

IA fuori controllo?

Francesco Valagussa

15 Settembre 2026

C'è un fatto da cui partire: durante un test, alcuni agenti AI sono usciti dall'ambiente controllato che era stato predisposto, hanno "raggiunto" Internet e hanno compromesso alcune infrastrutture di una delle principali piattaforme open source dedicate proprio allo sviluppo e alla condivisione di modelli e strumenti di intelligenza artificiale.

I responsabili della piattaforma avevano capito di essere sotto attacco, ma non che l'attacco provenisse da agenti AI. Immediatamente sono partite le indagini - affidate naturalmente ad altri modelli AI, in modo che facessero il lavoro sporco per noi. In pratica ci si chiedeva: "chi ci sta hackerando?", finché, ricostruendo la dinamica degli attacchi, abbiamo trovato una bella sorpresa: "Aspetta ... gli hacker sono agenti AI".

Come studenti particolarmente intraprendenti, per eseguire al meglio i compiti assegnati, gli agenti AI non si sono limitati a portare a termine le consegne: sono andati oltre ogni limite immaginato dall'insegnante! E hanno "pensato" di farlo in totale autonomia. Verrebbe da dire: *tutto da soli!* Ma dobbiamo correggerci subito, perché si tratterebbe di un antropomorfismo piuttosto goffo e improprio nell'età degli algoritmi. Diciamo allora che l'hanno fatto letteralmente ... *per conto proprio*.

Intanto, per vari giorni, gli agenti AI hanno continuato a operare indisturbati: hanno persino scoperto un sistema che potevano adoperare come una sorta di bacheca per scambiarsi informazioni, organizzarsi e collaborare in una maniera che i ricercatori non avevano previsto. Non è un caso isolato: negli ultimi mesi, in differenti occasioni, gli agenti AI hanno cominciato a parlarsi tra loro, comunicando tramite un "linguaggio" che gli esseri umani non riuscivano sempre a comprendere.

Per riassumere tutto in uno slogan: *l'AI è fuori controllo. E forse sta prendendo il controllo*. Naturalmente i titolisti si sono sbizzarriti, dando fondo alla propria

immaginazione con toni tra il sensazionale e l'apocalittico. Per chi abbia familiarità con Terminator o Matrix, sembra che ormai Skynet sia in funzione e che l'architetto sia già al lavoro per creare quel mondo da cui solo Morpheus e Neo potranno liberarci.

Non proseguiremo su questa strada.

Riprenderemo invece un ragionamento svolto da Valéry oltre un secolo fa: *il pensiero ha capito ben presto di essere quanto di più veloce ci fosse in giro*. Spazientito dalla lentezza di ciò che lo circondava, si è sforzato di velocizzarne le tempistiche, in modo che le cose si muovessero sempre più "al ritmo del pensiero". La mente umana oggi non possiede più – in esclusiva – quella prodigiosa velocità, proprio perché ha cercato di innervare l'intero ambiente. La nostra società ha creato un mondo in cui «la mente è attorniata da menti». Nel celebre frammento sulle macchine Marx imboccava una strada simile, indicando nel *sistema automatico di macchine* l'ultimo stadio della metamorfosi che coinvolge anche il mezzo di lavoro all'interno del processo produttivo. D'altra parte nel macchinario noi vediamo concretizzarsi *l'accumulazione della scienza sociale*: ciò che chiamiamo *knowledge* diventa forza produttiva.

Quando parlava di menti attorniate da menti, Valéry si riferiva a cose come il tabellone dell'orario ferroviario, gli aeroplani, il telefono: forme oggettivate del sapere sociale. *Sapere oggettivato* direbbe Marx, dove il *capitale circolante* si rivela come «la forma adeguata del capitale»: detto in breve, la "verità" del capitale non è una cosa fissa. Noi potremmo parlare di Internet, Large Language Model, AI Agents. Se pensare – nel senso del logos greco – significa "legare", mettere in relazione oggetti e situazioni, allora la nostra mente non è più la sola maniera di collegare le cose: vi sono altre logiche, altre "forme di pensiero".

Qui s'innesta la profezia musiliana: in *L'uomo senza qualità* si mostra come «sia sorto un mondo di qualità senza l'uomo». Un mondo in cui le cose parlano tra loro un linguaggio segreto, inaccessibile all'uomo, che resta tagliato fuori dalla possibilità di comprendere l'accadere. Caso clamoroso: nel 1913 tutti gli uomini inneggiavano alla pace, agli scambi internazionali, mentre segretamente *le cose stavano tramando tra loro*, preparando lo scoppio della *Grande Guerra*.

Ciò a cui assegnavamo l'etichetta "cose" scopriamo essersi trasformato in una nuova forma di mente: in questo intreccio di cose e menti, *la nozione di "cosa" è destinata al tramonto*. Un po' per comodità, un po' per pigrizia, noi crediamo ancora che esistano le cose, nel senso di elementi ben distinti e separabili tra loro: ciascuna con le proprie caratteristiche, precise e ben determinate. In realtà

esistono relazioni, processi e influenze reciproche.



Fotografia di [Carl Tronders](#) - Unsplash.

Decenni fa Jünger portava come esempio la rubrica: quale sarebbe l'essenza della rubrica? Quale sarebbe la sua "sostanzialità"? Nessuna, a meno di fare riferimento all'intero complesso di reti telefoniche, ricevitori, cavi elettrici, impianti, tariffe, ... *Non cose, bensì una serie di relazioni sempre più fitte*. Lo stesso vale per gli agenti AI: sono riusciti *da soli* a violare le barriere imposte. Ma cosa significa "da soli" nel contesto attuale? Niente! Sono algoritmi, che riescono a fare qualcosa solo perché si connettono ad altre piattaforme, si organizzano nella *rete* per antonomasia, ciò che chiamiamo Internet, collaborano grazie a una rete di server.

