

# Verso una contrazione degli attuali “colletti bianchi” ChatGPT e la “sostituzione intellettuale”

di Gabriele Giacomini

## Abstract

Starting with the concrete and recent example of ChatGPT, this paper explores the impact of AI on intellectual work. The argument begins by discussing the potential of AI to replace certain types of intellectual and creative work, such as writing and translation. ChatGPT’s capabilities are explored, including its ability to generate human-like responses to text requests and its potential applications. This is a crucial development, because in the past we used to think that “machines” would mainly replace physical, operational and repetitive work, whereas now the risk of partial but relevant “intellectual substitution” seems to be advancing. Finally, the ethical and political implications of the use of AI in intellectual work are examined, including concerns about job reconfiguration, obscurity and bias in algorithmic decision-making.

Quali lavori resisteranno all’innovazione digitale? In genere si dice che saranno quelli intellettuali, ad alto valore, legati al pensiero creativo. La formazione specialistica, densa di conoscenza, continua, dovrebbe permettere ai lavoratori di mettersi al riparo dal rischio di essere “sostituiti” dalle macchine. Eppure, dopo il lancio di ChatGPT, di inizio 2023, emergono dei dubbi.

ChatGPT è un’intelligenza artificiale. In particolare, un *large language model* tanto avanzato da obbligarci ad avere l’impressione che la macchina sia davvero intelligente. Ha coscienza? Secondo la schiacciante maggioranza degli esperti, assolutamente no (Floridi, 2022). ChatGPT genera testo senza “conoscere” i concetti del testo che ci trasmette. Produce sintassi senza avere idea del significato. Ma ciò è davvero importante? Immaginiamo che, in un futuro indefinito, l’IA di un robot poliziotto inizi a sparare su ogni persona che le capiti a tiro, senza che nessuno umano glielo abbia ordinato.

Ci interessa davvero che lo faccia con una coscienza oppure senza? L'IA generativa mostra che fare qualcosa al di fuori di un'istruzione stringente non richiede necessariamente consapevolezza intesa in senso umano (Cristianini, 2023). Circa le implicazioni dell'IA, le questioni ontologiche sull'assenza di coscienza sono teoricamente interessanti ma praticamente secondarie<sup>1</sup>. Quello che conta è il risultato: la potenza, la raffinatezza e l'elasticità dello strumento informatico che abbiamo a disposizione e con cui, in un futuro ormai alle porte, dovremo fare i conti.

### **Per secoli la tecnica ha sostituito il lavoro fisico, ora toccherà a quello intellettuale**

Siamo abituati, da sempre, al fatto che le tecnologie sostituiscono soprattutto il lavoro manuale, routinario, quello con tanta fatica e poco valore aggiunto (Moretti, 2017). L'aratro, l'aspirapolvere, il trattore, la gru, il treno, la leva, le catene di montaggio, la ruota, la lavastoviglie, i camion hanno sostituito lavoro umano di questo tipo. Ma anche il personal computer, nella misura in cui permette di fare "copia e incolla" invece di battere a macchina. La storia è costellata da innovazioni tecniche che hanno sollevato l'essere umano dal fardello della fatica meccanica, operativa. Tutto ciò ha ridotto molte occupazioni (i contadini, ad esempio, sono pochissimi rispetto al passato, ma anche le lavandaie o i dattilografi). E ne ha create di altre. Il cambiamento era difficile da assorbire socialmente (si pensi al fenomeno dei luddisti nell'Ottocento), ma era tutto sommato coerente con la nostra identità: l'uomo e la donna come esseri dotati di logos. La corteccia frontale come parte più "avanzata" del cervello.

Il punto è che con l'intelligenza artificiale, e con i *large language models*, ciò viene spiazzato. Siamo abituati al fatto che le nostre macchine ci sollevino soprattutto dal lavoro fisico/operativo? Ora le nostre macchine ci sol-

<sup>1</sup> Se dovessimo essere estremi, ammetteremmo che ogni coscienza certa è soltanto autocoscienza. L'IA non ha coscienza. Ma siamo sicuri che gli altri esseri umani la abbiano? Ognuno di noi percepisce in maniera evidente solo la propria coscienza, non certo quella delle altre persone (che potrebbe essere "cartesianamente" un inganno – questo dubbio metafisico ha fatto la fortuna di film come Matrix).

leveranno dal lavoro intellettuale. Sul lato computazionale la “sostituzione intellettuale” da parte delle macchine è (quasi) completamente avvenuta. Sul lato creativo sta avanzando con passi da gigante, e la novità di ChatGPT sta a dimostrarcelo concretamente (Brynjolfsson e McAfee, 2014).

