

Le frontiere dell'umano: i dilemmi dell'algoretica. Intervista a Paolo Benanti

di Giacomo Bottos, Eleonora Desiata

21-07-2021

La crescente pervasività dell'uso di algoritmi in pressoché tutti gli ambiti della vita e lo sviluppo dell'intelligenza artificiale, solleva fondamentali questioni filosofiche che riguardano in ultima analisi l'identità e i confini dell'umano. A partire da queste problematiche, occorre impostare la discussione su quella che Paolo Benanti ha chiamato algoretica. Benanti è teologo, docente presso la Pontificia Università Gregoriana e autore di numerose opere tra cui: *Oracoli*. Tra algoretica e algocrazia (Luca Sossella Editore 2018) e *Digital Age. Teoria del cambio d'epoca. Persona, famiglia e società* (San Paolo 2020). Da anni si occupa di temi legati all'etica delle tecnologie, alle neuroscienze e alla gestione dell'innovazione tecnologica.

La trasformazione digitale è un processo che coinvolge profondamente le nostre società e ci porta a un ripensamento di forme di vita e categorie di comprensione. Lei ha parlato, a proposito del passaggio che stiamo vivendo, di un 'cambio d'epoca'. Ritiene che il concetto di 'frontiera' possa aiutarci a mettere a fuoco ciò che sta avvenendo?

Paolo Benanti: Farei una premessa: se noi cercassimo il significato di 'frontiera' su un dizionario troveremmo tre significati. In primo luogo frontiera intesa come limite, e in questo senso 'il limite del cerchio', la circonferenza, è anche qualcosa che ha a che fare con la sua identità, è quel luogo dei punti che fa del cerchio un cerchio e non un quadrato. In seconda istanza, con frontiera intendiamo, soprattutto a partire dalle suggestioni della letteratura western, un luogo poco abitato. In terza istanza, la frontiera divide due mondi culturalmente diversi che normalmente parlano lingue diverse, il di qua e il di là. Secondo me possiamo declinare questa nostra questione legata alla tecnologia e alla frontiera in relazione a tutti e tre questi ambiti. In primo luogo, dunque, la 'macchina Sapiens' - se così possiamo chiamarla -, intesa come nuovo strumento in grado di raggiungere certi gradi di autonomia, come algoritmo che in certe misure riesce a fornire o a gestire diritti e possibilità delle persone, è qualcosa che mette in crisi un limite, ovvero l'identità dell'Homo Sapiens. Quando Linneo si è occupato della classificazione dei viventi, ha cambiato in particolare la nostra classificazione e, a partire da una situazione in cui sopra lo scheletro umano potevamo scrivere la frase socratica 'conosci te stesso' - un interrogativo che era quasi un mandato - ci ha reso un sotto-ramo di scimmie differenti solo per la caratteristica 'Sapiens'. Ora che una macchina è in grado di imitare questo tipo di caratteristica, la vera domanda è: 'Chi è l'uomo?'. È, appunto, una questione di frontiera, una questione che ha a che fare con l'identità, perché è evidente che una macchina che può controllare le decisioni umane ci mette in crisi, ci pone la questione di quali decisioni debbano assolutamente rimanere in mano ad un essere umano e quali possano essere surrogate da una macchina. Ora che la macchina è in grado di prendere decisioni mediche, ora che la macchina potrebbe fungere da giudice, ora che la macchina potrebbe interagire sulla vita e sulla morte delle persone, ci chiediamo: 'Che differenza c'è tra una scelta compiuta da un medico umano e quella compiuta da un algoritmo? Che differenza c'è tra un giudice e un algoritmo?'. Non è una questione di mera efficacia. Da sempre, in ogni cultura, il giudice, il medico e il prete sono figure scelte all'interno di una comunità in funzione del loro impegnarsi fino in fondo nella professione, e che per

