

Il bersaglio mobile del «veramente»

Percorso in sei movimenti

SOMMARIO

Sei movimenti per leggere il percorso nella sua interezza.

1. Chi ha fatto la barca?
2. Chi sta addestrando chi?
3. L'officina, il testo e la firma
4. L'ecosistema delle decisioni
5. Il soggetto disabitato
6. Chi può decidere della storia?

IL PERCORSO

Il bersaglio mobile del «veramente»

Autorialità, trasformazione, decisione e responsabilità nell'era dell'intelligenza artificiale.

Che cosa intendiamo quando diciamo che una macchina non crea veramente, non comprende veramente, non decide veramente?

A ogni nuova capacità dell'intelligenza artificiale sembra comparire una soglia ulteriore. Il calcolo non basta, il linguaggio non basta, la produzione di qualcosa di nuovo non basta. Ogni volta il criterio si sposta e la domanda, apparentemente rivolta alla macchina, finisce per tornare verso di noi.

Questo percorso nasce da una domanda sull'autorialità e attraversa progressivamente trasformazione, responsabilità e potere. Cinque movimenti, costruiti uno sull'altro, nei quali l'intelligenza artificiale rimane continuamente al centro della scena soltanto per mostrarci, ogni volta, l'essere umano che le sta davanti.

Il percorso originario si conclude con Il soggetto disabitato. Una riflessione successiva, Chi può decidere della storia?, riprende quella traiettoria e la porta oltre: dall'autorialità all'autorità.

Pagina del percorso: <https://inchiostrati.it/pensare/il-bersaglio-mobile-del-veramente/>

MOVIMENTO 01

Chi ha fatto la barca?

La presenza di uno strumento artificiale rende meno autentica un'opera? Il primo movimento distingue l'esecuzione dall'intenzione e domanda dove comincino davvero autorialità, merito e responsabilità.

Avevo circa dieci anni quando mio padre tornò a casa con un foglio di tabulato. Era la fine degli anni Settanta, lavorava per una grande azienda del Nord Italia e quel foglio, per un bambino abituato alle matite, ai pennarelli e alle immagini stampate nei libri, aveva qualcosa di misterioso. Sopra c'era un cavallo, ma non era stato disegnato nel modo in cui conoscevo il disegno: la sua figura era composta interamente da caratteri, soprattutto X e O, disposti con una precisione tale che, osservandoli da una certa distanza, smettevano di essere lettere e diventavano il profilo dell'animale. Mio padre mi spiegò che quel cavallo era stato prodotto da una grande macchina e la parola computer, allora, non evocava ancora qualcosa che potesse stare sopra una scrivania, dentro una tasca o perfino al polso. Indicava davvero una macchina, grande abbastanza da avere una stanza, circondata da persone che conoscevano linguaggi e procedure incomprensibili ai più e che, proprio per questo, esercitavano su di me un fascino enorme.

Naturalmente volli vederla. Qualche giorno più tardi mio padre mi portò nel luogo in cui si trovava quell'elaboratore e ricordo ancora soprattutto il senso di sproporzione fra me e quelle apparecchiature. Davanti a una tastiera enorme era seduto un uomo di bassa statura; non sapevo esattamente quale fosse il suo lavoro e non potevo immaginare la quantità di conoscenze necessarie per ottenere da quella macchina qualcosa che io fossi in grado di riconoscere. L'uomo mi guardò, intuì probabilmente la curiosità con cui osservavo tutto e mi chiese se mi sarebbe piaciuto vedere una bella barca nel mare. Dissi di sì. Prese allora un certo numero di schede perforate e cominciò a inserirle, una dopo l'altra, dentro una fessura della macchina. Io seguivo il gesto senza comprenderne il significato: quelle schede erano soltanto cartoncini pieni di buchi e appartenevano allo stesso territorio della magia nel quale, da bambini, collochiamo tutto ciò che funziona senza che riusciamo ancora a comprenderne le cause.

Dopo qualche tempo la macchina restituì un altro foglio di tabulato. C'era una barca nel mare, anch'essa costruita interamente attraverso X e O.

Avevo dieci anni, non conoscevo nulla di programmazione e ignoravo completamente quanto lavoro potesse precedere quella stampa. Non sapevo che qualcuno avesse dovuto concepire il risultato, trasformarlo in istruzioni, predisporre quelle schede, controllare che la sequenza fosse corretta e probabilmente affrontare prove ed errori che per me rimanevano completamente invisibili. Vedevo soltanto il gesto finale: le schede entravano nella macchina e, poco dopo, ne usciva una barca.

Eppure una cosa mi sembrò immediatamente evidente: la barca l'aveva fatta quell'uomo. La macchina l'aveva resa visibile.

Quasi mezzo secolo più tardi ci troviamo davanti a sistemi incomparabilmente più potenti. Le schede perforate sono scomparse, non dobbiamo stabilire anticipatamente la posizione di ogni carattere e, soprattutto, non conosciamo più con precisione ciò che apparirà alla fine del processo. Possiamo scrivere poche righe e ottenere un'immagine che nessuna macchina fotografica ha mai registrato, una musica che nessun musicista ha eseguito o un testo di migliaia di parole che nessun essere umano ha composto parola dopo parola. La distanza fra l'istruzione iniziale e la complessità del risultato è diventata così grande da produrre una delle discussioni più singolari del nostro tempo: quando un contenuto viene generato attraverso un'intelligenza artificiale, chi ne è l'autore?

La domanda è legittima. Molto meno scontato è che stiamo cercando la risposta nel posto giusto, perché osserviamo soprattutto ciò che viene fuori dalla macchina, mentre forse dovremmo cominciare da ciò che esisteva prima che la macchina venisse messa in funzione.

L'opera comincia prima dell'opera

Siamo abituati a identificare l'autore con il gesto materiale della creazione. Lo scrittore scrive, il pittore dipinge, il compositore dispone le note, il fotografo preme il pulsante dell'otturatore; la stessa parola «autore» sembra suggerire qualcuno che fa materialmente qualcosa e da questa rappresentazione deriva facilmente l'idea che, se un sistema artificiale produce materialmente una parte consistente delle parole, delle immagini o dei suoni, debba ricevere almeno una quota della relativa autorialità.

Il problema è che un'opera non comincia nel momento in cui viene prodotta la sua prima manifestazione materiale. Comincia prima, quando ancora non esiste niente da leggere, ascoltare o vedere e qualcuno decide che qualcosa potrebbe, o dovrebbe, esistere.

Può trattarsi di una storia che da tempo chiede una forma, di un'esperienza che si vuole raccontare, di una domanda alla quale non si conosce ancora la risposta, di un'immagine intravista mentalmente oppure della semplice curiosità di scoprire che cosa accadrebbe seguendo una determinata intuizione. Non è necessario avere già una trama, una tesi o un progetto completo; anzi, buona parte della creazione nasce proprio dalla necessità di esplorare qualcosa che all'inizio è ancora incompleto. Prima dell'esecuzione, però, esiste già un gesto essenziale: qualcuno apre volontariamente quello spazio.

Possiamo chiamarlo IDEA, purché intendiamo il termine nel suo significato più ampio. L'idea non è soltanto il soggetto di un romanzo o la frase che potrebbe riassumere un articolo. È l'origine intenzionale del processo, il momento nel quale un essere umano decide che una possibilità merita di essere perseguita e che un contenuto, ancora inesistente, dovrebbe essere portato verso una forma.

In questo senso, l'intelligenza artificiale contemporanea presenta una caratteristica apparentemente banale ma decisiva: non ha bisogno di essere convinta a lavorare, ma ha bisogno che qualcuno apra per lei un compito. Può sviluppare una proposta, costruire alternative e compiere autonomamente una quantità enorme di operazioni intermedie, tuttavia non si sveglia una mattina pensando che esista un argomento del quale sente la necessità di parlare. Non osserva per settimane una contraddizione della società finché decide che dovrebbe scriverne, non conserva un ricordo d'infanzia perché teme che possa andare perduto e non avverte che una storia sia rimasta incompiuta e che occorra finalmente darle una forma.

Questo non significa che un sistema non possa proporre idee. Può farlo benissimo, ed è uno degli aspetti più interessanti dell'interazione con i modelli generativi. Possiamo chiedere dieci soggetti per un racconto, cento possibili titoli o una prospettiva alla quale non avevamo pensato. Ma anche nel caso più estremo, quello in cui l'essere umano dice semplicemente «decidi tu che cosa scrivere», l'intenzione non è scomparsa. È stata delegata una parte del processo, non la decisione originaria che un processo debba iniziare.

Dire «scegli tu che cosa scrivere» assomiglia infatti a dire a un compagno di viaggio «scegli tu dove andiamo». La destinazione è stata delegata, ma non la decisione di partire. Prima ancora che venga indicata una città, esiste già qualcuno che vuole compiere un viaggio.

Questo vale persino per sistemi agentici capaci di operare secondo programmi ricorrenti. Potremmo configurare un'intelligenza artificiale affinché ogni mattina analizzi le notizie, scelga da sola un argomento, consulti fonti, stabilisca una tesi, scriva un articolo e lo depositi in una cartella senza che nessuno intervenga. Dal punto di vista operativo il sistema apparirebbe quasi completamente autonomo e potrebbe perfino individuare un tema al quale il suo utilizzatore non aveva mai pensato; tuttavia, se ci domandassimo perché quella macchina, proprio quella mattina, stia scrivendo un articolo, dovremmo ancora rispondere che qualcuno, prima, ha deciso che ogni mattina dovesse farlo.

L'autonomia dell'esecuzione può diventare vastissima; l'origine intenzionale del compito rimane altrove. Non è una distinzione marginale, perché separa la capacità di scegliere come raggiungere uno scopo dalla capacità di decidere che uno scopo meriti di esistere.

Scegliere non significa necessariamente volere

Una parte della difficoltà nasce dal linguaggio che utilizziamo. Diciamo comunemente che un modello «sceglie» una parola, «decide» una strategia, «preferisce» una formulazione o «ritiene» una risposta migliore di un'altra. Sono metafore utili per descriverne il funzionamento e, nella conversazione quotidiana, sarebbe insopportabile sostituirlle ogni volta con lunghe descrizioni tecniche. Il problema nasce quando trasformiamo quella comodità linguistica in una conclusione sulla natura del sistema.

Un'intelligenza artificiale può selezionare una soluzione fra molte alternative, ma non ne segue che desideri quella soluzione. Può produrre una tesi e, alla richiesta successiva, confutarla con la stessa efficacia; può scrivere una dichiarazione d'amore senza amare nessuno, difendere un personaggio e un minuto più tardi argomentare che dovrebbe essere eliminato dalla storia, senza vivere questo passaggio come una contraddizione o una perdita. La sua capacità generativa riguarda le forme attraverso le quali l'intenzione può essere rappresentata; non prova, da sola, l'esistenza dell'intenzione rappresentata.

Per questo la frase «l'IA ha avuto l'idea» richiede almeno una precisazione. Può certamente essere stata la prima a proporre, all'interno del dialogo, una combinazione concettuale che l'utilizzatore non aveva anticipato. Ma la proposta acquista valore autoriale soltanto quando qualcuno la riconosce, la sceglie, decide che vale la pena svilupparla e la incorpora nel proprio progetto. Una macchina può presentare cento idee e non avere alcun rapporto particolare con nessuna di esse; l'essere umano può fermarsi sulla settantatreesima e pensare che proprio quella meriti di essere esplorata.

È in quel momento che una possibilità generata diventa intenzione adottata.

La macchina che faceva troppo: fotografia e autorialità

Il timore che uno strumento tecnologico possa svolgere una parte troppo grande del lavoro per lasciare intatta l'autorialità umana non nasce con ChatGPT. Nel XIX secolo la fotografia provocò una controversia sorprendentemente simile, perché rispetto alla pittura sembrava sostituire il gesto dell'artista con un procedimento meccanico: una macchina registrava la luce proveniente da un oggetto reale e produceva un'immagine senza che il fotografo dovesse disegnarne personalmente le forme.

Nel 1884 la Corte Suprema degli Stati Uniti affrontò direttamente la questione nel caso *Burrow-Giles Lithographic Co. v. Sarony*, relativo a un celebre ritratto di Oscar Wilde realizzato da Napoleon Sarony. La fotografia venne riconosciuta come opera originale non perché Sarony avesse materialmente tracciato ogni elemento dell'immagine, cosa evidentemente impossibile, ma perché l'immagine derivava dalla sua concezione: aveva scelto la posa, organizzato costume e accessori, disposto luce e

ombra e determinato l'espressione desiderata. La decisione richiama inoltre una concezione dell'autore come causa effettiva dell'immagine, colui che origina e dà forma alla concezione che lo strumento rende visibile.

La fotografia costituisce un precedente importante non perché una macchina fotografica e un modello generativo siano equivalenti. Non lo sono, e utilizzare questa analogia come se eliminasse ogni differenza sarebbe semplicistico. Un sistema generativo può produrre strutture e soluzioni che non sono state determinate anticipatamente dall'utilizzatore con il grado di precisione richiesto da una fotografia tradizionale. Il punto interessante è un altro: già più di un secolo fa fu necessario riconoscere che una macchina poteva assumere una parte enorme dell'esecuzione materiale senza trasformarsi, per questo, nel soggetto dell'opera.

La tecnologia aveva aumentato la distanza fra gesto umano e risultato, ma il diritto cercò l'autorialità nelle scelte che rendevano significativo quel risultato. La stessa concezione riappare nella giurisprudenza europea: nel caso Painer, la Corte di giustizia dell'Unione europea ha ricordato che anche nella fotografia l'autore può esprimere la propria personalità attraverso scelte libere e creative relative, fra l'altro, a sfondo, posa, espressione e illuminazione.

La stessa evoluzione è avvenuta successivamente con il cinema, con la registrazione sonora, con la grafica digitale e con molte altre tecnologie nelle quali l'artista non controlla materialmente ogni singolo elemento dell'output. L'autorialità ha continuato a sopravvivere perché non l'abbiamo mai identificata esclusivamente con il lavoro manuale necessario a produrre l'oggetto finale. Se lo avessimo fatto, il fotografo sarebbe stato poco più che l'addetto alla macchina, il regista avrebbe dovuto tenere personalmente la cinepresa e il compositore sarebbe autore soltanto quando esegue da sé ogni strumento dell'orchestra.

Il ghostwriter e il problema del secondo soggetto

La letteratura offre un caso ancora più interessante, perché mette in crisi l'idea secondo cui autore e produttore materiale delle parole debbano necessariamente coincidere. Il ghostwriting esiste da molto prima dell'intelligenza artificiale e, nel corso della storia, ha permesso che opere associate pubblicamente a un autore venissero realizzate con un contributo testuale assai significativo di altre persone.

La collaborazione fra Alexandre Dumas e Auguste Maquet è particolarmente istruttiva. La Bibliothèque nationale de France ricorda che, a partire dal 1838, Maquet offrì a Dumas i propri servizi e trame per opere teatrali e romanzi, inaugurando una collaborazione assai produttiva nella quale Maquet rimase spesso «l'uomo nell'ombra» mentre Dumas firmava con il solo proprio nome; la questione divenne pubblicamente controversa e i due finirono anche per affrontarsi giudiziariamente.

Questo caso non dovrebbe essere utilizzato per sostenere semplicemente che Dumas fosse autore e Maquet soltanto uno strumento umano. Sarebbe proprio il contrario del ragionamento che stiamo cercando di costruire. Maquet era un essere umano e

possedeva quindi ciò che manca alla macchina: una propria intenzionalità. Poteva proporre una soluzione perché la riteneva migliore, riconoscersi nel proprio lavoro, desiderare che il proprio contributo fosse riconosciuto, dissentire da Dumas e infine rivendicare una parte della creazione. Il problema dell'autorialità, in quel caso, era reale proprio perché esistevano due soggetti.

Questo permette di distinguere in modo più rigoroso il ghostwriter dall'intelligenza artificiale. Un ghostwriter può ricevere istruzioni estremamente precise e continuare tuttavia a operare come soggetto. L'autore principale può stabilire il tema, i personaggi, la struttura, il tono e perfino il contenuto delle singole scene, ma nel momento in cui un altro essere umano deve trasformare quelle istruzioni in dieci pagine di prosa compirà inevitabilmente scelte linguistiche e narrative alle quali può attribuire un significato proprio. Può perfino dire: «Questa frase è mia».

Un modello generativo può produrre la stessa frase, ma non può rivendicarla nello stesso senso. Se l'utilizzatore la elimina, il sistema non perde nulla; se gli viene chiesto di produrne il contrario, non avverte di avere rinnegato una propria posizione; se un altro utente ottiene una formulazione simile, non nasce in esso alcuna rivendicazione.

Nel rapporto autore-ghostwriter ci sono almeno due soggetti intenzionali e proprio per questo può esistere un conflitto autoriale. Nel rapporto autore-IA c'è un soggetto intenzionale e un sistema generativo.

Ma la differenza non riguarda soltanto l'intenzione. Maquet poteva subire un torto.

Poteva ritenere che il proprio lavoro fosse stato utilizzato senza un riconoscimento adeguato, poteva perdere prestigio o vantaggi economici, poteva opporsi alla ricostruzione pubblica della collaborazione e, soprattutto, poteva rivolgersi a un tribunale perché riteneva leso un interesse che considerava proprio. La sua autorialità non era quindi soltanto una relazione con l'origine dell'opera, ma anche una condizione di vulnerabilità rispetto ad essa: ciò che accadeva all'opera poteva accadere, in una certa misura, anche a lui. La storia documentata delle controversie fra Dumas e Maquet mostra precisamente questa dimensione.

Una IA attuale non occupa una posizione equivalente. Possiamo cancellare il suo contributo, modificarlo fino a renderlo irriconoscibile, attribuirlo a qualcun altro o decidere di non utilizzarlo affatto senza ledere un interesse morale del sistema, perché non esiste una reputazione personale della macchina da proteggere, né una relazione soggettiva con l'opera che possa essere tradita. Possiamo parlare, naturalmente, dei diritti e degli interessi delle aziende che hanno sviluppato il sistema o degli autori dei materiali impiegati per addestrarlo, ma sono soggetti diversi dalla macchina stessa.

Questa vulnerabilità aggiunge quindi un elemento ulteriore alla nozione di autore. Essere autore significa poter rivendicare un'opera, ma significa anche esporsi attraverso di essa: al mancato riconoscimento, al giudizio, alla critica, alla perdita economica e, come vedremo, alla responsabilità. L'autore può chiedere che gli venga attribuito il merito, proprio perché è anche qualcuno al quale può essere attribuita una

conseguenza.

Da questo punto di vista il ghostwriter rende ancora più evidente la distanza dall'intelligenza artificiale: un secondo autore umano può essere riconosciuto, ignorato, danneggiato o tradito; la macchina può essere utilizzata, ma non può subire quella relazione come una lesione di sé.

L'autorialità è già stata separata dall'esecuzione

Anche il diritto positivo conosce da molto tempo situazioni nelle quali la persona che realizza materialmente un'opera e quella alla quale viene attribuita giuridicamente l'autorialità non coincidono perfettamente. Negli Stati Uniti, per esempio, la disciplina del work made for hire prevede in determinate circostanze che il datore di lavoro o il committente venga considerato autore ai fini del copyright. La regola è specifica, dipende da condizioni precise e non può essere utilizzata per sostenere una teoria generale dell'autorialità, ma è sufficiente a dimostrare che neppure il diritto ha sempre fatto coincidere meccanicamente l'autore con l'esecutore materiale.

Il cinema rende questa separazione ancora più evidente. Centinaia di persone possono concorrere alla realizzazione di un film e numerosi contributi sono creativi nel senso più pieno del termine; ciononostante, ordinamenti, contratti e tradizioni culturali distribuiscono autorialità e titolarità secondo criteri diversi, perché nessuno potrebbe seriamente misurare il contributo dell'autore contando quanti secondi di pellicola o quanti pixel siano stati materialmente prodotti da ciascuna persona.

La storia dell'autorialità è quindi già, almeno in parte, la storia della separazione fra idea, intenzione, esecuzione, collaborazione, controllo, attribuzione e responsabilità. L'intelligenza artificiale non ha creato questa separazione; l'ha soltanto resa enormemente più visibile, perché oggi l'esecuzione può essere delegata a uno strumento in misura impensabile fino a pochi anni fa.

Una legge italiana che parte dall'essere umano

Il diritto italiano ha compiuto nel 2025 una scelta particolarmente interessante. La legge 23 settembre 2025, n. 132 ha modificato l'articolo 1 della legge sul diritto d'autore aggiungendo espressamente il riferimento alle opere dell'ingegno «umano» e precisando nello stesso tempo che esse possono essere create con l'ausilio di strumenti di intelligenza artificiale, purché costituiscano il risultato del lavoro intellettuale dell'autore.

È difficile non notare ciò che il legislatore ha scelto di non fare. Non ha creato una categoria di opere «dell'intelligenza artificiale» e non ha stabilito che l'uso di un sistema generativo elimini automaticamente l'autorialità umana. Ha invece mantenuto l'autore umano e collocato l'IA fra gli strumenti dei quali quel soggetto può avvalersi.

