

Intelligenza Artificiale e intersoggettività

Giorgio Grignaffini

Università IULM, Milano

Introduzione. Semiotica e Intelligenza artificiale : alcune evidenze

Le tecnologie digitali più avanzate, nelle loro innumerevoli e sempre più diversificate applicazioni, spesso comprese sotto il termine generico Intelligenza Artificiale (IA), sono caratterizzate da alcuni elementi che sembrano radicalmente differenti rispetto alle macchine (computer, sistemi robotici per l'industria, ecc.) che li hanno preceduti¹.

Ad essere maggiormente sotto i riflettori della stampa, della politica oltre che dei settori più tradizionalmente interessati alla tecnologia sono in questo momento i cosiddetti LLM (Large Language Model), sistemi di apprendimento automatico specializzati, come si sa, nella comprensione e nella generazione di testo naturale e in grado di “dialogare” con l'uomo, rispondere a domande, tradurre in altre lingue con una straordinaria velocità, fornendo un'esperienza simile a quella dell'interazione umana. Altri dispositivi su cui si è sviluppata una forte attenzione e anche preoccupazione riguardano i cosiddetti *deepfakes*, foto, video e audio creati grazie a software che, partendo da contenuti reali (immagini e audio), riescono a modificare o ricreare, in modo estremamente realistico, le caratteristiche e i movimenti di un volto o di un corpo e a imitare fedelmente una determinata voce.

¹ “Le intelligenze artificiali sono quei sistemi tecnologici che, in vista del raggiungimento di obiettivi specifici, possono compiere azioni con un certo grado di autonomia per mezzo di un'analisi computazionale”, C. Voto, “Interpretare il cambiamento climatico attraverso l'intelligenza artificiale”, in A. Santangelo e M. Leone (eds.), *Semiotica e Intelligenza Artificiale*, Torino, Aracne, 2023. p. 267.

Questi sistemi sembrano promettere o minacciare in breve tempo il rimpiazzo di interi corpi sociali professionali, in un'oscillazione che va dall'entusiasmo assoluto per le prospettive che si aprirebbero per la specie umana fino ad arrivare alle proiezioni più apocalittiche che vedono addirittura rischi per la sopravvivenza della nostra civiltà². Come afferma Ferraro “molti scritti sull'IA, con artifici retorici talvolta anche banali ma comunque evidentemente efficaci, ne propongono una visione mitizzante, giocata sull'immagine di un futuro dal sapore fantascientifico (...) ; tanto l'immaginario fantascientifico si ispira agli avanzamenti tecnologici quanto questi ultimi prendono spunto, sempre più decisamente, dalle proposte dell'immaginario”³.

La semiotica ha immediatamente messo sotto i riflettori dell'analisi questo fenomeno allo stesso tempo tecnologico e sociale così ricco di implicazioni e lo ha fatto concentrandosi sui molteplici aspetti che lo caratterizzano. In particolare, molte delle analisi si sono interrogate sul significato stesso dell'espressione “intelligenza artificiale”, e sui due termini che la costituiscono. Tra le domande più frequenti, quelle riguardanti il concetto di “intelligenza” e come può essere applicato questo termine a qualcosa di non umano⁴, quali siano le sue potenzialità e i suoi limiti, se e in che modo chi l'ha progettata e la usa ha proiettato su essa la propria ideologia⁵; che cosa si intende per “artificiale” e in che rapporto sta la tecnologia con il “naturale”⁶; le modalità in cui l'IA opera e il modo in cui i suoi utilizzatori la usano; l'impatto sulle varie attività umane di questa tecnologia, in particolare riguardo alla capacità della tecnologia di imitare sia i comportamenti che i prodotti dell'attività creativa umana; i discorsi che su di essa circolano nel sociale⁷.

1. Un approccio strutturalista

Quello utilizzato in tutti i lavori appena menzionati è un approccio certamente interessante e in grado di aprire ulteriori piste di analisi, ma nella maggior parte dei casi si mette al centro dell'analisi o l'oggetto “intelligenza artificiale” o i suoi utilizzatori umani, analizzandoli “in sé” o, ancora, cercando di trovare somiglianze e differenze nel loro funzionamento. Quello che meno spesso si è fatto è invece mettere al centro dell'attenzione la *relazione* tra questi due termini. Come ricorda Landowski, “A la base de la théorie sémiotique de la signification, une des idées les moins contestées est que la relation prime sur les termes qui en sont

2 Su questi argomenti contrapposti cfr. A. Santangelo, “Essere umani al tempo dell'Intelligenza Artificiale. Narrazioni a confronto” in A. Santangelo e M. Leone (eds.), *op. cit.*

3 G. Ferraro, “Mito e realtà dell'AI. Uno sguardo semiotico”, in A. Santangelo e M. Leone (eds.), *op. cit.* p. 67.

4 E. Sadin, *L'Intelligence artificielle ou l'enjeu du siècle. Anatomie d'un antihumanisme radical*, Paris, L'échappée, 2018.

