

LA CULTURA

Benanti: perché gli algoritmi non riusciranno a cambiarci

PAOLO BENANTI - PAGINA 31

LE IDEE

Paolo Benanti

L'AI può manipolarci ma non persuaderci
cosa ho imparato da eticista della tecnologia

Per non soccombere alle macchine dobbiamo capire la differenza tra rischio e pericolo. Si intervenga sul contesto sociale per rinnovare la fiducia nella legge e nella mente che la produce

PAOLO BENANTI

Cosa significa guardare alla tecnologia con un approccio etico? Non è certo cercare le tavole dei Dieci Comandamenti della tecnologia, ma è chiedersi e chiedere quale forma di potere inietta un certo tipo di innovazione tecnologica all'interno della società.

Prendete Manhattan come la conosciamo oggi. È stata progettata dall'architetto e politico Robert Moses, secondo il quale nella parte migliore della città doveva esserci per la parte migliore della popolazione.

Nella mentalità di Moses, la parte migliore della popolazione era la classe media ricca e bianca. Così progettò ogni singolo aspetto della conformazione cittadina, dalle strade alle infrastrutture, con il preciso intento di raggiungere quell'obiettivo.

Ecco, l'etica della tecnologia considera la tecnologia come un costrutto socio-tecnico, cerca di mettere in luce che tipo di relazioni ci sono tra tecnologia e vita sociale. Rispetto all'intelligenza artificiale, dovremmo chiederci quale effetto essa abbia sulle nostre capacità decisionali.

Il problema era già presente tra la fine degli anni Quaranta e l'inizio degli anni Cinquanta, quando Claude Shannon progettò il prototipo di un nuovo sistema di comuni-

cazione considerato il precursore dell'AI: un topolino meccanico cui diede il nome di Teo, che riusciva a uscire dal labirinto sbattendo colpo dopo colpo sulle pareti cablate. Creò una macchina diversa da quelle delle rivoluzioni industriali, perché utilizzava le informazioni come strumento di controllo, quindi non era semplicemente il surrogato di un potere umano, ma una macchina che orientata a uno scopo e dotata dei mezzi per ottenerlo. Erano le origini della cibernetica.

Un'altra grande questione che ci si pone davanti è la capacità di persuasione della macchina. In tanti studi che leggo, trovo che, purtroppo, è diffusa una retorica simile a quella che abbiamo visto con il tabacco e con le armi: non si distingue tra persuasione e manipolazione, cosa che da Aristotele in poi abbiamo imparato a distinguere. Mi spiego. Io posso non farti cambiare idea, ma nell'immediato ti faccio agire in una certa maniera. È quello che ci dice la *Nudge Theory* (o spinta gentile), una delle teorie cardine delle scienze comportamentali, oggi così potente proprio perché viviamo in un contesto culturale che ha difficoltà con il rispetto delle norme.

Per dirla con un'immagine: nella vita di tutti i giorni noi passiamo dal semaforo (che in concreto impone un comportamento; un indicativo normativo esterno, direbbe Kant) - alla rotonda (che si ba-

sa sul principio "ti regoli come vuoi"). Ma come regolare una società se l'indicazione normativa non funziona?

Ad Amsterdam, all'aeroporto di Schiphol, hanno un problema: pulire i bagni degli uomini costa quasi il doppio rispetto a pulire quelli delle donne. Cosa fare? Un dispositivo normativo esterno avrebbe funzionato benissimo: «Si prega di farla dentro». Avrebbero potuto scriverlo grosso quanto volevano, ma la situazione non sarebbe cambiata. È bastato disegnare una mosca all'interno del water per far sì che il costo della pulizia dei bagni maschili arrivasse a essere quasi uguale a quello delle donne. Si è passati di fatto da una norma a un dispositivo informativo che spinge sull'uomo e cambia il suo comportamento.

Ed è qui che troviamo la tensione più forte possibile, secondo me, sullo stato di diritto. Ed è qui che si gioca il problema della legittimità nell'età moderna: cos'è che fa di una legge un qualcosa che è lecito?

