

Wittgenstein e il neuro-simbolismo: come ChatGPT e il *Tractatus* risolvono un antico rompicapo logico.

Abstract:

Recent debates on Artificial Intelligence (AI) have raised not only technical questions but also profound philosophical ones, as this emerging technology continues to transform society. While reducing the time required to perform tasks ranging from simple to complex, it also prompts significant ethical, ontological, and epistemological inquiries. Philosophers and researchers are actively working to clarify key aspects, such as the ethical, moral, and legal implications of AI, while exploring the implicit relationship between humans and technology. This paper aims to delve into fundamental issues such as data representation and demonstrate how Wittgenstein's philosophical ideas, particularly those in his *Tractatus Logico-Philosophicus*, can contribute to the contemporary debate on AI. Focusing on neuro-symbolic AI, we explore both its technical and theoretical aspects, comparing connectionist and representational data models like Vector Symbolic Architectures (VSA). The paper emphasizes the importance of interdisciplinary dialogue between philosophy and technology, offering a unique perspective on the relationship between human functions and their technological imitation. It also discusses Wittgenstein's theories of propositions, logical space, and proposition compatibility, drawing parallels to AI's use of multidimensional vectors in logical inference. The work ultimately argues for the relevance of Wittgenstein's ideas in advancing AI, providing philosophical insights for both researchers and philosophers engaged in this technological field.

Keywords:

Artificial Intelligence, Wittgenstein, Tractatus Logico-Philosophicus, Neuro-Symbolic AI, Data Representation, Vector Symbolic Architectures (VSA), Connectionism, Logical Space, Logical Inference, Philosophy and Technology, AI Ontology, AI Epistemology.

1) Introduzione

Negli ultimi tempi il dibattito sull'Intelligenza Artificiale sta sollevando questioni non solo di natura tecnica, ma anche filosofica. Questa nuova tecnologia sta profondamente cambiando la nostra società, riducendo drasticamente i tempi con cui portiamo a termine dal più semplice al più arduo incarico, ma ponendo grossi interrogativi rispetto la sua eticità, la sua ontologia ed epistemologia. Diversi ambiti della ricerca¹ allora, specialmente in ambito filosofico, sono ora all'opera per chiarirne alcuni punti fondamentali, quali le sue implicazioni etiche, morali e legali², chiarificando la relazione implicita tra umano e tecnologia al fine di offrire prospettive filosofiche concrete nella realtà dell'era digitale³. Ripensare l'umano nel digitale e il digitale nell'umano è ad oggi più che mai indispensabile⁴.

In questo testo ci proponiamo di sviscerare alcune questioni di base, come quella relativa al rappresentazione dei dati, e di dimostrare perché il pensiero filosofico di un autore come Wittgenstein possa risultare ad oggi terreno fertile per il dibattito su filosofia e tecnologia. A partire dall'analisi di alcuni punti cardine della sua filosofia contenuti nel *Tractatus Logico-Philosophicus*⁵, metteremo in luce come queste idee possano offrire un contributo fondamentale alla comprensione delle problematiche contemporanee legate all'IA. Rispetto a queste ultime, ci concentreremo in particolare sull'Intelligenza Artificiale neuro-simbolista, esplorandone gli aspetti tecnici e teorici. Alla fine del nostro contributo, il lettore saprà discernere quali sono le differenze tra la rappresentazione dei dati di tipo connessionista e quella di tipo rappresentazionale, detta VSA (Vector Symbolic Architecture). In breve, le VSA sono un modello computazionale, un'architettura informatica, in cui informazioni simboliche vengono rappresentate come vettori su cui eseguire operazioni logiche in spazi ad alta dimensionalità.

Il nostro obiettivo è tracciare un'analisi al confine tra filosofia e tecnologia, dimostrando come un dialogo interdisciplinare possa fornire risposte profonde alle questioni sollevate dallo sviluppo tecnologico; sviluppo tecnologico che cerca di avvicinarsi, indefinitamente, alla perfetta imitazione delle funzioni dell'essere umano.