5 Sui rapporti tra ideologie e produzione di immagini generate da AI, cfr. M. Leone, “L'idéologie sémiotique des deepfakes”, *Interfaces numériques*, 11, 2022.

6 G. Ferraro, “Mito e realtà dell'AI”, *art. cit.*, p. 88.

7 Su questi argomenti cfr. *Semiotica e Intelligenza Artificiale*, *op. cit.*

les aboutissants”⁸. Questo primato della relazione rispetto ai termini primari permette di riflettere sulle modalità di interazione che si instaurano tra questi due poli, per provare a definirne le peculiarità.

Quindi, da un punto di vista semiotico, nella relazione uomo / IA l’importante non è determinare tanto la natura dell’uno o dell’altra (la ricerca di che cos’è l’IA o di cos’è l’intelligenza umana), ma capire quale relazione si instaura tra essi, perché è nella relazione che si definiscono l’una rispetto all’altra.

Quella che si viene a costruire quando l’essere umano entra in contatto con un dispositivo tecnologico avanzato, non solo con i LLM, ma anche i sempre più numerosi oggetti di uso comune in cui vengono incorporati componenti elettroniche che funzionano grazie all’IA può essere definito qualcosa di più di un semplice “uso” di uno strumento. Rispetto alla pura strumentalità della macchina utensile, nel caso del rapporto con questi dispositivi quella che si realizza appare a tutti gli effetti un’interazione tra soggetti, salvo che uno dei due soggetti è una macchina. Infatti, come afferma Brandt,

C’est l’échange possible d’informations avec une entité physique — qui, grâce à l’installation de routines symboliques de dialogue, répond utilement à nos demandes, ou demande nos réponses — qui nous fait reconnaître dans cet artefact inanimé une sorte de partenaire capable de nous instruire et digne de notre attention intentionnelle.⁹

A questo proposito è molto importante considerare quali possono essere le interfacce più comuni con cui le macchine che incorporano l’IA si mettono in relazione con l’uomo. Al momento l’IA è integrata non solo nei sistemi informatici di cui abbiamo parlato in precedenza, legati alla generazione di testi o immagini fruibili attraverso gli schermi dei computer o dei telefoni cellulari, ma anche in strumenti con cui abbiamo la possibilità in quanto esseri umani di interagire con tutto il nostro corpo. Oltre ai veicoli a guida autonoma¹⁰ che sono tra i più promettenti prodotti su cui l’industria dell’automobile sta investendo, esistono altri dispositivi come i robot, termine generale il cui uso spazia dall’industria, in cui sostituiscono ormai sempre di più l’attività ripetitiva dell’operaio, all’agricoltura, alla medicina, in cui la tecnologia ha reso possibile l’esecuzione di interventi chirurgici con tecniche minimamente invasive, fino all’intera area della vita quotidiana, dove l’IA ha permesso la creazione di veri e propri assistenti che svolgono compiti come il servire a tavola, pulire i pavimenti, intrattenere le persone.

Ora, tutti questi strumenti tecnologici incorporano al loro interno l’IA che funziona permettendo al dispositivo di agire non in seguito a una serie di istruzioni date una volta per tutte, ma attraverso un apprendimento costante delle pratiche attraverso l’interazione con l’uomo e con l’ambiente circostante. Di qui

8 E. Landowski, “Aspects sémiotiques du changement”, *Acta Semiotica*, III, 6, 2023, p. 30.

9 P. Aa. Brandt, “Sens et machine. Vers une techno-sémiotique”, *Actes Sémiotiques*, 121, 2018.

10 I veicoli autonomi o le auto a guida autonoma funzionano grazie all’aiuto dell’intelligenza artificiale, che a sua volta consiste in algoritmi di autoapprendimento (<https://www.iris-display.com/info/veicoli-autonomi-e-ai.html>).

una costante dinamica, non prevedibile a priori, di azione e retroazione che si viene a creare tra macchina, uomo e ambiente. Per fare un esempio concreto, prendiamo proprio l'auto a guida autonoma. Esse tengono infatti sotto controllo il traffico e la segnaletica stradale, anticipano le frenate in caso di pericolo e possono integrare o sostituire del tutto la presenza umana ; inoltre, quando fungono da taxi interagiscono con il passeggero che sale a bordo, assecondandone desideri e necessità. Per fare tutto ciò utilizzano applicazioni IA in grado d'imparare e modificare il proprio comportamento con l'esperienza.