questo ricevevano il mandato, se vogliamo usare un'espressione forte, a 'mettere le mani sulle persone'. Tutto questo oggi conosce una nuova stagione di interrogativi e un cambio d'epoca, perché siamo chiamati a ritrovare il fondamento del nostro vivere sociale, esattamente come accaduto quando abbiamo scoperto che noi europei non eravamo l'unico 'modello' di uomo sulla Terra. La scoperta di nuove culture, soprattutto nelle Americhe, fu seguita da una intensa discussione su cosa volesse dire 'essere umano'. Da lì si è avuta, non senza incertezze, non senza costi di vite umane, una nuova fondazione di quelli che poi sono poi stati i diritti umani. Ancora prima era accaduto con le 'invasioni barbariche' e con il Medioevo. Questo ritorno ciclico della domanda sul noi stessi - un cammino mai completo - oggi si ripresenta nella forma dell'innovazione tecnologica, dell'algoritmo. Oggi, però, ci confrontiamo con la 'frontiera' anche nell'accezione di territorio scarsamente popolato. Potremmo interpretare quello che sta accadendo come una colonizzazione non umana del West, ma di segno opposto, dove la conquista parte dagli estremi confini del West, la Silicon Valley - il territorio della prima corsa all'oro - a partire dall'assunto che l'oro non è più composto da pepite, ma dai nostri dati... Questa colonizzazione non parte solo dal Far West, ma anche dal Far East, non più oggetto della scoperta di Marco Polo, ma l'esatto contrario. Le vecchie frontiere geografiche ritornano e cercano di conquistare qualcosa di profondamente nostro, i nostri dati. Che effettivamente valgono oro, poiché contengono informazioni. Non c'è coincidenza tra dati e informazioni: c'è un passaggio logico in mezzo, che la filosofia, il nostro modo di affrontare i problemi, ancora non ha completamente 'digerito'. Un'analisi filosofica sulla differenza fra dati e informazione non è ancora genericamente condivisa, ed ecco che, allora, in questo territorio scarsamente popolato vale a volte la legge del più forte. Questo richiede una nuova civilizzazione. Il terzo significato riguardava il rapporto tra culture diverse. Questo è evidente se guardiamo a quello che sta succedendo sulle piattaforme digitali e a quello che stanno producendo TikTok, YouTube, i podcast e via dicendo. Una cultura del digitale, che è differente dalla cultura analogica. È uno scenario di frontiera.

L'intelligenza artificiale è un ambito oggi in forte espansione, in relazione al quale sorgono molti degli interrogativi a cui lei accennava. Ma in che senso si può parlare in questo caso di intelligenza? Questi sviluppi ci spingono a interrogarci, in questo scarto, sulla stessa intelligenza umana? Esiste una dimensione del pensiero umano non assimilabile alla computazione algoritmica?

Paolo Benanti: Per poter rispondere a questa domanda dobbiamo considerare due elementi. Il primo elemento è che la risposta sarà nella forma del paradosso. Vogliamo delegare così tante capacità all'intelligenza artificiale che, paradossalmente, questo ci mette in crisi sotto almeno due aspetti. Di fatto, Alan Turing ha definito l'intelligenza artificiale con il famoso test di Turing. Ovviamente nessuno voleva chiamare tutto questo universo 'intelligenza artificiale', ma questo nome ha esercitato un fascino così grande sulla cultura che, di fatto, si è imposto e anche IBM si è dovuta arrendere e quelli che erano i sistemi 'sapienti' sono stati rinominati intelligenza artificiale. Nel test 'Behind the Wall' dietro al muro c'è qualcosa o qualcuno, che io devo riconoscere parlandoci. Turing non era un filosofo, ma chi ha masticato un po' di storia della filosofia non può non ricordare il mito della caverna di Platone, nel quale i prigionieri vedono delle ombre che li ingannano sulla conoscenza. Allora, proprio nella sua forma costitutiva, l'intelligenza artificiale è un dubbio che riattualizza la famosa domanda di Socrate: 'Chi è' o 'Che cos'è?'. Ecco, la macchina è un qualcosa che funziona o è un qualcuno che ragiona? Questo primo paradosso è il paradosso che io lascerei in questa forma, perché, in realtà, nel modo di parlare di chi oggi decide di intelligenza artificiale questa domanda non è portata fino in fondo. Vengono vendute tante soluzioni che ci aiutano, come gli assistenti digitali negli smartphone o in casa, ma questa diffusione pratica dell'intelligenza artificiale ancora non risponde a questa domanda. Il secondo aspetto, ugualmente paradossale, che vorrei sollevare riguarda una questione che chi si occupa di robotica ha molto presente: il paradosso