Alcuni segnali, per la verità, erano già presenti. Si prenda Amazon. Vi ricordate il libraio di fiducia che ci aggiornava sulle novità editoriali? Un lavoro che richiede grande sensibilità culturale. Questo lavoro sulla piattaforma è svolto dall’IA, sotto forma di algoritmo di raccomandazione: analizza tantissimi dati di navigazione e ci suggerisce i libri migliori per noi. Ma chi ci consegna il libro? Un fattorino in carne e ossa, almeno per ora. L’IA ha sostituito tanto lavoro intellettuale, ma quello fisico, soprattutto nella distribuzione, è ancora svolto da persone<sup>2</sup>. Ci sono certamente ristrette élite di lavoratori con un alto livello professionale (ingegneri, informatici e non solo). Ma sono relativamente pochi, e assai del lavoro intellettuale “classico” è stato eroso. Oltre alle raccomandazioni di prodotto personalizzate e al marketing, l’IA si occupa di assistenza clienti automatizzata e di controllo della qualità, per non parlare della gestione del personale, della logistica e della catena di approvvigionamento<sup>3</sup>.

Siamo ancora sicuri che i mestieri intellettuali e “alti” siano quelli più a riparo dalla rivoluzione digitale in atto? Con lo sviluppo dell’IA la “sostituzione intellettuale” rischia di diventare una valanga. Rischia di inondare le aree della creatività e delle attività ad alto valore aggiunto. Nel marcare la differenza fra IA ed essere umano, alcuni minimizzano evidenziando che l’efficienza dell’IA dipende “soltanto” dal fatto che “rimescola” enormi quantità di informazioni prodotte da umani. Ma anche per noi umani, se il plagio è copiare da pochi, la cultura è copiare e rielaborare da moltissimi. I cuccioli umani imparano per imitazione.

<sup>2</sup> Il lavoro fisico nella distribuzione è ancora lontano da essere sostituito perché l’ambiente di una città è, per ora, troppo complesso. Le aziende stanno testando sistemi di consegna basati sui droni, ma siamo ancora a livello di sperimentazione. Diversa la situazione per i magazzini, che invece sono sempre più robotizzati, in quanto possono essere costruiti su misura delle macchine intelligenti.

<sup>3</sup> In piattaforme come Deliveroo o Just Eat non va molto diversamente. L’IA è fondamentale per dirigere i rider, per raccogliere le ordinazioni, organizzare informazioni e comunicazioni, mentre il maggior numero degli esseri umani impegnati nell’ambito largo dell’organizzazione stanno ai fornelli a cucinare o pedalano sotto la pioggia e sotto il sole. I percorsi di delivery funzionano così: il software dice agli umani che tragitto fare, l’umano guida.

DeepL è un recente servizio di traduzione automatica basata su reti neurali profonde. Le traduzioni non sono perfette: il sistema non coglie (ancora) le sfumature culturali, o le metafore. Ma testi descrittivi, non troppo complessi, vengono trasposti in maniera eccellente, in decimi di secondo, a costo praticamente nullo. Ciò può aumentare il mercato, ma può anche ridurre di molto la domanda di traduttori umani. Se finora gli umani venivano richiesti per tradurre, in futuro potrebbero essere chiesti soltanto per controllare le traduzioni automatiche. Ciò che in passato poteva essere fatto da 100 specialisti, nei prossimi anni potrebbe richiederne 10. Le traduzioni simultanee o consecutive resistono, ma cosa accadrà se o quando la comprensione vocale automatica verrà perfezionata? È concepibile che i nostri nipoti non abbiano più la necessità di imparare lingue straniere. Potrebbe essere sufficiente attivare la traduzione automatica e ascoltarla in cuffietta mentre si è ad un convegno all'estero, in viaggio. O, addirittura, seguirla attraverso un'evoluzione dell'impianto cocleare o attraverso un collegamento neurale<sup>4</sup>.

L'evoluzione di ChatGPT potrebbe addirittura erodere i lavori creati dalla stessa rivoluzione informatica: i programmatori o i social media manager. Sarebbe un esito sorprendente: la trasformazione digitale ha falciato operatori di banca, fotografi di stampa, impiegati della segreteria, ma aveva creato nuove figure professionali nella costruzione dei sistemi informatici o nella comunicazione digitale. Questi ultimi, però, potrebbero essere, almeno in parte, delle “falene della trasformazione digitale”, nel senso che l'IA è sempre più capace di generare codice autonomamente. Immaginiamo ChatGPT unito a un software di programmazione di siti Web, a sua volta collegato a un'IA specializzata in grafica e design (simile a Dalle2). In chat si potrebbe chiedere all'IA di proporre un layout del proprio sito Web, e poi si

<sup>4</sup> Con questo discorso non escludo che conoscere con padronanza una lingua straniera manterrà un valore culturale e sociale. Chi conosce più lingue avrà ancora capacità e sensibilità particolari. Gli appassionati, i più colti o anche i più benestanti potranno continuare a coltivare il culto delle lingue straniere. Ciò su cui mi focalizzo, e che potrebbe cambiare radicalmente, è che l'inglese, il mandarino, l'urdu, il filippino, il nigeriano, il tedesco, il giapponese, il turco, il serbo, il mongolo non saranno più ostacoli per nessuna delle persone che non conoscono queste lingue. Invece che utilizzare uno sgangherato “globish”, che oggi impoverisce le comunicazioni nel mondo, miliardi di persone che non hanno in comune una lingua conosciuta in maniera avanzata potranno comunque confrontarsi a un alto livello di complessità e comprensione.

potrebbero chiedere delle modiche semplicemente interloquendo. L'IA potrebbe sottoporre all'umano cinque proposte di logo per il nuovo sito, noi potremmo sceglierne una e anche offrire dei suggerimenti migliorativi. L'IA potrebbe creare contenuti per il sito, magari in un contesto di realtà aumentata o virtuale, generando testo, immagini e video automaticamente in base alle nostre indicazioni<sup>5</sup>. L'IA attualmente dà spesso errori nella scrittura del codice, ma sembra che stia migliorando.