Quello che è più interessante però in ottica semiotica è osservarne il comportamento in relazione con il soggetto umano che si trova ad incontrarne una sulla sua strada : il pedone sulle strisce pedonali, quando improvvisamente si trova di fronte a un'auto senza guidatore, deve interagire con un oggetto inanimato che però adotta un comportamento totalmente equiparabile a quello umano, senza però, ad esempio, poter interpretare possibili reazioni fisiognomiche sul volto del guidatore. Si tratta quindi di un'interazione allo stesso tempo uguale e diversa rispetto a quella che avviene tra soggetti umani, e che, tuttavia, si può cercare di analizzare con gli stessi strumenti della semiotica interazionale¹¹.

2. Una relazione intersoggettiva

Perché si possa parlare di relazione intersoggettiva nel caso dei soggetti umani è necessario che vi sia una riconoscibilità reciproca tra essi : i modi con cui essi si confrontano o si affrontano si fondano sul presupposto che ciascuno veda in chi ha di fronte un soggetto dotato di una serie di caratteristiche che possano essere riconducibili al campo delle modalizzazioni (sapere, volere, potere, dovere).

Quando un soggetto opera nei confronti di un altro secondo il regime della *programmazione*, della *manipolazione*, dell'*aggiustamento* e, per certi versi anche dell'*alea*, mette in atto una serie di attività che si basano sul presupporre come la propria azione potrà influire sull'altro. Quando un soggetto, ad esempio, opera in modo "programmatorio" o "manipolatorio" o cerca di "aggiustarsi" ad un altro soggetto, postula che vi sia (almeno in linea teorica) una possibile reversibilità dei ruoli. Il manipolato può diventare un manipolatore ; nel caso dell'*aggiustamento* la reciprocità è strutturale, e anche il soggetto *programmato* può a sua volta, mutate le condizioni di base, agire come "programmatore" di un altro soggetto.

In particolare, nel caso della *programmazione* — il regime che più può essere avvicinato al modo in cui un soggetto "usa" una macchina — il soggetto attivo S1 trasforma il "sapere" e il "poter fare" del soggetto S2 in un "dover fare", piegando il "volere" di quest'ultimo al proprio. In questo modo l'azione del soggetto S1 tende a considerare l'altro, S2, come un partner che a fronte di determinati input darà origine a reazioni sempre prevedibili, come nel caso degli ordini impartiti da un comandante di un plotone ai suoi sottoposti. Viene annullato, almeno in

11 Ci riferiamo alla teoria sviluppata da E. Landowski. Cfr. *Rischiare nelle interazioni* (2005), Milano, FrancoAngeli, 2010.

linea teorica, ogni margine di interpretazione o retroazione di S2 nei confronti di S1. Nel caso delle macchine che incorporano l'AI, al contrario, abbiamo a che fare con un Soggetto S2 (la macchina) che ha la capacità di analizzare e interpretare (dando a questo termine il significato comune di “dedurre da indizi o da parole i pensieri e le intenzioni di una persona”) quanto gli viene comunicato e tradurlo in un'azione che a differenza delle macchine tradizionali, non è prevedibile totalmente a priori.

Un sistema che impara dalla propria esperienza a reagire, adattandosi ai contesti in cui agisce è paradossalmente più libero di un soggetto umano che, come S2 nel caso del regime della *programmazione*, è obbligato a ripetere sempre gli stessi gesti, applicando uno schema che gli viene indicato da S1. La catena di montaggio di stampo fordista annulla le modalizzazioni dell'operaio che si trova ad agire solo secondo il “dover fare” imposto da S1, alla stregua di una macchina che, programmata a compiere una serie di operazioni, anche molto complesse, non può (o non dovrebbe) deviare mai dal compito prefissato ; al contrario quando ci si trova di fronte a un sistema di IA (come un generatore di testi o di immagini) all'input fornito dal soggetto umano reagisce fornendo un output sempre diverso.

Nel caso dell'IA o di sistemi sofisticati come la guida autonoma si verifica quindi una dinamica interazionale peculiare. I due poli (umano e macchina) sono definiti da una relazione che da una parte è diversa da quella che intercorre tra uomo e macchina tradizionale, ma dall'altra sembra invece ricreare le condizioni di una relazione tra soggetti umani proprio a causa della (relativa) imprevedibilità della reazione della macchina. Infatti, maggiore è la sofisticazione dei sistemi con cui abbiamo a che fare, minore è il grado di comprensibilità dei meccanismi che li guidano e maggiore è il grado di inatteso che può nascondersi dietro alle risposte che essi danno all'interlocutore.

Questa componente di imprevedibilità che caratterizza il regime di interazione uomo macchina che sfugge alla definizione di *programmazione*, si fonda come abbiamo visto sul presupposto che in un'interazione tra soggetti umani, ci si rapporti con un soggetto di cui riconosciamo una comune matrice umana.