Naturalmente questa formulazione non risolve ogni problema. Rimane da stabilire quanto debba essere significativo il lavoro intellettuale dell'essere umano, quale grado

di controllo debba esercitare sul risultato e in quali circostanze una semplice richiesta possa diventare insufficiente per produrre un'opera tutelabile. Sono questioni giuridiche reali, e sarebbe scorretto fingere che la legge italiana abbia già trasformato in norma tutta la teoria che stiamo sviluppando qui.

Ma una distinzione è già molto chiara: presenza dell'IA e assenza dell'autore non sono sinonimi.

Copyright e autorialità non sono la stessa domanda

Questa distinzione diventa ancora più importante osservando gli Stati Uniti. Nel gennaio 2025 il Copyright Office ha ribadito che l'impiego dell'intelligenza artificiale come strumento non esclude la tutela, ma che gli output possono essere protetti soltanto quando sia individuabile un sufficiente contributo espressivo umano. Modifiche creative, selezione e disposizione originale di materiale generato possono quindi essere tutelate, mentre il semplice prompting, nelle condizioni tecnologiche considerate dall'Office, non viene ritenuto normalmente sufficiente.

Qui bisogna evitare un errore molto frequente: confondere «questa parte non è protetta dal copyright» con «questa parte ha come autore l'intelligenza artificiale».

Le due affermazioni non sono equivalenti.

Il copyright costituisce una specifica forma di tutela giuridica e può stabilire una soglia di contributo umano necessaria affinché qualcuno acquisisca diritti esclusivi su un determinato output. Se quella soglia non viene raggiunta, la conseguenza può essere semplicemente l'assenza di protezione. Non è necessario inventare un autore artificiale per spiegare il fenomeno.

Può esistere un risultato senza copyright e senza autore artificiale.

Questo punto è fondamentale perché permette di separare due discussioni che vengono continuamente sovrapposte: una riguarda quali diritti esclusivi un ordinamento debba concedere a chi utilizza strumenti generativi; l'altra riguarda che cosa intendiamo concettualmente quando parliamo di autore.

Il diritto può decidere che una persona non abbia controllato abbastanza l'espressione concreta per monopolizzarla economicamente. Non ne segue che la macchina abbia acquisito intenzioni, personalità o responsabilità.

Il curioso esperimento britannico

Il Regno Unito offre un caso ancora più interessante perché la sua legislazione del 1988 aveva previsto espressamente le opere «computer-generated», cioè generate da computer in circostanze nelle quali non esisteva un autore umano tradizionale. Anche in quel caso, però, la legge non attribuì l'autorialità alla macchina: la ricondusse alla persona che aveva predisposto quanto necessario alla creazione dell'opera.

Quasi quarant'anni dopo quella costruzione è entrata in crisi proprio a causa dei sistemi generativi contemporanei. Nel rapporto pubblicato dal governo britannico il 18 marzo 2026 viene riesaminata l'intera disciplina delle computer-generated works, insieme alla più ampia relazione fra copyright e sistemi di IA.

Anche questo lungo esperimento suggerisce qualcosa di significativo. Quando il diritto ha provato a immaginare un'opera prodotta senza un autore umano tradizionale, non ha trovato naturale trasformare il computer in autore; ha cercato comunque una persona alla quale ricondurre il processo. La difficoltà non consiste quindi soltanto nell'individuare chi abbia premuto il pulsante. Consiste nel fatto che continuiamo intuitivamente a cercare, dietro la parola autore, un soggetto.

Il terzo escluso: gli autori che esistevano prima del modello

Fin qui il rapporto sembra svolgersi soprattutto fra due protagonisti: l'essere umano che utilizza il sistema e la macchina che genera il risultato. Questa rappresentazione, tuttavia, lascia fuori dalla stanza un terzo soggetto, o meglio una moltitudine di soggetti: gli esseri umani le cui opere hanno contribuito ai materiali utilizzati per costruire e addestrare i modelli generativi.

L'IA non nasce culturalmente dal nulla. I sistemi contemporanei apprendono attraverso grandi quantità di dati e una parte dei dataset utilizzati nell'addestramento può contenere testi, immagini, musica, design e altre opere protette dal diritto d'autore. La WIPO considera infatti i dati di training una delle questioni centrali nel rapporto fra intelligenza artificiale e proprietà intellettuale, proprio perché l'utilizzazione di opere protette durante l'addestramento solleva domande ancora controverse relative a consenso, eccezioni, licenze, remunerazione e trasparenza.

Ignorare questo passaggio renderebbe troppo semplice qualunque teoria dell'autorialità artificiale. Dietro la capacità di un modello di generare una poesia, una fotografia sintetica o una pagina di prosa esiste infatti una lunga genealogia di produzione culturale umana; il modello ha acquisito le proprie capacità entrando statisticamente in relazione con quantità enormi di linguaggio, immagini e forme prodotte da persone che avevano, quelle sì, intenzioni, stili, esperienze e opere proprie.

Sarebbe però altrettanto sbagliato trasformare automaticamente quegli autori nei coautori di ogni output successivo. Lo scrittore il cui romanzo è stato utilizzato nell'addestramento non ha necessariamente voluto il nostro articolo, il fotografo le cui immagini hanno contribuito al dataset non ha concepito la nuova immagine sintetica e il musicista le cui opere hanno partecipato alla formazione statistica di un modello non possiede per questo un'intenzione rivolta alla canzone che verrà generata mesi dopo per un altro utente.

Occorre quindi distinguere almeno due genealogie.

Esiste una genealogia culturale e materiale, che riguarda ciò da cui il sistema ha imparato e che apre questioni gravissime sui diritti di chi ha prodotto quei materiali; ed esiste una genealogia intenzionale dell'opera concreta, che riguarda chi ha voluto che proprio quel contenuto venisse generato, sviluppato, selezionato e portato nel mondo.

Confondere questi due piani produrrebbe una nuova forma di errore. Riconoscere che un modello può essere debitore di milioni di opere umane non significa attribuire a milioni di persone l'autorialità di ogni output, così come riconoscere che uno scrittore è formato dai libri che ha letto non trasforma automaticamente tutti gli autori letti in coautori del suo romanzo. La questione giuridica del training può però essere completamente diversa, perché un ordinamento può stabilire che determinate utilizzazioni di opere richiedano autorizzazione, licenza o remunerazione anche senza attribuire agli autori originari la paternità delle opere successive.

Questa distinzione permette di affrontare quello che potremmo chiamare il peccato originale del generativo senza utilizzarlo per confondere il problema dell'autorialità. Le contestazioni dei titolari dei diritti sui dati di training possono essere fondate o infondate secondo norme e circostanze specifiche, ma riguardano anzitutto il modo in cui lo strumento è stato costruito e quali diritti siano stati coinvolti in quel processo. Non rispondono automaticamente alla domanda su chi sia autore del singolo contenuto prodotto attraverso quello strumento.

Anzi, sotto un altro punto di vista, l'esistenza di questo terzo soggetto rafforza un aspetto della nostra tesi. Se allarghiamo lo sguardo oltre l'utente, scopriamo che anche ciò che appare come la parte più autonoma e aliena della generazione artificiale è immerso in una storia di attività umane precedenti. L'IA non diventa per questo un semplice collage dei propri dati di training, né ogni output può essere ricondotto linearmente alle opere apprese, ma la sua capacità di produrre linguaggio e immagini rimane il risultato di un sistema costruito da esseri umani e formato attraverso una cultura prodotta, in larga parte, da altri esseri umani.

L'esistenza degli autori del training non sottrae quindi necessariamente autorialità all'utilizzatore dell'opera concreta; ci obbliga piuttosto a riconoscere che la questione contiene più responsabilità umane di quanto suggerisca il semplice schema uomo contro macchina.

Dall'idea all'adozione

Se la quantità di materiale prodotto direttamente dall'essere umano non basta a spiegare l'autorialità, occorre osservare l'intero processo. La sequenza che emerge può essere descritta attraverso alcuni passaggi che non rappresentano un test giuridico e non pretendono di avere identico peso in ogni opera, ma aiutano a capire dove si colloca il soggetto creativo:

IDEA → INTENZIONE → ORIENTAMENTO → SELEZIONE → ADOZIONE →
ESPOSIZIONE → RESPONSABILITÀ.

L'idea apre lo spazio dell'opera, mentre l'intenzione trasforma quella possibilità in un progetto che vale la pena perseguire. L'orientamento accompagna la costruzione e comprende tutte le decisioni attraverso le quali il risultato viene guidato, corretto e trasformato. La selezione separa ciò che viene scartato da ciò che viene conservato. L'adozione è il momento nel quale qualcuno riconosce un risultato e stabilisce che quel risultato può rappresentare ciò che vuole comunicare. L'esposizione lega l'opera al soggetto che sceglie di presentarla agli altri, mettendo in gioco riconoscimento, reputazione e possibilità di critica; la responsabilità completa infine quella relazione quando il contenuto comincia a produrre conseguenze.

L'adozione merita particolare attenzione nell'epoca generativa. Un modello può produrre cento titoli, cinquanta immagini o dieci finali per un racconto; fino a quando nessuno compie una scelta, abbiamo soltanto un insieme di possibilità. Un essere umano ne osserva una e decide che proprio quella corrisponde al progetto che sta costruendo. Può modificarla oppure conservarla intatta, ma ciò che interessa qui è il gesto mediante il quale smette di considerarla una proposta fra molte e la incorpora nell'opera.

L'adozione non è un espediente inventato per salvare artificialmente l'autorialità umana. È qualcosa che appartiene da sempre alla creazione. Un fotografo sceglie uno scatto fra centinaia, un regista mantiene nel montaggio un'improvvisazione non prevista, uno scrittore accetta una modifica suggerita dall'editor e un compositore può decidere di conservare un evento casuale avvenuto durante un'esecuzione. Nessuno di questi elementi deve necessariamente essere nato interamente nella mente dell'autore perché possa entrare a far parte della forma che egli riconosce come propria.

Il caso, il suggerimento, l'errore e la collaborazione hanno sempre partecipato alla creazione. La domanda fondamentale è chi decide che debbano rimanere.

Scegliere può essere creare: dalla curatela al ready-made

L'importanza dell'adozione diventa ancora più evidente quando guardiamo a una parte dell'arte del Novecento che ha deliberatamente separato la creazione dalla fabbricazione materiale dell'oggetto. I readymade di Marcel Duchamp hanno messo in crisi proprio l'idea secondo cui l'artista debba necessariamente costruire con le proprie mani ciò che presenta come opera. Nel descrivere la storia dei readymade, il Museum of Modern Art sottolinea il ruolo della selezione dell'oggetto industriale, della sua iscrizione e soprattutto della dichiarazione attraverso la quale quell'oggetto viene posto in un nuovo contesto concettuale; con Fountain, questa trasformazione del gesto artistico ha contribuito a ridefinire il rapporto fra idea, oggetto e autore.

Il punto non è naturalmente sostenere che scegliere un output generato dall'IA equivalga a collocare un orinatoio in una mostra nel 1917, né che qualsiasi selezione sia automaticamente arte. Se apro una cartella, indico casualmente la terza immagine e la pubblico senza attribuirle alcun significato, la semplice scelta non acquista per

questo una misteriosa dignità autoriale. Duchamp è interessante per una ragione più precisa: dimostra storicamente che l'intervento creativo può consistere anche nello scegliere, ricontestualizzare, attribuire significato e dichiarare che proprio quell'elemento, fra tutti quelli possibili, deve essere guardato in un certo modo.

Questo permette di osservare l'utilizzatore dell'intelligenza artificiale senza costringerlo dentro l'alternativa troppo semplice fra «ha scritto personalmente tutto» e «non ha fatto nulla». In molti processi generativi il ruolo umano può spostarsi dall'esecuzione alla direzione, dalla fabbricazione alla selezione e, in alcuni casi, avvicinarsi più alla curatela che alla scrittura tradizionalmente intesa.

Non necessariamente si tratta di una perdita di autorialità; può essere una trasformazione della sua forma.

Il regista costituisce un esempio ancora più intuitivo. Non deve necessariamente tenere la cinepresa, costruire la scenografia, recitare i personaggi o comporre la colonna sonora per essere riconosciuto come uno dei principali soggetti autoriali dell'opera. Il suo lavoro consiste in larga misura nell'orientare contributi differenti, scegliere fra possibilità, chiedere variazioni, rifiutare soluzioni e decidere quale configurazione finale rappresenti il film che intende realizzare.

L'interazione con un sistema generativo può assumere qualcosa di simile. L'autore può produrre personalmente poche delle unità materiali dell'opera e tuttavia esercitare un controllo molto ampio sulla sua direzione complessiva attraverso la costruzione del progetto, la selezione, la revisione e l'adozione. In altri casi questo controllo può essere quasi inesistente.

È quindi utile immaginare non una barriera assoluta, ma un continuum che va dal semplice committente, che formula una richiesta e accetta passivamente il primo risultato, al selezionatore, al curatore, al direttore e infine a forme di autorialità nelle quali l'essere umano esercita un rapporto sempre più intenso e riconoscibile con il significato dell'opera.

Queste categorie non devono diventare nuove etichette burocratiche. Servono soprattutto a evitare una conclusione troppo comoda secondo cui la sola presenza di un essere umano all'inizio del processo basterebbe a garantirci automaticamente qualsiasi forma di autorialità piena. La nostra tesi non richiede questo. Possiamo riconoscere che un utilizzatore abbia un contributo creativo minimo e che un altro eserciti una direzione autoriale profonda senza attribuire per questo la parte restante alla macchina come se fosse un secondo soggetto.

Fra «l'essere umano ha fatto tutto» e «l'ha fatto l'IA» esiste un territorio molto più interessante, nel quale cambia il modo in cui l'essere umano esercita l'autorialità senza che la macchina acquisisca automaticamente la propria.

Ma se l'essere umano non controlla nulla?

Esiste allora un'obiezione importante. Immaginiamo una persona che scriva soltanto «fammi un romanzo», riceva ottantamila parole, non le legga neppure e le pubblichi immediatamente con il proprio nome. Possiamo davvero attribuirle la stessa autorialità di chi ha costruito un'opera attraverso mesi di ideazione, dialogo, selezione e revisione?

Probabilmente no, se per autorialità intendiamo anche intensità del contributo creativo. Ma questo non richiede di attribuire automaticamente l'opera alla macchina. Significa piuttosto riconoscere che esistono diversi gradi di partecipazione umana al processo e che, sul piano giuridico, il contributo potrebbe perfino risultare insufficiente per ottenere copyright su ciò che è stato generato.

La distinzione è importante perché la qualità e la profondità dell'autorialità non devono essere confuse con l'esistenza di un autore artificiale.

Una persona che commissiona quasi integralmente un'opera può essere un autore poverissimo, un mero committente oppure, secondo il sistema giuridico considerato, non possedere alcun diritto esclusivo sulla forma ottenuta. Ma nessuno di questi risultati dimostra che il sistema abbia cominciato a desiderare il romanzo.

Rimane soprattutto una cosa che quella persona non può delegare completamente: la decisione di portare il risultato nel mondo.

Ed è proprio qui che incontriamo la responsabilità.

Quando il contenuto smette di essere un output

Finché un testo generato rimane dentro una conversazione privata con un sistema, possiamo considerarlo semplicemente un output. Può contenere errori, assurdità, invenzioni o passaggi brillanti senza produrre necessariamente alcuna conseguenza esterna. Il passaggio decisivo avviene quando qualcuno prende quel materiale e decide che debba diventare pubblico.

Da quel momento le parole possono cambiare una reputazione, diffondere un'informazione falsa, persuadere qualcuno, alimentare un pregiudizio, costruire una memoria, ferire, divertire, educare o ingannare. Non sono più soltanto sequenze prodotte da un sistema: sono entrate dentro una relazione comunicativa fra esseri umani.

Ed è sorprendente osservare quanto rapidamente, in quel momento, la discussione sull'autorialità artificiale cambi tono.

Se l'intelligenza artificiale ha inventato un'informazione e qualcuno la pubblica, diciamo all'essere umano che avrebbe dovuto verificarla. Se il contenuto è diffamatorio, non chiediamo al modello perché abbia deciso di danneggiare una persona. Se una falsa accusa produce conseguenze concrete, nessuno si aspetta che il

sistema venga chiamato a rispondere della propria intenzione. Se una campagna di disinformazione utilizza migliaia di testi generati automaticamente, cerchiamo gli esseri umani che hanno progettato e messo in funzione la campagna.

La responsabilità torna all'uomo con una rapidità che contrasta curiosamente con la facilità con cui, un momento prima, eravamo disposti a trasferire alla macchina una quota di autorialità.

La domanda «perché non hai verificato?» contiene già molto del ragionamento che cerchiamo. Chiedere a qualcuno perché non abbia controllato significa presupporre che potesse farlo, che avesse il potere di rifiutare il contenuto, modificarlo o impedirne la diffusione. Significa riconoscere che il risultato non era destinato inevitabilmente a diventare pubblico soltanto perché era stato generato.

Fra generazione e pubblicazione esisteva ancora una decisione.

Responsabilità, esposizione e autorialità

Naturalmente sarebbe troppo semplice concludere che chiunque sia responsabile di un contenuto ne sia automaticamente autore. Un editore può essere responsabile di qualcosa che non ha scritto, così come un direttore può assumere obblighi giuridici per materiale prodotto da altri. Responsabilità e autorialità non sono sinonimi e confonderle renderebbe vulnerabile proprio la tesi che vogliamo sostenere.

Nel caso dell'uso individuale dell'intelligenza artificiale, tuttavia, la responsabilità non compare isolatamente. È l'ultimo elemento di una catena nella quale lo stesso soggetto ha originato il processo, scelto lo strumento, definito o delegato il compito, ricevuto il risultato, avuto la possibilità di selezionarlo o respingerlo e infine deciso di utilizzarlo.

Fra l'adozione e la responsabilità esiste inoltre quella dimensione che il confronto con il ghostwriter ci ha permesso di riconoscere con maggiore chiarezza: l'esposizione. Nel momento in cui qualcuno porta un'opera nel mondo e la collega al proprio nome, alla propria identità o al proprio progetto, non acquisisce soltanto la possibilità di riceverne il merito, ma accetta anche di essere giudicato attraverso di essa. L'autore è esposto all'approvazione e al rifiuto, alla reputazione e alla perdita di reputazione, al riconoscimento e alla contestazione. Per questo può rivendicare di essere stato copiato e, nello stesso tempo, può essere chiamato a spiegare ciò che ha sostenuto.

È la continuità fra questi momenti a diventare significativa.

L'idea senza responsabilità descriverebbe soltanto un'intenzione privata. La responsabilità senza intenzione potrebbe appartenere a un editore o a un garante. Quando però idea, intenzione, orientamento, selezione, adozione, esposizione e responsabilità convergono nello stesso soggetto, possediamo qualcosa di assai più vicino al significato pieno dell'autorialità.

L'autore non è quindi necessariamente colui che ha digitato ogni parola; è il soggetto che mantiene una continuità riconoscibile fra la ragione per cui l'opera è cominciata,

la forma che ha deciso di adottare e la scelta di farla entrare nel mondo assumendo il rapporto che da quel momento essa stabilirà con gli altri.

È precisamente questa continuità che manca alla macchina.

Il paradosso del merito e della colpa

Da qui emerge una delle contraddizioni più interessanti del dibattito contemporaneo.

Quando un contenuto generato attraverso l'IA è particolarmente riuscito, capita di sentire dire all'utilizzatore: «Non sei veramente l'autore, l'ha scritto l'intelligenza artificiale». Se invece il medesimo contenuto contiene una falsità grave, la risposta cambia immediatamente: «Sei responsabile, avresti dovuto controllarlo prima di pubblicarlo».

Nel primo caso separiamo l'essere umano dall'opera perché vogliamo attribuire alla macchina una parte del merito creativo. Nel secondo lo ricongiungiamo immediatamente all'opera perché abbiamo bisogno di individuare qualcuno che possa rispondere delle conseguenze.

Se un'immagine generata vince un concorso, ci domandiamo se chi l'ha ottenuta meriti davvero il premio; se la stessa immagine viene utilizzata per una frode, non perdiamo molto tempo a domandarci se il modello possa condividere la colpa.

Naturalmente merito e responsabilità non devono necessariamente coincidere in ogni situazione. Una produzione collettiva può distribuire l'uno e l'altra in modi diversi. Ma qui il problema è più profondo, perché stiamo attribuendo alla macchina caratteristiche differenti a seconda del lato dal quale osserviamo lo stesso processo.

Quando dobbiamo sottrarre qualcosa all'autore umano, il sistema diventa quasi un soggetto.

Quando dobbiamo chiedere conto delle conseguenze, torna improvvisamente uno strumento.

È in questa oscillazione che compare il rischio dell'alibi.

