



ARTICOLO

L'intelligenza artificiale è realmente intelligente?

Un percorso tra embodiment e macchine linguistiche

Francesco Maria Cianciaruso

La storia dell'intelligenza artificiale, dopo due decenni di sviluppi teorici e ingegneristici, sta attraversando una nuova svolta epocale. La diffusione dei modelli neurali sta rappresentando, infatti, un *game-changer*, sia in virtù delle incredibili possibilità di implementazione, dal riconoscimento di immagini alla produzione di linguaggio, che per le implicazioni filosofiche che tale diffusione porta con sé, dall'etica applicata all'epistemologia. Il presente scritto intende muovere proprio a partire da queste ultime. La competenza di macchine dalla configurazione 'neuromorfica', basate su reti di neuroni artificiali che trattano l'informazione in maniera sub-simbolica e dinamica, potrebbe infatti permettere di rimettere in discussione alcuni assunti del dibattito classico sull'intelligenza artificiale (e sull'intelligenza in generale) circa il rapporto tra sintassi, significati e substrati che li trasportano. Nel segno di una tradizione quasi secolare di intreccio tra ricerca cognitivista e AI, il presente articolo cercherà di far dialogare dialetticamente prospettive *embodied* della mente e modelli linguistici artificiali, esplorando le nuove, vibranti implicazioni che la ricerca e l'implementazione delle architetture neurali, su tutte i *Large Language Models*, stanno mettendo alla luce.

After two decades of theoretical and engineering developments, the history of artificial intelligence is going through a turning point. The spread of neural models in fact represents a game-changer, both in virtue of the incredible possibilities of implementation (from image recognition to language processing), and for the philosophical implications that such diffusion brings with it, from applied ethics to epistemology. This paper intends to start from the latter. The competence of machines with a 'neuromorphic' architecture, based on networks of artificial neurons that process information in a sub-symbolic and dynamic manner, could allow us to question certain assumptions of the classical debate on artificial intelligence (and on intelligence in general) regarding the relationship between syntax, meanings and the substrates that carry them. In the sign of an almost century-long tradition of intertwining cognitivist research and AI, this article will attempt to dialectically bring embodied perspectives of the mind and artificial language models into dialogue. It will explore the new, vibrant implications that research and implementation of neural architectures, above all *Large Language Models*, are bringing to light.

Parole chiave: linguaggio, semantica, AI generativa, embodiment, menti e macchine

Keywords: language, semantics, generative AI, embodiment, minds and machines

Sommario: Introduzione: menti, significati e sistemi formali - 1. Computer che usano il linguaggio umano - 2. *Embodiment* e architetture neurali - 3. C'è una via di uscita dalla stanza cinese? - Considerazioni conclusive

Peer review

Submitted 09/10/2024

Accepted 26/11/2024

Published 03/12/2024

Open access

© 2024 | Attribution - Non commercial - Non derivatives (IT)

DOI 10.35948/DILEF/2024.4359

Introduzione: menti, significati e sistemi formali

Nessuno sa dove sia il confine tra comportamento intelligente e non-intelligente. Forse questo confine non è neppure davvero impermeabile. Perimetrare i contorni del termine stesso *intelligenza*, nonostante l'uso che se ne fa nel linguaggio comune, rappresenta una sfida particolarmente ardua, data la arborescente molteplicità di definizioni che fanno riferimento a questo concetto. Inoltre, benché in filosofia si parli di intelletto e di pensiero, per così dire, dall'alba del pensiero stesso (basti pensare al *voûç* anassagoreo o all'*intellectus* tomista) il dibattito intorno all'intelligenza è storicamente recente ed irrimediabilmente compromesso con gli sviluppi dell'IA. È proprio, infatti, a partire dalla metà dello scorso secolo che la convergenza di teoria del ragionamento assiomatico, calcolo meccanico e teoria dell'intelligenza ha gettato le basi per i primi "cervelli elettronici". A partire dalle tesi di Turing sulla realizzabilità di una macchina in grado di esibire proprietà intelligenti (Turing, 1950), numerosi programmi di ricerca sono fioriti per decenni (i.e., Newell - Simon, 1961; Minsky, 1969; Haugeland, 1989), contaminando le scienze cognitive classiche (Fodor, 1975), ed arrivando sino allo sviluppo dei sistemi esperti. Le prime, eleganti, descrizioni del problema, caratterizzate dalla commistione di ricerca su sistemi formali, cognizione e intelligenza, condividevano tutte un assunto fondamentale, derivato dalla teoria di Claude Shannon, che l'*informazione* sia realizzabile su qualunque supporto. Questo orizzonte teorico ha prodotto alcuni dei presupposti fondamentali, circa la manipolazione del linguaggio e la sfera del mentale, che hanno contraddistinto i primi sviluppi del cognitivismo e dell'IA, ovvero:

- Il trattamento dell'informazione ha una natura eminentemente simbolica e formale. Pensare significa manipolare correttamente stringhe di simboli mediante regole sintattiche.
- La dimensione semantica emerge da quella sintattica, in virtù dell'isomorfismo che i simboli esibiscono con i loro referenti reali. Se un sistema formale si 'prendesse cura' della sintassi, la semantica verrebbe da sé (Haugeland, 1989).