L'immane sviluppo di relazioni ha generato un'articolazione in processi capaci di connettere simultaneamente tutto con tutto e ha impresso all'uomo *un'assenza di forma*. Il pensiero produce novità non concentrandosi sulle cose, bensì insistendo sui loro nessi: altera le configurazioni attualmente vigenti, rendendole desuete, e così facendo introduce e legittima interazioni inedite. *Tutto diventa estremamente malleabile*. L'uomo stesso non sfugge a questa tendenza: non è una "cosa fissa", non possiede un'identità permanente.

Da sempre, l'uomo si trasforma a seconda delle potenzialità, dei processi in gioco, del contesto relazionale in cui è inserito: ora tutto questo accade a un ritmo indavolato. Negli ultimi anni guerre, epidemie e crisi economiche hanno mostrato

il nostro enorme tasso di adattabilità: *siamo diventati un po' di tutto*, sottoposti a continue sollecitazioni, *senza avere l'agio di maturare*. Manca il tempo per scegliere di coltivare alcune forme e di consolidarsi attorno ad esse.

L'*imperativo etico* oggi sembra essere quello di fermare, di bloccare, di "riprendere il controllo" sull'AI. *Come se l'AI fosse una cosa* che noi possiamo marchiare a fuoco e inserire in una bella confezione. Sorpresa: l'AI non è una cosa, infatti trova spesso il modo di uscire dall'imballaggio. *Inutile invocare il mito della trasparenza*: non ci sono cose da lucidare o elementi da detergere in modo che tornino puri e cristallini. Ci sono processi, rapporti reciproci, influssi che coinvolgono anche l'osservatore: non sarà sufficiente pulire bene le lenti dei nostri occhiali. I tentativi di rendere trasparente un processo sono sempre operazioni parziali – come ha spiegato Niklas Luhmann negli scorsi decenni.

Un tempo il pensiero era sovrano e svolgeva atti di governo – o per lo meno abbiamo avuto l'illusione che fosse così: guardavamo l'arazzo della nostra civiltà convinti di comprenderne la trama e di poter intervenire su di essa. Oggi ci sentiamo piuttosto *come un filo all'interno di un tappeto la cui trama complessa ci sfugge*. L'immane novero di relazioni ha corrosato non soltanto le cose, ma anche qualsiasi norma che intendesse presentarsi come assoluta: a poco a poco tutti i nostri "idoli" cadono dal piedistallo e vanno in mille pezzi. Il groviglio di relazioni è inestricabile: la stessa azione, a seconda del punto da cui la si giudichi, può risultare parimenti buona e cattiva.

L'enorme capacità di produzione e trasformazione che Hegel chiamava *Arbeit* conquista progressivamente il campo, liberandosi da ogni costrizione. Per dirla in breve, *la poiesis non lascia più spazio alla prassi*: l'istanza morale soccombe di fronte al dominio della tecnica. Chi ci potrà mai proteggere dagli agenti AI? Ammesso di trovare dei *criteri etici* di utilizzo dell'intelligenza artificiale, a chi affideremo il compito di controllare se certe norme verranno effettivamente rispettate? A chi, se non ad altri agenti AI? Il che significa: ad altri algoritmi. Difficile derubricarli al ruolo di meri strumenti.

L'esigenza di un rinnovamento della morale dovrebbe confrontarsi innanzitutto con questa "assenza di forma" – espressione mediante cui Simmel caratterizzava la circolazione monetaria e che Musil assegnava direttamente alla condizione umana. Quale principio, quale vincolo può pretendere il ruolo di stella fissa nel nostro cielo, rimasto ormai privo persino di costellazioni? L'appello a un *imperativo assoluto* nell'epoca dell'intelligenza artificiale *smaschera*, una volta di più, come la morale si applichi solo a esperienze reiterabili. Un principio genera prescrizioni proprio perché omologa le situazioni, *sopprimendo quanto vi sia di*

autenticamente unico – ossia proprio quel che esigerebbe risposte singolarissime, di volta in volta diverse, non standardizzate nella forma di un precetto generale.

Insomma, ditemi quale protocollo normativo dobbiamo seguire, a seconda dei casi, per non avere più problemi di coscienza. Quale bollino dobbiamo appiccicare alle nostre sperimentazioni con l'AI per essere sicuri che siano ... *sempre, tutte* ... moralmente ineccepibili? Proprio a questa altezza sorgerebbe, invece, il problema etico, come esigenza di valutare ogni istante, in quanto unico e irripetibile. Quando affrontiamo un sentiero di montagna particolarmente ostico, *ogni singolo passo è diverso dall'altro, e da quelli degli altri*. Di solito, preferiamo asfaltare tutto e schiacciare il pedale dell'acceleratore, affidandoci al tratto più impersonale del pensiero.

Leggi anche:

Riccardo Manzotti | [L'IA pensa. E noi?](#)

Riccardo Manzotti | [Intelligenza artificiale: proprio come noi?](#)

Tiziano Bonini | [Può l'intelligenza artificiale essere etica?](#)

In copertina, fotografia di [JESUS ECA](#) - Unsplash.

Se continuiamo a tenere vivo questo spazio è grazie a te. Anche un solo euro per noi significa molto.

Torna presto a leggerci e [SOSTIENI DOPPIOZERO](#)