Pur nell'imprevedibilità dei comportamenti dei singoli, ciascuno di noi apprende gli schemi di azione e reazione dell'essere umano a livello non solo cognitivo, ma anche patemico, una dimensione che è estranea ai sistemi IA ; inoltre, gli schemi di azione e le passioni nel caso dei soggetti umani sono inseriti in una corporeità che li indirizza e li condiziona. Pensiamo ad esempio alla relazione intersoggettiva che si instaura nel regime della *manipolazione* : l'azione di S1 è finalizzata a un far fare a S2 una determinata azione attraverso strumenti modali complessi, determinati anche dalla corporeità attraverso cui si articolano, e socialmente inquadrati in schemi a cui sono stati dati i nomi di seduzione, minaccia, promessa ecc. E in modo ancora più significativo nel caso del regime dell'*aggiustamento* tra soggetti umani la possibilità di trovare un accordo passa attraverso il riconoscere da parte di S1 nell'altro una dimensione anche corporea o comunque un agire coordinato in S2 su cui S1 si sincronizza.

Persino nel regime dell'*alea*, in cui a dominare è la totale imprevedibilità del caso, abbiamo comunque una relazione che avviene con un soggetto di cui riconosciamo fisionomia e modalità di azione : nell'*alea* è l'incontro tra S1 e S2 che è inaspettato, ma la relazione che si instaura viene gestita dal punto di vista di un soggetto che si trova a interagire con questa dimensione e cerca di governarla.

Nel caso dell'IA invece come abbiamo visto il soggetto umano S1 interagisce mettendo in atto un comportamento che è modellato su quello dell'interazione tra umani (con Chat GPT si conversa, si pongono domande, si attendono risposte ; con un'auto a guida autonoma si danno indicazioni se si è a bordo, oppure ci si confronta come se a guidarla fosse un essere umano se la si incontra sulla strada). Quindi a questo livello si tratta a tutti gli effetti di una vera e propria interazione intersoggettiva. Ma con questi sistemi che in realtà non sono umani e non hanno corporeità (anche quando sono applicati a robot si tratta di una corporeità simulata), il soggetto umano si trova nella ambigua situazione di interagire con le stesse modalità dell'interazione tra umani, pur sapendo che è una macchina, cioè con un dispositivo con cui da sempre siamo abituati a relazionarci in termini di perfetta causalità di stimolo (il nostro input) e risposta (l'output della macchina stessa).

La relazione si instaura in questo caso con un apparato che agisce su nostra sollecitazione o reagendo a un nostro stimolo, ma che dà risposte non programmate e a volte apparentemente casuali o errate : è quello che accade ad esempio quando a fronte di una domanda il sistema LLM risponde dando luogo alle cosiddette allucinazioni, cioè risposte date dal sistema IA che non rispondono alla verità storica o scientifica, come ad esempio biografie di personaggi conosciuti che vengono ricostruite inserendo molti dati falsi o comunque riconducibili a logiche che non sono conoscibili (a detta degli stessi programmatori) ; o quello che è accaduto recentemente a San Francisco quando una donna investita da un'auto con conducente e sbalzata contro un taxi senza guidatore, è rimasta intrappolata sotto il veicolo governato dall'IA che l'ha trascinato per metri invece di fermarsi. In questi casi non si tratta in realtà di un'*alea*, di un semplice casuale inconveniente tecnico, ma di un funzionamento inatteso, non prevedibile dal soggetto umano. Sembra venir meno il principio base del rapporto uomo-macchina, perché ora queste macchine sembra possano agire autonomamente e non soltanto "retroagire" ad uno stimolo fornito dall'uomo e anzi, come nel caso dell'auto, non compiere la manovra per far uscire la vittima dell'incidente da sotto l'auto, manovra che i soccorritori arrivati sul posto richiedevano.

Quello che emerge è quindi che la relazione uomo-macchina legata al modello deterministico stimolo-risposta, sembra lasciare il posto a una nuova modalità di interazione. Essa si fonda sul fatto che a queste nuove macchine che incorporano l'IA si riconosca una competenza "quasi umana" perché simulano il comportamento umano (guidano, scrivono o producono immagini come gli esseri umani) e nello stesso tempo "sovrumana" (la macchina riesce a fare con velocità e precisione incredibile quello che già fanno gli umani o che gli umani non potrebbero fare per i loro limiti fisici), ma anche in un certo modo

si presentino come “diversamente umane”, nel senso che sembrano ragionare secondo schemi che non riusciamo a cogliere fino in fondo e che proprio per questo motivo a volte, inaspettatamente, sbagliano, anche quando il soggetto umano ha fornito loro tutti gli input richiesti in modo corretto.

3. Il livello modale

Abbiamo parlato di competenza, di un presupposto all’azione che la macchina sembra condividere con l’umano : ma a livello modale a quale competenza ci si riferisce ?