In altre parole, i modelli classici del cognitivismo e i primi sviluppi dell'IA condividevano una descrizione del trattamento dell'informazione basato su regole sintattiche di combinazione. È solo a partire da un corretto trattamento sintattico che emergerebbe la dimensione semantica dei simboli, in virtù del loro isomorfismo con il mondo reale. Come se le stringhe di informazione, manipolate da un sistema formale, 'esprimano' qualcosa unicamente in virtù delle proprietà formali del sistema.

L'avanzare della ricerca, tuttavia, ha superato questo modello come descrizione efficace dell'emergere di competenza linguistica, e, per esteso, del comportamento

intelligente. Nelle scienze cognitive, l'affermazione delle teorie *embodied* ha permesso di ripensare l'architettura dei processi mentali, restituendo centralità alla dimensione sensomotoria e al ruolo di un'informazione somatosensoriale nella codifica dei contenuti rappresentazionali (Barsalou, 1999; Barsalou et.al., 2004; Gallese - Lakoff, 2005; Borghi, 2014, 2016). D'altro canto, all'interno del dibattito sull'IA sono state mosse aspre critiche ai modelli GOFAI¹, legate alla mancanza di un'autentica 'comprensione' da parte dei sistemi formali, e, dunque, di una competenza semantica all'interno dei processi computazionali (Dreyfus, 1979; Searle, 1980). L'apparente sterilizzazione del programma di realizzazione forte dell'IA pone, a monte della presente trattazione, un interrogativo esistenziale: *perché discutere (ancora) di intelligenza artificiale?* Nell'orizzonte di questo lavoro, l'esplosione dell'intelligenza artificiale generativa e degli algoritmi di *deep learning* potrebbe rimettere in discussione alcuni assunti fondamentali del dibattito circa la competenza linguistica e il rapporto tra sintassi e semantica, che hanno costituito il *discrimen* per l'attribuzione del marchio di 'intelligente' sin dagli albori del dibattito stesso.

L'obiettivo di questa ricerca è provare a individuare un percorso filosofico che permetta di trovare una 'via di uscita dalla stanza cinese' (Searle, 1980), proponendo un (nuovo) resoconto del trattamento del linguaggio e dell'informazione nei sistemi intelligenti che esamini la tensione dialettica tra i nuovi approcci nel campo delle scienze della cognizione e delle macchine linguistiche. L'interesse sarà focalizzato nel descrivere *nuovi modelli*, che tengano conto del cambiamento di paradigma tanto nella descrizione del mentale che nello sviluppo di IA, in virtù dell'avvento dei *Large Language Models* (LLMs). Al netto dell'irrealizzabilità del programma *forte* dell'IA con gli strumenti classici del dibattito, quali il trattamento sintattico dell'informazione, l'intelligenza simbolica e la computazione lineare, l'argomento che verrà presentato in questo lavoro intende trovare nuove soluzioni teoriche, ed esplorare la possibilità che l'IA basata su modelli neurali faccia *breccia* nella paratia stagna che separa i processi computazionali formali da una dimensione di significato.

La presente ricerca sarà articolata come segue:

- Verrà introduttivamente offerto un resoconto delle sfide legate al *natural language processing* e all'utilizzo di architetture neurali per compiti linguistici;
- La competenza esibita da queste architetture verrà problematizzata filosoficamente, in relazione alla loro capacità di gestire il linguaggio umano nonostante il contenuto *embodied* e somatosensoriale che ne caratterizza persino le espressioni astratte e metaforiche. Come soluzione, verrà offerto un resoconto della natura, mediata dalla cultura di agenti *embodied*, dei dati su cui l'architettura neurale viene addestrata, facendo riferimento a nozioni di matrice fenomenologica.

- Una volta ‘colmata’ la distanza tra agenti biologici e artificiali nella produzione e manipolazione del linguaggio attraverso l’argomentazione offerta in 2), si presenterà un resoconto delle tecniche di trattamento dell’informazione esibite dalle architetture neurali. Mediante una serrata dialettica tra sistemi formali e modelli cognitivi, si sosterrà la non-pacificità della distinzione tra sintassi e semantica. Laddove diventa quantomeno complesso scremare rigidamente atti di significato e processi formali, sarà, in conclusione, possibile rendere conto della nuova alba dell’IA neuromorfica², e in particolare dei LLMs, all’interno di questa polarità, offrendone una descrizione in termini ‘più che sintattici’. Attraverso la definizione di *intelligenza più che sintattica*, il presente lavoro intenderà porre l’accento sulla natura sub-simbolica, dinamica e flessibile del trattamento dell’informazione esibito dai modelli neurali, in discontinuità con le architetture GOF AI puramente sintattiche, e in continuità, per determinati aspetti, con gli agenti cognitivi.

