oppure le macchine che riteniamo - erroneamente - nostre schiave stanno iniziando a raccomandarci le azioni escludendoci dal giudizio

sulle loro previsioni? Stiamo essenzialmente delegando la nostra fiducia a un'entità che non è assolutamente in grado di comprendere quello che comporta ogni sua azione e che non può esserne responsabile. Stiamo cercando, attraverso le macchine, di deresponsabilizzarci, quando invece dovremmo essere più responsabili per evitare rischi sistemici e imprevedibili. Perché ci comportiamo così? Alcune cause di questo comportamento sono storiche: le persone connesse in rete sono lontane nello spazio, mentre i tempi di comunicazione si accorciano; altre sono tecniche ed economiche: la facilità di ottenere risposte su tutto in real time; altre ancora sono sociali: non ci fidiamo più dell'altro, e neanche di noi stessi. Se questo è vero, allora stiamo cambiando il mondo alla cieca: siamo ciechi davanti alla mole di dati, ciechi di fronte ai pregiudizi contenuti in dati e modelli, ciechi perché non sappiamo dove stiamo andando. Eppure, ci stiamo affidando alle macchine più che agli esperti, perché *lasima* sembra (ma non è) agnostica e affidabile. Seguendo l'IA, che fine faranno le nostre scelte? Nessuna scienza può rispondere, solo la filosofia può orientarci in un tale giudizio, ed è quindi necessario un nuovo umanesimo per capirlo. L'IA è inevitabile, ma non la delega umana. Sarebbe un errore presumere che questo trasferimento di delega comporti una simultanea assoluzione dalle responsabilità.

È possibile andare oltre il sistema basato sulla correlazione per introdurre forme di IA in grado di impiegare il concetto di causalità?

Massimo Chiriatti: La macchina individua correlazioni che però possono essere del tutto casuali. Si parla allora di correlazioni spurie, ossia di relazioni matematiche in cui due o più eventi o variabili sono associati per una coincidenza o per la presenza di un terzo fattore invisibile, ma non sono correlati causalmente. Le correlazioni non causali di questo tipo si rilevano empiricamente e possono sembrare anche consistenti. Qualsiasi relazione causale, per definizione, è unidirezionale. Questo è un principio fondamentale per distinguere tra correlazione e causalità, poiché il processo decisionale richiede una comprensione causale. In alcune condizioni la correlazione non è computabile, sia per la mancata capacità di comprendere i nessi causali, sia per l'abbondanza di informazioni, come per esempio nel linguaggio, dove non è possibile capire il significato solo seguendo le istruzioni: il caso della stanza cinese lo spiega bene. Pur con tanti investimenti sulla potenza computazionale, a *lasima* manca la capacità prettamente umana di capire il rapporto causa-effetto[1]. Siamo bravi a usare le macchine per far emergere le correlazioni, ma non riusciamo ancora a fare granché con le relazioni causali. È solo grazie al nostro Sistema 2 che possiamo elaborare le relazioni causa-effetto, operazione al momento impossibile per *lasima*. Questa rimane una impareggiabile caratteristica del pensiero umano.

In che modo l'IA ridefinisce il rapporto soggetto-oggetto?

Massimo Chiriatti: Come ha luogo il processo decisionale quando qualcuno ha già deciso per noi è un tema politico, perché riguarda sia chi ha il potere di decidere, sia le basi su cui è stato scelto. Ma cosa accade quando qualcosa sta per decidere per noi è un problema filosofico, soprattutto nel momento in cui quel qualcosa sta diventando qualcuno; quell'oggetto sta diventando soggetto. Siamo le nostre scelte. Viviamo nello spazio delle scelte. Se riduciamo questo spazio, perché la macchina ha già deciso di restituirci una versione ridotta e arbitraria dello spettro delle possibilità, non esercitiamo un'effettiva scelta di valore. Sarebbe allora come cadere in un imbuto dove i nostri valori, fatalmente, si annullano.

Quali rischi pongono i modelli di business delle Big Tech?

Massimo Chiriatti: Le Big Tech sono entità che ricordano fortificazioni medievali e che definiamo più prosaicamente *walled garden* ("giardini recintati"), ossia piattaforme chiuse, ma la cui chiusura non ha lo scopo di ostacolare tanto l'accesso, quanto l'uscita. I muri fisici sono antistorici su questo

piccolo pianeta, e per fortuna la tendenza, almeno sul medio-lungo periodo, sembra quella di abatterli. Purtroppo, però, stiamo erigendo muri immateriali, che sono anche più pericolosi. Le grandi piattaforme digitali, dato che più dipendono dalle leggi, dai costumi o dalla lingua locale, meno sono socialmente scalabili, dopo aver svuotato il potere degli Stati nazionali le piattaforme devono per forza diventare entità sovranazionali. Hanno immense possibilità economiche, perché hanno dalla loro parte (il portafoglio di) miliardi di persone e i loro dati. Insomma, il loro non è un business model, ma uno State model, di successo e, al momento, a prova di futuro.

Cosa può voler dire in questo contesto mettere l'uomo al centro?

Massimo Chiriatti: C'è un punto molto importante da considerare: ogni previsione di IASIMA è una valutazione quantitativa, mai qualitativa, mentre per noi umani una scelta non è quasi mai un semplice calcolo. Noi decidiamo in base a valori incommensurabili e quindi incomputabili. Siamo noi gli insegnanti delle macchine. Lo siamo implicitamente quando assimilano i dati che noi creiamo, quando costruiscono il modello e ci forniscono le risposte. Lo siamo esplicitamente quando diamo loro istruzioni su come svolgere un lavoro. Per queste ragioni dobbiamo porre attenzione a come apprendono, perché nel farlo evolveranno. Questo è il tema: se IASIMA sa valutare meglio di noi, e impara più di noi, siamo automaticamente tagliati fuori dal loop? No, il ruolo dell'essere umano in questa evoluzione resta essenziale e sempre più importante, perché non dobbiamo confondere l'intervento con l'osservazione e voltarci dall'altra parte: non dobbiamo intervenire nel processo automatico di analisi, ma supervisionarlo. IASIMA, non avendo un'esperienza soggettiva, lavora senza coscienza esponendoci a rischi imprevedibili.

Cosa occorre fare affinché si realizzi uno scenario positivo di utilizzo di queste tecnologie?

Massimo Chiriatti: Il tema dell'IA - un tema globale, politico, tecnico e filosofico - è così ampio che avanzare proposte è molto ambizioso. Tuttavia, nel testo ne sono umilmente indicate alcune. L'IA sta generando contenuti più che reali, ma anche problemi sociali, etici, geopolitici e di sostenibilità. Abbiamo visto che le persone e le macchine hanno differenti capacità e debolezze, ma se le studiamo, le riconosciamo e le coordiniamo, lavorando insieme potremo affrontare anche i problemi al momento impensabili. Occorrono sforzi comuni, nella formazione e nell'investimento, altrimenti lo sviluppo non solo è insostenibile, ma addirittura incompatibile con i valori dell'umanesimo. Finora abbiamo costruito macchine che giocano contro altre macchine o contro di noi, ma ora abbiamo bisogno di altro: la cooperazione tra noi e con IASIMA. Dobbiamo democratizzare l'IA e dobbiamo farlo tutti insieme. Pertanto, dobbiamo preparare gli antidoti per non sprecare l'intelligenza umana in modo incosciente. Il futuro ci giudicherà.

[1] Cfr., Judea Pearl e Dana Mackenzie, *The Book of Why. The New Science of Cause and Effect*, Penguin Books, Londra 2018.