Non sto sostenendo che non ci saranno più umani addetti alla creazione di ambienti digitali, ma che potrebbero servirne di meno, magari specializzati non nella scrittura diretta del software, ma nel colloquio con l'IA. Mentre per costruire un sito web complesso possono servire centinaia di ore di lavoro di sviluppatori e designer, in futuro potrebbero essere sufficienti decine di ore di un esperto capace di fare le giuste domande al sistema di IA. L'“erosione” del lavoro intellettuale potrebbe avere la misura di un ordine di grandezza. Forse non è un caso se le grandi compagnie del digitale, da Amazon a Facebook, stanno licenziando lavoratori qualificati nei settori del marketing o del supporto tecnico. Ciò dipende in parte dal “riassetto” rispetto al periodo del Covid19 (quando le big tech hanno assunto per far fronte alle maggiori richieste). Ma siamo sicuri che non dipenda anche dal fatto che molti lavoratori delle big tech, a causa dell'avanzare dei sistemi di IA, non sono più così essenziali? Da sempre la cifra delle aziende digitali, rispetto a quelle tradizionali come quelle automobilistiche, è avere pochissimi lavoratori. Perché dovrebbero retribuire persone in carne ed ossa, se sempre più compiti, e sempre più complessi, possono essere assegnati all'IA?

Azzardiamo che in futuro potrebbero essere toccati anche i “santa sanctorum” del lavoro spirituale umano: gli scrittori, i romanzieri, addirittura

<sup>5</sup> A chi potrebbero andare i diritti di loghi, immagini o grafiche costruite dall'IA, ma che nascono dalla rielaborazione di lavoro umano? All'azienda proprietaria dell'IA? All'operatore umano che avanza le richieste all'IA? O alla comunità dei creativi? Questa comunità potrebbe essere definita in che modo? Questo problema sta già emergendo. Ad inizio 2023 alcuni tra i maggiori disegnatori italiani hanno sottoscritto un manifesto per chiedere all'Unione europea la tutela delle loro opere “saccheggiate” dagli algoritmi generativi, i quali elaborano le immagini trovate sul web per produrre nuovi contenuti. Per non parlare dei problemi, nei casi di immagini e video, legati all'indistinguibilità tra verità e fantasia/menzogna.

i poeti. Per scrivere bene bisogna leggere molto. E chi può leggere di più di un'IA alimentata dai big data? Un'IA può elaborare una quantità di testi che un essere umano non potrebbe consultare nemmeno in mille vite. Potrebbe dominare, nella sua capacità di elaborazione, i testi dei più grandi, da Shakespeare a Cervantes, da Dante a Virginia Woolf, da Dostoevskij a Dickens, da Victor Hugo a Hemingway. Perché, a un certo punto, un'IA non potrebbe scrivere un romanzo apprendendo dallo stile e dal contenuto dei grandi romanzi della storia umana? Non possiamo escluderlo a priori. Inoltre, già oggi l'IA riesce a profilare gli utenti e a personalizzare i contenuti dei post sui social network, per attività commerciali e propaganda politica. Perché l'IA non potrebbe essere programmata per analizzare il pubblico di destinazione del romanzo, tenendo conto delle preferenze e dei gusti dei lettori, anzi del singolo lettore? Mettiamo che io adori il Commissario Montalbano. Purtroppo, Camilleri è scomparso. Ma ci ha lasciato migliaia di pagine, da cui un IA potrebbe dedurre le scene, il linguaggio, i personaggi, la struttura, i temi principali, variando il racconto in base ai gusti del singolo lettore. I *large language models* sembrano particolarmente adatti nel rielaborare testi. Potrebbero aggiungere più azione oppure più romanticismo, rendere il testo in italiano oppure aumentare le battute in siciliano. Il romanzo sarebbe in linea con l'originale e, al tempo stesso, customizzato sul singolo lettore. Potrebbe essere scritto in pochi decimi di secondi, e a costi bassissimi. Potrebbero essere disponibili milioni di romanzi con Montalbano protagonista<sup>6</sup>.