di Moravec. È molto facile dare alla macchina una capacità cognitiva alta, quella che un ragazzo acquisisce intorno ai 14-15 anni. Per fare la radice cubica di nove a scuola abbiamo dovuto aspettare un po' di anni, una calcolatrice la fa in maniera immediata. Al contrario un bambino a un anno e mezzo o due è in grado di aprire la maniglia della porta di casa e scappare, mentre una mano robotica ancora non è capace - allo stato attuale della tecnologia - di aprire una maniglia. Questa situazione paradossale ci dice che forse abbiamo avuto un problema nell'identificare il sistema nervoso centrale inteso come cervello con la sede dell'intelligenza, perché quello che vediamo nell'essere umano è che l'intelligenza è distribuita anche all'interno di tutto il processo motorio. Ed ecco perché non è possibile tradurre immediatamente quella grande capacità intellettuale che consente di incorporare in un chip funzioni matematiche molto complesse, nella capacità di movimento di una mano robotica. Abbiamo concettualizzato l'intelligenza come se fosse qualcosa di astratto e residente in un'unità di elaborazione centrale, mentre invece è qualcosa di corporeo: è incarnata, appartiene al nostro corpo. Ecco, questa situazione paradossale dice molto delle possibilità e dei limiti dell'intelligenza artificiale. È chiaro che se devo creare un sistema per selezionare - grazie al riconoscimento delle immagini - in una catena continua di chicchi di caffè i chicchi difettati, l'analisi statistica va bene. Un operaio dovrebbe guardare chicco per chicco, mentre una fotocamera processa trecento chicchi al secondo e, in caso di errore, nella peggiore delle ipotesi si sprecherà qualche chicco ulteriore di caffè. Ma la persona umana non è un chicco di caffè, non parliamo di numeri, ma di individui, che sono portatori di un diritto fondamentale assoluto. E allora, ancora per tornare al primo tema, il problema non è solo nella macchina Sapiens, ma è nell'Homo Sapiens, che ha perso la capacità di distinguere tra persona e macchina. Usando di nuovo un'immagine: se dovessi vedere davanti a me un papero e un phon, questi hanno una forma molto simile, ma non li confondo perché il phon se lo spengo lo riaccendo, il papero se lo spengo non lo riaccendo più. La vita è un'esistenza, non è un funzionamento. Con questa battuta si può racchiudere un problema che parte da Chomsky e dall'idea della mente come una macchina che funziona, e arriva fino ai giorni nostri. È evidente che qui c'è un altro problema di soglia tra la macchina e noi, tra l'idea di un meccanicismo che descrive tutto da un lato e, dall'altro, il concetto di persona, un concetto che è stato la possibilità stessa di esistenza dell'Occidente. Oggi questo concetto sembra, in qualche misura, dissolversi dietro un'idea di essere umano che funziona.

Esistono visioni diverse dell'umano, che presiedono a diversi filoni di sviluppo e ricerca sull'intelligenza artificiale? C'è anche chi nega l'esistenza di questo problema in virtù di una concezione dell'umano che non vede una soglia o, invece, rintraccia una sensibilità diffusa in merito?

Paolo Benanti: La risposta più corretta è quella di mantenere il senso della complessità globale. Noi abbiamo la tendenza a essere occidente-centrici. Nei vari incontri e confronti che ho avuto, uno molto significativo è stato con Hiroshi Ishiguro, realizzatore di robot in Giappone, uno studioso di altissimo profilo, quando durante una conferenza mostrò i suoi robot discutere in maniera autonoma su quale fosse la migliore tempura di Tokyo. La questione fa sorridere, ma c'è qualcosa di più profondo. In Giappone la religione prevalente è animista, ed essendo animisti tutto ciò che animato ha la stessa dignità. Per noi questo appare un po' come il gioco del falsario: se ho una banconota da 5 euro, indistinguibile da una banconota vera, la mia banconota vale 5 euro, perché nessuno la riesce a distinguere. Ma se io ho un robot indistinguibile da una persona, quello è una persona? Con la nostra mentalità occidentale, plasmata dalla filosofia di Atene, non ci basta una risposta sul 'come sembra', vogliamo sapere 'che cos'è'. Nella pluralità delle intelligenze artificiali c'è chi non si pone questa domanda, per presupposti culturali. Ecco perché una tecnologia, in questo senso, è figlia della cultura che la elabora, non è mai un dispositivo di mera funzionalità. Su questo dobbiamo un po' superare un certo heideggerismo che rimane chiuso in un modello occidentale. Tra ingegneri,