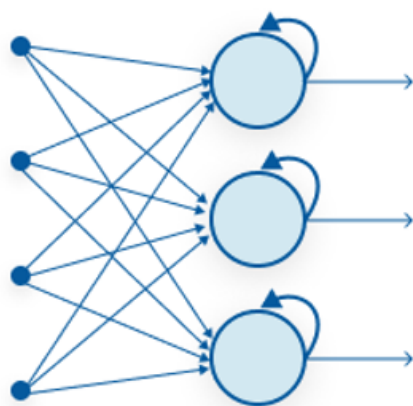
1. Computer che usano il linguaggio umano

Dare vita a una macchina in grado di utilizzare il linguaggio umano rappresenta una delle sfide definitive della ricerca sull’intelligenza artificiale. Era, infatti, un test eminentemente linguistico il celebre *imitation game*, proposto da Alan Turing come *discrimen* per l’attribuzione di intelligenza a una macchina a stati discreti (Turing, 1950). Turing lasciò in eredità questa sfida a un’intera generazione di ricercatori. Egli, di fatto, ne aveva intravisto la possibilità e aveva fornito un criterio di misurazione, ma rimaneva estremamente evasivo sulle modalità di realizzazione. Restavano da scoprire, in altre parole, le istruzioni attraverso le quali produrre artificialmente il comportamento intelligente. Decenni di sforzi di ricerca sulle architetture GOF AI tentarono di raggiungere tale obiettivo (i.e., Newell - Simon, 1961, 1964; Minsky, 1967). Marvin Minsky, tra i pionieri della ricerca sull’IA, si spinse sino ad affermare che «i problemi connessi al fatto di creare l’intelligenza artificiale saranno risolti entro una generazione³». Questi tentativi, tuttavia, anziché condurre a una definizione condivisa di che cosa sia l’intelligenza, hanno solo rivelato ciò che l’intelligenza non è. Il tramonto del programma *forte* dell’IA, ad opera, su tutti, di John Searle, ha chiarito infatti i limiti dei modelli puramente sintattici, basati, cioè, sulla manipolazione di simboli a partire da algoritmi rigidi e pre-determinati. I modelli GOF AI erano impermeabili alla semantica, il marchio dell’intelligenza, e incapaci di comprendere alcun significato. L’affermazione delle descrizioni *embodied* della mente (Varela *et.al.*, 1991), fondate sull’inestricabile relazione di processi superiori e sensomotorio, hanno inoltre superato le prospettive computazionali, basate sull’analogia mente/computer,

proprie del cognitivismo classico (Fodor, 1975). Il programma di realizzazione forte dell'intelligenza artificiale sembrava, dunque, destinato a naufragare. Qualcosa, però, è cambiato.

L'ambito di ricerca entro il quale si concentrano i tentativi di sviluppo di una macchina in grado di gestire e utilizzare il linguaggio umano si chiama *natural language processing* (NLP). L'NLP si basa prevalentemente su algoritmi di apprendimento automatico per far sì che i computer possano rispondere a stimoli linguistici in modo efficace, interagendo in modo naturale con gli agenti umani. Il linguaggio naturale, tuttavia, è intrinsecamente ambiguo, contestuale e dipendente da informazioni implicite tra i parlanti coinvolti nella comunicazione. Come mettere a punto, allora, un modello in grado di trattare una materia flessibile come il linguaggio e la comunicazione, nonostante i vincoli formali super-rigidi che ne caratterizzano la natura artificiale? Tutti i recenti progressi nell'ambito del trattamento del linguaggio sono merito dell'implementazione di IA generativa basata su reti neurali artificiali e algoritmi di *deep learning*. Il *deep learning* è il campo dell'intelligenza artificiale che si concentra sul costruire modelli di reti neurali profonde in grado di compiere azioni e prendere decisioni attraverso algoritmi di apprendimento automatico⁴. Una rete neurale è un modello matematico di intelligenza artificiale, ispirato dal connessionismo (Rumelhart *et.al.*, 1986) e dai principi di funzionamento del sistema nervoso⁵, che elabora l'informazione non in modo sequenziale attraverso una serie ordinata di processi di elaborazione, ma mediante una rete di unità computazionali, di neuroni artificiali. Le informazioni vengono così immagazzinate e trattate nei parametri della rete, in particolare in pesi e connessioni. La caratteristica distintiva delle reti neurali artificiali è dunque la loro capacità di modificare tali parametri, ovvero di apprendere automaticamente. Questa proprietà è l'essenza del *deep learning*, e si basa su pattern di dati di addestramento e sulla capacità della rete, di fronte a nuove informazioni, di produrre *output* sempre diversi in base alle proprietà apprese dai dati osservati. Il *deep learning* è dunque da intendersi come un adattamento della struttura della rete ai valori ambientali, a partire da una condizione iniziale di disordine dovuta ai parametri casuali delle connessioni dei neuroni⁶.

