Educare le macchine in una società che disimpara la relazione

Michele Scaffidi, | 17 settembre 2026

Tra le professioni cresciute con la diffusione dell'intelligenza artificiale ce ne sono alcune che contengono un piccolo paradosso: persone chiamate ad addestrare i sistemi di IA, valutarne le risposte, correggerne gli errori e indicare quali comportamenti siano preferibili.

AI trainer, annotatori, valutatori umani, esperti di human feedback: le definizioni cambiano e sarebbe improprio equiparare il loro lavoro all'educazione di una persona. Una macchina non cresce né costruisce un'identità. Eppure, quando un essere umano segnala che una risposta è preferibile a un'altra, contribuisce a orientarne il comportamento.

È questa immagine – un essere umano impegnato a insegnare a una macchina come rapportarsi agli esseri umani – che merita attenzione. Accade mentre le relazioni sembrano diventare più fragili, selettive e costruite intorno alla contrapposizione. Mentre proviamo a insegnare alle macchine a riconoscere l'altro, una parte delle nostre società sembra sempre più impegnata a stabilire da chi difendersi.

Le macchine non imparano dall'umanità in astratto

L'intelligenza artificiale viene spesso raccontata attraverso ciò che sa fare: scrive, sintetizza, traduce, organizza informazioni, risponde a domande. La sensazione è quella di trovarsi di fronte a un sistema che "sa". Ma quel sapere non nasce dal nulla.

Nei principali modelli linguistici oggi disponibili, all'addestramento su grandi quantità di contenuti umani si aggiungono procedure nelle quali intervengono valutazioni, regole e feedback. Non ogni conoscenza viene "insegnata" da un formatore, ma è altrettanto fuorviante pensare a un'intelligenza che costruisca autonomamente il proprio universo morale e culturale.

Dietro una risposta apparentemente neutra ci sono testi, classificazioni, criteri e scelte. La macchina non impara dall'umanità in astratto, ma da ciò che l'umanità ha prodotto e da ciò che alcuni esseri umani, dentro determinate istituzioni e culture, hanno deciso di considerare preferibile.

Quale risposta è rispettosa? Quando una generalizzazione diventa uno stereotipo? Che cosa significa essere imparziali? Quando è necessario dichiarare un'incertezza? Non sono domande puramente tecniche, ma sociali, culturali ed etiche.

L'"allineamento" dei sistemi di intelligenza artificiale – il tentativo di orientarne il comportamento verso valori compatibili con quelli umani – contiene una questione che precede la tecnologia: compatibili con quali valori umani? E definiti da chi?

Alcune aziende hanno reso espliciti questi criteri, arrivando a definire vere e proprie "costituzioni" destinate a orientare il comportamento dei modelli. Scegliere i valori di una macchina attribuisce un potere rilevante a chi compie quella scelta.

La società che sta dietro lo schermo

È qui che il paradosso diventa più evidente. Chiediamo alle macchine di evitare generalizzazioni e discriminazioni mentre la comunicazione pubblica procede spesso attraverso categorie nette, appartenenze e confini; di distinguere un fatto da un'affermazione infondata mentre opinioni, identità e conoscenze vengono facilmente poste sullo stesso piano; di essere prudenti proprio mentre la comunicazione politica e digitale premia la risposta immediata e la semplificazione.

Non significa descrivere una società priva di solidarietà. Ma alcuni segnali sono difficili da liquidare come semplici percezioni.

Nel 59° Rapporto Censis, pubblicato nel 2025, il 63% degli italiani intervistati sostiene che i flussi migratori debbano essere limitati, il 59% associa una forte presenza di immigrati al degrado del quartiere e il 54% li considera un pericolo per l'identità e la cultura nazionali. La Commissione europea contro il razzismo e l'intolleranza del Consiglio d'Europa ha inoltre descritto il discorso pubblico italiano come sempre più xenofobo e divisivo soprattutto verso migranti e altre minoranze.

Questi dati non consentono di riassumere un'intera società nella formula "gli italiani sono chiusi". Dicono però che la paura della diversità, la rappresentazione dell'altro come minaccia e il bisogno di tracciare confini identitari non sono fenomeni marginali.

Ed è dentro questa stessa società che stiamo costruendo sistemi ai quali chiediamo di essere inclusivi, non discriminatori e capaci di comprendere punti di vista differenti. Stiamo tentando di incorporare nelle macchine qualità relazionali che non possiamo dare per acquisite tra gli esseri umani che le costruiscono.

Quale "giusto" insegniamo?

Si potrebbe obiettare che proprio per questo servono regole e sistemi di controllo. Un'intelligenza artificiale non deve riprodurre automaticamente tutto ciò che trova nella società, ma essere progettata per correggere stereotipi, discriminazioni e comportamenti dannosi.

Ma il problema si sposta soltanto di un livello. Per correggere un pregiudizio dobbiamo prima riconoscerlo; per stabilire che una risposta è discriminatoria dobbiamo condividere una definizione, almeno minima, di discriminazione; per chiedere a una macchina di rispettare la pluralità dobbiamo considerare la pluralità un valore. La tecnologia non elimina questo lavoro: lo rende più visibile.

Un sistema può essere sottoposto a migliaia di valutazioni per evitare espressioni offensive, mentre nello spazio pubblico si normalizzano linguaggi che trasformano gruppi interi in problemi o minacce. Alla macchina chiediamo di dichiarare quando non sa e di distinguere tra spiegare una posizione e aderirvi; tra esseri umani, invece, l'incertezza viene spesso letta come debolezza e il confronto con una posizione diversa come una sua legittimazione.