Nell’azione della macchina è implicito certamente un “saper fare” e un “poter fare” come accade in ogni dispositivo tecnico. La macchina da sempre esiste perché rende possibile un’azione che l’uomo non saprebbe o non potrebbe fare nel modo accurato o veloce o faticoso che la macchina riesce invece a svolgere senza problemi. Ma come abbiamo visto questo tipo di competenza è quella che presiede alle macchine che rientrano dal punto di vista dei regimi interazionali sotto quello della *programmazione*. Si tratta quindi delle modalità attualizzanti secondo la riflessione greimasiana, alle quali si affiancano però quelle virtualizzanti rappresentate dal “volere” e dal “dovere”.

Queste ultime nello schema narrativo canonico sono quelle che spingono il soggetto all’azione, la quale viene resa possibile dalla successiva applicazione delle modalità attualizzanti. Nel caso del rapporto uomo-macchina tradizionale, all’uomo sono riservate le modalità virtualizzanti : l’azione si compie perché l’uomo “vuole” o “deve” fare qualcosa, mentre la macchina serve a rendere “attuali” la necessità o il desiderio.

Nel caso dei dispositivi contenenti l’AI questa separazione non è più così evidente.

La macchina di questo tipo infatti mantiene al suo interno una parte di imprevedibilità e inconoscibilità dei suoi meccanismi : “L’intérieur de la machine, dans ce cas pourtant bien simple, peut en principe rester entièrement inconnu pour l’usager, comme une ‘boîte noire’, sauf pour un spécialiste de son entretien et de sa réparation”¹². All’interno di questa scatola nera, sembra poter risiedere una sorta di “voler fare” che non dovrebbe appartenere a una macchina : e questo “voler fare” (o il “non voler fare”) appare particolarmente difficile da integrare nella nostra esperienza.

Contrairement à l’outil, qui ne réalise convenablement sa tâche qu’étroitement tenu en main, la machine est par construction un partenaire auquel on peut dans une grande mesure “lâcher la bride” : n’a-t-elle pas été conçue de telle sorte que précisément, une fois mise en mouvement, elle puisse fonctionner d’elle-même, sur le mode de l’automatisme ? La faire marcher, c’est la *laisser faire* ce qu’elle est faite pour exécuter. Ce qu’elle “sait” faire, ou plutôt ce qu’elle peut faire — ce pour quoi elle est programmée — lui assure une sorte de domaine et de temps réservés à l’intérieur desquels il faut en somme la laisser “libre” d’agir. Evidemment pas “à

12 P. Aa. Brandt, “Sens et machine. Vers une techno-sémiotique”, *art. cit.*

sa guise” mais conformément à ses algorithmes de fonctionnement. Si cela n’en fait pas un sujet autonome, cela suffit du moins pour lui conférer le statut d’un *presque-sujet* doté de rôles normés (dits “thématiques”) commandant l’accomplissement de performances prédéfinies, et en principe constantes.¹³

Riconoscere a un dispositivo tecnologico una volontà, significherebbe dotarlo di una natura di attante paritario rispetto all’uomo (ma in realtà con una forza molto superiore in quanto sapere e potere sono sovrumani per la velocità di risposta e la capacità di elaborazione). Ma nello stesso tempo, renderebbe reversibile una relazione che fino ad oggi è stata monodirezionale: l’uomo che “vuole” o “deve” realizzare un’azione si affida alla macchina per ottenere il risultato, ma ora le due modalità virtualizzanti sembrano diventare appannaggio anche dei dispositivi che incorporano l’IA. Quindi a fronte di un soggetto umano che “vuole” o “deve” fare, si oppone un soggetto macchina dotato di un suo “anti-volere” non prevedibile o non riconducibile a un’istruzione che l’uomo ha dato, un volere che sembra essersi generato autonomamente.

Dans divers secteurs de pointe, surtout ceux, de plus en plus étendus, où domine l’électronique, le coup de main ponctuel, technique, bien délimité, tend ainsi à céder la place à une prise en main globale et durable, la machine-adjuvant se transformant peu à peu en un destinataire autorégulé, capable de jugement et d’initiative.¹⁴

Ora, con l’avvento dell’AI, questa preoccupazione che a inizio millennio era ancora a livello di suggestione sta diventando sempre più realtà.

E oltre al “volere” quello che sembra creare ancora più timore è la possibilità che una macchina possa essere dotata anche di un “dovere”, cioè che possa trovare delle “motivazioni” nell’azione che risiedono in un qualche input che non è stato imposto dall’uomo. Come osserva Santangelo: “chi svolge la funzione narrativa che Greimas attribuisce al Destinante, vale a dire di indicare la direzione verso cui andare, il valore delle cose e delle azioni da compiere? Sono gli esseri umani che dettano il proprio punto di vista alle macchine o è il contrario?”¹⁵. Siamo di fronte a scenari forse fantascientifici, all’immagine del robot che come in universi distopici rappresentati in romanzi e film, prende decisioni sulla base non dell’interesse dell’uomo ma di un proprio volere e necessità, un attante che non ha più un destinante umano ma un proprio autodesinante.