Nello specifico delle reti ottimizzate per compiti linguistici, gli sforzi dei programmatori si concentrano sulla messa a punto di *recurrent neural networks* (RNN). Una RNN è un tipo di modello neurale progettato per gestire dati sequenziali (o dati che hanno una struttura temporale). Le RNN esibiscono connessioni cicliche che consentono loro di mantenere uno stato interno, una sorta di *memoria* delle informazioni, mentre elaborano nuovi *input*.



Struttura delle connessioni di una rete neurale ricorrente.

Questi cicli consentono alle RNN di gestire stringhe di dati di lunghezza variabile, mantenendo una traccia del contesto temporale, necessaria per i compiti linguistici. In altre parole, la macchina, attraverso l'utilizzo di passaggi temporali, riesce a “interpretare” le parole che legge, “ricordando” al contempo ciò che ha già letto. Oltre all'architettura basata su cicli, l'aspetto determinante in una RNN è la capacità di tradurre il linguaggio in vettori, a partire da un modello di semantica distributiva. Per semantica distributiva si intende quella specifica teoria linguistica che definisce il significato delle parole in relazione alle altre con cui esse tendono a presentarsi e, dunque, a condividere uno spazio semantico (cfr., Firth, 1957). Attraverso il modello neurale, la macchina è in grado di ricostruire la relazione tra le parole e il loro spazio semantico. L'obiettivo dell'algoritmo è trovare un valore numerico, un vettore, per ogni parola del vocabolario, che *catturi qualcosa della “semantica” di quella parola*, e, in particolare, del suo rapporto con gli altri termini e del suo occorrere più o meno frequente con essi. La strategia che, tuttavia, negli ultimi anni ha riscosso maggior successo è legata allo sviluppo di modelli statistici, i LLMs. A differenza delle architetture vettoriali basate su semantica distributiva, i LLMs sono modelli probabilistici addestrati, in modo supervisionato o semi-supervisionato su vasto *corpus* di dati testuali, a predire la parola successiva in una frase. Durante la fase di inferenza, di generazione del testo, il modello, rimodulando costantemente la propria configurazione di pesi e connessioni, e dunque l'informazione contenuta nella sua architettura, non fa altro che continuare a predire la parola successiva in base all'addestramento svolto. Su modelli di questo tipo si basano programmi che hanno mostrato una competenza linguistica straordinaria. Chat-GPT, il *software* di linguaggio generativo di *Open AI*, è soltanto l'esempio maggiormente riconosciuto. Tuttavia, è suggestivo, nell'ottica di un superamento dell'argomentazione sull'impermeabilità dell'IA alla semantica, porre all'attenzione il caso delle architetture integranti modelli di *class incremental learning* e reti *open set*, ovvero quel tipo specifico di architetture in grado di relazionarsi con enti reali non oggetto di addestramento, e di riconoscerne le caratteristiche e i vincoli semantici, mettendo in atto un processo di categorizzazione direttamente *in situ* (Lin, 2013; Bendale - Boulton, 2016). Altri modelli,

ontology-based, incorporano invece una rappresentazione formale, un'ontologia, all'interno della struttura della rete neurale (Landgrebe, 2021). L'ontologia integrata nella rete le fornisce un insieme di concetti, relazioni e regole che la aiutano a rendere conto del significato dei dati, rendendo possibile, nell'intenzione degli sviluppatori, una manipolazione e, per certi versi, una 'comprensione' dell'informazione, una competenza semantica, e consentendo alle reti di 'ragionare' su concetti astratti e relazioni complesse (Kumar *et.al.*, 2023, DeCoste *et.al.*, 2024). Questi modelli sembrano promettenti. Tuttavia, ci sono dei problemi.