John Rawls, nella sua *Teoria della Giustizia*, ci suggerisce alcuni criteri formali: una legge, per essere conforme al-



lo stato di diritto, deve rispettare tre condizioni: essere conoscibile, universale e generale. Ora, come si fa a rendere legittime le regole di funzionamento delle macchine intelligenti? Proviamo a vedere se rispettano quelle condizioni. Queste regole sono conoscibili? Si potrebbe rendere il codice open, cosicché tutti lo leggano. Ma già Ken Thompson, vincitore del premio Turing, in un suo paper del 1982 *Trusting the Trust*, ha dimostrato che non possiamo essere certi che l'algoritmo faccia solo quello che detta il codice; inoltre chiunque potrebbe alterare il codice. Seconda condizione: sono universali? Beh, l'algoritmo profila. Terza condizione: sono generali? Le macchine intelligenti obbediscono solo al padrone dei server. Alla partenza di un 2024 che vedrà la Francia, gli Stati Uniti e l'Europa votare, ci si deve chiedere come il lato umano della relazione tra uomo e macchina può metabolizzare, gestire e armonizzare questo strumen-

to all'interno dei processi che scrivono il contratto sociale.

Possiamo aver fiducia in queste macchine intelligenti? Quando prendo un aereo, ho fiducia nel fatto che qualcuno ha controllato le gomme, che i motori sono stati revisionati, che il pilota ha fatto un training specifico. Tutto questo possiamo chiamarlo contratto sociale.

Una macchina di Ai che lavora senza sosta con quelle modalità di cui abbiamo parlato rischia di erodere la fiducia che sta alla base del contratto sociale: tutti gli studi sulla polarizzazioni lo dimostrano. Perché una piattaforma non è un soggetto neutro: è un soggetto che monetizza rispetto a questa polarizzazione. Ovvero estrae valore rispetto a un qualcosa che è un valore comune e condiviso, che è la collettività sociale. Non è bene cominciare a definire e misurare la manipolazione digitale, che è cosa diversa dalla persuasione?

Perché, se è vero, come dimostrano i dati, che le persone

esposte ai sistemi di intelligenza artificiale non cambiano idea nel lungo termine, è altrettanto vero che questi sistemi inducono a comportamenti nell'immediato (per esempio all'acquisto). Ed è questa differenza tra manipolazione e persuasione che va rimessa al centro del dibattito pubblico.

In conclusione, penso che mantenere la centralità del lato umano della relazione uomo-macchina non sia solo una questione educativa o formativa: è una questione di tolleranza del sistema democratico. Occorre reintrodurre probabilmente una categoria che già esiste nel diritto: la differenza tra pericolo e rischio. Maneggiamo ogni giorno sostanze pericolose (si pensi alla benzina), ma lo facciamo usando una serie di dispositivi di protezione.

Dunque non è un problema il potere della macchina, ma come lo inoculiamo in un contesto sociale e come gestiamo e monitoriamo il contesto sociale. —

© RIPRODUZIONE RISERVATA

Barbara Gallavotti



“

La morale evolve con l'essere umano, ma non spetta alla scienza rinnovarla. Dal metodo scientifico si può mutuare il principio di autorità

Giorgio Vallortigara



“

I valori morali sono certamente un prodotto culturale ma dentro di noi c'è un'intelaiatura biologica che li riconosce e ce li fa assimilare

LA LACTIO

Il consigliere di Papa e governo per l'AI

Paolo Benanti, consigliere di Papa Francesco sull'AI, professore alla Pontificia Università Gregoriana e Presidente della Commissione AI per l'informazione Algoritica (al posto di Giuliano Amato, dimessosi nei giorni scorsi), ha tenuto una lectio a Camogli, al Forum dell'Intelligenza Artificiale, organizzato da Frame-Festival della Comunicazione, tenutosi il mese scorso. Il testo qui pubblicato è un estratto del suo intervento. —