In un certo senso stiamo chiedendo alle macchine una coerenza che noi stessi fatichiamo a praticare. Per insegnare alla macchina come vorremmo che si comportasse siamo costretti a dichiarare quale idea di relazione consideriamo desiderabile. E quella dichiarazione finisce inevitabilmente per parlare di noi.

La macchina torna nella relazione

La questione sarebbe già rilevante se l'intelligenza artificiale rimanesse esterna alla nostra vita sociale. Sta accadendo il contrario. La utilizziamo per studiare, informarci, scrivere, interpretare documenti, elaborare dati e orientarci in una scelta; progressivamente entrerà sempre più nel lavoro, nell'educazione, nei servizi, nella sanità e nell'amministrazione.

La questione assume una consistenza particolare quando questi sistemi entrano nei servizi alla persona e nel lavoro sociale. Qui le categorie contribuiscono al modo in cui un bisogno e, attraverso di esso, una persona vengono letti. Il lavoro sociale conosce da tempo il rischio del labeling e di servizi organizzati in silos attorno a categorie sociali e demografiche: "migrante", "persona con disabilità", "povero", "detenuto" possono essere definizioni utili, ma non esauriscono mai la storia di una persona.

L'intelligenza artificiale può aiutare a organizzare conoscenze e mettere in relazione informazioni. Proprio per questo diventa decisivo interrogarsi sui criteri sociali che incorpora: il riconoscimento dei diritti fondamentali, il rispetto dell'autodeterminazione, la capacità di non ridurre la persona alla categoria attraverso cui viene letta. Se apprende da una società attraversata dagli stessi stereotipi e dalle stesse gerarchie che il lavoro sociale cerca di non riprodurre, la questione riguarda anche la qualità sociale dello sguardo che contribuiamo a incorporare nella macchina.

La relazione conserva qui una funzione difficilmente riducibile a un insieme di dati. Nel lavoro con le persone non è soltanto il luogo nel quale applicare saperi già disponibili: è spesso quello in cui emergono elementi che prima dell'incontro non erano conoscibili. La macchina può aiutarci a organizzare ciò che sappiamo; più difficile è delegarle ciò che prende forma soltanto nella relazione.

Il processo non procede quindi in una sola direzione. Gli esseri umani producono conoscenze, linguaggi e valori; le macchine apprendono attraverso quel patrimonio e le valutazioni che ne orientano il funzionamento; successivamente rientrano nelle nostre vite e partecipano al modo in cui conosciamo, scegliamo, comunichiamo e interpretiamo gli altri.

L'uomo forma la macchina. La macchina torna negli ambienti attraverso cui l'uomo forma se stesso.

Ciò che incorporiamo oggi nei sistemi di intelligenza artificiale potrebbe contribuire domani a rendere normale un determinato modo di leggere la realtà. Non perché la macchina possieda un progetto morale, ma perché ogni mediazione ripetuta diventa parte dell'ambiente sociale. Se sempre più persone vi si rivolgeranno per spiegazioni e consigli, il modo in cui quei sistemi rappresentano differenza, conflitto o fragilità non sarà irrilevante.

La qualità della relazione tra esseri umani diventa così una delle premesse dello sviluppo tecnologico.

Non possiamo delegare ciò che non sappiamo più fare

La promessa più seducente dell'intelligenza artificiale riguarda la delega: farle svolgere attività ripetitive, elaborare quantità di informazioni che non potremmo gestire, assisterci nelle decisioni, colmare alcuni limiti delle nostre capacità. In molti ambiti può produrre benefici enormi.

La contraddizione emerge quando cominciamo ad aspettarci che la tecnologia compensi non soltanto i nostri limiti operativi, ma anche quelli relazionali e sociali. Possiamo progettare sistemi che non amplifichino discriminazioni e conflitti; non possiamo però delegare loro il lavoro necessario a costruire una società meno discriminatoria, il desiderio di incontrare chi non ci assomiglia o la capacità di attraversare un conflitto senza trasformare l'altro in un nemico.

La relazione non è quindi un residuo destinato agli esseri umani dopo che le macchine avranno imparato a fare il resto. È il contesto nel quale continuiamo a decidere che cosa consideriamo conoscenza, quale significato attribuiamo alla differenza e quali problemi riteniamo possibile affrontare insieme.

Forse dovremmo allora invertire almeno per un momento la domanda che domina il dibattito pubblico. Ci chiediamo continuamente che cosa l'intelligenza artificiale sarà capace di fare al posto nostro. Potremmo domandarci anche che cosa dobbiamo continuare a essere capaci di fare noi perché l'intelligenza artificiale possa davvero esserci utile.

Tra queste capacità c'è quella, elementare e complessa, di riconoscere l'altro: accettare che non tutto possa essere semplificato, che una società possa contenere differenze senza trasformarle in minacce e che comprendere una posizione diversa non significhi aderirvi. Significa anche riconoscere l'incertezza come parte di ogni conoscenza che voglia restare aperta.

Forse l'intelligenza artificiale ci pone allora una questione educativa diversa da quella che immaginiamo. Non soltanto: come dobbiamo educare le macchine? Ma: quali esseri umani e quali società sono necessari per poterle educare?

Perché le macchine destinate ad assisterci, consigliarci e produrre conoscenza insieme a noi non arriveranno da un luogo esterno alla società. Impareranno da ciò che abbiamo prodotto e dai criteri attraverso cui avremo scelto di orientarle.

Prima di affidarci all'intelligenza artificiale per risolvere problemi sempre più complessi, dovremmo preoccuparci della società dalla quale le stiamo chiedendo di imparare. Tornare a investire seriamente nelle relazioni tra esseri umani non è soltanto una necessità sociale: sta diventando anche una condizione per costruire le macchine alle quali vogliamo affidare una parte crescente della nostra vita.