Nella prima parte forniremo una breve panoramica storica sull'Intelligenza Artificiale di tipo neuro-simbolista, in modo tale che il lettore prenda confidenza con questo complesso sistema e possa agevolmente distinguerla da altri approcci, come quello connessionista. Al fine di illustrare i passaggi logici che la macchina esegue, poi, illustreremo passo passo come un famosissimo enigma

¹Per uno sguardo ampio sulle discipline umanistiche che ad oggi si occupano di Intelligenza Artificiale rimandiamo ad un'accreditata collettanea come Luca Taddio e Gabriele Giacomini (a cura di), *Filosofia del digitale*, Milano, Mimesis, 2020.

²Rimandiamo a Mauro Bellini e Maria Teresa Della Mura, *Umano digitale: verso un'etica dell'innovazione*, Milano, In Dialogo, 2023, dove i due autori, esperti in innovazione e tecnologia, si occupano proprio dell'etica dell'Intelligenza Artificiale, esplorandone implicazioni pratiche e morali.

³Offre appunto questa prospettiva Giacomo Pezzano, *Pensare la realtà nell'era digitale: una prospettiva filosofica*, Roma, Carocci, 2023, che cerca di riformulare il ruolo del pensiero critico nella società della trasformazione digitale.

⁴Ripensare l'umano in questo clima pregno di innovazione è quanto si è proposto M. Ferraris in *Documanità. Filosofia del mondo nuovo* (Roma-Bari: Laterza, 2021).

⁵Sarebbe ambizioso da parte nostra pensare di poter rileggere tutto il *Tractatus* alla luce della ricerca sull'IA. Ciò che faremo sarà analizzare e commentare assieme, attraverso confronti con la macchina intelligente, una "zona" specifica del *Tractatus* che riguarda la rappresentazione delle proposizioni nello spazio logico. Nel dettaglio, citeremo le proposizioni 2.18, 2.181, 2.19, 2.202 e infine 2.2, tutte inerenti alla sistematizzazione della conoscenza nel nostro "mondo logico".

logico può essere risolto utilizzando la logica dei predicati, che ha lo stesso impianto della logica usata dalla rete neuro-simbolista; descriveremo infatti le proprietà di rappresentazione per inferenze logiche e non mancheremo di applicarle a dei problemi concreti.

La seconda parte del testo è dedicata alla filosofia di Ludwig Wittgenstein, con particolare attenzione alla sua teoria delle proposizioni. Presenteremo alcuni concetti chiave del *Tractatus Logico-Philosophicus* utili alla nostra analisi, soffermandoci su due nozioni ben precise: lo *spazio logico* e la *compatibilità delle proposizioni*. Attraverso un confronto concettuale, mostreremo come queste idee trovino un'eco nei modelli matematici dell'Intelligenza Artificiale. Lo spazio logico wittgensteiniano, inteso come il dominio in cui ogni possibilità logica è contenuta, non comprende solo ciò che accade, ma anche tutto ciò che potrebbe accadere. Nello spazio logico risiedono i fatti e le possibilità di accadere dei fatti: nulla di illogico, in questo senso, accade o può accadere. Guideremo il lettore nel comprendere come questa struttura metafisica possa essere avvicinata allo spazio matematico con cui l'IA rappresenta gli oggetti —sotto forma di vettori multidimensionali— e genera inferenze logiche. Inoltre, analizzeremo il concetto di compatibilità tra proposizioni, ovvero la proprietà di condivisione di un medesimo fondamento di verità. Mostreremo come questa compatibilità, all'interno dello spazio logico wittgensteiniano, sia analoga alla vicinanza probabilistica tra vettori multidimensionali in uno spazio matematico complesso, così come viene utilizzato nei modelli di Intelligenza Artificiale.