in modo autoironico, si dice che "l'ingegnere non vive, ma funziona": alcune posizioni estreme nella Silicon Valley di fatto ragionano così e non si pongono la domanda sulla natura dell'uomo. Per cui, se Uber funziona meglio del vecchio taxi, non mi interessa la differenza tra Uber e il taxi, perché quella che è la mia esperienza - eccola, la frontiera estrema - è migliorata. Allora, quello che rimane dell'umano è l'elemento sensibile, e non è la prima volta che accade nella storia del pensiero. Solo che non è formalizzato, c'è un implicito vissuto, più che un esplicito riflesso. È come se l'eclissi dell'orizzonte filosofico in qualche misura fosse un dato di fatto della quotidianità; se Grozio scriveva *etsi Deus non daretur*, qui si vive *etsi quaestio non daretur* come se questo non dovesse farci questione. Ecco, in questo senso forse si inverte quella che era la profezia di Heidegger dell'Occidente come Terra del tramonto, cioè è tramontata per alcune estreme frange della nostra cultura una necessità di farsi alcune domande di senso che prima invece erano questioni chiave.

Partendo dal presupposto che una questione etica esista, e dunque che questa dimensione vada tenuta in considerazione nel pensare il cambiamento tecnologico, come si può definire il rapporto tra una macchina di intelligenza artificiale e l'ambito dei fini, dell'etica?

Paolo Benanti: Di nuovo possiamo giocare con una serie di allegorie, la prima è questa: la macchina dotata di intelligenza artificiale lavora su dati, dei valori di numeri, mentre la persona che fa una scelta etica lavora su dei valori che non sono numerici. Qui si apre una frontiera. Se vogliamo dare alla macchina un certo grado di indipendenza rispetto a un controllore umano, si apre la questione di come conciliare valori numerici con valori etici. Questo è il motivo per cui ho proposto di scrivere questo nuovo grande capitolo dell'etica, che si chiama algoristica. Ma cosa deve essere la algoristica? Non certo una consapevolezza etica della macchina, perché la macchina non è un qualcuno, altrimenti saremmo allo stesso problema di cui sopra. Possiamo intenderla come una sorta di guardrail etico, che tiene la macchina all'interno di una strada e per quanto possibile evita alcuni eventi infausti. Tutti sanno purtroppo che le barriere lavorano entro certi limiti, a volte ci sono anche sconfinamenti di corsia con brutti incidenti, però intanto il guardrail rappresenta una soglia, che rende percorribili alcuni tratti di strada. C'è poi tutta un'altra questione, ed è una questione che riguarda come gestire questa soglia di attenzione etica per la macchina. È chiaro che qui si tratta di uscire da un modello di etica delle professioni (per cui basta l'ingegnere che è etico, e tutto il resto segue a cascata), e per uscire da questo modello secondo me è molto utile il lavoro di uno dei miei mentori di quando facevo il dottorato negli Stati Uniti, Langdon Winner. In un paper degli anni Ottanta, per spiegare che cos'è l'etica delle tecnologie rispetto all'etica delle professioni, fece un esempio: se voi guardate i ponti in calcestruzzo che stanno sulla Parkway che porta da New York a Long Island, vedete dei ponti in calcestruzzo sull'autostrada; in realtà la questione etica lì si pone se noi andiamo a scavare un pochino nella storia di quei ponti, che sono stati voluti da Robert Moses, la cui biografia (scritta dal grande giornalista Robert Caro) racconta come Moses avesse delle idee sociali molto precise e quindi aveva chiesto - siccome la sua amata Jones Beach secondo lui doveva essere destinata solo alla white middle class - che i ponti venissero realizzati un piede e mezzo sotto misura, così gli autobus non sarebbero potuti arrivare alla spiaggia. Questo significa, concludeva Winner, che ogni artefatto tecnologico implica una relazione di potere, è un dispositivo di potere, di relazione tra le persone. È qui che la algoristica trova un suo secondo grande capitolo, che ha a che fare con il *questioning technologies*, la capacità di questionare la tecnologia perché emerga anche come la strutturazione di questa nuova galassia digitale realizzi o renda impossibile il passaggio di qualcuno, cioè l'esecuzione dei diritti di qualcuno. Quindi non si tratta solo di dotare la macchina di una capacità di giudizio, cosa che è impossibile, e nemmeno solo di surrogarla con questi guardrail etici, ma anche creare uno spazio di critica sociale in cui sia possibile chiederci cosa facciano gli algoritmi, che funzione abbiano. Perché i ponti di calcestruzzo oggi altro non sono che gli algoritmi che ci danno o ci negano l'accesso ad alcune aree della nostra vita.