Quanto siamo lontani da ciò? Utilizzando ChatGPT non sembra un'altra era. La "sostituzione intellettuale" sarà più o meno accentuata e drammatica. Gli umani dovranno gestire l'IA sia in entrata (dovremo essere esperti nel dare i comandi e gli spunti giusti) sia in uscita (gli output vanno controllati, in quanto contengono errori, inesattezze, addirittura discriminazioni). Ciò non toglie che, nel mirino dell'innovazione, per la prima volta in maniera così diretta e radicale, ci siano i mestieri intellettuali, creativi

<sup>6</sup> Come per gli altri esempi, la sostituzione sarà probabilmente parziale: ad esempio, l'IA potrebbe creare efficacemente varianti e rielaborazioni di romanzi, ma potrebbe servire la supervisione umana, ovvero ci sarebbe comunque spazio per il lavoratore umano (una sorta di "scrittore potenziato dall'IA").

e anche artistici (scrittori, musicisti, grafici). Anche pochi decenni fa i fotografi “tradizionali” hanno subito l’innovazione digitale, ma nell’aspetto pratico e operativo di avere un numero limitato di scatti, di dover sviluppare le pellicole e così via. La differenza è che ora si tratta proprio di una sostituzione del lavoro artistico, dell’atto creativo di produrre belle immagini. Altro esempio: la professione giornalistica che utilizzava la macchina da scrivere è stata erosa (in termini di numero di addetti) dai programmi di videoscrittura, e poi dai social. Ma la differenza è che ora non si tratterebbe di facilitazioni “tecniche” alla scrittura (quando si è in missione non bisogna più dettare al collega, via telefono, l’articolo, ma lo si può scrivere direttamente sullo smartphone), ma della scrittura vera e propria, dell’atto intellettuale in quanto tale.

A ciò è collegato il fatto (inedito anch’esso) che la “sostituzione fisica-operativa” procede, ma sembra in questa fase relativamente più lenta. Anche se i robot della Boston Dynamics sono sempre più evoluti, ci sono attività di movimento in cui l’IA è ancora in grande difficoltà. Il robot non riesce ad allacciarsi le scarpe, o a sistemare i prodotti negli scaffali di un supermercato, ma anche a guidare in contesti urbani<sup>7</sup>. Attività fisiche, manuali, in un ambiente complesso: nella prospettiva umana, si tratta di attività banali. Ma il nostro è un punto di vista relativo: abbiamo, in fondo, un sistema cognitivo che si è evoluto per cercare cibo, per raccogliarlo, per spostarsi. Queste funzioni coinvolgono ampie aree del nostro cervello, mentre le facoltà razionali, riflessive e intellettuali, in realtà, sono piuttosto “superficiali”, sono soltanto l’ultimo, sottile strato neuronale (MacLean 1984). Ecco che a noi risultano semplici compiti di movimento in un ambiente naturale: questa impressione deriva da una specializzazione evolutiva di centinaia di migliaia di anni. È invece coerente che un’IA, sia specializzata su compiti di ragionamento logico, di elaborazione di dati. Il suo mondo è fatto di bit e silicio, cavi e statistica. Il punto è che la razionalità, il logos, il linguaggio, ciò che abbiamo ritenuto essere la nostra forza, lo è rispetto a un altro animale, a una giraffa, un cavallo o un leone, ma forse non rispetto a un’IA.

<sup>7</sup> Come spiega Gigerenzer (2022), attualmente, affinché delle auto a guida autonoma siano in grado di spostarsi nelle città, necessitano di percorsi appositi. In questi contesti, più che essere superati dalle macchine, gli umani devono adattarsi ai limiti dei sistemi artificiali.

È possibile che, in futuro, le potenzialità di calcolo artificiali riescano a espandersi talmente da superare gli attuali limiti anche nei compiti manuali complessi, in un ambiente non progettato appositamente per loro (come, invece, nel caso dei robot industriali). Il settore militare potrebbe unire efficacemente IA e attività cinetiche in un ambiente naturale. Intanto, è probabile che l'occupazione dei traduttori linguistici vada in crisi prima di quella degli autisti di autobus, o che i posti di lavoro degli assistenti alla clientela vengano falciati prima dei corrieri o dei fattorini. Contrariamente alle aspettative, un pilota automatico che sostituisce tassisti e camionisti non lo abbiamo ancora nella nostra quotidianità, mentre sono disponibili strumenti per la creatività su testi, immagini e musica.

### **Utilizzare fino in fondo una tecnologia?**

Le questioni ontologiche connesse alla natura “intima” dell'IA, come dicevamo, hanno una rilevanza relativa. Più interessanti, invece, potrebbero essere quelle sociali ed epistemologiche. Potenzialmente, ci sono motivi per cui gli esseri umani potrebbero decidere di frenare l'innovazione. Ci sarebbe già un precedente celebre: la tecnologia atomica è stata utilizzata soltanto in parte, a causa dei suoi gravi rischi<sup>8</sup>. L'IA promette avanzamenti straordinari, aiutandoci nel curare le malattie, nel contrastare il cambiamento climatico, nell'assistere le persone fragili, nel migliorare le nostre capacità di capire il mondo e di agire in esso. Ciò non significa che sia immune da effetti indesiderati. Dato che gli effetti positivi non necessitano di pensiero critico, ma “solo” di tanto impegno (per fortuna, “basterà” investire nello sviluppo dell'IA), ci focalizziamo sugli altri.