In conclusione, illustreremo le ragioni per cui un autore come Wittgenstein merita di essere indagato nell'ambito dell'Intelligenza Artificiale, sia dai filosofi che dagli scienziati, offrendo spunti bibliografici sulle prospettive filosofiche che, a partire dal suo pensiero, risultano tuttora rilevanti. Il nostro percorso toccherà dapprima il contributo di Wittgenstein alla psicologia e alla filosofia della *Gestalt*, ambito in cui opere come *Osservazioni sulla Filosofia della Psicologia* sono divenute punti di riferimento per le teorie della percezione. Successivamente, analizzeremo le applicazioni pratiche delle sue idee nei campi della tecnologia e dell'informatica. Dal lavoro di Margaret Masterman sui testi wittgensteiniani, infatti, sono stati sviluppati modelli utili per diversi algoritmi di analisi del linguaggio e rappresentazione informatica, come quelli di K. S. Jones e Peter Gärdenfors. Questi modelli offrono infatti un supporto significativo alle architetture di Intelligenza Artificiale che nel testo analizzeremo. Un ulteriore esempio è dato dai sistemi di IA interpretabili sviluppati da Tim Miller, basati proprio sulla filosofia del contesto e del significato esplorata da Wittgenstein.

L'analisi che vogliamo qui intraprendere mira ad offrire degli strumenti di comprensioni avanzati che servano non solo ad un concettualismo astratto, ma ad un vero e proprio modo di applicazione della filosofia wittgensteiniana nel campo dell'ingegneria dei modelli matematici, al fine di supportare lo sviluppo delle tecniche di rappresentazione dei dati nell'Intelligenza Artificiale. Speriamo che queste prospettive diletino non solo i filosofi che si interessano alla filosofia di questo acuto pensatore austriaco, ma anche ai ricercatori impegnati in questo momento nel campo dello sviluppo tecnologico. È proprio a loro che vogliamo rivolgerci, nella speranza di aiutarli ad adottare un approccio interdisciplinare fruttuoso ad entrambi i campi del sapere.

2) La questione neuro-rappresentazionale

Nel dibattito sulla possibilità di un pensiero dell'Intelligenza Artificiale somigliante a quello umano, si discute spesso sul valore epistemico dei dati che essa immagazzina, ponendosi domande fondamentali come la seguente: in che misura questi dati possono costituire conoscenza autentica? L'IA elabora informazioni o le comprende realmente? Per quanto riteniamo di estrema importanza questi quesiti, preferiamo qui approfondire il problema della rappresentazione, non meno rilevante nel campo dell'indagine filosofica sulle nuove tecnologie.

Perché risulta così importante interrogarsi sulla rappresentazione dei dati? Perché determina non solo ciò che può essere conosciuto, ma anche come questo si manifesta. La struttura dei dati influisce direttamente sull'applicazione del sistema: affinché un'IA sia davvero utile, il suo modello di rappresentazione deve essere adeguato ai suoi fini. Un database disorganizzato ha un valore limitato. Dunque, ignorare la rappresentazione e sistematizzazione dei dati, siano questi epistemicamente validi o meno, significa lasciare la macchina senza direzione, riducendola a un mero insieme di numeri.

Alcuni potrebbero sostenere che, finché la valenza epistemica dei dati non sia stata chiarita - ossia finché non si sia decretato se questi dati possano giustificare, sostenere o generare conoscenza autentica - risulti inutile approfondire il livello rappresentazionale. A questi obiettiamo con la tesi secondo cui l'IA può essere vista come un'estensione fisica - a livello hardware - di un sistema rappresentazionale fisico-matematico. Facciamo un esempio: quando svolgiamo operazioni algebriche, non includiamo nella notazione una vasta gamma di proprietà degli oggetti rappresentati, eppure i nostri segni sono sufficienti per ottenere un risultato universalmente riconosciuto come valido. Analogamente, anche se un sistema informatico non può, per il momento, integrare ai dati i qualia derivanti dall'esperienza sensibile, esso può comunque essere apprezzato al pari di una formula matematica, poiché in grado di produrre conclusioni logicamente coerenti e utili al nostro scopo.

In parole semplici, il risultato dell'operazione "ho due mele, ne mangio una" sarà sempre "mi resta una mela", indipendentemente dal colore della mela o dalle sue proprietà sensibili. Dunque, anche se i dati immagazzinati sono considerati epistemicamente poco attendibili, la loro rappresentazione resterebbe comunque valida e operativamente significativa.

Chiarito il motivo per cui il livello rappresentazionale può essere discusso indipendentemente da quello epistemico dei dati, procediamo alla nostra esposizione dei modelli di Intelligenza Artificiale.