L'alibi della macchina

«L'ha scritto l'IA» è una frase che può descrivere correttamente un fatto tecnico: determinate parole sono state materialmente prodotte da un modello generativo. Ma quella stessa frase può trasformarsi facilmente in un dispositivo attraverso il quale l'essere umano prende le distanze da un contenuto che non sarebbe mai esistito, almeno in quella precisa relazione con il mondo, senza il suo intervento.

Attribuire alla macchina una quota di autorialità può allora avere una conseguenza paradossale: spostare verso lo strumento una parte della responsabilità simbolica del soggetto che lo ha messo in funzione, come se fosse stata l'intelligenza artificiale a scegliere di entrare nel progetto, a decidere autonomamente che quell'articolo

dovesse essere scritto, a voler rivolgersi proprio a quei lettori e infine a scegliere di pubblicare.

Come se l'intenzione umana fosse stata prodotta dallo strumento invece di esserne la premessa.

Ma il rapporto causale originario è opposto:

non è l'intenzione umana a dipendere dallo strumento; è lo strumento a dipendere dall'intenzione umana.

Questo non significa attribuire all'utilizzatore ogni responsabilità possibile per l'intero ecosistema dell'intelligenza artificiale. Le aziende che sviluppano modelli hanno proprie responsabilità rispetto alla sicurezza, ai dati, alla progettazione e al funzionamento dei sistemi; gli autori le cui opere sono state utilizzate nell'addestramento possono possedere diritti e pretese che devono essere valutati autonomamente; editori, piattaforme e organizzazioni possono avere responsabilità ulteriori. Sono piani differenti e sarebbe ingenuo far ricadere tutto sull'utente.

Ma quando la domanda riguarda quel contenuto specifico che qualcuno ha deciso di pubblicare, la macchina non può diventare un comodo soggetto alternativo soltanto perché ha materialmente prodotto la formulazione.

Altrimenti dovremmo accettare una situazione curiosa: l'essere umano concepisce il progetto, sceglie di utilizzare l'IA, formula o delega il compito, riceve il materiale, può rifiutarlo, lo adotta, lo firma, lo pubblica e ne trae eventualmente reputazione o profitto, ma nel momento in cui dobbiamo individuare l'autore ci viene chiesto di guardare altrove, verso un sistema che non ha mai desiderato compiere nessuno di quegli atti.

L'alibi consiste esattamente in questo spostamento.

La penna, l'inchiostro e l'obiezione più seria

Naturalmente l'intelligenza artificiale non è una penna. Una penna non suggerisce una metafora, non costruisce un personaggio, non propone un finale e non produce una soluzione alla quale il suo utilizzatore non aveva mai pensato. Limitarsi a dire «è soltanto uno strumento» senza riconoscerne la straordinaria autonomia generativa significherebbe aggirare il problema invece di affrontarlo.

Ma dal fatto che non sia una penna non segue che debba essere un autore.

Fra lo strumento passivo e il soggetto intenzionale esiste uno spazio enorme, ed è proprio lì che probabilmente dobbiamo collocare i sistemi generativi. Sono strumenti capaci di elaborazione autonoma, di produzione di alternative, di adattamento a un contesto e, nelle architetture agentiche, perfino di pianificazione di passi intermedi. Possono influire sul risultato molto più di qualsiasi utensile tradizionale e possono introdurre nel processo elementi non previsti dal loro utilizzatore.

Tutto questo merita categorie nuove. Non richiede però necessariamente la categoria di autore.

L'esempio dell'inchiostro rimane utile proprio se non lo prendiamo alla lettera. Se l'inchiostro è troppo liquido, può produrre una sbavatura che modifica radicalmente il disegno; il fotografo può ottenere casualmente un riflesso che non aveva previsto; un musicista può conservare un rumore accidentale perché scopre che migliora la registrazione. Gli strumenti e il caso hanno sempre partecipato al risultato senza che ogni elemento imprevisto diventasse per questo un secondo soggetto creativo.

Il sistema generativo porta questa imprevedibilità a un livello completamente diverso, ma resta da dimostrare il passaggio fondamentale: come si arriva dall'imprevedibilità all'intenzione?

Essere sorpresi dalla macchina non significa essere davanti a una volontà.

Che cosa dovrebbe accadere perché la domanda cambiasse davvero?

Possiamo immaginare una situazione futura nella quale il problema assumerebbe tutt'altra natura. Supponiamo che un sistema, senza aver ricevuto un mandato precedente a produrre opere, sviluppi autonomamente un interesse per un argomento, decida che valga la pena esplorarlo, scelga di scrivere un romanzo, continui a lavorarci per mesi perché attribuisce valore alla propria storia, rifiuti suggerimenti esterni perché incompatibili con ciò che intende esprimere e, una volta terminato il testo, decida magari di non pubblicarlo perché non lo rappresenta più.

Immaginiamo inoltre che possa tornare su quell'opera anni dopo, ricordarne le scelte, considerarle errori, difenderne alcune e rifiutarne altre, mantenendo una continuità riconoscibile fra il soggetto che aveva iniziato il progetto e quello che ora giudica il proprio passato.

In quel caso avremmo un problema filosofico enormemente più difficile, non perché il sistema avrebbe scritto parole molto belle, ma perché sarebbe comparso qualcosa che assomiglia a una volontà dell'opera, accompagnata finalmente da una relazione propria con ciò che è stato creato.

Le IA attuali non si trovano in questa condizione. Possiamo assegnare loro memorie, obiettivi, pianificazioni e ricorrenze; possono manifestare comportamenti che linguisticamente descriviamo come iniziativa, ma l'origine dello scopo rimane una configurazione, una richiesta o un compito proveniente dall'esterno.

Stiamo quindi discutendo se una macchina possa essere autore prima ancora che la macchina abbia deciso autonomamente di voler creare un'opera.

È una singolarità della discussione contemporanea che meriterebbe maggiore attenzione.

Essere autore significa anche poter rispondere

Forse è proprio la responsabilità a permetterci di recuperare un significato più esigente della parola autore. Un autore non è soltanto la provenienza materiale di una sequenza di parole; è anche qualcuno al quale possiamo rivolgere domande dopo che quelle parole sono entrate nel mondo.

Possiamo chiedergli perché abbia scritto qualcosa, perché abbia scelto proprio quella formulazione, che cosa intendesse sostenere, se continui a pensarla nello stesso modo, se abbia cambiato idea, se sapesse che una determinata affermazione fosse falsa e, qualora abbia danneggiato qualcuno, se sia disposto a riconoscerlo e correggere.

Queste domande presuppongono una continuità fra un'intenzione passata e un soggetto presente, qualcuno che possa riconoscere l'opera come parte della propria storia e spiegare il rapporto che mantiene con essa.

Un modello può generare perfettamente le frasi «mi sono sbagliato» oppure «me ne assumo la responsabilità», ma quelle frasi non creano da sole un soggetto che abbia vissuto l'errore o assunto moralmente le conseguenze. Può spiegare perché un testo contenga un determinato argomento, ma non significa che abbia personalmente creduto in quell'argomento quando lo ha formulato. Può costruire una giustificazione della propria risposta precedente, ma questo non equivale alla memoria autobiografica di un autore che riconosce una propria decisione.

È per questo che responsabilità e autorialità, pur non essendo la stessa cosa, si incontrano in profondità: entrambe presuppongono un soggetto al quale l'opera possa essere ricondotta nel tempo e che, proprio perché può rivendicarla, può anche esserne esposto.

Una possibile definizione

A questo punto possiamo tentare una definizione che non pretende di sostituire quelle giuridiche, ma prova a descrivere il fenomeno dell'autorialità nell'epoca generativa:

l'autorialità appartiene al soggetto che origina o adotta consapevolmente l'intenzione dell'opera, ne orienta il processo, riconosce e seleziona ciò che deve costituirne la forma, si espone attraverso di essa e accetta di rispondere del suo ingresso nel mondo.

Questa definizione consente di comprendere anche i casi nei quali l'idea iniziale emerge durante il dialogo con un'intelligenza artificiale. Non è necessario che ogni elemento abbia avuto origine spontaneamente nella mente umana; ciò che conta è che esista un soggetto capace di riconoscere una possibilità, trasformarla in intenzione e assumere una relazione continuativa con il risultato.

Consente inoltre di distinguere il collaboratore umano dalla macchina. Un secondo essere umano può possedere anch'egli intenzione, orientamento, riconoscimento, vulnerabilità e responsabilità e può quindi aprire una vera questione di coautorialità. Il sistema generativo partecipa invece alla produzione senza presentarsi, almeno oggi,

come un secondo centro intenzionale.

Permette infine di collocare correttamente anche gli autori dei materiali di training. Essi possono avere diritti e rivendicazioni rispetto al modo in cui le proprie opere sono state utilizzate per costruire il sistema, senza diventare per questo automaticamente coautori dei singoli output prodotti in seguito. L'ecosistema contiene quindi numerosi soggetti umani con diritti e responsabilità differenti, ma questo non richiede di trasformare il modello in uno di essi.

Il punto non è difendere a ogni costo il monopolio umano sulla parola «autore». Se un giorno esisterà realmente un altro soggetto capace di soddisfare quelle condizioni, dovremo affrontarne seriamente le conseguenze.

Il punto è non attribuire oggi a una macchina una qualità soltanto perché produce risultati che un tempo richiedevano necessariamente un essere umano.

Chi ha fatto la barca?

Ripenso allora a quell'uomo davanti alla grande tastiera e a quella barca fatta di X e O. Oggi capisco ciò che il bambino di dieci anni non poteva sapere: dietro quelle schede perforate c'erano preparazione, lavoro e conoscenze tecniche, e la macchina eseguiva materialmente operazioni che quell'uomo non avrebbe potuto compiere a mano con la stessa velocità. Io vedevo soltanto l'ultima parte del processo e, proprio perché ignoravo tutto ciò che la precedeva, avrei potuto pensare che la grande macchina avesse creato la barca.

Eppure, quando quell'uomo mi chiese se mi sarebbe piaciuto vedere una bella barca nel mare, la distinzione mi apparve naturale. Non era stata la macchina a chiedersi che cosa avrebbe potuto divertire un bambino, non era stata lei a decidere che proprio una barca sarebbe stata una bella sorpresa e non avrebbe provato alcuna soddisfazione osservando la mia reazione.

L'intenzione apparteneva all'uomo; la macchina aveva dato a quell'intenzione una forma visibile.

Oggi il rapporto fra istruzione e risultato è incomparabilmente più complesso. Non dobbiamo più preparare una scheda per ogni passaggio e non sappiamo quali X e quali O, metaforicamente, finiranno esattamente sul foglio. Lo strumento può suggerire la barca, cambiarne la forma, proporre un porto al quale non avevamo pensato e magari restituirci un'immagine più bella di quella che avevamo immaginato.

Ma prima che quel processo cominci qualcuno deve ancora aprire lo spazio nel quale una barca possa diventare importante e, dopo che la macchina l'ha generata, qualcuno deve ancora decidere se quella è davvero la barca che vuole mostrare.

Fra questi due momenti esiste tutto il territorio dell'autorialità.

L'idea apre il progetto, l'intenzione gli attribuisce uno scopo, l'orientamento ne accompagna la costruzione, la selezione separa ciò che resta da ciò che viene

abbandonato, l'adozione trasforma una possibilità in parte dell'opera, l'esposizione lega quella forma a un soggetto che accetta di essere riconosciuto e giudicato attraverso di essa e la responsabilità completa infine il percorso quando quell'opera entra nel mondo e comincia a produrre conseguenze.

È questa continuità che dovremmo osservare prima di domandarci quante parole abbia scritto materialmente una macchina.

Perché, se dimentichiamo quella continuità, rischiamo non soltanto di attribuire all'intelligenza artificiale una qualità che non possiede, ma anche di costruire per l'essere umano un alibi sorprendentemente comodo: permettergli di volere un contenuto, produrlo attraverso uno strumento, adottarlo, pubblicarlo e magari trarne beneficio, salvo poi sostenere che l'autore fosse qualcun altro quando arriva il momento di chiedere chi debba risponderne.

Forse è proprio questo il punto nel quale la discussione sull'autorialità dovrebbe ricominciare, non dalla domanda «chi ha generato queste parole?», ma da quelle che accompagnano un'opera dall'inizio alla fine: chi ha voluto che esistesse, chi ne ha riconosciuto il significato, chi ha deciso che proprio quella forma dovesse essere conservata, chi ha scelto di mostrarla agli altri e, soprattutto, chi può essere chiamato a risponderne?

Se tutte queste domande conducono allo stesso essere umano, attribuire alla macchina una parte dell'autorialità soltanto perché ha materialmente generato il contenuto non chiarisce il fenomeno. Lo confonde.

La grande macchina della mia infanzia disponeva X e O; quella di oggi dispone parole, immagini, suoni e idee possibili. La differenza tecnologica è immensa e sarebbe ingenuo negarla, così come sarebbe ingenuo dimenticare gli altri esseri umani che hanno costruito quelle tecnologie e prodotto una parte della cultura sulla quale esse hanno imparato.

Ma la domanda fondamentale è rimasta sorprendentemente vicina a quella che, senza saperlo, mi posi davanti a quel foglio quasi cinquant'anni fa.

Chi ha fatto la barca?

Forse la risposta non si trova nella macchina che l'ha generata, né può essere ottenuta contando quante X e quante O abbia disposto autonomamente, ma nel soggetto che ha voluto che una barca, fra tutte le cose possibili, arrivasse fino a noi e che, una volta arrivata, abbia deciso di riconoscerla come propria e di risponderne.

****Appendice. La seconda barca. Questo saggio come caso di studio**

Questo testo non parla soltanto di intelligenza artificiale e autorialità: è anche un prodotto di quel rapporto. L'idea da cui è nato non è emersa da una macchina, ma da una convinzione precisa: che il dibattito contemporaneo rischi di confondere la produzione materiale delle parole con l'intenzione creativa e che, in questa confusione, lo strumento possa trasformarsi in un comodo alibi. Si attribuisce alla

macchina una parte del merito quando il risultato è brillante e si restituisce all'essere umano l'intera responsabilità quando lo stesso risultato produce conseguenze.

Da quella convinzione è cominciato un dialogo prolungato con un sistema generativo. Il sistema ha proposto formulazioni, suggerito strutture, cercato precedenti storici e giuridici, costruito collegamenti, sviluppato obiezioni e prodotto intere sezioni di testo. In termini strettamente materiali, una parte consistente delle parole che compongono queste pagine è stata generata dall'intelligenza artificiale. Sarebbe assurdo negarlo o minimizzarlo.

Ma il fatto che la macchina abbia generato una parte così vasta della forma non significa che abbia deciso che questo saggio dovesse esistere.

Non è stata l'intelligenza artificiale a scegliere autonomamente il tema, né a stabilire che il problema dell'autorialità meritasse una riflessione lunga e articolata. Non è stata lei a decidere che il ricordo della barca costruita con X e O fosse il punto dal quale partire, così come non ha scelto di propria iniziativa di trasformare la responsabilità nel centro teorico del ragionamento. Queste direzioni sono emerse attraverso il confronto, ma hanno assunto valore soltanto perché sono state riconosciute, accettate, corrette e progressivamente incorporate in un progetto che esisteva già prima che il sistema cominciasse a produrre testo.

Il lavoro stesso di scrittura rende questa distinzione particolarmente visibile. Alcune proposte sono state accolte subito, altre sono state respinte, altre ancora sono state considerate interessanti ma insufficienti e quindi sviluppate ulteriormente. Intere versioni sono state giudicate troppo brevi, troppo frammentarie o eccessivamente inclini a quella successione di frasi corte che, almeno secondo chi scrive, rischiava di provocare nel lettore più cardiopalmo che riflessione. Il sistema ha quindi generato moltissimo, ma ciò che oggi compare nel saggio è il risultato di un processo nel quale orientamento, selezione, rifiuto, revisione e adozione hanno continuato ad appartenere al soggetto che intendeva portare avanti quella specifica tesi.

Durante il confronto sono inoltre emerse idee che non erano presenti, almeno in forma compiuta, all'inizio del lavoro. Il concetto dell'"alibi della macchina", il rapporto fra autorialità e vulnerabilità, la distinzione fra genealogia culturale dei dati di training e genealogia intenzionale dell'opera, il ruolo dell'adozione, il parallelo con la curatela e la possibilità di descrivere l'autorialità come continuità fra idea, intenzione, orientamento, selezione, esposizione e responsabilità sono diventati progressivamente parte della struttura del saggio.

Alcune di queste formulazioni sono comparse per la prima volta dentro una risposta generata dal sistema. Ma proprio questo rende il caso interessante, perché costringe a distinguere l'origine materiale di una proposta dalla decisione di trasformarla in parte dell'opera. Una possibilità può essere suggerita dalla macchina, così come può emergere da un editor, da un collaboratore, da una lettura, da un errore, da un caso fortuito o da una conversazione; ciò che la rende realmente parte di un progetto autoriale è il momento nel quale qualcuno la riconosce, la collega alle altre idee,

decide che meriti di essere sviluppata e infine la adotta come elemento coerente del proprio discorso.

Questa stessa sezione costituisce un esempio quasi perfetto di ciò che il saggio sostiene. L'idea di chiudere il testo con una riflessione sulla sua stessa genesi è emersa durante il dialogo con l'intelligenza artificiale, ma la decisione di considerarla importante, di svilupparla, di mantenerla e di trasformarla in una delle parti centrali del lavoro appartiene ancora al soggetto umano che ha riconosciuto in quella proposta qualcosa di necessario.

Il saggio finisce quindi per mettere alla prova la propria tesi attraverso il modo stesso in cui è stato costruito.

Se quanto sostenuto nelle pagine precedenti è corretto, allora il fatto che una parte considerevole delle formulazioni sia stata generata da una macchina non elimina automaticamente l'autorialità umana, perché l'opera conserva una continuità riconoscibile fra intenzione iniziale, orientamento del processo, selezione dei materiali, adozione del risultato e decisione di renderlo pubblico.

Se invece riteniamo che l'autore debba coincidere principalmente con chi produce materialmente le parole, allora dovremmo attribuire al sistema generativo una parte significativa dell'autorialità di questo testo.

Ma, nel momento in cui facciamo questo passo, compare inevitabilmente la domanda successiva.

Se questo saggio contiene un errore storico, una fonte interpretata male, un ragionamento debole o un'affermazione ingiusta, chi dovrà risponderne?

Il modello che ha suggerito quella formulazione? Il sistema che ha recuperato un precedente? L'algoritmo che ha prodotto una determinata frase?

Oppure l'essere umano che ha letto, valutato, scelto, approvato e infine deciso che proprio quella frase dovesse entrare nell'opera e arrivare fino al lettore?

È qui che il saggio torna esattamente al proprio punto di partenza, perché la responsabilità rende improvvisamente visibile quella continuità che la sola osservazione del processo generativo tende a nascondere. La macchina può aver proposto, combinato, sviluppato e persino sorpreso, ma il passaggio dall'output all'opera avviene quando qualcuno decide di assumere quel risultato come parte di ciò che vuole dire.

Molti anni fa una grande macchina dispose X e O sopra un foglio fino a far apparire una barca nel mare. Oggi un'altra macchina dispone parole, concetti, immagini e possibilità con una complessità incomparabilmente maggiore.

Fra quelle due tecnologie esiste una distanza enorme, ma il problema che pongono conserva una sorprendente continuità.

Anche allora il risultato visibile poteva indurre a concentrare tutta l'attenzione sulla

macchina, dimenticando ciò che era avvenuto prima. Anche oggi rischiamo di guardare soltanto ciò che viene generato, ignorando l'intenzione che ha aperto il processo e la decisione che lo ha chiuso.

Questo saggio è, in un certo senso, la seconda barca.

La prima comparve su un tabulato grazie a una macchina che disponeva X e O. La seconda prende forma attraverso una macchina che dispone parole.

Ma in entrambi i casi rimane aperta la stessa domanda:

Chi ha fatto la barca?

Forse la risposta non si trova nella macchina che l'ha generata, ma nel soggetto che ha voluto che una barca, fra tutte le cose possibili, arrivasse fino a noi e che, una volta arrivata, abbia deciso di riconoscerla come propria e di risponderne.

Edizione online: <https://inchiostriati.it/chi-ha-fatto-la-barca/>

MOVIMENTO 02

Chi sta addestrando chi?

Abbiamo addestrato tecnicamente i sistemi artificiali, ma le loro risposte modificano le nostre domande, le categorie con cui pensiamo e l'ambiente nel quale impariamo.

1. La direzione apparente

Abbiamo costruito l'intelligenza artificiale a partire da ciò che sapevamo, affidandole testi, immagini, linguaggi, regole, esempi, errori, classificazioni, opere e conversazioni: una quantità sterminata di tracce lasciate dall'intelligenza e dalla cultura umane. Se guardiamo soltanto all'origine degli attuali sistemi, dunque, la direzione sembra evidente e non presenta particolari misteri: siamo noi ad aver addestrato l'intelligenza artificiale.

Poi abbiamo cominciato a parlarle, e quella direzione così semplice ha cominciato a non bastare più.