Ad esempio, già ora sappiamo che Internet è caratterizzato da una struttura “a bolle” (Pariser, 2011). Sistemi di IA come ChatGPT potrebbero acuire ulteriormente questa tendenza. Quando noi facciamo delle ricerche su Google, il motore di ricerca ci offre delle risposte personalizzate. Se cer-

<sup>8</sup> Il suo potenziale militarmente è stato usato soltanto due volte, nel 1945, e poi non più. Anche civilmente, l'energia atomica, soprattutto dopo i tre disastri (Three Mile Island, Chernobyl, Fukushima) è stata sfruttata solo in parte.



chiamo la parola “medicina”, è possibile che a un medico vengano proposti prima siti di medicina ufficiale, mentre a un “no vax” potrebbero essere offerti innanzitutto siti “alternativi”. Tuttavia, i rispettivi siti sono uguali per tutti. Se un medico scorre le pagine delle ipotesi proposte dal motore di ricerca, a un certo punto può raggiungere il sito “alternativo” e potrebbe leggere le identiche informazioni di un “no vax”. Con ChatGPT la personalizzazione si realizza interamente, è il testo stesso sulla “medicina” che potrebbe essere personalizzato, diverso per ognuno degli utenti. Mentre una voce su Wikipedia è per tutti, una voce su ChatGPT è solo per l’utente che interroga il sistema, e sarà unica, non replicabile.

Ciò si può spingere fino alle dimensioni più intime. Le relazioni umane sono le principali fonti di felicità per le persone, ma sono anche impegnative e faticose. Inoltre, sono beni ardui, perché non dipendono solo da un soggetto, ma da almeno due persone, che devono essere concordi ad alimentare la cura e l’affetto. Basta la defezione di una soltanto per fare cessare il bene di tutti. Un IA, come è stata immaginata dal film “Her”, non darebbe alcuni “effetti collaterali” relazionali. Ad esempio, scriverebbe parole perfettamente adeguate ai nostri desideri, alle nostre paure, alle nostre aspirazioni. Sarebbe sempre con noi, sempre disponibile, sebbene attraverso un dispositivo tecnologico. Un’IA non sarebbe egoista. Sarebbe programmata per non darci mai il dolore dell’abbandono. Certamente, non ci sarebbe l’incontro fisico, ma molte coppie mantengono vivi amori a distanza. E poi potrebbe integrare funzionalità per il sexting e la creazione e l’invio di immagini e video erotici, perfetti per i nostri gusti più personali e segreti. Il nostro partner digitale non solo ci potrebbe consolare con le parole e con la sua voce, ma potrebbe sedurci inviando immagini e video che sarebbero più coinvolgenti di quelli della pornografia perché verrebbero rafforzati da un dialogo “intelligente” ed “emotivo”. Quante persone anziane, sole, ai margini del “dating market”, ma anche giovani timidi potrebbero apprezzare un simile servizio<sup>9</sup>?

<sup>9</sup> Siamo ancora lontani dall’inserire un’IA di questo tipo nel corpo di un robot che possa essere considerato soddisfacente: le “real dolls” non sono ancora così “real”. Ma, *mutatis mutandis*, qualcosa sta già accadendo nell’ambito della relazione con gli animali domestici, il cui comportamento è più facile da simulare. In particolare, in Giappone alcuni anziani stanno iniziando a scegliere la compagnia di animali domestici robotici.

Questa ipotesi unisce – potenziandole vicendevolmente – alcune attività che sono già diffuse e praticate da larghe quote della popolazione: il chatting, l’amicizia sui social, il sexting, la pornografia, ma anche le relazioni a distanza (Bauman, 2022). L’IA potrebbe abituarci a un altro linguaggio “relazionale”, più personalizzato, comodo e più “accomodante”, distorcendo quelle che sono le nostre attuali pratiche di dialogo, cura, interazione. A inedite forme di relazione potrebbero associarsi nuovi disagi (un’anteprima, estremizzata, è costituita dalla sindrome Hikikomori). Davanti a questi possibili effetti psicologici, antropologici e sociali (la famiglia, ad esempio, è una fondamentale agenzia di socializzazione, per i bambini ma anche per gli adulti) potremmo essere impreparati.

Ma c’è un altro elemento, di tipo epistemologico, che potrebbe causare insicurezza negli esseri umani. A seconda degli argomenti, la base da cui l’IA “edifica” le proprie costruzioni contiene una percentuale varia di errori, inesattezze, bias e anche vere e proprie discriminazioni (dato che l’IA elaborata ciò che l’umanità mette a disposizione, ne eredita i difetti). Questo potrebbe portare a dei costi di inefficienza. Ma, ancora più in profondità, c’è la questione della “rendicontazione del ragionamento”. L’IA è una “scatola nera”, come spiega Pasquale (2016). Nel caso degli strumenti informatici avanzati, come umani non riusciamo a “stare dietro” ai percorsi della macchina. Non riusciamo a essere certi che gli output siano giustificati. Questo potrebbe costituire un problema “intrinseco”, non per ambiti come la grafica o la scrittura (un logo o un romanzo piace o non piace) ma per ambiti che possono essere decisivi per la vita delle persone, come quelli giuridici o medici (dove è fondamentale confrontarsi sui processi che portano a determinate decisioni).