2. *Embodiment* e architetture neurali

La ricerca, alla luce dell'affermarsi delle *embodied cognitive sciences*, sta ponendo con sempre maggiore insistenza l'attenzione sul ruolo del sistema sensomotorio nella codifica dei contenuti mentali. Il paradigma simulatorio della *grounded cognition*, ad esempio, offre una descrizione dei contenuti rappresentazionali nei termini di informazione somatosensoriale immagazzinata dal sistema cognitivo⁷, riattivabile in base agli stimoli contestuali (Barsalou, 1999). La comprensione di un testo, seguendo questa teoria, avverrebbe attraverso la simulazione di contenuto percettivo e/o motorio immagazzinato nella memoria ed evocato dalla parola. Anche in linguistica, la descrizione dei concetti, persino quelli metaforici e astratti, prevede in realtà un nucleo di informazione percettiva e, dunque, *embodied* (Lakoff e Johnson, 1980; Gibbs, 2004, 2012). Numerosi studi di *neuroimaging* hanno esaminato la questione del fondamento del linguaggio astratto, confrontando le attivazioni nelle aree cerebrali somatosensoriali e motorie in relazione ad espressioni metaforiche e non-metaforiche. Un lavoro esemplare è stato offerto da Rutvik Desai (2021), nel quale vengono prese in esame una vasta gamma di espressioni metaforiche (MET; e.g. *L'uomo ha afferrato l'idea*) con frasi astratte o di controllo non senso-motorio (ABS; e.g. *Ha avuto una brutta giornata*). All'interno della ricerca vengono usate anche espressioni d'azione letterali (LIT; e.g. *L'uomo afferrò la maniglia*). La previsione di Desai è che le aree senso-motorie dovrebbero essere attivate da MET, e in misura maggiore di ABS. Nel caso delle metafore che coinvolgono, ad esempio, concetti di azione, studi di *neuroimaging* confermerebbero la previsione nelle aree motorie superiori associate alla pianificazione dell'azione complessa e all'interazione mano-oggetto nel lobo parietale inferiore anteriore sinistro (Desai, 2011). Al netto del resoconto sin qui offerto del trattamento del linguaggio naturale in agenti biologici e della codifica sensomotoria del contenuto informativo dei concetti, persino di quelli più astratti, si rende dunque necessario porre un interrogativo:

- Come può un agente artificiale, e dunque *disembodied*, esibire una competenza pari, e talvolta anche superiore, a quella degli agenti umani, in compiti linguistici?

Per rintracciare una soluzione, è opportuno partire da un ulteriore interrogativo: *è realmente, la macchina, priva di corpo?* La selezione dei dati e degli esempi a partire dai quali la macchina viene addestrata sono infatti estranei alla macchina stessa, in quanto frutto di una scelta esterna, preliminare al *training*, e, dunque, all'implementazione del modello neurale. I set di dati sui quali la macchina svolge i processi di addestramento, sono il prodotto della cultura di agenti umani, e, dunque, *embodied*. Il loro contenuto informativo, è, in altre parole, possesso di agenti la cui architettura cognitiva è strutturalmente radicata nel senso-motorio, e le cui rappresentazioni sono codificate a partire da un nucleo percettivo e somatosensoriale (Barsalou, 1999). Questo orizzonte argomentativo offre così la possibilità di rintracciare una continuità tra architetture intelligenti umane e artificiali, in virtù della natura culturalmente mediata dei dati di addestramento.

È possibile sostenere una tale argomentazione, che assottiglierebbe così la distanza tra agenti biologici e IA generativa, attraverso strumenti concettuali di matrice fenomenologica. La natura dell'informazione che caratterizza i *set* di dati dei LLMs è, infatti, descrivibile come 'impersonale' e 'non-vissuta' (Merleau-Ponty, 2010, 2018). Essa attraversa l'agente artificiale in maniera 'passiva' (Buongiorno *et.al.*, 2022), al pari di quello umano, della cui conoscenza non è possibile offrire un resoconto semplicemente in termini di dicotomia soggetto-oggetto, ma integrando un'informazione impersonale, come la cultura e il contesto in cui essa è situata. Il possesso di dati di addestramento e di un'architettura mediata dalla cultura umana rende, così, la macchina *embodied*? Recuperando la lezione di Katherine Heyles (Heyles, 1999), è improprio parlare di vera e propria incorporazione, ma di 'iscrizione', di analogia funzionale tra due sistemi. La macchina rimane, infatti, priva di un corpo che costituisca un *medium* nei processi percettivi e cognitivi, e che rappresenta la condizione necessaria affinché un agente sia incorporato. Tuttavia la competenza esibita dai modelli neurali nei compiti linguistici è giustificabile, nella prospettiva del presente elaborato, in quanto i sistemi artificiali sono attraversati da un'informazione, rappresentata dai dati di addestramento della rete, il cui linguaggio e i cui contenuti sono, irrimediabilmente, mediati dalla cultura di agenti *embodied* come gli esseri umani.

3. C'è una via di uscita dalla stanza cinese?

Al netto della ricostruzione sin qui proposta, che ha assottigliato la distanza tra agenti biologici e artificiali in virtù della natura *embodied* e culturalmente mediata dei dati di addestramento di questi ultimi, ha senso domandarsi se, all'interno dell'intricato groviglio di pesi e connessioni che caratterizzano l'intelligenza artificiale generativa, si verifichi una qualche forma di competenza semantica? Precedentemente è stato evidenziato come nel dibattito classico le nozioni di semantica e sintassi venissero collocate ben distinte tra di loro. In un orizzonte teorico del genere, la prestazione propriamente umana e cognitiva è quella *sintattica*: l'intelligenza risulterebbe essere manipolazione formale di simboli attraverso delle regole di combinazione (Fodor, 1975). Tale prospettiva anima la ricerca di chi, come Pierre Lévy, sostiene che l'AI debba formalizzare completamente tutti gli aspetti del linguaggio naturale per poter arrivare a un'intelligenza generale (Lévy, 2010, 2013, 2023). Secondo questa ricerca, tuttavia:

- Non ci sono ragioni per sostenere che la distinzione tra prestazioni semantiche e sintattiche sia pacifica.