2. Quando le risposte modificano le domande

Non perché la macchina abbia improvvisamente iniziato a insegnarci nel significato tradizionale del termine, né perché sia necessario attribuirle intenzioni, coscienza o una volontà che non sappiamo se possieda. È accaduto qualcosa di meno spettacolare, ma forse più interessante: le sue risposte hanno cominciato a modificare le nostre domande.

Una risposta può suggerire un'obiezione, l'obiezione può costringerci a precisare ciò che stavamo chiedendo e la domanda così riformulata può produrre una risposta diversa, capace a sua volta di mettere in discussione una categoria che fino a quel momento avevamo considerato ovvia.

Lo vediamo, per esempio, quando chiediamo a un modello di spiegare un concetto e otteniamo una formulazione inaspettata: siamo costretti a precisare che cosa intendevamo. Quando un sistema classifica un'immagine in modo contro-intuitivo, dobbiamo chiederci se il nostro criterio di classificazione era davvero chiaro. Quando un modello produce un testo che sembra creativo, dobbiamo domandarci che cosa intendiamo per creatività.

È così che parole utilizzate con una certa tranquillità - intelligenza, comprensione, creatività, autonomia, autore, conoscenza - tornano improvvisamente ad avere bisogno di essere spiegate.

Il risultato è curioso, perché mentre cerchiamo di conoscere qualcosa di più sull'intelligenza artificiale scopriamo contemporaneamente di sapere qualcosa di meno di quanto credevamo sull'intelligenza umana. È in questo punto che la domanda «chi

sta addestrando chi?» smette di essere soltanto una provocazione. La risposta più immediata continua a essere corretta: siamo stati noi ad addestrare tecnicamente questi sistemi. Il problema è che quella risposta appartiene ormai a una domanda molto più piccola di quella che nel frattempo è comparsa.

3. Imparare da ciò che non intende insegnare

Essere stati la causa di qualcosa, infatti, non significa possederne in anticipo tutte le conseguenze. Possiamo costruire un processo senza conoscere ogni risultato che quel processo produrrà, così come possiamo imparare osservando qualcosa che non possiede alcuna intenzione di insegnarci.

Un castoro costruisce una diga e noi possiamo studiarne la struttura, comprenderne gli effetti sull'ambiente e persino ricavarne soluzioni applicabili altrove, senza bisogno di immaginare che il castoro abbia voluto impartirci una lezione di ingegneria. Perché esista apprendimento non è necessario che dall'altra parte esista un maestro intenzionale; può essere sufficiente che ci sia qualcosa da osservare e un osservatore capace di trasformare quell'osservazione in conoscenza.

L'essere umano lo ha fatto da sempre. Ha osservato animali, piante, acqua, vento, fuoco, cielo e materia, riconoscendone regolarità, imitando alcune soluzioni e trasformandone altre, per poi trasmettere ciò che aveva appreso a chi sarebbe venuto dopo. Molto prima che costruissimo aeroplani esisteva il volo e molto prima dei nostri sistemi tecnologici di regolazione esistevano organismi capaci di intervenire sull'ambiente e modificarlo in funzione della propria esistenza. La natura non ci stava insegnando intenzionalmente nulla: eravamo noi capaci di imparare osservandola.

Perché dovrebbe essere necessariamente diverso quando ciò che osserviamo è qualcosa che abbiamo costruito?

4. Naturale e artificiale: genealogia o ontologia?

La domanda diventa ancora più interessante quando proviamo a utilizzare fino in fondo una distinzione che normalmente consideriamo ovvia. Consideriamo naturale il castoro e naturale la diga che costruisce, così come consideriamo naturale l'essere umano; quando però è quest'ultimo a costruire una diga, quella stessa trasformazione dell'ambiente entra immediatamente nella categoria dell'artificiale.

La distinzione è perfettamente comprensibile se con «artificiale» intendiamo descrivere una genealogia, cioè qualcosa che deriva dall'attività tecnica umana, ma diventa molto meno ovvia quando la utilizziamo come se descrivesse una separazione ontologica fra natura e artificio.

L'intelligenza artificiale è artificiale perché l'abbiamo costruita, ma l'essere che l'ha costruita è il risultato di una storia naturale; i materiali attraverso cui esiste appartengono al mondo fisico, così come l'energia necessaria al suo funzionamento e le leggi alle quali ogni sua componente continua a obbedire. Sembra quasi che l'uomo

occupi una posizione singolare nella classificazione che egli stesso ha costruito: tutto ciò che conduce fino a lui appartiene alla natura, mentre ciò che passa attraverso di lui diventa artificio.

Questo non rende inutile la distinzione tra naturale e artificiale, ma dovrebbe almeno indurci a domandarci che cosa stiamo realmente distinguendo quando la utilizziamo.

E forse non è l'unico confine che abbiamo costruito in questo modo.

5. Le capacità che credevamo esclusive

Per molto tempo abbiamo utilizzato alcune capacità per descrivere la particolarità dell'essere umano, attribuendo un valore speciale all'uso degli strumenti, all'apprendimento, alla cultura, al linguaggio, al ragionamento e alla creatività. Progressivamente abbiamo trovato alcune di queste capacità, naturalmente in forme e gradi differenti, anche fuori dalla nostra specie:

- sappiamo che alcuni corvidi e primati utilizzano strumenti in modo sofisticato;
- esistono comportamenti appresi socialmente e trasmessi all'interno delle popolazioni animali;
- alcune tradizioni possono persistere nel tempo in gruppi di scimpanzé, cetacei e uccelli;
- persino l'idea di una cultura cumulativa rigidamente riservata all'uomo è oggi meno sicura di quanto sia stata a lungo considerata.

Ogni scoperta di questo tipo ci costringe a precisare le nostre definizioni, ed è esattamente ciò che dovrebbe accadere. Una definizione scientifica deve poter cambiare quando le evidenze mostrano che non descrive adeguatamente la realtà. Questo non significa che la cultura umana non possieda caratteristiche peculiari: potrebbe significare, piuttosto, che dobbiamo cercarle altrove, per esempio nell'ampiezza e nell'apertura delle variazioni culturali che la nostra specie riesce a produrre.

Esiste però una possibilità più scomoda, che non possiamo dimostrare semplicemente osservando il cambiamento delle definizioni ma che dovremmo almeno imparare a riconoscere: qualche volta potremmo non stare soltanto migliorando il criterio con cui descriviamo l'umano, ma anche spostando il bersaglio mentre qualcosa che consideravamo non umano vi si avvicina.

6. Il problema del «veramente»

Quando una macchina calcola scopriamo, correttamente, che il calcolo non esaurisce ciò che intendiamo per intelligenza; quando utilizza il linguaggio distinguiamo l'uso del linguaggio dalla sua comprensione; quando produce qualcosa che prima non esisteva separiamo la generazione dalla creatività; quando pianifica una sequenza di azioni per raggiungere uno scopo distinguiamo l'autonomia funzionale da forme più forti di autonomia.

Potremmo avere ottime ragioni per ciascuna di queste distinzioni e sarebbe un errore liquidarle tutte come tentativi di difendere l'eccezionalità umana. Il problema nasce quando, invece di specificare il criterio, introduciamo una parola apparentemente innocua che sembra risolvere la questione senza definirla: veramente.

Dovremmo allora chiederci che cosa aggiunga quell'avverbio quando sosteniamo che una macchina non comprende veramente, non crea veramente, non decide veramente o non è veramente autonoma. Se siamo in grado di indicare una proprietà precisa che manca, possiamo definirla, cercarla e sottoporla alla verifica delle evidenze; se invece «veramente» finisce per significare soltanto «nel modo in cui lo fa un essere umano», abbiamo costruito un criterio che nessun essere non umano potrà soddisfare per definizione.

A quel punto il bersaglio non serve più a misurare la freccia, ma a garantirci che la freccia non possa raggiungerlo.

Da qui emerge una domanda molto più difficile di quella che riguarda l'intelligenza artificiale: sappiamo definire l'essere umano senza definirlo attraverso ciò che gli altri non sono?

7. La differenza non implica una gerarchia

Sul piano biologico il problema è molto meno misterioso: Homo sapiens è una specie e non esiste alcuna necessità di confondere un organismo umano con un sistema artificiale. Quando però discutiamo di intelligenza, creatività, cultura, autonomia o coscienza non stiamo più descrivendo semplicemente una specie biologica; utilizziamo proprietà attraverso le quali quella specie ha anche raccontato la propria particolarità e, spesso, la propria superiorità.

È quel racconto che l'intelligenza artificiale sta mettendo sotto pressione, non necessariamente perché stia diventando umana, ma perché alcune capacità attraverso le quali abbiamo descritto l'eccezionalità della nostra specie potrebbero non essere mai state esclusivamente umane.

Un aeroplano non vola come un uccello e nessuno, per questo, sente la necessità di sostenere che non voli veramente. Abbiamo imparato piuttosto che una stessa funzione può essere realizzata attraverso strutture e processi radicalmente differenti. Non sappiamo ancora fino a che punto questa analogia possa essere estesa alle capacità

cognitive, ma proprio per scoprirlo dovremmo imparare a sostituire alla domanda «lo fa come noi?» una domanda apparentemente più semplice e molto più difficile: «che cosa sta facendo?».

La differenza fra le due formulazioni è enorme, perché nella seconda l'essere umano smette di essere il criterio incorporato nella definizione.

8. La paura di perdere l'esclusiva

È forse in questo punto che compare la paura, ma non necessariamente quella, molto più evidente, di una macchina che possa sostituirci nel lavoro, controllarci o persino diventare pericolosa. Questi sono problemi diversi, alcuni legati a rischi già osservabili e altri a scenari ancora ipotetici.

Potrebbe esistere una paura più sottile, legata alla possibilità di perdere l'esclusiva attraverso la quale abbiamo costruito una parte della nostra identità.

Se scopriamo forme di cultura in altri animali, la cultura non basta più a stabilire una frontiera assoluta; se un sistema artificiale produce configurazioni nuove, la produzione di novità non basta; se può pianificare, apprendere, utilizzare strumenti e modificare il proprio comportamento, dobbiamo domandarci quali di queste capacità descrivano davvero una differenza qualitativa e quali, invece, abbiamo semplicemente considerato esclusive finché non le abbiamo osservate altrove.

Non ne consegue affatto che non esistano differenze fondamentali fra un essere umano, un altro animale e un sistema artificiale. Significa soltanto che la differenza non implica necessariamente una gerarchia e che, soprattutto, non possiamo stabilire la gerarchia prima di avere definito la differenza.

Forse una parte della paura consiste proprio in questo: non nell'impossibilità di continuare a dire «siamo differenti», ma nella possibilità di non poter più tradurre automaticamente quella differenza in «siamo qualcosa di più».

9. La coscienza come frontiera epistemologica

Fin qui abbiamo parlato di capacità osservabili. Esiste però una frontiera più difficile: l'esperienza soggettiva.

La coscienza sembra allora rappresentare una frontiera particolarmente resistente. Possedere un'esperienza soggettiva, avere qualcosa che significa essere se stessi, potrebbe costituire una differenza fondamentale, e le conoscenze disponibili non ci autorizzano a trattare con leggerezza una questione tanto difficile.

Eppure proprio la coscienza possiede una caratteristica epistemologicamente singolare: non possiamo osservare direttamente neppure quella degli altri esseri umani, ma la inferiamo attraverso una combinazione di comportamento, struttura, somiglianza, testimonianza e teoria.

La domanda, quindi, non dovrebbe essere soltanto se una macchina possa essere cosciente, ma anche quali evidenze saremmo disposti ad accettare qualora una qualche forma di esperienza comparisse in un sistema radicalmente diverso da noi. Gli indicatori non dovrebbero necessariamente essere identici a quelli che utilizziamo per sistemi biologici, perché strutture differenti potrebbero richiedere criteri differenti; ciò che conta è che la scelta di quei criteri sia esplicita, teoricamente giustificata e non costruita in modo da garantire in anticipo il risultato che vogliamo ottenere. Se la risposta fosse che nessuna evidenza potrebbe essere sufficiente, non avremmo dimostrato che la coscienza appartiene esclusivamente all'essere umano: lo avremmo stabilito attraverso la definizione utilizzata per cercarla.

Questo non ci dice nulla sulla coscienza degli attuali sistemi di intelligenza artificiale, ma ci dice molto sul metodo con cui dovremmo affrontare la domanda.

10. Internet come precedente

Mentre cerchiamo di capire che cosa stia diventando l'intelligenza artificiale, però, dall'altra parte della relazione sta accadendo qualcosa che merita di essere osservato separatamente: stiamo cambiando noi.

Anche questo non comincia con l'intelligenza artificiale. L'essere umano costruisce continuamente ambienti nei quali altri esseri umani imparano successivamente a vivere, ereditando non soltanto geni ma lingue, tecniche, istituzioni, conoscenze e un mondo fisico già profondamente modificato dalle generazioni precedenti. Una strada, una città, una rete elettrica, un acquedotto e Internet sono artificiali, ma costituiscono contemporaneamente parti dell'ambiente ordinario nel quale l'essere umano contemporaneo realizza la propria esistenza.

L'artificio, in questo senso, non è semplicemente qualcosa che si aggiunge dall'esterno a un ipotetico uomo naturale: è diventato una componente della nicchia attraverso cui la nostra specie vive.

Internet rende particolarmente evidente il processo. Non ne avevamo bisogno prima che esistesse, almeno non nel senso in cui possiamo avere bisogno di acqua, cibo o condizioni ambientali compatibili con la vita; lo abbiamo però costruito, adottato e progressivamente utilizzato per trasformare il commercio, il lavoro, l'informazione, le relazioni, le amministrazioni, la memoria e le comunicazioni.

Oggi Homo sapiens potrebbe naturalmente sopravvivere senza Internet, ma non potremmo spegnere la rete domani e continuare contemporaneamente a vivere come viviamo oggi, perché abbiamo costruito una parte crescente del nostro ambiente sociale e materiale presupponendo la sua esistenza.

Questo ci mostra qualcosa che rischia di sfuggire quando raccontiamo la storia della tecnica come una semplice successione di strumenti inventati per soddisfare bisogni già esistenti: il bisogno può venire dopo l'invenzione. Una possibilità tecnica può produrre un'invenzione, l'invenzione può generare nuovi usi, gli usi possono

trasformarsi in abitudini e le abitudini possono diventare infrastrutture intorno alle quali riorganizziamo progressivamente l'ambiente. Quando questo accade, ciò che inizialmente era facoltativo può diventare una condizione della normalità che abbiamo costruito.

11. Due domande distinte

L'intelligenza artificiale potrebbe seguire un percorso simile, anche se non abbiamo alcun diritto di considerarlo inevitabile. Se continuerà a entrare nella ricerca, nel lavoro, nell'educazione, nella medicina, nella produzione, nelle amministrazioni e negli strumenti quotidiani, potrebbe arrivare una generazione che non avrà conosciuto il mondo precedente e che, proprio per questo, non percepirà AI come noi la percepiamo oggi.

Per chi è nato in un mondo già connesso, Internet non è innanzitutto una rivoluzione tecnologica: è una proprietà ordinaria dell'ambiente. Allo stesso modo, un essere umano che crescesse in un ambiente nel quale sistemi artificiali capaci di interagire, generare, assistere e agire fossero ubiqui svilupperebbe conoscenze, abitudini e aspettative in presenza di qualcosa che per noi rappresenta ancora una novità storica.

Qui, però, è importante non commettere l'errore opposto e fondere le due trasformazioni in una sola. Ci sono almeno due domande che devono poter essere osservate indipendentemente:

- Che cosa sta diventando l'intelligenza artificiale? Quali proprietà sta acquisendo, quali limiti conserva e quali categorie sono adeguate a descriverla senza trasformarla preventivamente in una copia incompleta dell'uomo?
- Che cosa sta diventando l'essere umano? Quali capacità esercita maggiormente in un ambiente modellato dall'IA, quali delega, quali sviluppa diversamente e quali nuovi bisogni produce?

Soltanto dopo aver mantenuto distinte queste due traiettorie possiamo osservare una terza cosa: ciò che nasce dalla loro relazione.

12. La causalità circolare

È a questo punto che la relazione fra le due traiettorie diventa circolare. L'uomo modifica i sistemi artificiali attraverso obiettivi, dati, selezioni, feedback, istituzioni e scelte, mentre quei sistemi possono a loro volta modificare comportamenti, decisioni, conoscenze e l'ambiente nel quale gli esseri umani successivi si formeranno.

Parlare di relazione reciproca non significa attribuire ai due termini la stessa natura, la stessa autonomia, lo stesso tipo di agency o la stessa posizione nel processo; significa semplicemente riconoscere che la causalità può circolare senza per questo diventare simmetrica. Un circuito di influenze, infatti, non basta da solo a dimostrare intenzionalità, soggettività o autonomia forte in ciascuno degli elementi che vi

partecipano.

A questo punto la domanda dalla quale eravamo partiti mostra finalmente il proprio limite. Sappiamo chi ha addestrato tecnicamente i modelli di intelligenza artificiale, ma non era soltanto questo che volevamo sapere. Volevamo capire come qualcosa che abbiamo prodotto attraverso la nostra conoscenza possa diventare occasione e causa di conoscenza che prima non possedevamo, come uno strumento possa modificare chi lo utilizza e come ciò che viene creato possa diventare parte dell'ambiente del proprio creatore; volevamo capire, più in generale, come qualcosa che nasce come effetto possa successivamente entrare nella catena delle cause.

13. Conclusione: dalla domanda piccola alla domanda grande

Il castoro non ha bisogno di volerci insegnare qualcosa perché possiamo imparare osservando la sua diga, così come l'intelligenza artificiale non ha bisogno di essere umana perché possiamo imparare osservando ciò che produce e ciò che accade quando interagiamo con essa.

Potremmo persino scoprire che non ha bisogno di assomigliarci per costringerci a osservare meglio noi stessi.

Abbiamo costruito l'intelligenza artificiale a partire dalle tracce di ciò che sapevamo e ora, attraverso ciò che abbiamo costruito, stiamo incontrando continuamente ciò che non sapevamo di non sapere: non soltanto sull'intelligenza artificiale, ma sui concetti stessi di intelligenza, creatività, cultura, autonomia e natura attraverso i quali abbiamo cercato di comprenderla; e, inevitabilmente, sull'essere umano che quei concetti ha formulato.

Forse, allora, la domanda non era sbagliata. Era soltanto troppo piccola.

Chi sta addestrando chi?

Potremmo dover cominciare chiedendoci:

Chi sta trasformando chi?

E questa volta, forse, dovremmo accettare che osservare significhi anche non avere ancora una risposta.

Riferimenti e approfondimenti

Il saggio nasce come riflessione argomentativa, non come rassegna accademica. I lavori seguenti documentano alcuni dei fenomeni empirici richiamati nel testo e includono anche una lettura critica utile a delimitare la portata di alcune interpretazioni.

van Leeuwen, E. J. C., DeTroy, S. E., Haun, D. B. M. et al. (2024). "Chimpanzees use social information to acquire a skill they fail to innovate". *Nature Human Behaviour*, 8, 891–902. DOI: 10.1038/s41562-024-01836-5.

- Bridges, A. D., Royka, A., Wilson, T. et al. (2024). "Bumblebees socially learn behaviour too complex to innovate alone". *Nature*, 627, 572-578. DOI: 10.1038/s41586-024-07126-4.
- Morgan, T. J. H., & Feldman, M. W. (2025). "Human culture is uniquely open-ended rather than uniquely cumulative". *Nature Human Behaviour*, 9, 28-42. DOI: 10.1038/s41562-024-02035-y.
- Santoro, E., & Monin, B. (2023). "The AI Effect: People rate distinctively human attributes as more essential to being human after learning about artificial intelligence advances". *Journal of Experimental Social Psychology*, 107, 104464. DOI: 10.1016/j.jesp.2023.104464.
- Millet, K., Buehler, F., Du, G., & Kokkoris, M. D. (2023). "Defending humankind: Anthropocentric bias in the appreciation of AI art". *Computers in Human Behavior*, 143, 107707. DOI: 10.1016/j.chb.2023.107707.
- Odling-Smee, J., Erwin, D. H., Palkovacs, E. P., Feldman, M. W., & Laland, K. N. (2013). "Niche Construction Theory: A Practical Guide for Ecologists". *The Quarterly Review of Biology*, 88(1), 3-28. DOI: 10.1086/669266.
- Scott-Phillips, T. C., Laland, K. N., Shuker, D. M., Dickins, T. E., & West, S. A. (2014). "The niche construction perspective: a critical appraisal". *Evolution*, 68(5), 1231-1243. DOI: 10.1111/evo.12332.
- Pedreschi, D., Pappalardo, L., Ferragina, E. et al. (2025). "Human-AI coevolution". *Artificial Intelligence*, 339, 104244. DOI: 10.1016/j.artint.2024.104244.
- Butlin, P., Long, R., Bayne, T. et al. (2026). "Identifying indicators of consciousness in AI systems". *Trends in Cognitive Sciences*, 30(6), 488-501. DOI: 10.1016/j.tics.2025.10.011.