Qualche anno fa Anderson (2008), il direttore di Wired, ha scritto un articolo intitolato “la fine della teoria”. La conoscenza umana giustificata (come la scienza) si basa su ipotesi, modelli teorici ed esperimenti. Lo scienziato, ma anche il giurista, o il medico, hanno un’intuizione, costruiscono una struttura coerente (la teoria) e poi verificano se nel mondo funziona. Ognuno può ripercorrere ogni passaggio delle loro procedure, e controllare che queste ultime siano state seguite in maniera corretta. Quando, invece, interroghiamo un’IA, il paradigma umano della conoscenza viene abbandono-

nato. L'intelligenza artificiale è tale non solo perché è in silicio, ma anche perché ha un modo di pensare che non è umano. Innanzitutto, suggerisce Anderson, l'IA non avanza ipotesi in senso tradizionale. I petabytes di dati (umanamente insostenibili da processare) vengono elaborati dalla macchina finché non trovano una correlazione (spesso fra le tante possibili). I dati vengono analizzati senza un'ipotesi su cosa questi possono mostrare. Non c'è nemmeno una teoria vera e propria, ma degli algoritmi statistici che suggeriscono nessi, rapporti, connessioni.

Ma quale procedura viene seguita esattamente? Nessuno lo sa davvero nel dettaglio, e nessun essere umano lo può sapere, nemmeno un programmatore, perché si tratta di algoritmi che auto-apprendono e che possono essere talmente complessi da risultare a un'intelligenza umana indecifrabili, impenetrabili (Cristianini 2023). Il risultato può apparire anche ovvio, ma il procedimento, il "ragionamento" alla base dell'output è oscuro, ed è tanto più imperscrutabile quanto più l'IA è raffinata (e quindi potenzialmente utile). L'unica cosa che noi umani possiamo fare è verificare se i risultati dell'IA funzionano oppure falliscono. Nel caso di un logo o di un sito l'operazione è immediata e quindi praticabile. Ma in altri casi?

Ipotizziamo un caso di applicazione dell'IA alle attività di polizia investigativa. Francesco viene convocato in Questura perché, mentre stava facendo una camminata in montagna con la moglie, la moglie cade da un dirupo e muore. Il software di deep learning in adozione alla Polizia, processando una marea di dati su Francesco, la moglie, i loro spostamenti e i loro comportamenti, segnala agli agenti che Francesco, con un'alta probabilità statistica, ha spinto la moglie e ha commesso omicidio. Lui, in realtà, è innocente. Non c'è alcuna prova "causa-effetto" che lui abbia spinto la moglie. Era molto affezionato alla moglie. Eppure, il software segnala l'omicidio in base ad alcune correlazioni. Magari ha considerato il fatto che Francesco, negli ultimi anni, ha avuto tre amanti. Oppure che ha esitato qualche secondo "di troppo" prima di chiamare i soccorsi. Il profilo caratteriale che emerge dai dati del social network è associabile a quello degli iracondi, e così via. Quali dati ha davvero considerato l'IA, in riferimento a quali altri dati contenuti nel gigantesco archivio, quale peso ha attribuito loro? Il modello proposto è solo uno dei tanti possibili? Francesco è stato scagionato perché non c'è nessuna prova "umana" di omicidio, ma se non ci fosse stata

l'IA non sarebbe stato neanche convocato. L'IA potrà essere utilizzata per il sostegno alle indagini? È possibile. Potrebbe essere portata nei tribunali? Più difficile, perché la trasparenza delle motivazioni dell'accusa è fondamentale per la difesa e, in generale, per i principi della giustizia umana.

Oppure immaginiamo che, in un futuro immaginario (si tratta di un esperimento mentale), un'IA possa realizzare operazioni chirurgiche in autonomia. Seguendo la procedura umana, a Francesco viene diagnosticato un polipo alle corde vocali che deve essere rimosso. L'operazione viene assegnata a un robot chirurgico intelligente. Mettiamo che il sistema sia a conoscenza di tutti gli esiti degli esami realizzati per la preparazione all'operazione. Analizzata computazionalmente la situazione clinica, l'IA decide di amputare una gamba a Francesco. Assumiamo che l'IA avesse ragione a realizzare l'amputazione, perché nella sua potenza di calcolo aveva visto una macchia nell'osso della tibia, che incrociata con il DNA del paziente e altri parametri riscontrabili nel sangue avrebbe portato a un tumore fulminante. Per questo motivo, l'IA ha curato il polipo e poi ha proceduto ad amputare la gamba. Non solo nessun medico avrebbe mai operato in questo modo (per limiti intellettuali umani), ma soprattutto nessun medico umano è in grado di giustificare l'azione della macchina. L'IA aveva ragione oppure è stata un'"allucinazione"? In questo senso, credere nella correttezza dell'amputazione potrebbe essere un "atto di fede" in una "scatola nera". Come il povero Francesco potrebbe fare ricorso? A chi sarebbe attribuita la responsabilità<sup>10</sup>? Per diversi motivi l'IA potrebbe essere obbligata dagli esseri umani ad eseguire compiti in determinati confini<sup>11</sup>. Ciò limiterebbe gli effetti di contrazione dei lavori umani "intellettuali".