Se la macchina è solo programmabile non fa paura, perché la relazione che si instaura si limita alla *programmazione* o in qualche caso a forme di *aggiustamento* come nel caso del guidatore di una vettura con la quale si instaura un’interazione che ha forti somiglianze con quella del rapporto con l’essere umano; come ricorda Landowski quando mi accade non solo di “usare” un’auto ma di “praticarla”, “je lui donne la possibilité de déployer ses potentialités — du moins certaines

13 E. Landowski, *Avoir prise, donner prise, Actes Sémiotiques*, 112, 2009 (III.II.1); tr. it., *Lexia*, 3-4, 2009.

14 E. Landowski, “La part des choses”, *Protée*, 29, 1, 2001, p. 4.

15 A. Santangelo, “Essere umani al tempo dell’Intelligenza Artificiale”, *art. cit.*, p. 241.

d'entre elles, à proportion de ma propre aptitude à les deviner, à les solliciter, à les laisser s'exprimer"¹⁶.

In questo caso al massimo nell'interazione con essa si può incorrere in un incidente perché si è sbagliato a programmarla (anche nell'*aggiustamento* : se guidando mi fido troppo della mia sensibilità e non presto attenzione posso rischiare di uscire di strada). Ma in fin dei conti è un errore a cui si può porre rimedio adottando strategie di mitigazione dell'errore : “si vede bene infatti a cosa sono dovuti questi insuccessi : o a qualche quiproquo che ha portato alla scelta di un regime di interazione inadeguato alla circostanza, o, una volta scelto il regime di interazione adeguato, a un errore tecnico nell'esecuzione del progetto”¹⁷.

Se invece la macchina non risponde e apparentemente può decidere (“volontariamente” ?) di dare un feedback imprevisto al nostro input allora finisce che ad essa attribuisco una dimensione quasi animista.

4. Problemi di simulazione e di fiducia

A questo punto il problema può essere visto ad altri due livelli : la simulazione dell'interazione e la fiducia nell'interlocutore.

Infatti, questa trasformazione dello statuto modale della macchina che passa da essere esecutore dotato di un “sapere” e un “potere”, a diventare, agli occhi di qualcuno, un soggetto potenzialmente autodestinante, dotato quindi di un “volere” e di un “dovere”, nasce dal tipo di interazione che l'uomo attiva con essa e risulta essenzialmente di un'interpretazione superficiale se non illusoria dei principi di funzionamento del sistema.

Quest'ultima si fonda su un principio di simulazione : la macchina viene costruita dall'uomo perché simuli l'attività e anche in molti casi le sembianze umane, o meglio perché l'uomo si possa relazionare con essa in un modo quanto più “naturale” possibile.

Nei sistemi LLM l'uomo simula di avere a che fare con un interlocutore umano a cui può rivolgere domande, chiedere di scrivere testi, tradurre da lingue straniere, usando il linguaggio naturale nel modo più semplice. E il risultato che si chiede alla macchina non è più un modello standard, ma un output ogni volta diverso. Se ad esempio si chiede diverse volte al dispositivo di scrivere un riassunto di un testo o di rispondere a una domanda, quello che il sistema restituisce è ogni volta differente. Quindi l'interazione non si basa su una generazione standardizzata di risultati e infatti non è necessario ricorrere a linguaggi di programmazione specifici ma al linguaggio naturale. E allo stesso modo anche quando si ha a che fare non con un modello linguistico ma con un dispositivo dotato di un'interfaccia che si relaziona in modo corporeo all'uomo, come ad esempio con un'auto a guida autonoma o con un robot più o meno antropomorfo, il tentativo dei progettisti è di realizzare delle apparecchiature che rendano l'interazione uomo-macchina il più possibile simile a quella uomo-uomo.

16 E. Landowski, “Voiture et peinture : de l'utilisation à la pratique”, *Galáxia*, 24, 2012, p. 244.

17 *Rischiare nelle interazioni*, op. cit., p. 65.

Di qui discendono alcune conseguenze, in primo luogo l'idea che dietro l'indubbia innovazione in certo modo rivoluzionaria di questo sistema tecnologico, vi sia in fondo anche una dimensione discorsiva fondata appunto su una simulazione dell'attività umana e di conseguenza anche una retorica della capacità sostitutiva dell'uomo da parte della macchina. Ricreare delle condizioni di interazione uomo-macchina così abilmente e sorprendentemente simili a quelle umane, ci stimola a vedere nella macchina un qualcosa che va oltre la semplice dimensione tecnica (appunto quella relativa alle modalità attualizzanti) e che arriva invece a adombrare l'esistenza di una dimensione volitiva e deontica fin qui mai sperimentata in un meccanismo tecnico creato dall'uomo. Perfettamente coerente con questa retorica simulativa è anche il termine stesso "Intelligenza Artificiale" che, mettendo sullo stesso piano una qualità prettamente umana quale l'intelligenza e una invece di pertinenza non naturale come è per sua definizione il termine "artificiale", ci spinge ad ipotizzarne una corrispondenza inedita e per questo allo stesso tempo rivoluzionaria e potenzialmente minacciosa.