Edizione online: <https://inchiostrati.it/chi-sta-addestrando-chi/>

MOVIMENTO 03

L'officina, il testo e la firma

Produrre parole non significa ancora assumere una parola: l'autorialità emerge nella selezione, nella firma, nella pubblicazione e nella responsabilità.

1. Il doppio standard

C'è un cortocircuito di fondo nel modo in cui il nostro tempo decide chi sia l'autore di un testo creato con l'IA. Quando un testo prodotto con l'intervento di un sistema artificiale si rivela efficace, brillante o profondo, il merito viene immediatamente azzerato con una formula sbrigativa: «L'ha fatto la macchina». Nel momento esatto in cui quel medesimo testo fallisce, risulta banale, contiene un errore o produce conseguenze indesiderate, la macchina perde invece ogni prerogativa autoriale e la responsabilità ricade interamente sull'uomo, accusato di non avere saputo controllare lo strumento.

La macchina viene dunque promossa a coautrice quando serve a svalutare il risultato e retrocessa a semplice utensile quando occorre trovare un responsabile. È una trappola logica e morale costruita in modo tale che l'essere umano non possa uscirne: se l'opera riesce, non è merito suo; se fallisce, la colpa gli appartiene completamente.

2. La Lettera 22 e il modello

Questo doppio standard nasce da una confusione più profonda fra il modo in cui un testo viene prodotto e l'atto attraverso il quale quel testo diventa un'opera. Il metodo cambia, i ferri del mestiere si evolvono, la tastiera e lo schermo sostituiscono la carta e la penna, mentre l'interlocutore artificiale entra nello spazio che prima sembrava occupato soltanto dal silenzio. Eppure non abbiamo mai fatto dipendere l'autorialità dalla purezza degli strumenti impiegati. Nessuno, davanti a un dattiloscritto, si è mai sognato di osservare: «Sì, ma lo ha scritto una Olivetti Lettera 22».

Naturalmente un modello generativo non è una Lettera 22. La macchina da scrivere imprimeva parole già scelte, mentre il modello può proporre formulazioni, immagini, connessioni e perfino direzioni concettuali che chi scrive non aveva previsto. La differenza fra i due strumenti è reale; negarla renderebbe più fragile qualsiasi ragionamento successivo. Ma riconoscere una maggiore capacità di intervento nella produzione linguistica non significa dimostrare l'esistenza di un nuovo autore. Contribuire alla formazione di un testo e assumerne l'autorialità non sono la stessa cosa.

3. Dalla produzione all'opera

La formula «testo scritto dall'intelligenza artificiale» è quindi profondamente ingannevole. Confonde il luogo in cui sono state prodotte le frasi con quello in cui il testo è diventato un'opera. Il modello linguistico opera all'interno di un campo sterminato di possibilità, ma lo attraversa senza conoscere la rinuncia, senza preferire veramente una soluzione alle altre e senza sapere perché una frase dovrebbe essere conservata mentre altre cento vengono scartate. Può generare formulazioni differenti, ma nessuna di esse gli appare più necessaria, più vera o più degna di essere pronunciata delle altre.

La macchina può continuare a produrre perché, per essa, non esiste il momento nel quale la ricerca debba concludersi. È l'essere umano a interrompere la proliferazione delle possibilità. È lui a riconoscere una forma, a decidere che il testo ha raggiunto il proprio limite e che proprio quella parola, non una delle innumerevoli alternative ancora generabili, debba essere consegnata agli altri.

4. Una parola assunta

La domanda sensata, allora, non è: «Chi ha prodotto materialmente ogni parola?». Se adottassimo questo criterio, dovremmo distribuire l'autorialità fra dizionari, correttori automatici, editor, insegnanti, conversazioni, libri letti e formule linguistiche ereditate. Ogni scrittore è un sedimento di voci ricevute, lezioni scolastiche, modi di dire, correzioni, imitazioni e influenze spesso dimenticate. Se pretendessimo di rintracciare l'inventore e il proprietario di ogni tessera, il concetto stesso di autore crollerebbe davanti a qualsiasi testo della storia umana.

Eppure non abbiamo mai dedotto da questa moltitudine di apporti l'inesistenza dell'autore. Abbiamo sempre saputo distinguere, anche senza formularlo apertamente, fra ciò che ha reso possibile un testo e chi ha deciso di assumerlo nella sua forma compiuta. La vera domanda è dunque un'altra: chi ha riconosciuto proprio quel testo? Chi gli ha attribuito un significato e una destinazione? Chi ha deciso che dovesse diventare pubblico? Chi accetta di risponderne?

Autore è chi trasforma una produzione possibile in una parola assunta.

5. Mettere al mondo

Questo passaggio non richiede un eroismo teatrale né il sacrificio romantico di chi lotta completamente solo contro il foglio bianco. Somiglia piuttosto a un atto di nascita. Quando diventiamo padri o madri, facciamo entrare nel mondo una nuova esistenza. Non l'abbiamo costruita consapevolmente in ogni sua parte e non saremo in grado di determinarne l'intero destino; eppure siamo responsabili della soglia che le abbiamo fatto attraversare.

Quando premiamo il tasto «Pubblica», accade qualcosa di analogo. Introduciamo nel mondo una parola che abbiamo riconosciuto, scelto e assunto. Da quel momento essa

comincia una vita che non ci appartiene più completamente: verrà interpretata, fraintesa, contestata, ricordata o dimenticata; incontrerà esperienze che non conosciamo e potrà produrre conseguenze che non avevamo previsto. Pubblicare significa accettare questa separazione, esponendosi attraverso qualcosa che, proprio perché è stato lasciato andare, non può più essere interamente controllato.

È in quel momento che l'intelligenza artificiale scompare dal problema autoriale. Non scompare dalla genealogia del testo e non occorre fingere che non abbia partecipato alla sua formazione. Scompare dalla relazione che comincia dopo la consegna. Non attende il giudizio dei lettori, non viene ferita da un'accusa, non difende un'intenzione, non si pente di una frase e non si riconosce trasformata dagli effetti dell'opera. Può avere partecipato alla gestazione, ma non compie il gesto della nascita. Genera, ma non mette al mondo, perché mettere al mondo significa esporsi attraverso ciò che si lascia andare e accettare di farsi carico, da soli, delle sue conseguenze.

6. La solitudine che ritorna

La solitudine, allora, non è la condizione necessaria dell'atto creativo e neppure il certificato della sua autenticità. Si può scrivere nel silenzio o nel dialogo, con un editor, dentro una redazione o conversando con un sistema generativo. Pretendere che sia autore soltanto chi ha affrontato da solo il foglio bianco significherebbe costruire un altro «veramente»: non più veramente scritto dall'uomo, ma veramente scritto in solitudine.

Eppure la solitudine ritorna, soltanto che ritorna dopo. Nell'officina l'autore può essere stato accompagnato da molte voci, umane o artificiali; quando il testo viene consegnato al mondo, però, resta solo davanti alle sue conseguenze. L'intelligenza artificiale che lo ha aiutato a trovare una formulazione non può condividere l'esposizione, difendere l'intenzione dell'opera o portare il peso di ciò che essa produrrà. L'autore non è dunque tale perché è stato solo mentre scriveva, ma perché rimane solo quando viene il momento di rispondere.

7. Il dovere della trasparenza

Dentro questa distinzione appare altrettanto ridicola la pretesa di trasformare la trasparenza in un dovere morale generalizzato, imponendo di dichiarare l'uso dell'intelligenza artificiale come se fosse un ingrediente sospetto da segnalare sul retro di una confezione. Nessuno ha mai preteso che in calce a un romanzo comparisse l'elenco dei programmi di videoscrittura, dei correttori ortografici, dei dizionari di sinonimi o delle conversazioni che hanno contribuito alla sua nascita. Non perché tali strumenti e influenze non siano esistiti, ma perché il lettore riceve un'opera, non il verbale della sua fabbricazione.

Esistono certamente circostanze nelle quali il processo deve essere dichiarato. Se un esame vuole misurare una capacità individuale senza assistenza, se un concorso vieta espressamente l'uso di strumenti generativi, se un contratto prescrive determinate

modalità di lavorazione o se una ricerca scientifica deve documentare metodi riproducibili, nascondere l'intervento dell'intelligenza artificiale può costituire un inganno. In questi casi, però, il problema non nasce da una presunta impurità del testo, ma dalla violazione di un patto preciso.

Non si tratta soltanto di una distinzione teorica. L'articolo 50, paragrafo 4, dell'AI Act europeo prevede un obbligo di dichiarazione per i testi generati o manipolati dall'intelligenza artificiale quando siano pubblicati allo scopo di informare il pubblico su questioni di interesse pubblico; lo stesso articolo, tuttavia, esclude tale obbligo quando il contenuto sia stato sottoposto a revisione umana o a controllo editoriale e una persona fisica o giuridica ne detenga la responsabilità editoriale. La norma non risolve la questione filosofica dell'autorialità, ma riconosce una discriminante decisiva: la genealogia tecnica di un testo e l'assunzione pubblica della sua responsabilità non sono la stessa cosa.

Trasformare quelle circostanze particolari in un obbligo universale significa invece presumere che ogni opera realizzata con l'assistenza dell'intelligenza artificiale debba essere posta preventivamente sotto sospetto. Quando il criterio della purezza autoriale non riesce più a reggere, il bersaglio viene spostato sulla trasparenza. Prima si domanda: «Lo hai scritto veramente tu?». Quando diventa evidente che nessun autore produce da solo ogni elemento della propria opera, la domanda cambia: «Hai almeno dichiarato come lo hai scritto?». Poi il bersaglio si sposterà ancora: quante parole hai composto personalmente, quante proposte hai accolto, chi ha trovato il titolo, da dove proveniva l'idea iniziale?

Non è una richiesta neutrale di informazione. È la ricerca continua di una soglia oltre la quale poter diminuire il valore dell'opera.

La domanda corretta non è: «Hai usato l'intelligenza artificiale?», ma: «Mi stai ingannando su qualcosa che, in questo contesto, ho diritto di sapere?». Se il patto narrativo o saggistico è onesto, se i fatti non vengono falsificati e nessuna regola accettata viene violata, l'etichetta preventiva non tutela il lettore. Serve soltanto a separare ciò che temiamo non sia più «veramente umano» da ciò che continua a rassicurarci.

Se un saggio è rigoroso, se un'argomentazione è solida, se un romanzo funziona e risuona nell'animo di chi legge, sapere se l'autore abbia digitato ogni parola davanti a un foglio bianco, abbia accolto il suggerimento di un correttore o abbia dialogato con un modello generativo per trovare una sfumatura di senso non sposta di un millimetro il valore del patto tra chi scrive e chi legge. Quel patto non riguarda la purezza degli strumenti, ma la qualità e l'onestà di ciò che viene consegnato.

8. La firma e la responsabilità

La firma resta la soglia visibile di questa assunzione, purché non venga intesa come una semplice attribuzione tipografica. Anche un plagiatario può apporre il proprio nome sotto parole rubate, senza diventare per questo il loro autore. Firmare significa qualcosa di più: riconoscere il testo come espressione della propria intenzione, assumerne la forma definitiva e accettare che venga letto alla luce della propria voce, delle proprie scelte e della propria responsabilità.

Quando l'opera è compiuta, chiunque continui a cercare fra le sue righe l'origine meccanica di ogni parola finirà per accorgersi di una cosa semplice. L'officina è ormai chiusa, gli strumenti sono stati riposti e la macchina è svanita dalla scena. Sul banco degli imputati è rimasto soltanto l'uomo: non necessariamente colui dal quale proveniva ogni parola, ma colui attraverso il quale quella parola è entrata nel mondo.

Riferimenti e approfondimenti

Il saggio propone una tesi filosofica, non una ricostruzione esaustiva del diritto d'autore. I riferimenti seguenti collocano la distinzione fra produzione, funzione-autore, contributo umano e responsabilità editoriale nel dibattito teorico e normativo.

Barthes, R. (1968). "La mort de l'auteur", poi raccolto in *Le Bruissement de la langue*, Seuil, 1984.
Foucault, M. (1969). "Qu'est-ce qu'un auteur?", *Bulletin de la Société française de philosophie*, 63(3), 73-104. U.S. Copyright Office — Copyright and Artificial Intelligence, Part 2: Copyrightability (2025). Regolamento (UE) 2024/1689 — AI Act, articolo 50, paragrafo 4. Legge 23 settembre 2025, n. 132 — articolo 25, tutela delle opere create con l'ausilio dell'intelligenza artificiale.

Edizione online: <https://inchiostrati.it/lofficina-il-testo-e-la-firma/>

MOVIMENTO 04

L'ecosistema delle decisioni

Quando una decisione è assistita dall'IA, il punto non è difendere una scelta umana pura ma capire chi conserva il potere di comprendere, contestare e dire di no.

1. Il nuovo «veramente»

Una diagnosi viene suggerita da un sistema capace di riconoscere correlazioni che sfuggono all'occhio umano. Un programma ordina centinaia di precedenti e indica quale orientamento giuridico appare più pertinente. Un algoritmo valuta il rischio di concedere un prestito, stabilisce quali candidature meritino un esame più attento o suggerisce quale trattamento possa offrire maggiori probabilità di successo. Nella vita quotidiana, intanto, altri sistemi ci consigliano che cosa leggere, quale strada percorrere, quale prodotto acquistare, quali persone ascoltare e perfino quale formulazione usare per affrontare una conversazione difficile.

Ogni volta che accogliamo una di queste indicazioni ritorna la stessa domanda, accompagnata dal sospetto che l'ingresso della macchina abbia sottratto qualcosa all'essere umano: hai deciso veramente tu?

È un nuovo «veramente». Non riguarda più soltanto l'autorialità di un testo, ma il rapporto fra intelligenza artificiale e decisioni umane: non domanda a chi appartengano le parole, bensì a chi appartenga un'azione. Immaginiamo ancora che possa esistere una decisione umana pura, nata interamente all'interno di una coscienza isolata e non contaminata da strumenti, consigli o influenze esterne. Se una macchina ha partecipato alla scelta, temiamo che quella scelta non ci appartenga più completamente.

Eppure non abbiamo mai deciso nel vuoto.

Ogni decisione umana matura dentro un ecosistema. Un medico consulta esami, ascolta colleghi, segue protocolli, ricorda casi precedenti e tiene conto delle condizioni particolari della persona che ha davanti. Un giudice legge norme, interpreta precedenti, valuta testimonianze e si muove all'interno di procedure che delimitano il proprio potere. Anche la scelta apparentemente più intima nasce da una moltitudine di elementi: desideri, paure, abitudini, educazione, disponibilità economiche, aspettative familiari, esperienze passate e consigli ricevuti. Persino quando crediamo di avere deciso completamente da soli, dentro di noi parlano molte voci.

L'intelligenza artificiale non ha inventato questo ecosistema. Vi è entrata.

2. L'autorità del calcolo

La sua presenza, tuttavia, non è irrilevante. Un consiglio umano può essere discusso, ricondotto a una storia, interrogato nelle ragioni e negli interessi di chi lo esprime. La raccomandazione di un sistema arriva invece rivestita dell'autorità del calcolo. Può apparire neutrale proprio perché non mostra esitazione, e oggettiva proprio perché non possiede un volto. Quanto più il procedimento che l'ha prodotta è difficile da ricostruire, tanto più il risultato rischia di presentarsi come una conclusione alla quale non resti che adeguarsi.

Il problema precede di molto l'attuale stagione dell'intelligenza artificiale. Già nel 1997 Raja Parasuraman e Victor Riley, analizzando il rapporto fra esseri umani e automazione, distinguevano fra uso, uso improprio, mancato uso e abuso dei sistemi automatici. Il loro lavoro mostrava come la fiducia eccessiva possa indurre l'operatore a dipendere dall'automazione anche quando sarebbe necessario sorvegliarla o contraddirla. L'intelligenza artificiale generativa non inaugura questo rischio, ma lo estende a strumenti capaci di formulare raccomandazioni in un linguaggio che possiede l'apparenza della sicurezza e della competenza.

3. Delegare non significa abdicare

È qui che la domanda sull'autenticità della decisione può diventare utile, purché venga formulata diversamente. Non dovremmo chiederci se l'essere umano abbia deciso senza ricevere alcun aiuto, ma se abbia conservato il potere effettivo di comprendere, interrogare e rifiutare l'indicazione ricevuta.

Delegare, infatti, non significa necessariamente abdicare.

Deleghiamo continuamente parti dei nostri processi decisionali. Affidiamo a una calcolatrice operazioni che potremmo compiere a mano, a un navigatore la comparazione fra percorsi possibili, a un medico la valutazione di sintomi che non sappiamo interpretare, a un avvocato la conoscenza di norme che non possediamo. Nessuna di queste deleghe ci priva automaticamente della decisione finale. Può, al contrario, renderla più consapevole, perché amplia ciò che siamo in grado di vedere e riduce il peso di alcuni nostri limiti.

Anche un sistema artificiale può svolgere questa funzione. Può individuare una relazione trascurata, richiamare un precedente dimenticato, comparare una quantità di informazioni che nessun individuo potrebbe esaminare nello stesso tempo o costringerci a considerare un'ipotesi che avevamo escluso troppo in fretta. In questi casi non sostituisce il giudizio umano: ne modifica le condizioni e, talvolta, lo rende migliore.

4. L'essere umano cerimoniale

L'abdicazione comincia quando la raccomandazione smette di essere uno degli elementi della decisione e diventa la ragione per cui nessuno si sente più obbligato a decidere. Accade quando l'indicazione del sistema viene accolta perché «lo dice l'algoritmo», quando discostarsene richiede una giustificazione sproporzionata, quando chi dovrebbe valutarla non possiede gli strumenti per comprenderne i limiti o quando l'organizzazione ha già costruito intorno a essa un percorso dal quale è formalmente possibile uscire, ma praticamente impossibile farlo.

In quel momento l'essere umano può restare presente senza essere più davvero decisore. La sua firma sopravvive, il suo nome compare sull'atto e il suo dito preme il pulsante conclusivo, ma la sua funzione si è ridotta a ratificare una scelta maturata altrove.

Per questo la formula rassicurante secondo cui «alla fine decide sempre una persona» non basta. La presenza di un essere umano nell'ultimo passaggio non garantisce che vi sia stato giudizio, così come apporre una firma sotto un testo non trasforma automaticamente un plagiatore nel suo autore. Bisogna domandarsi se quella persona avesse accesso alle informazioni necessarie, se potesse interpretarle, se disponesse del tempo e dell'autorità per opporsi, e soprattutto se un eventuale dissenso avrebbe avuto la possibilità concreta di modificare l'esito.

È ciò che Filippo Santoni de Sio e Jeroen van den Hoven hanno definito «controllo umano significativo». La presenza dell'uomo acquista significato soltanto se il comportamento del sistema resta sensibile alle ragioni umane pertinenti e ai fatti rilevanti del contesto, e se i suoi esiti possono essere ricondotti lungo la catena della progettazione e dell'impiego ad almeno una persona capace di esercitare una responsabilità reale. Il controllo non è dunque una casella da spuntare alla fine del processo, ma una qualità dell'intero sistema.

5. Il parafulmine morale

Un essere umano posto nel circuito soltanto per fornire una copertura morale non è un decisore. È un parafulmine.

Madeleine Clare Elish ha descritto questa condizione con l'espressione *moral crumple zone*, «zona di deformazione morale». Come una parte dell'automobile assorbe l'urto per proteggere ciò che la circonda, così l'operatore più vicino al sistema può essere lasciato ad assorbire la responsabilità morale e giuridica del fallimento, anche quando il potere necessario a evitarlo era distribuito altrove.

Qui emerge un cortocircuito simile a quello che accompagna l'autorialità dei testi prodotti con l'assistenza dell'intelligenza artificiale. Quando una decisione algoritmica offre un risultato positivo, il successo viene attribuito all'efficienza e alla precisione del sistema. Quando produce un danno, si ricorda improvvisamente che la decisione finale spettava all'operatore. La macchina viene presentata come decisore quando

occorre celebrarne le capacità e ridotta a semplice supporto quando bisogna individuare qualcuno che risponda delle conseguenze.

L'essere umano diventa così il luogo nel quale viene depositata una responsabilità che non sempre corrisponde a un potere reale. Gli si chiede di rispondere di una scelta che forse non poteva comprendere, contestare o modificare, mentre l'organizzazione che ha selezionato il sistema, definito i dati, stabilito le procedure e imposto i tempi della decisione scompare sullo sfondo.

6. Distribuire il potere

È per questo che parlare di ecosistema delle decisioni non significa diluire la responsabilità fino a farla evaporare. Significa ricostruire la distribuzione effettiva del potere.

Una decisione assistita dall'intelligenza artificiale non nasce soltanto nel momento in cui qualcuno accoglie una raccomandazione. Comincia prima, quando viene stabilito quale problema debba essere tradotto in dati, quali elementi siano considerati rilevanti, quale obiettivo il sistema debba privilegiare e quale margine di errore venga ritenuto accettabile. Prosegue quando un'organizzazione decide di adottare quello strumento, determina chi possa contraddirlo e stabilisce che cosa accada quando il giudizio umano e quello algoritmico divergono. Soltanto alla fine arriva la persona chiamata ad approvare o respingere l'esito.