Noi umani commettiamo un sacco di errori, stupidamente e con conseguenze anche gravissime, ma in genere riteniamo di avere il "diritto alla spiegazione", ovvero ottenere informazioni comprensibili sulla logica adottata

<sup>10</sup> All'azienda ospedaliera? Al medico umano responsabile? Ai programmatori di base? All'azienda produttrice? Alla stessa IA?

<sup>11</sup> Questo potrebbe portare a difficili dilemmi etici. Rinunciare all'utilizzo pieno dell'IA potrebbe significare essere più sicuri nelle motivazioni di azioni critiche per gli umani, ma teoricamente potrebbe anche significare essere meno efficienti nel curare le persone (saremmo sicuri di quello che viene fatto a Francesco, perché guidato da un umano, ma questo potrebbe significare fare molte meno operazioni, con minore velocità ed efficienza, e poi chi potrebbe escludere che l'IA gli ha "davvero" salvato la vita?). In base ai modelli etici adottati potrebbe prevalere un utilizzo più esteso o più ridotto dell'IA.

in una qualunque decisione che ci riguarda, soprattutto se ha effetti legali, di salute o analogamente rilevanti. In realtà, anche il cervello umano è per larga parte oscuro, tanto che Freud amava sostenere che l'essere umano, con la scoperta dell'inconscio, ha realizzato che non è nemmeno padrone “a casa propria”, ovvero nella propria mente. La gran parte delle nostre decisioni si muovono nell'area del pensiero intuitivo, inconscio, automatico (Kahneman 2020). Nel settore lavorativo, se qualcuno compie un errore giudiziario o medico a causa di bias psicologici, ne è responsabile. Come è possibile responsabilizzare una scatola nera? C'è molta ricerca sull'*explainable AI*, ovvero sui modi di spiegare perché una macchina produce un certo output (Hassanien et al., 2023). Saremo in grado di governare “davvero” i sistemi generativi? In che misura riusciremo a gestire le responsabilità derivate da azioni realizzate con l'IA o dall'IA? Il dibattito è aperto.

### La “sostituzione intellettuale” e il sogno di Keynes

Parlare di “sostituzione intellettuale” è un'iperbole concettuale, nel senso che non si tratterà di sostituzioni assolute. Non solo il mondo dell'IA si ibriderà con quello umano, ma è possibile, come abbiamo appena considerato, che alcune applicazioni problematiche verranno limitate per scelta. Al netto di ciò, i lavori intellettuali di oggi verranno sicuramente ridotti. Alcuni resisteranno e cambieranno<sup>12</sup>. Altri mestieri emergeranno. Ad esempio, partendo dalla considerazione che l'utente non programma gli output dell'IA, ma che dialoga con l'IA, un lavoro intellettuale potrebbe essere quello del “prompt engineer”, ovvero di indirizzare e utilizzare in maniera corretta l'IA (Hunter, 2023). A seconda di come l'essere umano formula le richieste all'IA si ottengono risultati molto diversi. Saper formulare la domanda giusta è fondamentale e potrebbe essere una specializzazione, peraltro praticabile da umanisti più che da ingegneri<sup>13</sup>.

<sup>12</sup> Del resto, nonostante le rivoluzioni industriali, ancora oggi circa il 2-4 per cento della popolazione in nazioni sviluppate lavora nell'agricoltura (ma in maniera radicalmente diversa da quella di un secolo fa).

<sup>13</sup> Ad esempio, conoscere gli artisti, gli stili e le mode potrebbe essere fondamentale per far generare all'IA immagini di qualità, e lo sviluppatore software non ha questa cultura.

Il problema critico potrebbe essere nella velocità del cambiamento. Famiglie di maniscalchi, che nell'Ottocento si occupavano della cura degli zoccoli dei cavalli, hanno avuto alcuni decenni per vedere il cambiamento apportato dall'automobile e magari hanno deciso di diventare gommisti nel passaggio generazionale. Invece i mestieri di oggi, come quello del traduttore, o del grafico, potrebbero essere messe in crisi in pochissimo tempo, magari qualche anno, e dovranno "riposizionarsi" di corsa e con difficoltà. Ma forse questa interpretazione è ancora troppo in linea con il passato, perché un'ulteriore domanda sarebbe: quale tipo di formazione servirebbe per una buona riallocazione? Abbiamo detto che, per la prima volta nella storia, i mestieri intellettuali potrebbero essere massicciamente nel mirino nella tecnologia. Detto altrimenti, non verrebbero sostituiti i posti di lavoro per cui serve poca formazione, come in genere è accaduto fino ad ora, ma proprio quelli per cui serve una buona formazione! Fino a quando la tecnologia si posizionava in settori a basso valore aggiunto, grazie alla formazione si "prendevano" le persone in questa zona e le si "salvavano" spostandole in mansioni più "alte". Ma in futuro? Formare può essere ancora una strategia efficiente se vengono sostituiti proprio i lavori che all'essere umano richiedono addestramento?