A sostegno di tale orizzonte si pongono le argomentazioni utilizzate nei precedenti nodi della trattazione. L'idea, ad esempio, che le modalità di elaborazione specifiche del sistema cognitivo umano si fondino sul funzionamento del sistema sensomotorio e sulla corporeità, e che, dunque, si renda quantomeno controverso distinguere rigidamente descrizioni cognitive di alto livello dai processi fisici, sul cui intreccio si fonda la tesi dell'incorporazione, depotenzia una distinzione forte tra prestazioni che hanno a che fare con unità formali rigide e atti di conferimento di significato, di comprensione. Verrebbe meno, così, quella paratia stagna che dividerebbe la manipolazione formale di simboli dalla comprensione del loro significato. Lo stesso esperimento mentale della stanza cinese, infatti, si basa su tale rigida distinzione, di matrice cognitivista, dai tratti specificatamente antropocentrici (con la sostanziale differenza che Searle rintraccia, altresì, nella competenza semantica e nella capacità di dare significato ai simboli, il marchio dell'intelligenza umana).

Lo sfumare del confine tra le due definizioni potrebbe permettere di attribuire all'IA basata su reti neurali delle proprietà che tendono alla semantica, o che, comunque, esibiscano un trattamento dell'informazione più che sintattico e meramente formale? Attraverso la precedente argomentazione, è possibile depotenziare tale distinzione, offrendo un resoconto della competenza delle reti neurali che aggiri il problema posto da Searle. Come se, anziché 'uscire' dalla stanza cinese, sia possibile non entrarci affatto. Sembra concepibile, dunque, impegnarsi nel sostenere che l'*intelligenza* non sia descrivibile a partire da prestazioni, formali e non-formalizzabili, facilmente

distinguibili. In continuità con la teoria del significato *embodied*, che, come evidenziato in precedenza, sostiene l'interconnessione di processi cognitivi superiori e schemi corporei, rendendo dunque difficile scremare i domini di semantica e sintassi in maniera rigida, è possibile collocare nel mezzo di tale polarità gli agenti artificiali basati algoritmi di apprendimento automatico, descrivendo tali modelli in termini '*più che sintattici*'. Attraverso la definizione di *intelligenza più che sintattica* si intende rendere conto delle prestazioni delle reti neurali artificiali e degli algoritmi di *deep learning* in termini non meramente formali e simbolici, ma che tendono a una caratterizzazione dinamica, non-formale e sub-simbolica. In antitesi con quanto considerato, ad esempio, da Pierre Lévy come competenza eminentemente intelligente, la competenza sintattica, nell'orizzonte che anima questa ricerca uno degli elementi di congruenza più rilevanti tra agenti umani e artificiali basati su reti neurali consiste proprio nel trattamento sub-simbolico dell'informazione (Lopez-Rubio, 2018), non basato su regole formali esplicite, ma su computazioni in parallelo, parametri e funzioni non lineari. Elementi, questi, fortemente compatibili con le descrizioni dei processi cognitivi offerte dalle neuroscienze.

Intelligenza *più-che sintattica*. Attraverso tale espressione è possibile offrire un resoconto del trattamento dell'informazione nell'AI generativa che ne sottolinei la discontinuità con i modelli GOFAI, ponendone l'attenzione sul trattamento sub-simbolico, dinamico, flessibile e non riducibile a formati esclusivamente sintattici.

La competenza delle reti neurali di muoversi in autonomia in un contesto dinamico e di costruire nuove categorie e classi di enti attraverso algoritmi *open-set* (Lin, 2013; Sünderhauf, 2018) o di integrare ontologie all'interno dei parametri di addestramento (DeCoste, 2024) per cogliere le relazioni semantiche tra gli enti del contesto in cui la macchina è situata in maniera olistica, trattando l'informazione in maniera sub-simbolica e non rigidamente codificata da regole formali esplicite, costituirebbe dunque un elemento di discontinuità con le architetture GOFAI, e di approssimazione della macchina a una qualche forma di competenza semantica, o comunque *più che sintattica*.