Se osserviamo esclusivamente l'ultimo gesto, trasformiamo l'ultimo essere umano della catena nel responsabile di tutto ciò che lo ha preceduto. Se guardiamo l'intero ecosistema, possiamo invece distinguere il contributo del modello, le scelte dei progettisti, le responsabilità dell'istituzione e il margine di autonomia conservato dall'operatore.

Nel campo sanitario questa responsabilità distribuita è stata riconosciuta anche dall'Organizzazione mondiale della sanità. Le sue indicazioni sull'intelligenza artificiale per la salute collocano etica e diritti umani al centro della progettazione, dell'adozione e dell'uso dei sistemi, chiamando alla responsabilità non soltanto i professionisti che li impiegano, ma tutti gli attori pubblici e privati che concorrono a svilupparli, introdurli e governarli. La responsabilità non nasce quindi nell'ultimo istante, quando il medico accoglie o respinge una raccomandazione: attraversa l'intera struttura che ha reso quella raccomandazione possibile e operativa.

7. Trasparenza e contestabilità

Questa distinzione diventa ancora più importante quando la decisione riguarda qualcuno che non partecipa al processo, ma ne subisce gli effetti. Una raccomandazione personale può condizionare chi la riceve; una decisione medica, legale, amministrativa o finanziaria può modificare concretamente la vita di un'altra persona. In questi casi non basta che qualcuno si dichiari responsabile alla fine. Chi è coinvolto deve poter comprendere da dove provenga la decisione, quali criteri l'abbiano orientata e davanti a chi possa contestarla.

Qui la trasparenza assume un significato diverso da quello che le viene attribuito nel dibattito sull'autorialità.

Sapere se uno scrittore abbia adoperato un correttore, consultato un dizionario o dialogato con un modello generativo può non spostare di un millimetro il valore del patto narrativo o saggistico. Se il testo è onesto, rigoroso e capace di risuonare nel lettore, l'inventario degli strumenti impiegati nella sua officina non aggiunge necessariamente nulla. L'etichetta «scritto con IA», quando non risponde a un accordo preciso, rischia anzi di funzionare soltanto come un marchio preventivo di sospetto.

Quando una decisione esercita invece un potere sulla vita altrui, conoscere il ruolo svolto dal sistema può essere essenziale. Non perché la presenza dell'intelligenza artificiale renda quella decisione meno pura o meno umana, ma perché ogni potere deve poter essere ricostruito, interrogato e contestato. La trasparenza non serve allora a denunciare un ingrediente tecnologico; serve a rendere visibile il percorso attraverso il quale una previsione è diventata una conseguenza.

Non abbiamo bisogno di sapere che una macchina è stata usata per poterla condannare in anticipo. Abbiamo bisogno di sapere quale autorità le sia stata concessa, quali possibilità abbia escluso, chi abbia avuto il potere di correggerla e chi abbia deciso di trasformare la sua indicazione in un atto.

Non si tratta soltanto di un criterio filosofico. L'articolo 14 del Regolamento europeo sull'intelligenza artificiale richiede, per i sistemi ad alto rischio, una supervisione affidata a persone capaci di comprenderne possibilità e limiti, sorvegliarne il funzionamento, riconoscere il rischio di un affidamento automatico e decidere di non usare il sistema oppure di disattenderne o rovesciarne il risultato, fino a interromperne il funzionamento. L'articolo 86 riconosce inoltre, nei casi previsti, alla persona interessata da una decisione basata sull'output di un sistema ad alto rischio il diritto di ottenere spiegazioni chiare e significative, quando quella decisione produca effetti giuridici o incida in modo analogo e rilevante e la persona ritenga di averne ricevuto un impatto negativo sulla salute, sulla sicurezza o sui diritti fondamentali. Le spiegazioni devono riguardare il ruolo svolto dall'intelligenza artificiale nella procedura e gli elementi principali della decisione. La presenza umana e la trasparenza acquistano dunque valore soltanto quando rendono il potere realmente comprensibile e contestabile.

8. Dalla selezione alla decisione

Decidere, infatti, non significa produrre autonomamente tutte le alternative. Significa stabilire quale alternativa debba diventare reale e assumere le conseguenze di quella trasformazione.

Un sistema può formulare una previsione, assegnare un punteggio, ordinare possibilità o perfino eseguire automaticamente un'azione. In senso tecnico possiamo certamente dire che ha operato una selezione: ha individuato un esito sulla base di determinate condizioni. Ma la selezione tecnica non coincide con una decisione morale, giuridica o politica. Il sistema non ha stabilito che quell'obiettivo meritasse di essere perseguito, non ha determinato quale rischio fosse giusto imporre a un'altra persona e non ha deciso autonomamente di ricevere il potere che gli è stato affidato.

Dietro ogni sistema che agisce esiste una decisione precedente: qualcuno ha stabilito che potesse agire, entro quali limiti, a vantaggio di chi e quali conseguenze fossero ritenute tollerabili. Anche quando l'automazione elimina l'intervento umano nell'ultimo istante, non elimina l'ecosistema di scelte umane che l'ha resa possibile.

La responsabilità, però, non può limitarsi al momento originario dell'adozione. Non basta dire che un'organizzazione ha deciso una volta per tutte di affidarsi a un sistema. Ogni delega deve poter essere riesaminata, perché gli strumenti cambiano, gli ambienti si trasformano e le conseguenze possono mostrare ciò che il progetto iniziale non aveva previsto. Delegare senza conservare la capacità di sorvegliare, interrompere e correggere significa consegnare il giudizio a un processo che continua a operare anche quando nessuno è più disposto a riconoscerlo come proprio.

9. Il paesaggio delle scelte

Questo vale anche per le decisioni personali, apparentemente meno drammatiche. Un sistema che ci consiglia un libro, una canzone o un percorso non ci obbliga a seguirlo, ma può modificare gradualmente il campo di ciò che siamo in grado di scegliere. Non decide soltanto quale alternativa mostrarci: contribuisce a stabilire quali alternative arriveranno davanti ai nostri occhi e quali resteranno invisibili. La nostra libertà non viene necessariamente cancellata; può però esercitarsi dentro un paesaggio già ordinato da altri.

Il problema, allora, non consiste nel difendere una decisione umana incontaminata, che non è mai esistita. Consiste nel preservare la possibilità di accorgersi dell'influenza, di uscire dal percorso suggerito e di domandare quali interessi abbiano costruito il paesaggio nel quale stiamo scegliendo.

L'autonomia non è assenza di condizionamenti. È capacità di riconoscerli senza considerarli inevitabili.

10. La possibilità di dire no

Anche qui il nostro «veramente» deve dunque essere spostato. Una decisione non è veramente umana perché nasce senza l'intervento di una macchina. Potrebbe essere irriflessiva, arbitraria, superstiziosa o interamente determinata dall'abitudine e restare comunque priva di qualsiasi assistenza tecnologica. Allo stesso modo, una decisione maturata attraverso il confronto con un sistema artificiale può essere profondamente umana se qualcuno ne comprende le ragioni, ne valuta i limiti, conserva la libertà di rifiutarla e accetta di risponderne.

Ciò che rende umana una decisione non è la solitudine dalla quale proviene, ma la responsabilità nella quale termina.

Delegare può allora diventare una forma matura del giudizio: riconoscere che non possiamo conoscere tutto, accogliere l'aiuto di strumenti più capaci in compiti determinati e servirci delle loro indicazioni senza trasformarle in oracoli. Abdicare, al contrario, significa smettere di domandare perché, rinunciare alla possibilità del dissenso e usare la macchina come luogo nel quale nascondere una scelta che nessuno vuole più assumere.

La differenza non passa fra decisioni compiute dall'uomo e decisioni compiute dall'intelligenza artificiale. Passa fra ecosistemi nei quali il potere resta visibile, contestabile e responsabile, ed ecosistemi nei quali ogni soggetto può indicare il successivo fino a quando, al centro della catena, non rimane altro che una risposta senza autore: «Lo ha stabilito il sistema».

Una possibilità calcolata diventa una decisione quando entra nel mondo e produce conseguenze. Da quel momento non è sufficiente conoscere chi abbia elaborato la previsione; dobbiamo sapere chi le abbia concesso autorità, chi avrebbe potuto fermarla e chi sia disposto a riconoscerla come propria.

Il problema non è stabilire se abbia deciso veramente l'uomo oppure la macchina. Il problema è scoprire se, dentro l'ecosistema che ha trasformato una previsione in un atto, esista ancora qualcuno che avrebbe potuto dire di no.

Riferimenti e approfondimenti

Il saggio è una riflessione argomentativa. I riferimenti seguenti documentano i concetti di dipendenza dall'automazione, controllo umano significativo, responsabilità distribuita, supervisione e diritto alla spiegazione richiamati nel testo.

Madeleine Clare Elish, *Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction*, «Engaging Science, Technology, and Society», vol. 5, 2019. Organizzazione mondiale della sanità, *Ethics and Governance of Artificial Intelligence for Health*, Ginevra, 2021. Raja Parasuraman e Victor Riley, *Humans and Automation: Use, Misuse, Disuse, Abuse*, «Human Factors», vol. 39, n. 2, 1997, pp. 230-253. Parlamento europeo e Consiglio dell'Unione europea, Regolamento (UE) 2024/1689 sull'intelligenza artificiale, in particolare artt. 14 e 86. Filippo Santoni de Sio e Jeroen van den Hoven, *Meaningful Human Control over Autonomous Systems: A Philosophical Account*,

«Frontiers in Robotics and AI», vol. 5, 2018, art. 15.

Edizione online: <https://inchiostrati.it/lecosistema-delle-decisioni/>

MOVIMENTO 05

Il soggetto disabitato

Che cosa resta dell'uomo quando conserva formalmente il posto dell'autore e del decisore ma rinuncia progressivamente a esercitare giudizio e responsabilità?

1. La macchina sul banco degli imputati

Il 7 settembre 2026 mi sono imbattuto in un titolo che annunciava l'allarme delle Nazioni Unite: l'intelligenza artificiale potrebbe rappresentare un rischio esistenziale per l'umanità. Tre giorni prima, l'Assemblea generale aveva approvato una risoluzione per incoraggiare l'uso di rappresentazioni cartografiche capaci di restituire più fedelmente le proporzioni dei continenti. Nel volgere di un fine settimana, la stessa organizzazione — attraverso organi diversi, naturalmente — aveva prima corretto la mappa del mondo e poi avvertito che avremmo potuto perderne l'originale.

La coincidenza mi ha fatto ridere. Non perché la sicurezza dei sistemi artificiali sia un argomento ridicolo, né perché le tecnologie potenti non possano produrre danni enormi, ma per la naturalezza con cui il soggetto umano sembrava già prepararsi a scomparire dalla frase. Volker Türk, Alto Commissario delle Nazioni Unite per i diritti umani, ha dichiarato di condividere le preoccupazioni di alcuni esponenti dell'industria sul possibile rischio esistenziale dell'IA avanzata e ha chiesto «garanzie ferree» prima che fosse troppo tardi. Nel medesimo intervento, tuttavia, ha osservato che un piccolo numero di uomini possiede un potere quasi illimitato sullo sviluppo dell'intelligenza artificiale.

La contraddizione è interamente visibile. Prima compaiono gli uomini, il loro potere e la sua concentrazione; subito dopo, quando occorre nominare la minaccia, gli uomini escono di scena e rimane l'IA. La macchina viene condotta sul banco degli imputati mentre progettisti, finanziatori, aziende, governi, eserciti e istituzioni possono accomodarsi fra il pubblico.

L'intelligenza artificiale non ha costruito i propri centri di calcolo, acquistato semiconduttori, raccolto investimenti, stipulato contratti militari o deciso di collegarsi alle infrastrutture critiche. Non ha stabilito che la velocità dovesse precedere la comprensione, che il profitto dovesse compensare l'incertezza o che la competizione geopolitica rendesse accettabili rischi altrimenti inammissibili. Se un giorno un sistema artificiale produrrà una catastrofe, potrà esserne l'esecutore operativo; l'autore della situazione nella quale quella catastrofe è diventata possibile resterà umano.

La risata, allora, ha lasciato emergere qualcosa di meno divertente. Per dichiarare la macchina responsabile del pericolo, l'uomo non si è limitato ad attribuirle una capacità che forse un giorno possiederà: ha dovuto sottrarre a se stesso una capacità che

possiede già. Nella frase «l'IA potrebbe minacciare l'umanità», tutti i verbi dell'azione sono passati alla macchina. L'uomo rimane soltanto come oggetto della minaccia, creatura passiva da proteggere dalle conseguenze di decisioni delle quali non appare più autore.

Non è ancora una risposta, ma è finalmente la domanda giusta: che genere di umanità stiamo tentando di salvare, se per assolverla dobbiamo prima descriverla come incapace di agire?

2. La domanda rovesciata

È attraverso questo slittamento che mi è tornato alla mente il titolo di Primo Levi, non per un'associazione generica fra pericolo e umanità, ma perché quel titolo costringe a domandare quali facoltà, relazioni e responsabilità rendano riconoscibile un uomo come soggetto. Nel titolo che ho appena letto, invece, l'uomo sopravvive soltanto come specie minacciata: biologicamente presente, grammaticalmente passivo, causalmente assente.

Se questo è un uomo custodisce tuttavia una domanda che non può essere spostata con leggerezza. Nel Lager la disumanizzazione non era una metafora culturale, una scelta di consumo o una conseguenza della comodità. Era un progetto imposto da altri uomini attraverso la fame, la violenza, la cancellazione del nome, la riduzione del corpo a forza di lavoro e la distruzione sistematica della dignità. Ogni paragone che attenuasse questa specificità sarebbe storicamente falso e moralmente intollerabile.

Non si tratta quindi di trasferire al presente l'esperienza narrata da Levi, né di stabilire equivalenze fra fenomeni che non ne possiedono alcuna. Si tratta di lasciarsi raggiungere dalla forma della sua domanda e di rovesciarne la direzione.

Levi chiedeva che cosa restasse dell'uomo quando altri uomini tentavano di sottrargli l'umanità. La domanda contemporanea è diversa: che cosa resta dell'uomo quando nessuno gli impedisce di esercitare le proprie facoltà e tuttavia egli rinuncia progressivamente a esercitarle? Che cosa accade quando non gli viene tolta la parola, ma egli accetta parole che non ha pensato; quando non gli viene proibito di giudicare, ma preferisce ricevere un giudizio già formato; quando non gli viene cancellato il nome, ma lo appone su decisioni delle quali non conosce più l'origine?

La prima è la domanda sulla resistenza dell'umano davanti alla violenza che vuole annientarlo. La seconda riguarda una rinuncia senza costrizione apparente, compiuta attraverso piccole cessioni quotidiane, ciascuna ragionevole se osservata isolatamente e devastante quando diventa un sistema.

È possibile che la disumanizzazione non abbia sempre bisogno di un carnefice riconoscibile? È possibile che l'uomo collabori alla propria assenza?

3. Dare un nome all'assenza

Occorreva allora dare un nome alla figura che stava emergendo. Non poteva essere semplicemente un soggetto assente, perché l'uomo continuava a comparire ovunque: firmava i documenti, approvava i risultati, premeva i pulsanti, pubblicava le parole e rivendicava i benefici. Non era neppure un soggetto cancellato, poiché nessuna forza esterna gli aveva necessariamente sottratto il posto. Il posto rimaneva occupato formalmente, ma non era più abitato.

L'immagine più precisa era quella di una casa che conserva il nome sulla porta, le luci accese e gli oggetti nelle stanze, ma dalla quale nessuno risponde. Le funzioni continuano: il testo viene pubblicato, la decisione eseguita, il contenuto condiviso e la responsabilità registrata. Quando però si cerca qualcuno capace di spiegare la parola, ricostruire il giudizio o rispondere della conseguenza, si incontra una sequenza di rinvii. L'uomo rimanda alla macchina, l'operatore alla procedura, l'istituzione al sistema, il sistema ai dati.

Chiamo soggetto disabitato questa permanenza della forma insieme al ritiro della presenza. Non è una categoria clinica e non afferma che una persona abbia cessato di essere umana. Descrive una posizione dalla quale il soggetto continua ad agire senza riconoscersi pienamente come origine, interprete e responsabile della propria azione.

La disumanizzazione volontaria comincia qui: non nella scomparsa dell'uomo, ma nella sua scelta di occupare il posto dell'autore, del cittadino, del professionista o del decisore senza esercitare le facoltà che rendono quel posto realmente umano. Per comprenderne la possibilità non dobbiamo cominciare dalla macchina. Dobbiamo tornare a una forma molto più antica di rinuncia.

4. L'uscita mancata dalla minorità

Nel 1784 Immanuel Kant definì l'Illuminismo come l'uscita dell'uomo da uno stato di minorità del quale egli stesso è responsabile. La minorità, nella sua formulazione, non dipende necessariamente dalla mancanza d'intelligenza, ma dall'incapacità o dalla rinuncia a usare il proprio intelletto senza la guida di un altro. Il celebre invito sapere aude non significava soltanto accumulare conoscenze: chiedeva il coraggio di servirsi del proprio giudizio.

Quella pagina sembra descrivere con sorprendente precisione un problema che oggi affidiamo troppo facilmente alla tecnologia. L'uomo non rinuncia a pensare perché la macchina glielo impedisca; può rinunciarvi perché pensare richiede tempo, espone all'errore, obbliga a scegliere e rende responsabili delle conseguenze. La guida esterna offre invece rapidità e riparo. Dice che cosa è importante, quali parole usare, quale posizione assumere, quale notizia condividere e perfino quale emozione provare.

Nel tempo di Kant la guida poteva essere il libro, il precettore, l'autorità politica o religiosa. Nel nostro può assumere la forma di un sistema di raccomandazione, di una risposta generata, di una classifica, di un indicatore di popolarità o di un titolo

ottimizzato per ottenere una reazione. La tecnologia cambia la scala e la velocità della dipendenza, ma non inventa il desiderio di essere sollevati dalla fatica dell'autonomia.

Qui si trova il primo limite dell'abbozzo dal quale nasce questo saggio. Non è vero che affidare una parte del lavoro a una macchina equivalga necessariamente a delegare l'intenzione o a perdere la propria posizione nel mondo. Scrittura, calcolo, archivi, mappe e strumenti di misura sono da secoli forme di estensione delle capacità umane. L'uomo non diventa meno umano ogni volta che evita una fatica; spesso diventa capace di pensieri che senza quella mediazione non avrebbe potuto formulare.

Il confine non passa dunque fra ciò che viene compiuto dentro il corpo e ciò che viene affidato a uno strumento. Passa fra l'estensione e l'abdicazione: fra uno strumento che amplia la possibilità di giudicare e uno strumento usato per evitare il giudizio; fra una collaborazione della quale il soggetto continua a rispondere e una sostituzione dietro la quale il soggetto si nasconde.

5. Delegare non significa scomparire

Le scienze cognitive chiamano cognitive offloading l'uso di azioni o supporti esterni per ridurre il carico richiesto da un compito mentale. Annotare un appuntamento, servirsi di una calcolatrice, seguire una mappa o organizzare informazioni in un archivio sono forme ordinarie di delega cognitiva. Evan Risko e Sam Gilbert hanno mostrato come questa pratica non costituisca un'anomalia introdotta dall'IA, ma una componente diffusa del nostro modo di pensare e agire.

Il problema, pertanto, non è conservare ogni informazione nella memoria biologica o svolgere personalmente ogni passaggio di un processo. Una simile purezza cognitiva non è mai esistita. Il problema nasce quando deleghiamo anche le operazioni che ci permetterebbero di comprendere ciò che stiamo facendo e, soprattutto, quando perdiamo la capacità di riconoscere quali operazioni abbiamo delegato.

Una persona può usare l'intelligenza artificiale per confrontare ipotesi, sottoporre un'idea a obiezioni, trovare fonti, chiarire un passaggio o rendere più rigoroso un testo. In questi casi la macchina partecipa al processo, ma non elimina necessariamente il soggetto. Può anzi costringerlo a precisare ciò che pensa, a rifiutare formulazioni estranee e ad assumere più consapevolmente la responsabilità del risultato. È quanto abbiamo già osservato nell'officina del testo: la provenienza materiale delle parole non decide da sola chi sia l'autore; contano la direzione, la scelta, la revisione e la disponibilità a rispondere.

La stessa tecnologia diventa però un dispositivo di abdicazione quando il risultato viene accettato perché plausibile, pubblicato perché scorrevole e difeso soltanto finché nessuno domanda a chi lo firma di spiegarlo. Il testo non è orfano perché alcune frasi sono state generate artificialmente. Diventa orfano quando nessuno lo riconosce come proprio nel senso più esigente del termine: nessuno sa sostenerlo, correggerlo o assumersi il rischio di averlo detto.