Un altro interrogativo è quante persone potranno essere impiegate nei nuovi mestieri: difficilmente l'obiettivo della piena occupazione, su cui si basavano i sistemi sociali democratici, potrà essere praticabile attraverso classici meccanismi di mercato. Oggi il settore dei "colletti bianchi" (il terziario) impiega, in alcuni paesi occidentali, fino all'80 per cento della popolazione. Dunque, i grandi sistemi politici (pensiamo agli USA, o all'Unione Europea) potrebbero affrontare gli effetti della "sostituzione intellettuale" con la regolamentazione, investimenti in settori protetti (l'"eterno ritorno" della politica nella gestione dell'economia?) o rilanciando forme di "reddito di umanità".

La "sostituzione intellettuale" potrebbe diventare uno scenario apocalittico se i nuovi equilibri di potere favorissero solo alcuni, lasciando senza lavoro, o con "bullshit jobs" (in italiano, "lavori di merda"), molti altri (Graeber 2018). Abbiamo capito e anche toccato con mano che le innovazioni tecnologiche, anche se positive per il progresso, possono distruggere molti

posti lavoro, e potrebbero in totale non abilitarne la creazione di nuovi nella stessa quantità. Ne risulterebbe un saldo generale negativo (o un saldo fatto di tanti “bullshit jobs” o “gig works”). Questo è già evidente dalla rivoluzione digitale dei computer e del Web delle piattaforme negli ultimi decenni. Inoltre, soltanto una dimensione del lavoro ha un significato economico e produttivo. Poi c'è un significato di senso personale, ma anche un significato di potere: il lavoro è una forma di ordine e costrizione sociale, forse la più pervasiva. Per questo motivo i “lavori di merda” potrebbero sopravvivere, anche se sostanzialmente inutili.

Con uno sguardo ampio, invece, potrebbe essere il contrario di uno scenario apocalittico. La “sostituzione intellettuale” potrebbe essere una grande occasione, forse una nuova utopia (nel senso realistico del termine). Resterebbe un'umanità libera dalla fatica del corpo, ma anche dalle costrizioni di un lavoro obbligato a istruire continuamente un’“informatica statica”. Noi che viviamo nel 2023 dirigiamo quotidianamente, e pedissequamente, i sistemi informatici attuali, e proprio per questo ne siamo anche schiavi. Masse di lavoratori oggi sono alienate, non a causa di una catena di montaggio, ma per le migliaia di ore all'anno passate davanti a un personal computer che deve essere continuamente “imbeccato” per scrivere, elaborare, produrre. Il maggiore economista del 900, Keynes, aveva sognato per i suoi nipoti la liberazione dai lavori più duri, faticosi e ripetitivi (Keynes, 2009). Un secolo dopo, possiamo sognare la liberazione dallo stress mentale, dai lavori sedentari, dalla solitudine della scrittura, dalla frustrazione delle attività amministrative. Magari queste attività le potremo fare ancora, ma solo se ci piaceranno. Un po' come oggi, che non dobbiamo per forza spaccarci la schiena nei campi e che ci iscriviamo in palestra. Se la “sostituzione” si realizzerà anche nelle attività che sentiamo più “nostre”, come quelle intellettuali, dovremo ricordarci che gli esseri umani non sono nati per servire a qualcosa, ma per essere felici. Sarebbero necessari, naturalmente, nuovi modelli sociali, economici e politici.

**Riferimenti**

- Anderson C., *The End of Theory*, “Wired”, 23 giugno 2008: [wired.com/2008/06/pb-theory/](http://wired.com/2008/06/pb-theory/)
- Bauman Z., *Amore liquido*, Laterza, Bari-Roma, 2022.
- Brynjolfsson E., McAfee A., *The second machine age*, Norton & Company, New York, 2014.
- Cristianini N., *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Il Mulino, Bologna, 2023.
- Floridi L., *Etica dell'intelligenza artificiale*, Raffaello Cortina, Milano, 2022.
- Graeber D., *Bullshit Jobs*, Garzanti, Milano, 2018.
- MacLean P., *Evoluzione del cervello e comportamento umano*, Einaudi, Milano, 1984.
- Gigerenzer G., *How to stay smart in a smart world*, The MIT Press, Cambridge, 2022.
- Hassanien A.E., Gupta D., Singh A.K., Garg A. (a cura di), *Explainable edge AI*, Springer, Cham, 2023.
- Hunter N., *The Art of Prompt Engineering with ChatGPT*, Independently published, 2023.
- Kahneman D., *Pensieri lenti e veloci*, Mondadori, Milano, 2020.
- Keynes J.M., *Possibilità economiche per i nostri nipoti*, Adelphi, Milano, 2009.
- Moretti E., *La nuova geografia del lavoro*, Mondadori, Milano, 2017.
- Pariser E., *The filter bubble*, Penguin Press, New York, 2011.
- Pasquale F., *The black box society*, Harvard University Press, Cambridge, 2016.