Un'altra conseguenza dell'aver costruito le macchine legate a questa nuova dimensione tecnologica in termini di somiglianza con le facoltà umane e quindi dell'aver postulato la possibilità di un'interazione in qualche misura paritaria tra uomo e macchina, è relativa alla dimensione fiduciaria.

Quest'ultima si basa sull'aspettativa da parte di S1 di un comportamento di S2 che sia prevedibile e aderisca a quanto abbiamo stabilito più o meno esplicitamente con esso preliminarmente. Anche in questo caso ci rifacciamo alla teoria delle modalità nel senso che nel rapporto fiduciario S1 si attende che l'agire di S2 sia sovradeterminato da una volontà e una necessità che non si modificano rispetto a quanto pattuito. Se come docente affido un compito da svolgere a uno studente mi posso fidare di lui quando il mio volere che il compito sia svolto coincide con la sua volontà di svolgerlo e allo stesso tempo il suo dovere lo spinge a mantenere questo impegno. Se invece non posso entrare in contatto con la "scatola nera" della macchina e non sono sicuro che essa aderisca o meno a un sistema di valori o che agisca in base a interessi razionalmente calcolabili o invece sia in preda a qualche pulsione o desiderio, non posso fare ipotesi sul suo comportamento, nemmeno errate.

È naturalmente possibile ingannarsi sulla buona fede di una persona con cui interagiamo, ma almeno a posteriori riusciamo a ricostruire se il tradimento delle aspettative è stato causato da un calcolo utilitaristico o invece da un imprevisto movimento patemico che ha portato il nostro interlocutore a infrangere il patto che avevamo con lui. Ma nel caso di una macchina di cui non conosciamo lo statuto "morale" (anche ammettendo che ci sia) e che almeno a priori non è dotato di passioni, a cosa attribuire la *défaillance* ? A qualche errore di programmazione ? ma anche i programmatori stessi molto spesso non sanno dove intervenire per correggerli perché le macchine sono troppo complesse e soprattutto perché apprendono da sole e quindi diventano progressivamente diverse da come sono state programmate. O in modo più inquietante ci sarebbe, come dicono alcuni, qualcosa di assimilabile ad una volontà che sfuggirebbe all'uomo

e che progressivamente potrebbe andare a costituire la macchina come Soggetto indipendente dal suo creatore ?

Conclusioni

Il problema è quindi di fiducia, nella macchina e allo stesso tempo nell'uomo che l'ha creata e che però rischia di perderne il controllo. Il mito di Frankenstein o del Golem aleggia in queste visioni apocalittiche, ma come abbiamo notato all'inizio, l'IA come tutti i fatti sociali è in realtà sempre di natura discorsiva e come tale non esiste in sé, prima o al di là delle costruzioni narrative e ideologiche che le danno esistenza sociale. Infatti, non c'è bisogno di antropomorfizzare la macchina, attribuendola coscienza e volontà, per ammettere che ella possa, in un certo senso, tecnicamente oltrepassare le capacità cognitive individuali. È pertanto necessario distinguere tra ciò che pertiene al modo stesso di funzionamento di un sistema complesso i cui principi specificamente semiotici devono ancora essere identificati e quello che ne consegue in termini ideologici e anche passionali di superficie.