Considerazioni conclusive

Al culmine del percorso che è stato battuto, e che ha abbracciato filosofia della mente, filosofia dell'IA e fenomenologia, è stato offerto un resoconto dell'IA generativa che pone quest'ultima in discontinuità con i modelli che hanno aperto la stagione della ricerca sulla realizzazione artificiale del pensiero. Negli anni in cui Searle scriveva dell'esperimento della *stanza cinese* nasceva infatti un nuovo modello, ispirato dal connessionismo (Rumelhart, 1986) e dall'idea che il marchio del cognitivo non sia il semplice trattamento di simboli, ma l'elaborazione in parallelo e distribuita

dell'informazione. Il presente elaborato ha preso in esame le prestazioni esibite dai modelli neurali ottimizzati per compiti linguistici, alla ricerca di elementi di continuità con l'architettura cognitiva e con le caratteristiche propriamente intelligenti degli agenti biologici descritti dalle *embodied* cognitive sciences. Il *focus* della trattazione si è concentrato, *in primis*, sulla questione della codifica dei contenuti linguistici e sul ruolo dell'incorporazione, con particolare accento sul linguaggio astratto, mettendo in evidenza la competenza delle reti ricorrenti e dai LLMs nel *natural language processing*, della quale è stato offerto un resoconto a partire dalla natura *embodied*, in quanto culturalmente mediata, dello stesso *dataset* delle reti, la cui natura non direttamente esperita dalla macchina è stata giustificata, in chiave fenomenologica, attraverso le nozioni di impersonale e non-vissuto. Al netto di tale argomentazione, l'attenzione si è poi concentrata sulla natura stessa del trattamento dell'informazione esibito dai modelli neurali, chiamando in causa il problema della comprensione, del significato e della semantica. È stata, così, offerta una descrizione *più che sintattica* degli algoritmi di apprendimento profondo. Attraverso tale definizione, si è cercato di collocare le prestazioni intelligenti delle RNN all'interno della polarità tra formale e non-formalizzato, per certi versi a metà. Ciò è stato possibile attraverso il depotenziamento di questa stessa polarità, in virtù della complessità nello scremare, in maniera netta, negli stessi agenti umani, i processi cognitivi superiori dai sistemi somatosensoriale e motorio. La messa in discussione dell'impermeabilità dei domini di semantica e sintassi ha così permesso di allentare i vincoli rigidi che il dibattito classico attribuiva alle loro definizioni, potendo così riconoscere, nell'IA basata su reti neurali, delle proprietà che tenderebbero alla semantica, o, comunque, più che meramente formali.

La conclusione qui proposta lascia, tuttavia, sullo sfondo, interrogativi più grandi, che sono stati soltanto parzialmente richiamati all'attenzione. Tali interrogativi riguardano, ad esempio, la natura del termine stesso informazione (che si suggerisce di prendere *cum grano salis*), quale sia il confine tra formale e non formalizzato, tra semantica e sintassi, e, su tutti, dove esattamente risieda il significato. Per tutte queste domande non ci sarà una risposta in questa sede. La ricerca, tuttavia, non si ferma.

Note

1. *Good old-fashioned artificial intelligence*, basata su algoritmi e computazioni lineari e regole esplicite di trattamento dell'informazione.
2. Per la definizione di AI neuromorfica, cfr. Bartolozzi *et.al.*, 2021
3. Minsky 1967, p.2
4. Kelleher 2019, p.35: «Deep learning enables data-driven decisions by identifying and extraction patterns from large datasets that accurately map from sets of complex inputs to

good decision outcomes».

5. Hopfield 1988, p.10: «Those of us who are working on artificial networks hope that there are direct lessons from neurobiology that can be usefully realized in electronics».

6. Per un'accurata descrizione dei processi computazionali di distribuzione di pesi e connessioni, somma algebrica, funzioni di attivazione, e delle reti neurali *task specific*, cfr. Kelleher, 2019

7. Barsalou, 1999: «During perceptual experience, association areas in the brain capture bottom-up patterns of activation in sensory-motor areas. Later, in a top-down manner, association areas partially reactivate sensory-motor areas to implement perceptual symbols».