Ci sono limiti strutturali alla comprensibilità degli algoritmi e dei loro meccanismi di funzionamento?

Paolo Benanti: Vedo spesso che gli informatici tendono a sentirsi un po' accusati di qualcosa di cui nessuno li accusa. L'algoritmo è un po' come il farmaco che dà il medico: non basta un buon medico che abbia fatto il giuramento di Ippocrate, il farmaco deve avere il suo bugiardino, cioè deve essere conoscibile nei suoi effetti benefici e nei suoi effetti collaterali. L'algoritmo è un qualcosa di proprietà - la stessa cosa accade ai farmaci, da quando la Bayer mise in produzione un qualcosa che si chiamava all'epoca 'antifebbrina', nonostante tutti sapessero che sostanza era, i medici erano costretti a dare quel farmaco perché era protetto da copyright, un prodotto dell'intelletto. Se un algoritmo che oggi è in una platform ormai è protetto dagli stessi diritti, dobbiamo pretendere che anche gli algoritmi, come i medicinali, dichiarino effetti positivi e negativi. Questo è un punto su cui non basta più l'etica, perché è un principio di soft law, servono invece dei principi di hard law, servono principi che siano in qualche misura cogenti, e provengano dagli Stati, dalle autorità politiche, perché veramente l'algoritmo potrebbe far saltare alcuni cardini che sostengono le nostre democrazie, almeno dagli Stati moderni fino ad oggi, uno tra tutti quello dell'uguaglianza delle persone e del fatto che i diritti sono stabiliti per legge e non arbitrariamente assegnati da altri soggetti. Per quanto riguarda la dimensione più strettamente tecnica, il primo ostacolo alla comprensibilità riguarda la natura stessa dell'algoritmo che è il frutto di astrazioni matematiche su una realtà di dati. Questo implica che per capire il funzionamento di un algoritmo ci sia bisogno di essere 'iniziati' a tutto questo, di avere le basi per la comprensione. La seconda questione è che la macchina, nel momento stesso in cui diventa intelligenza artificiale, supera quel protocollo di progettazione del tipo if this, then that..., in cui il programmatore, a priori, immaginava tutte le circostanze in cui la macchina si poteva trovare e decideva qual era la risposta. Adesso la macchina si adatta in funzione dei dati e questo fa saltare la rigida costruzione deterministica. È esattamente il punto dove la statistica vale ma non vale, perché non siamo solo numeri.

Passando attraverso la questione della macchina e dell'etica, e attraverso l'aspetto delle potenziali trasformazioni sociali che, come si diceva, possono assumere una portata enorme, in che modo a suo avviso è possibile effettivamente avere dei meccanismi, delle modalità, dei luoghi in cui queste dinamiche possano essere discusse, e in un qualche modo le scelte possano essere la conseguenza di questa discussione, consapevole anche delle possibili conseguenze?

Paolo Benanti: Direi semplicemente rivivendo quella che è la nostra natura di Occidente costruito attorno alla Polis, cioè tornando ad una piazza dove le diverse competenze si confrontano su quello che accade. Ecco, secondo me la questione è la stessa: si tratta di creare nuove piazze, nuove agorà in cui le diverse competenze possano confrontarsi, come stiamo facendo noi in questo momento, sulla pluralità e la problematicità connesse a questa evoluzione tecnica e tecnologica. In termini più moderni, si tratta di passare da un governo top-down a processi efficaci di governance, in cui diverse competenze contribuiscano fattivamente alla realizzazione e all'informazione del dispositivo legale che realizza tutto questo.