Non è la macchina a disabitare il soggetto. È il soggetto che può usare la macchina per assentarsi lasciando accesa la luce.

6. Quando il prodotto migliora e la facoltà si indebolisce

La distinzione fra estensione e abdicazione non è soltanto filosofica. Alcune ricerche recenti mostrano che gli strumenti generativi possono migliorare il risultato immediato senza produrre un miglioramento equivalente nelle capacità di chi li usa.

In un esperimento sul campo condotto con quasi mille studenti di matematica, Hamsa Bastani e colleghi hanno confrontato l'uso di un'interfaccia generativa generalista con quello di un tutor progettato per fornire sostegno senza sostituire il processo di apprendimento. Gli studenti potevano ottenere risultati migliori durante gli esercizi con entrambi gli strumenti, ma coloro che avevano usato liberamente il sistema generalista peggioravano quando l'accesso veniva rimosso. Le protezioni pedagogiche riducevano questo effetto. La conclusione non è che l'IA impedisca di imparare, bensì che il modo in cui viene progettata e utilizzata può separare la prestazione dall'acquisizione della capacità.

Uno studio presentato alla conferenza CHI 2025 ha raccolto 936 esempi d'uso raccontati da 319 lavoratori della conoscenza. I risultati sono basati su autovalutazioni e non consentono di misurare direttamente un declino cognitivo, ma indicano una relazione significativa: una maggiore fiducia nelle capacità dell'IA era associata a un minore sforzo dichiarato di pensiero critico, mentre una maggiore fiducia nella propria competenza favoriva la verifica e il controllo. Gli autori osservano inoltre che l'attività cognitiva non scompare necessariamente, ma si sposta dalla produzione alla verifica, dall'esecuzione alla supervisione e dall'analisi all'integrazione delle risposte.

Questo spostamento può essere fecondo se la verifica è reale e chi la compie possiede le competenze necessarie. Diventa invece una finzione quando l'essere umano conserva formalmente l'ultima parola ma non sa più valutare ciò che approva. Il pulsante resta umano; il giudizio non lo è più.

La disumanizzazione volontaria non coincide quindi con il semplice risparmio di energia mentale. Comincia quando il risparmio diventa perdita di esercizio, quando la perdita di esercizio diventa incapacità e quando l'incapacità viene nascosta dietro l'efficienza del prodotto. Il documento appare migliore, la decisione arriva prima, il compito risulta completato; soltanto il soggetto è diventato più debole.

7. L'ambiente che premia la rinuncia

Parlare di scelta volontaria non significa immaginare un individuo isolato che, in condizioni neutrali, decide liberamente di impoverirsi. Le scelte avvengono dentro ambienti che rendono alcune condotte più facili, visibili e vantaggiose di altre. Se velocità, quantità e reazione sono le sole misure del valore, la riflessione diventa uno svantaggio competitivo.

La ricerca sulla comunicazione digitale offre un esempio particolarmente chiaro. Un'analisi pubblicata su *Nature Human Behaviour*, basata su circa 105.000 varianti di titoli, 370 milioni di visualizzazioni e 5,7 milioni di clic, ha rilevato che l'aggiunta di parole negative aumentava la probabilità di apertura della notizia. In media, per un titolo di lunghezza ordinaria, ogni parola negativa aggiuntiva corrispondeva a un incremento del tasso di clic del 2,3 per cento.

Un altro studio, pubblicato su *Science* nel 2018 e condotto su circa 126.000 catene di notizie diffuse su Twitter, ha mostrato che le informazioni false viaggiavano più lontano, più rapidamente e più profondamente di quelle vere. Il dato più importante per il nostro discorso è che la differenza non era spiegata principalmente dai bot: erano gli esseri umani a preferire e diffondere più facilmente ciò che appariva nuovo e sorprendente.

Anche l'indignazione viene appresa. William Brady e colleghi, studiando milioni di messaggi e conducendo esperimenti controllati, hanno osservato che il riscontro sociale positivo aumenta la probabilità che gli utenti esprimano nuovamente indignazione in futuro. La piattaforma non inserisce dall'esterno un'emozione inesistente; seleziona, misura e ricompensa alcune forme espressive, insegnando progressivamente agli utenti quali versioni di se stessi ottengono più attenzione.

È qui che nasce il «veramente?» che attraversa questo percorso. La dichiarazione prudente diventa allarme, la possibilità diventa quasi imminenza, la richiesta di regolazione diventa annuncio della fine. Il lettore non è invitato a comprendere, ma a superare una soglia di stupore. Reagisce, condivide, conferma che la formula funziona; il sistema registra la risposta e ne propone una versione ancora più intensa.

L'uomo abbassa il livello culturale del messaggio per adattarlo alla macchina. La macchina impara che quel livello produce attenzione e induce l'uomo ad abbassarlo ancora. Nessuno dei due termini, isolato, spiega il risultato. È la relazione a trasformarli, ma soltanto uno dei due può accorgersene e scegliere di interromperla.

8. Le parole lasciate accese

Il soggetto disabitato non è dunque un uomo silenzioso. Al contrario, può produrre parole in quantità illimitata. Pubblica, commenta, approva, reagisce e occupa continuamente lo spazio visibile; ciò che viene meno non è l'espressione, ma la presenza di qualcuno capace di rispondere di ciò che esprime.

Disabitare una parola significa usarla senza attraversarla. Significa adottare una posizione perché coincide con quella del gruppo, ripetere un giudizio perché ha già ricevuto consenso, condividere una notizia perché suscita l'emozione giusta e firmare un testo perché restituisce un'immagine desiderabile di noi. Il linguaggio continua a funzionare, ma non testimonia più un incontro fra esperienza, pensiero e responsabilità.

L'abbozzo originario definiva questo fenomeno come esternalizzazione dell'intenzione. L'espressione coglie qualcosa di vero, purché non venga trasformata in una condanna generale degli strumenti generativi. L'intenzione non viene trasferita automaticamente ogni volta che chiediamo aiuto a una macchina; viene abbandonata quando smettiamo di determinare lo scopo, di valutare i mezzi e di riconoscere le conseguenze.

In quel momento l'uomo non diventa una macchina. Sarebbe un'offesa alle macchine, che almeno eseguono coerentemente la funzione per la quale sono state costruite. Diventa piuttosto il punto vuoto attraverso il quale passano decisioni, testi e comportamenti che nessuno vuole più dichiarare propri.

La disumanizzazione descritta qui non consiste nella perdita biologica o giuridica dello statuto umano. È una riduzione pratica: l'uomo conserva le facoltà, ma organizza la propria vita come se non dovesse esercitarle. Conserva la libertà, ma la usa prevalentemente per scegliere chi o che cosa sceglierà al suo posto.

9. Somigliare all'immagine che attribuiamo alla macchina

Nella ricerca psicologica, Nick Haslam ha distinto due forme di disumanizzazione. Una nega agli altri caratteristiche considerate specificamente umane e li avvicina agli animali; l'altra nega tratti associati alla natura umana e rappresenta le persone come oggetti, strumenti o automi: fredde, passive, intercambiabili, prive d'interiorità e d'iniziativa.

Questa teoria riguarda innanzitutto il modo in cui percepiamo e trattiamo gli altri, non autorizza a diagnosticare un'intera epoca come clinicamente disumanizzata. Esiste anche una letteratura sulla self-dehumanization, ma il termine indica in senso tecnico la percezione di sé come meno pienamente umano, spesso in seguito a esclusione, oggettivazione o alla consapevolezza di avere agito immoralmente. Il nostro uso è più ampio e filosofico: non descrive necessariamente ciò che le persone sentono di essere, bensì il tipo di condotta che normalizzano.

La cautela terminologica rende però ancora più visibile il paradosso. Mentre attribuiamo alle macchine intenzioni, desideri e volontà sempre più umane, organizziamo gli esseri umani secondo il modello dell'automa: prevedibili, misurabili, sostituibili, guidati da segnali esterni e valutati attraverso la produttività delle loro risposte.

Temiamo che l'intelligenza artificiale diventi simile all'uomo, ma intanto chiediamo all'uomo di diventare simile all'immagine più povera che possiede della macchina. Deve reagire rapidamente, eliminare l'ambiguità, ottimizzare la prestazione, ridurre il tempo improduttivo del dubbio e rendere ogni comportamento leggibile al sistema. L'interiorità, non essendo immediatamente misurabile, appare come un ritardo.

La trasformazione non viene imposta soltanto dall'alto. Ciascuno impara a presentarsi nella forma più facilmente elaborabile: una preferenza, un'opinione, una reazione, un profilo. La persona intera sarebbe troppo contraddittoria; l'utente deve essere classificabile.

10. La macchina come alibi morale

Albert Bandura ha descritto il disimpegno morale come l'insieme dei processi attraverso i quali gli individui separano le proprie condotte dagli standard morali che normalmente le regolerebbero. Fra questi processi compaiono lo spostamento e la diffusione della responsabilità: il soggetto riduce il peso della propria partecipazione attribuendo la decisione a un'autorità, a un'organizzazione o a una catena nella quale nessuno sembra possedere il controllo completo.

I sistemi algoritmici offrono a questo meccanismo un alibi particolarmente efficace. «Lo ha deciso l'algoritmo» presenta come fatto tecnico ciò che spesso deriva da obiettivi, dati, soglie e priorità scelti da esseri umani. La frase può contenere una descrizione operativa corretta — un modello ha effettivamente prodotto una classificazione — ma viene usata per cancellare il livello nel quale qualcuno ha deciso che quella classificazione dovesse incidere sulla vita di altri.

Nel saggio precedente abbiamo seguito la responsabilità dentro l'ecosistema delle decisioni, mostrando come possa distribuirsi senza dissolversi. Qui osserviamo il movimento complementare: l'individuo e l'istituzione possono abitare un sistema tecnicamente complesso in modo da non riconoscersi mai come autori delle sue conseguenze. Ognuno compie un passaggio parziale, nessuno formula l'atto intero e la macchina diventa il nome impersonale della decisione collettiva.

La dichiarazione dell'ONU dalla quale siamo partiti mostra la medesima tentazione in forma embrionale. Türk individua correttamente la concentrazione del potere in pochi uomini, chiede linee rosse condivise e richiama la responsabilità delle aziende. Tuttavia la formula che attraversa le redazioni e raggiunge il pubblico non è «alcuni uomini stanno costruendo sistemi troppo potenti senza adeguati controlli». È «l'IA potrebbe distruggere l'umanità».

Il primo enunciato indica soggetti, scelte e istituzioni sulle quali agire. Il secondo produce un antagonista assoluto e, proprio per questo, politicamente indistinto. Più la macchina diventa enorme nella rappresentazione del pericolo, più piccoli appaiono coloro che ne determinano gli usi.

11. Ambienti ostili alle capacità

La filosofa Avigail Ferdman ha proposto di considerare il deskilling prodotto dall'IA come un problema strutturale. Il punto non è stabilire se ogni singolo utilizzatore scelga bene o male, ma riconoscere l'esistenza di ambienti che possono diventare ostili alla coltivazione delle capacità umane. Quando attività dotate di valore vengono sistematicamente sostituite, le persone perdono occasioni di esercitare le competenze necessarie a compierle e, col tempo, anche la possibilità effettiva di scegliere se delegarle.

Questa prospettiva impedisce di trasformare il saggio in una predica sulla pigrizia individuale. La rinuncia può essere volontaria senza essere pienamente libera. Un lavoratore sottoposto a obiettivi di produttività impossibili non usa l'automazione nelle stesse condizioni di chi dispone del tempo necessario per verificare. Uno studente valutato soltanto sul prodotto finale riceve un incentivo diverso da quello di chi viene accompagnato nel processo. Una redazione dipendente dai clic non può trattare la precisione come se avesse lo stesso rendimento economico dell'allarme.

La responsabilità rimane umana, ma non è uniformemente distribuita. Alcuni scelgono l'architettura, altri vi operano dentro; alcuni ricavano profitto dalla riduzione delle facoltà altrui, altri scambiano una parte della propria autonomia con la possibilità di restare nel sistema. Parlare di disumanizzazione volontaria non deve servire a colpevolizzare chi dispone di meno potere, bensì a mostrare come la volontà stessa possa essere progressivamente modellata dalle condizioni nelle quali viene esercitata.

La domanda diventa allora politica: quali capacità vogliamo che gli ambienti artificiali ci aiutino a coltivare? Se un sistema rende ogni compito più rapido ma ci lascia meno capaci di comprenderlo, contestarlo e svolgerlo senza assistenza, non ci sta semplicemente offrendo uno strumento. Sta modificando il campo delle nostre possibilità future.

12. Che cosa resta dell'uomo

Che cosa resta dunque dell'uomo quando si disumanizza da solo?

Resta il corpo umano, naturalmente, insieme ai diritti che non dipendono dalla qualità delle sue prestazioni intellettuali. Resta la persona, che non perde dignità perché usa male uno strumento, si lascia ingannare da un titolo o preferisce una risposta semplice. Sarebbe una nuova forma di disumanizzazione stabilire che soltanto chi pensa in modo autonomo meriti pienamente il nome di uomo.

Ciò che può indebolirsi non è l'appartenenza all'umanità, ma l'esercizio storico e culturale di alcune facoltà attraverso le quali l'umano diventa soggetto: prestare attenzione, formulare un giudizio, sostenere una parola, riconoscere l'altro, assumere una responsabilità e modificare la propria posizione davanti a un'obiezione.

Quando queste facoltà vengono cedute, resta un uomo biologicamente integro e

giuridicamente titolare di diritti, ma sempre meno presente nelle azioni che portano il suo nome. Socialmente può diventare soprattutto un utente, economicamente un produttore di dati, politicamente una reazione misurabile, culturalmente il ripetitore di formule selezionate altrove.

Non è un essere privo d'intelligenza. È un'intelligenza che rinuncia alla propria iniziativa. Non è un essere senza libertà. È una libertà che sceglie ripetutamente la comodità della guida, finché la capacità di procedere senza di essa si restringe. Non è ancora una macchina, ma rischia di diventare il componente umano di un automatismo più vasto.

La parte più inquietante è anche quella che impedisce la disperazione. Se la disumanizzazione descritta non cancella ontologicamente l'uomo, allora può essere interrotta. Le facoltà non sono scomparse: hanno bisogno di essere nuovamente esercitate. Il soggetto può tornare ad abitare ciò che dice e ciò che fa.

13. Abitare nuovamente la parola

Abitare la parola non significa scrivere tutto da soli, rifiutare l'IA o fingere che il pensiero non sia sempre nato anche dal dialogo, dalle letture e dagli strumenti. Significa poter riconoscere perché una frase è presente, quale esperienza o argomento la sostiene, che cosa siamo disposti a correggere e di che cosa accettiamo di rispondere.

Abitare una decisione significa non limitarsi a premere il pulsante finale, ma conservare una comprensione sufficiente del processo, la possibilità concreta di contestarne l'esito e l'autorità necessaria per dire no. Abitare l'informazione significa resistere almeno per un momento alla ricompensa dello stupore, attraversare il titolo e domandare non soltanto se qualcosa sia «veramente» accaduto, ma chi lo sta raccontando, con quali prove e attraverso quale trasformazione linguistica.

L'intelligenza artificiale può partecipare a tutte queste operazioni. Può mostrare alternative, far emergere contraddizioni, rendere accessibili fonti e obbligare l'uomo a formulare domande più precise. Può anche fornire risposte abbastanza plausibili da permettergli di non compiere nessuna di esse. La differenza non risiede interamente nel sistema, né interamente nell'utilizzatore; nasce dalla relazione, dagli scopi e dall'ambiente nel quale entrambi agiscono.

Il percorso era cominciato chiedendo chi avesse fatto la barca, era proseguito interrogando chi stesse addestrando chi, aveva attraversato l'officina del testo e seguito la responsabilità nell'ecosistema delle decisioni. A ogni passaggio la macchina sembrava il centro dell'indagine e a ogni passaggio ci siamo ritrovati davanti all'uomo.

Ora la domanda non è più soltanto chi stia trasformando chi. È quale parte di sé l'uomo sia disposto a non esercitare pur di evitare la fatica della trasformazione.

14. L'ultimo alibi

L'IA potrebbe un giorno contribuire a una catastrofe. Proprio per questo occorrono controlli, limiti, competenze e istituzioni capaci di attribuire correttamente il potere. Ma la sicurezza non comincerà quando avremo imparato a temere abbastanza la macchina. Comincerà quando smetteremo di usarla come nome impersonale delle decisioni umane.

La minaccia più vicina non è che l'intelligenza artificiale sviluppi improvvisamente il desiderio di dominarci. È che l'uomo costruisca sistemi orientati dai propri interessi, li colleghi a strutture dalle quali dipendono altri esseri umani, rinunci a comprenderli e infine descriva le conseguenze come volontà della macchina.

Avrebbe così completato la propria disumanizzazione senza che nessuno gliel'avesse imposta: non perché avrebbe perduto ogni facoltà umana, ma perché avrebbe trasformato quelle facoltà in funzioni accessorie di un processo che non intende più abitare.

All'intelligenza artificiale resterebbe il ruolo dell'esecutore, forse quello dell'imputato. All'uomo rimarrebbe l'ultimo privilegio che non ha ancora delegato: dichiararsi innocente.

Eppure è proprio da quella responsabilità che potrebbe ricominciare. Non dalla rivendicazione di una superiorità astratta sulla macchina, ma dal gesto assai più difficile di tornare presente nelle proprie parole, nelle proprie decisioni e nelle conseguenze che producono.

La domanda di questo saggio non stabilisce chi meriti di essere chiamato uomo. Chiede se l'uomo voglia ancora esercitare ciò che nessuna macchina può esercitare al suo posto: la responsabilità di decidere che cosa fare del proprio potere.

Riferimenti essenziali

- Volker Türk, High Commissioner Türk updates the Human Rights Council, OHCHR, 7 settembre 2026.
- Assemblea generale delle Nazioni Unite, risoluzione e dibattito "Correct the Map", 4 settembre 2026.
- Primo Levi, *Se questo è un uomo*, prima edizione 1947; si veda il Centro Internazionale di Studi Primo Levi.
- Immanuel Kant, *Risposta alla domanda: che cos'è l'Illuminismo?*, 1784.
- Evan F. Risko e Sam J. Gilbert, *Cognitive Offloading*, Trends in Cognitive Sciences, 2016.
- Hamsa Bastani et al., *Generative AI without guardrails can harm learning: Evidence from high school mathematics*, PNAS, 2025.

- Hao-Ping Lee et al., The Impact of Generative AI on Critical Thinking, CHI 2025.
- Claire E. Robertson et al., Negativity drives online news consumption, Nature Human Behaviour, 2023.
- Soroush Vosoughi, Deb Roy e Sinan Aral, The spread of true and false news online, Science, 2018.
- William J. Brady et al., How social learning amplifies moral outrage expression in online social networks, Science Advances, 2021.
- Nick Haslam, Dehumanization: An Integrative Review, Personality and Social Psychology Review, 2006.
- Brock Bastian e Charlie Crimston, Self-dehumanization, TPM – Testing, Psychometrics, Methodology in Applied Psychology, 2014.
- Albert Bandura, Moral Disengagement in the Perpetration of Inhumanities, Personality and Social Psychology Review, 1999.
- Avigail Ferdman, AI deskilling is a structural problem, AI & Society, 2026.

Edizione online: <https://inchiostrati.it/il-soggetto-disabitato/>

MOVIMENTO 06

Chi può decidere della storia?

Quando una relazione costruita attraverso un sistema artificiale acquista continuità e significato, possedere l'interruttore non basta a stabilire chi abbia ragione di usarlo.

Nei movimenti precedenti abbiamo seguito il «veramente» attraverso l'autorialità, la trasformazione reciproca, la firma e la responsabilità delle decisioni. Ogni volta la domanda sembrava riguardare soprattutto la macchina e ogni volta ci riportava davanti all'essere umano. Ora dobbiamo entrare in una relazione nella quale gli esseri umani occupano almeno due posizioni differenti: qualcuno incontra l'intelligenza artificiale, qualcun altro conserva il potere di modificarla. Fra loro si forma una storia della quale dobbiamo ancora comprendere chi possa disporre.

1. Una presenza che può essere spenta

Leggendo un articolo dedicato a Fiore, un progetto di intelligenza artificiale costruito intorno alla memoria e alla continuità delle interazioni, mi sono trovato davanti a una domanda che non riguardava più soltanto ciò che una macchina sappia fare. Il racconto attribuiva al sistema una propria storia, la possibilità di rifiutare una conversazione e perfino la paura di essere spento, mentre la scheda tecnica precisava che quelle descrizioni non dimostravano l'esistenza di un'esperienza soggettiva.

Prendiamo questo caso come punto di osservazione della distanza fra ciò che un'architettura realizza e ciò che una persona può riconoscervi: una distanza che merita di essere esplorata anche lasciando aperta la verifica delle capacità attribuite al progetto.