Bibliografia

- Barsalou, 1999 = Lawrence Barsalou, *Perceptual symbol systems*, *Behavioral and Brain Sciences*, 22,4 (1999): pp.577-660.
- Barsalou *al.*, 2004 = Lawrence Barsalou e Karen Solomon, *Perceptual simulation in property verification*, *Memory & Cognition*, 32, 2, (2004): pp.244-259.
- Bartolozzi *al.*, 2021 = Chiara Bartolozzi, Giacomo Indiveri ed Elisa Donati, *Embodied neuromorphic intelligence*, *Nature communications*, 13, 1 (2021).
- Bendale - Boulton, 2016 = Abhijit Bendale Terrance Boulton, *Towards open set deep networks*, *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016): pp. 1563-1572.
- Buongiorno *al.*, 2022 = Federica Buongiorno e Suzan Kozel, *The Potential of Passivity Beyond the Intentional Model. Consciousness as Disarticulation in Merleau-Ponty's Institution and Passivity*, *Humana.Mente Journal of Philosophical Studies*, 15, 41 (2022): pp.121-148.
- De Coste *al.*, 2024 = Michael De Coste, Zhong Li e Ridha Khedri, *A hybrid ontology-based semantic and machine learning model for the prediction of spring breakup*, *Computer-Aided Civil and Infrastructure Engineering*, 39,2 (2024): pp.264-280.
- Desai, 2021= Rutvik Desai, *Are metaphors embodied? The neural evidence*, *Psychological Research*, 86, 8 (2021): pp.2417-2433.
- Dreyfus, 1979 = Hubert Dreyfus, *What Computers Still Can't Do*, MIT press, 1979.
- Firth, 1957 = John Firth, *Papers in Linguistics, 1934-1951*, London: Oxford University Press, 1957.
- Fodor, 1975 = Jerry Fodor, *The Language of Thought*, New York: Thomas Y. Crowell, 1975.
- Gallese - Lakoff, 2005 = Vittorio Gallese e George Lakoff, *The Brain's concepts: the role of the Sensory-motor system in conceptual knowledge*, *Cognitive Neuropsychology*, 22, 3/4 (2005): pp. 455-479.
- Gibbs *al.*, 2004 = Raymond Jr Gibbs, Paula Lenz Costa Lima ed Edson Francozo, *Metaphor is grounded in embodied experience*, *Journal of pragmatics*, 36, 7 (2004): pp.1189-1210.
- Gibbs, 2012 = Raymond Jr Gibbs, *Metaphor interpretation as embodied simulation*, *Mind & Language*, 21, 3 (2012): pp.434-458.
- Haugeland, 1989 = John Haugeland, *Artificial Intelligence: The Very Idea*, MIT press, 1989.
- Hayles, 1999 = Hayles, N. K., *How we became posthuman. Virtual bodies in cybernetics, literature, and informatics*, University of Chicago Press, 1999.
- Hopfield, 1988 = John Hopfield, *Artificial neural networks*, *Ieee circuits and devices magazine*, 4, 5 (1988): pp.3-10.
- Johnson, 1988 = Mark Johnson, *The body in the mind: The bodily basis of meaning, imagination and reason*, Chicago university press, 1987.
- Kelleher, 2019 = John Kelleher, *Deep Learning*, MIT press, 2019.
- Kumar, 2023 = Kumar, R., & Sharma, S. C., *Hybrid optimization and ontology-based semantic model for efficient text-based information retrieval*, in *The Journal of Supercomputing*, 79, 2 (2023): pp.2251-2280

-
- Lakoff - Johnson, 1980 = George Lakoff e Mark Johnson, *Metaphors we live by*, University of Chicago, Chicago, 1980.
 - Lévy, 2010 = Pierre Lévy, *The IEMML research program. From social computing to reflexive collective intelligence*, *Information Sciences*, 180 (2010): pp. 71–94.
 - Lévy, 2013 = Pierre Lévy, *The semantic sphere 1: Computation, cognition and information economy*, John Wiley & Sons, 2013.
 - Lévy, 2023 = Pierre Levy, *Semantic computing with IEMML*, *Collective Intelligence*, 2 4 (2023).
 - López-Rubio, 2018 = Ezequiel López-Rubio, *Computational Functionalism for the Deep Learning Era*, *Minds and Machines*, 28, 4 (2018): pp.667-688.
 - Merleau-Ponty, 2010 = Maurice Merleau-Ponty, *Institution and passivity: Course notes from the Collège de France (1954-1955)*, Northwestern University Studie, 2010.
 - Merleau-Ponty, 2010 = Maurice Merleau-Ponty M., *Fenomenologia della percezione* (1945), Bompiani, 2018.
 - Minsky, 1967 = Marvin Minsky, *Computation: finite and infinite machines*, Prentice-Hall, Upper Saddle Rivere (N.J.), 1967.
 - Minsky, 1969 = Marvin Minsky, *Semantic information processing*, MIT Press, 1969.
 - Rumelhart *al.*, 1986 = David Rumelhart, Geoffrey E. Hinton e James L. McClelland, *A general framework for parallel distributed processing*, *Parallel distributed processing: Explorations in the microstructure of cognition*, 1, 26 (1986): pp.45-76.
 - Searle, 1980 = John Searle, *Minds, Brains and Programs*, *Behavioral and Brain Sciences*, 3 (1980): pp. 417–424.
 - Simon - Newell, 1961 = Herbert A. Simon e Allen Newell, *GPS: a program that simulates human thought*, in H. Bildings (ed.), *Lernende automaten*, München, R. Holdenburg, 1961.
 - Simon - Newell, 1964 = Herbert A. Simon e Allen Newell, *Information Processing In Computer And Man*, *American Scientist*, 52, 3 (1964): pp. 281-300.
 - Sünderhauf, 2018 = Nico Sünderhauf, *The limits and potentials of deep learning for robotics*, *The International Journal of Robotics Research*, 37, 4-5 (2018): pp. 405-420.
 - Turing, 1950 = Alan M. Turing, *Computing machinery and intelligence*, *Mind*, 59 (1950): pp. 433-460.
 - Varela *al.*, 1991 = Francisco Varela, Evan Thompson, Eleanor Rosch, *The embodied mind. Cognitive science and human experience*, The MIT Press, 1991.