Bibliografia

- Brandt, Per Aage, “Sens et machine. Vers une techno-sémiotique”, *Actes Sémiotiques*, 121, 2018.
- Ferraro, Guido, “Mito e realtà dell'AI : uno sguardo semiotico”, in A. Santangelo e M. Leone (eds.), *Semiotica e Intelligenza Artificiale*, Torino, Aracne, 2023.
- Landowski, Eric, “La part des choses”, *Protée*, 29, 1 (*La société des objets*), 2001 ; tr. it. in E. Landowski e G. Marrone (eds.), *La società degli oggetti*, Roma, Meltemi, 2002.
- *Rischiare nelle interazioni* (2005), Milano, FrancoAngeli, 2010.
- *Avoir prise, donner prise*, *Actes Sémiotiques*, 112, 2009 ; tr. it., *Lexia*, 3-4, 2009.
- “Voiture et peinture : de l'utilisation à la pratique”, *Galáxia*, 24, 2012 ; tr. it., “Vettura e pittura : dall'utilizzo alla pratica”, in S. Jacoviello, T. Lancioni et al. (eds.), *Texture. Scritti seriosi e schizzi scherzosi per Omar Calabrese*, Siena, Protagon, 2009.
- “Aspects sémiotiques du changement”, *Acta Semiotica*, III, 6, 2023.
- Leone, Massimo, “L'idéologie sémiotique des deepfakes”, *Interfaces numériques (Images, mensonges et algorithmes : La sémiotique au défi du Deep Fake)*, 11, 2022.
- Sadin, Eric, *L'Intelligence artificielle ou l'enjeu du siècle. Anatomie d'un antihumanisme radical*, Paris, L'échappée, 2018.
- Santangelo, Antonio, “Essere umani al tempo dell'Intelligenza Artificiale. Narrazioni a confronto”, in A. Santangelo e M. Leone (eds.), *Semiotica e Intelligenza Artificiale*, Torino, Aracne, 2023.
- Santangelo, Antonio e Massimo Leone (eds.), *Semiotica e Intelligenza Artificiale*, Torino, Aracne 2023.
- Voto, Cristina, “Interpretare il cambiamento climatico attraverso l'intelligenza artificiale”, in A. Santangelo e M. Leone (eds.), *Semiotica e Intelligenza Artificiale*, Torino, Aracne, 2023.

Résumé : L'article analyse le thème de l'Intelligence Artificielle et de ses rapports avec l'homme en partant du principe fondamental de la sémiotique structurale : la priorité de la relation sur les termes. D'où la focalisation sur les relations, ici considérées comme intersubjectives, qui s'établissent entre utilisateurs et machines comportant des dispositifs d'IA. A la différence des machines traditionnelles, elles vont au-delà de la simple exécution des tâches pour lesquelles

elles ont été programmées. Elles semblent rétroagir face à l'acteur humain en mettant en œuvre des compétences modales qui dépassent le savoir et le pouvoir faire et impliquent aussi le devoir et le vouloir, modalités jusqu'à présent regardées comme spécifiquement humaines. Ces caractéristiques rendent la problématique interactionnelle de Landowski pertinente pour la description de ces relations complexes.

Mots clefs : confiance, intelligence artificielle, régimes d'interaction, sujet.

Resumo : O artigo analisa o tema da Inteligência Artificial e de suas relações com o usuário a partir do princípio fundamental da semiótica estrutural : a prioridade das relações sobre os termos. Por isso, o estudo focaliza as relações, aqui consideradas como intersubjetivas, que se estabelecem entre homem e máquinas dotadas de dispositivos de IA. Essas últimas, diferentemente das máquinas tradicionais, vão além da simples execução das tarefas pelas quais foram programadas. Parecem retroagir face ao ator humano mediante competências modais que ultrapassam o saber e o poder fazer e envolvem também o dever e uma forma do querer, modalidades vistas até hoje como especificamente humanas. Essas características tornam o modelo interacional de Landowski relevante para analisar a complexidade do caso.

Abstract : The article analyzes the topic of Artificial Intelligence and its relations with humans, starting from the principle underlying structural semiotics, namely the priority of the relation over the terms. So the focus of investigation is placed on the intersubjective relationships that are built between humans and machines equipped with Artificial Intelligence devices. Compared with traditional machines, these are characterised by going beyond the pure performance of tasks for which they have been programmed : they seem to retroact with respect to the human subject by displaying a modalisation that goes beyond “knowing how to do” and “being able to do”, involving also “duty” and “will”, modalities considered as specifically human until now. Hence the relevance of Landowski's interactional model for the analysis of this intriguing type of human / machine relationships.

Riassunto : L'articolo analizza il tema dell'Intelligenza Artificiale e dei suoi rapporti con l'uomo partendo dal principio che sta alla base della semiótica strutturale, cioè la priorità della relazione rispetto ai termini che la compongono. Viene così messo al centro dell'indagine il tema dei rapporti intersoggettivi che si vengono a costruire tra uomo e macchine dotate di dispositivi di Intelligenza Artificiale. Queste ultime, rispetto alle macchine tradizionali, si caratterizzano per andare oltre il puro svolgimento dei compiti per i quali sono state programmate : esse sembrano retroagire rispetto al soggetto umano mettendo in mostra una modalizzazione che va oltre il sapere e il poter fare, coinvolgendo anche il dovere e il volere, modalità finora tipicamente umane. In questo modo è possibile applicare alla relazione uomo / macchina anche il modello interazionale di Landowski.

Auteurs cités : Per Aage Brandt, Guido Ferraro, Eric Landowski, Massimo Leone, Eric Sadin, Antonio Santangelo, Cristina Voto.

Plan :

Introduzione. Semiotica e Intelligenza artificiale : alcune evidenze

1. Un approccio strutturalista
2. Una relazione intersoggettiva
3. Il livello modale
4. Problemi di simulazione e di fiducia

Conclusioni

Recebido em 15/03/2024.

Aceito em 14/05/2024.