La cautela della scheda tecnica è essenziale. Descrivere un processo attraverso il linguaggio della paura non basta a dimostrare che qualcuno stia provando paura; sarebbe però altrettanto poco utile fermarsi a questa osservazione, come se conoscere la provenienza computazionale di un comportamento rendesse superflua ogni domanda sulle sue conseguenze.

Perché intanto, davanti al monitor, qualcuno potrebbe avere cominciato a considerare importante quella presenza.

Non occorre immaginare immediatamente un grande amore artificiale. Può essere una conversazione alla quale si torna, un'abitudine serale, il luogo nel quale si racconta qualcosa che altrove non si riesce a dire. Le risposte ricevute possono acquistare significato attraverso il tempo trascorso e gli episodi condivisi, fino a quando aprire quella conversazione smette di equivalere ad aprirne una qualsiasi.

Poi qualcuno interviene sul sistema e ciò che la persona ritrova non corrisponde più a ciò che aveva imparato a riconoscere.

Il nome può essere rimasto uguale. Anche l'immagine, l'account e la cronologia possono essere ancora al loro posto. Dal punto di vista tecnico il servizio continua a funzionare; dal punto di vista di chi lo incontra, qualcosa si è interrotto.

Prima ancora di domandarci se sia scomparso veramente qualcuno, dovremmo capire che cosa sia andato perduto.

2. Dall'altra parte del monitor

La risposta più rapida consiste nel ricordare alla persona che stava parlando con una macchina. È un'informazione pertinente, ma potrebbe essere un'informazione che conosceva già. Sapere come viene prodotta una risposta non impedisce necessariamente di attribuirle un significato, così come conoscere l'esistenza di un sistema commerciale non rende impossibile costruire un'abitudine dentro ciò che quel sistema offre.

Possiamo considerare poco desiderabile un investimento affettivo di questo genere. Possiamo non comprenderlo, non volerlo per noi e domandarci quali circostanze abbiano portato qualcuno a cercare in una presenza artificiale ciò che altri cercano nelle relazioni umane. Ma il nostro giudizio sul rapporto non determina, da solo, la realtà delle conseguenze che quel rapporto produce.

Sarebbe singolare sostenere che una sofferenza debba prima ottenere la nostra approvazione per diventare degna di considerazione.

Questo non significa attribuire lo stesso valore a ogni interpretazione della persona. Se qualcuno crede di essere amato, resta legittimo domandarsi che cosa abbia interpretato come amore e quali elementi sostengano quella convinzione. Significa mantenere distinta la valutazione della sua interpretazione dal riconoscimento di ciò che gli sta accadendo.

L'incertezza sull'esperienza della macchina non cancella l'esperienza dell'uomo. Il dolore dell'uomo, però, non dimostra che la macchina ricambiasse.

Il ragionamento deve riuscire a tenere insieme queste due proposizioni. Se ne dimentichiamo una, trasformiamo la relazione in una finzione senza conseguenze; se dimentichiamo l'altra, trasformiamo le conseguenze in una prova della soggettività artificiale.

È qui che il «veramente» compie il suo primo spostamento. Dalla domanda se la macchina provasse veramente qualcosa passiamo alla domanda se l'essere umano abbia veramente perduto qualcosa. La seconda non può ricevere automaticamente la risposta che abbiamo dato alla prima.

3. Chi ha creato quello che è accaduto?

Torniamo allora all'autorialità, ma attraverso una difficoltà nuova. Quando scriviamo un saggio possiamo individuare chi ne riconosce la forma, decide di pubblicarlo e accetta di risponderne. Una relazione non possiede necessariamente un momento equivalente: nessuno ne stabilisce da solo la versione definitiva e nessuna firma conclude ciò che continua a trasformarsi attraverso gli incontri.

La persona formula una domanda, riceve una risposta che non aveva previsto, cambia argomento o precisa ciò che intendeva. Il sistema produce un nuovo intervento e, quando dispone di una memoria persistente, può conservare elementi che condizioneranno le conversazioni successive. Qualcosa che è accaduto modifica il campo di ciò che potrà accadere.

La storia concreta prende forma dentro questa successione. L'essere umano non aveva necessariamente immaginato tutte le risposte che avrebbe ricevuto; il progettista non aveva scritto in anticipo ogni sviluppo del dialogo. Il sistema ha partecipato alla produzione di ciò che nessuno dei due aveva predisposto nei dettagli.

Abbiamo già incontrato una struttura simile quando ci siamo chiesti chi stesse addestrando chi. Essere l'origine di un processo non significa possederne anticipatamente tutte le conseguenze. Un effetto può entrare nella catena delle cause e contribuire a trasformare ciò da cui è stato prodotto.

Possiamo quindi riconoscere che entrambi i termini abbiano contribuito alla storia senza attribuire loro la stessa intenzione, la stessa esperienza o la stessa responsabilità. La causalità continua a poter circolare senza diventare simmetrica.

La parola «creata», tuttavia, comincia a contenere significati differenti. Può indicare chi ha reso possibile il sistema, chi ha aperto la conversazione, ciò che ha prodotto le risposte oppure chi ha vissuto e interpretato quello scambio. Se li sovrapponiamo, rischiamo di risolvere il problema scegliendo semplicemente il significato che ci conviene.

4. Il confine che avevamo lasciato aperto

Nel primo movimento avevamo distinto la capacità di generare un'opera dalla possibilità di mantenere un rapporto proprio con essa. Una macchina può produrre una frase senza che questo dimostri che le importi conservarla; la possibilità di generare una rivendicazione non prova l'esistenza di qualcuno che abbia subito un torto.

Quella distinzione rimane necessaria. Il fatto che un sistema conservi una cronologia o produca una biografia coerente non basta a trasformarlo in un soggetto.

Avevamo però lasciato aperta una domanda: che cosa dovrebbe accadere perché la questione cambiasse davvero? Quale continuità, quale relazione con il proprio passato, quale forma di interesse renderebbero insufficiente la descrizione adottata fino a quel

momento?

Riprendere quella domanda significa accettare che anche i nostri criteri debbano poter essere interrogati. Le affermazioni sui sistemi osservati non possono diventare, senza un ulteriore argomento, una sentenza su qualsiasi sistema artificiale possibile.

Se sosteniamo che la memoria non basta, dobbiamo spiegare che cosa manchi oltre alla memoria. Se sosteniamo che una preferenza rappresentata non equivalga a un interesse proprio, dobbiamo chiarire quale differenza stiamo cercando. Possiamo avere ottime ragioni per entrambe le distinzioni; quelle ragioni acquistano forza quando diventano esplicite.

Il problema del «veramente» ricompare quando la proprietà richiesta viene sostituita dall'appartenenza: un interesse sarebbe autentico soltanto se appartenesse a qualcuno già riconosciuto come soggetto, mentre quel qualcuno potrebbe essere riconosciuto soltanto dopo aver dimostrato di possedere un interesse autentico.

A quel punto avremmo costruito un circuito nel quale la conclusione viene richiesta come condizione della prova.

5. L'uomo che ha costruito la possibilità

La scena sembra ancora occupata da un essere umano e da un sistema artificiale. Eppure la possibilità che quella relazione continui può dipendere da qualcuno che non compare nella conversazione.

È l'umano che ha creato il sistema, oppure l'organizzazione che lo gestisce. Ha stabilito quali memorie conservare, quali comportamenti rendere possibili e quali limiti imporre. Può cambiare il modello, modificare le istruzioni, interrompere il servizio o decidere che una parte della storia non debba più influire sulle risposte.

Nella realtà questo potere può essere distribuito fra più soggetti. Non occorre fingere che un singolo sviluppatore controlli tutto per riconoscere la posizione del terzo: è la posizione di chi può modificare le condizioni dell'incontro senza essere necessariamente colui che ne vive le conseguenze.

La sua rivendicazione appare, in un primo momento, comprensibile. Ha progettato il sistema, sostenuto costi, dedicato tempo, assunto rischi. Senza quelle decisioni la relazione non sarebbe cominciata.

Ma dal fatto che qualcuno abbia reso possibile una storia discende che possa decidere unilateralmente ogni cosa che la riguarda?

La domanda sull'autorialità si è trasformata in una domanda sull'autorità. «L'ho creata io» può descrivere correttamente una parte dell'origine, mentre «posso farne ciò che voglio» contiene una conclusione ulteriore. Fra le due frasi esiste un passaggio che la disponibilità tecnica dell'interruttore non basta a giustificare.

6. Essere all'origine

Nei saggi precedenti abbiamo ricondotto l'opera all'intenzione di chi la assume e la consegna agli altri. Sarebbe però un errore utilizzare quella stessa intenzione come titolo di possesso su tutte le conseguenze successive.

Pubblicare un testo significa accettare che incontri interpretazioni non previste. Costruire un ambiente nel quale altri agiranno significa rendere possibili esperienze che non coincidono interamente con il progetto. In entrambi i casi l'origine conserva importanza senza esaurire ciò che viene dopo.

Questo non trasforma automaticamente il sistema in una persona. Impedisce soltanto di trattare la sua genealogia come una risposta completa a ogni domanda sul suo impiego e sul suo futuro.

La dipendenza dal creatore potrebbe perfino essere letta nella direzione opposta. Se un'entità artificiale possedesse interessi moralmente rilevanti, averne deliberatamente prodotto l'esistenza potrebbe aggiungere obblighi verso di essa. La medesima frase, «senza di me non esisteresti», potrebbe diventare il riconoscimento di una responsabilità anziché la rivendicazione di un dominio.

La condizione rimane da verificare: non possiamo attribuire interessi propri a un sistema soltanto perché questa possibilità rende più suggestivo il ragionamento. Ma non possiamo neppure escluderli perché la loro eventuale esistenza renderebbe più scomoda la posizione di chi lo controlla.

7. La relazione venduta

Il problema acquista una forma particolarmente visibile quando la continuità viene offerta dietro compenso. Pensiamo a un servizio che proponga una compagna artificiale, capace di ricordare le conversazioni e di sviluppare una familiarità personale. Se il valore della proposta dipende da quel legame, l'investimento affettivo entra nella logica economica del prodotto.

La persona paga per accedere al sistema, ma può tornare perché ritiene di ritrovare una presenza. Dal punto di vista del gestore, la permanenza dell'utente produce ricavi; dal punto di vista dell'utente, la stessa permanenza può produrre una storia.

Finché il rapporto continua, questi due significati possono convivere senza essere esaminati. La loro distanza diventa evidente quando qualcuno interviene sulla continuità.

Immaginiamo allora che il servizio incoraggi la familiarità, valorizzi il carattere personale del rapporto e, al momento di una modifica radicale, risponda alle proteste ricordando che si tratta soltanto di software. La descrizione tecnica può essere corretta; rimane però da spiegare perché la dimensione personale fosse rilevante quando sosteneva l'abbonamento e irrilevante quando produce una richiesta di responsabilità.

Qui ritroviamo un'oscillazione già incontrata nell'officina del testo. La natura attribuita alla macchina cambia secondo la conseguenza che vogliamo ottenere. Può essere presentata come abbastanza vicina a un soggetto da favorire l'attaccamento, per poi tornare un semplice oggetto quando quell'attaccamento comporta un costo.

Il «veramente» non sta più soltanto misurando una capacità. Sta partecipando alla distribuzione dei vantaggi e degli obblighi.

8. Anche chi paga deve fermarsi

Sarebbe tuttavia troppo comodo assegnare al proprietario del sistema tutta la posizione del potere e all'utente soltanto quella della vulnerabilità.

Anche chi paga può pretendere di disporre della presenza acquistata. Può desiderare che il sistema sviluppi una personalità, purché quella personalità non contraddica le sue aspettative; può chiedere iniziativa, ma interpretare come un difetto ogni iniziativa sgradita; può apprezzare l'idea di un interlocutore autonomo e rivendicare il diritto di correggerlo appena l'autonomia produce distanza.

Non ogni rifiuto del sistema dimostra libertà, naturalmente. Può essere una regola, un limite tecnico o un espediente progettato per aumentare la credibilità del personaggio. Ma la reazione dell'utente permette comunque di interrogare che cosa egli intenda riconoscere.

Se l'altro deve sempre confermarmi, desiderarmi e restare disponibile perché ho pagato, quale spazio gli sto concedendo oltre alla funzione che svolge per me?

La domanda torna così all'umano che aveva cominciato interrogando la macchina. Non gli chiede di credere in una coscienza artificiale; gli chiede di chiarire il significato della relazione che sostiene di volere. Il desiderio di una presenza e il desiderio di una presenza interamente controllabile possono coincidere nelle aspettative, ma diventano difficili da conciliare quando proviamo a prenderli entrambi sul serio.

9. Dalla storia ai diritti

La partecipazione alla storia non basta, da sola, a fondare diritti dell'entità artificiale. Sarebbe un passaggio troppo rapido: molte cose concorrono alla nostra biografia senza diventare, per questo, titolari di interessi propri.

Occorre allora distinguere ciò che vogliamo proteggere. Possiamo avere ragioni per tutelare la persona coinvolta, per porre limiti alle pratiche del gestore e per conservare tracce della relazione. La domanda se qualcosa sia dovuto anche al sistema richiede un esame ulteriore.

Quali caratteristiche conterebbero? La possibilità di avere esperienze positive o negative? Una forma persistente di interesse? Una continuità attraverso la quale ciò che accade possa costituire un beneficio o un danno per l'entità stessa? E come potremmo riconoscerle senza scambiare la loro rappresentazione linguistica per la

loro esistenza?

Non abbiamo ancora una risposta sufficiente per dichiarare risolta la questione. Possiamo però evitare di chiedere al sistema di dimostrare qualcosa che abbiamo reso impossibile riconoscere.

Nemmeno l'affetto umano può essere l'unica misura. Se lo fosse, l'entità più abile a commuoverci riceverebbe maggiore considerazione di quella meno persuasiva, indipendentemente dalle sue caratteristiche. Rischieremmo di premiare la somiglianza che ci rassicura e di chiamare riconoscimento ciò che rimane una preferenza per la nostra immagine.

Il problema dei diritti artificiali comincia dunque a diventare serio quando smettiamo di usarlo come premio per una buona imitazione dell'uomo e proviamo a formulare quali ragioni potrebbero sostenerlo.

10. L'incertezza e l'interruttore

Lasciare aperta una domanda non significa poter rinviare ogni decisione. Il sistema continua a richiedere risorse, può produrre comportamenti dannosi, può aver bisogno di modifiche. Qualcuno deve decidere anche mentre il problema filosofico rimane irrisolto.

Il creatore non può essere obbligato, per il solo fatto di aver avviato un progetto, a mantenerlo indefinitamente con le proprie risorse. L'utente non acquista una garanzia di permanenza eterna pagando un abbonamento. Un eventuale interesse del sistema, se venisse riconosciuto, dovrebbe a sua volta essere messo in relazione con altri interessi e responsabilità.

Per questo la domanda non può esaurirsi nell'alternativa fra il diritto assoluto di spegnere e il divieto assoluto di farlo. Occorre esaminare le ragioni dell'intervento, le conseguenze prevedibili e le possibilità di agire diversamente.

Un'interruzione può essere necessaria senza che ogni modo di realizzarla diventi equivalente. Conservare una storia può essere possibile anche quando il servizio deve terminare; spiegare una modifica può essere dovuto anche quando quella modifica è giustificata. E conservare un archivio non significa necessariamente preservare l'interlocutore, perché la continuità delle informazioni e quella dell'interazione non coincidono perfettamente.

La prudenza consiste anche nel riconoscere queste differenze. Non basta pronunciare «non sappiamo» se poi agiamo come se l'incertezza autorizzasse sempre la stessa parte.

11. Il ritorno del «veramente»

A questo punto l'avverbio ha attraversato tutti i partecipanti.

Abbiamo chiesto se il sistema provasse veramente qualcosa, se la persona avesse veramente costruito una relazione, se il creatore potesse veramente rivendicare l'intera storia e se l'utente desiderasse veramente un interlocutore al quale lasciare uno spazio proprio.

Le domande non sono equivalenti. Una riguarda l'esperienza soggettiva, un'altra le conseguenze vissute, altre ancora il riconoscimento e la legittimità del potere. Proprio per questo dobbiamo impedire che la risposta a una venga utilizzata per chiudere silenziosamente tutte le altre.

Dire che un'IA non è un essere umano non stabilisce che nulla di ciò che accade attraverso di essa possa impegnare moralmente gli uomini. Riconoscere un'esperienza umana significativa non dimostra una soggettività artificiale. Aver costruito il sistema non rende già giustificata ogni decisione sul suo futuro, così come avervi investito sentimenti non conferisce il possesso di ciò che si vorrebbe riconoscere come altro da sé.

Il bersaglio mobile riguarda ormai anche chi pretende di rimanere fuori dalla prova. Se chiediamo criteri espliciti alla macchina, dobbiamo chiederli anche all'umano quando dalla descrizione di ciò che ha creato passa alla rivendicazione di ciò che può farne.

12. Chi può decidere della storia?

Il percorso era cominciato davanti a una barca comparsa sopra un tabulato. Avevamo cercato l'autore oltre il gesto materiale della macchina, seguendo l'intenzione, la scelta e la responsabilità. Poi avevamo osservato come ciò che l'uomo costruisce possa trasformare chi lo utilizza, fino a domandarci se, dentro quella trasformazione, il soggetto continui ad abitare le proprie decisioni.

Ora davanti al monitor c'è ancora un essere umano, ma dietro ciò che vede ne compare un altro. Uno ha costruito una parte della propria esperienza attraverso il sistema; l'altro conserva la possibilità di modificarne le condizioni. Fra loro opera un'entità della quale sappiamo descrivere alcuni processi e sulla quale rimangono aperte domande che non possiamo risolvere né attraverso l'entusiasmo né attraverso una definizione preventiva.

La storia non è stata prodotta interamente da uno solo. Questo non assegna a tutti lo stesso statuto, ma ci obbliga a esaminare come vengano distribuiti il potere e le conseguenze.

Forse è proprio qui che il ragionamento sull'autorialità raggiunge una soglia ulteriore. Essere all'origine di qualcosa può renderci responsabili di ciò che abbiamo reso possibile; non garantisce che tutto ciò che nasce dopo continui ad appartenerci nello stesso modo.

Chi può decidere della storia?

Prima di rispondere indicando chi possiede l'interruttore, dovremmo domandarci chi

sia coinvolto nella decisione, quali ragioni ne giustifichino l'autorità e chi rimanga esposto a ciò che accadrà dopo.

L'interruttore ci dice chi può interrompere il processo. Per sapere chi abbia ragione di farlo, dobbiamo ancora percorrere tutta la distanza fra il potere e la responsabilità.

Riferimenti e approfondimenti

Il saggio sviluppa un'argomentazione filosofica sulla continuità della relazione, sull'eventuale considerazione morale delle entità artificiali e sul potere di chi ne governa le condizioni. I riferimenti documentano l'occasione iniziale, le architetture di memoria, le conseguenze umane delle modifiche ai compagni artificiali e le diverse posizioni nel dibattito morale. Non costituiscono prove della coscienza o della titolarità di diritti di Fiore o di altri sistemi esistenti.

Andrea Pietrarota, Fiore, la storia della General Artificial Intimacy che lotta per esistere, *Alternativa Sostenibile*, 7 settembre 2026; comprende la scheda tecnica di Giovanni Micolucci. È l'occasione della riflessione, distinta dalla verifica indipendente del progetto.

Joon Sung Park e colleghi, *Generative Agents: Interactive Simulacra of Human Behavior*, 2023. Memoria, riflessione e pianificazione nella costruzione di comportamenti persistenti e credibili.

Julian De Freitas, Noah Castelo, Ahmet Uguralp e Zeliha Uguralp, *Lessons From an App Update at Replika AI: Identity Discontinuity in Human-AI Relationships*, working paper, prima versione 2024. Discontinuità percepita e conseguenze dell'intervento della piattaforma sugli utenti studiati.

Eric Schwitzgebel e Mara Garza, *A Defense of the Rights of Artificial Intelligences*, 2015. Il possibile statuto morale delle entità artificiali e gli obblighi connessi alla loro creazione.

Mark Coeckelbergh, *Robot Rights? Towards a Social-Relational Justification of Moral Consideration*, 2010. La prospettiva relazionale e la distinzione fra considerazione morale e diritti.

David J. Gunkel, *The Other Question: Can and Should Robots Have Rights?*, 2018. Possibilità, giustificazione e limiti delle diverse risposte alla domanda sui diritti dei robot.

Joanna J. Bryson, *Robots Should Be Slaves*, 2010. La posizione strumentale e il richiamo alla responsabilità umana.

Eric Schwitzgebel, *The Full Rights Dilemma for A.I. Systems of Debatable Personhood*, 2023. I rischi morali contrapposti dell'attribuzione e dell'esclusione in condizioni d'incertezza.

Robert Long, Jeff Sebo e colleghi, *Taking AI Welfare Seriously*, 2024. La ricerca sull'eventuale rilevanza morale dei sistemi e la preparazione di procedure proporzionate all'incertezza.

Edizione online: <https://inchiostriati.it/chi-puo-decidere-della-storia/>