2.1) Reti connessioniste vs. reti neuro-simboliche

Le principali scuole di pensiero che si pongono la questione della rappresentazione dei dati delle reti artificiali sono due: quella connessionista e quella neuro-simbolista. Il connessionismo è un particolare approccio alle scienze cognitive che, in termini riduzionisti, vuole spiegare il funzionamento della mente umana tramite l'uso di reti neurali artificiali. Lo fa sviluppando iper semplificazioni informatiche di apparati recettivi e cognitivi del cervello.

La proposta di sviluppare reti neurali connessioniste viene elaborata nel 1986 da tre psicologi statunitensi: J. L. McClelland, D. E. Rumelhart e G. E. Hinton. Particolarmente rilevante il loro lavoro *The Appeal of Parallel Distributed Processing* che si pone sin dall'inizio il seguente

interrogativo: “What makes people smarter than machines?”⁶. Rispondendo a questa domanda, i tre studiosi affermano che un uomo possa dirsi più intelligente di una macchina poiché “[the people’s brain] employs a basic computational architecture that is more suited to deal with a central aspect of the natural information processing tasks that people are so good at it”⁷. Se dunque si potesse ricreare un modello pari o superiore in *complessità* a quello umano, le reti artificiali sarebbero adatte al compito di rappresentazione così come lo è il cervello biologico. In tutto e per tutto.

Queste reti connessioniste costituiscono quindi un paradigma informatico imitativo della struttura del cervello umano, dove la conoscenza non è codificata simbolicamente, ma distribuita in una rete di unità semplici (neuroni artificiali, ovvero funzioni matematiche) collegate tra loro. Il loro funzionamento si basa sull’apprendimento statistico: in una prima fase la rete connessionista viene “addestrata” a riconoscere schemi o relazioni nei dati modificando i pesi delle connessioni interne. I pesi non sono altro che dei valori utili, all’interno dell’algoritmo, a fare sì che le funzioni su cui sono impostati diano risposte più soddisfacenti. Il processo di elaborazione riguarda delle operazioni statistiche, dove le risposte più attendibili hanno una probabilità più alta di venire fornite, mentre quelle meno rilevanti sono sempre più ridotte a valore zero. Sebbene la loro precisione nel fornire le risposte stia diventando, con l’implementazione di modelli di addestramento avanzati, sempre più precisa, esiste ancora la possibilità di *allucinazione* - termine tecnico che indica l’errore commesso dalla rete neurale - e pone inoltri grossi problemi dal punto di vista dell’interpretabilità della risposta.

Differentemente, le reti neuro-simboliche rappresentano un tentativo di superare i limiti ora illustrati del connessionismo puro, combinando l’apprendimento distribuito delle reti neurali con la struttura formale della logica simbolica. In questo paradigma informatico ibrido, il trattamento statistico dei dati tipico delle reti neurali viene integrato con rappresentazioni simboliche esplicite, come regole logiche o inferenze. Questo consente di mantenere l’adattabilità del connessionismo, ma senza rinunciare alla trasparenza, alla spiegabilità e la manipolabilità logica dei sistemi simbolici. L’obiettivo è quello di costruire modelli computazionali che apprendano dai dati attraverso il paradigma connessionista, ma che ragionino al pari di un sistema logico.

Approfondendo le differenze tra i due sistemi di rete, dobbiamo menzionare come per memorizzare un’informazione all’interno di una struttura rappresentazionale non si utilizzi un neurone singolo ma, come avviene nella mente umana, un ogni singolo neurone è destinato a ricevere un gran numero di stimoli informativi diversi, ed ha il compito di elaborare idee anche molto distanti fra loro: l’informazione è distribuita. Un rapido sguardo a questa tecnologia ci permette di comprendere come questa rappresentazione distribuita consenta alla macchina di catturare la complessità dei dati, rendendo queste reti perfette per compiti che richiedono l’estrazione e la comprensione di caratteristiche complesse da dati grezzi: riconoscimento di immagini, elaborazione del linguaggio naturale, sintesi vocale e comprensione semantica.

L’importanza di questa particolare distribuzione dei dati risiede nel fatto che gli oggetti della rappresentazione possono essere riferiti a *simboli e regole logiche*. La conoscenza appresa è rappresentata esplicitamente, sotto forma simbolica. A questi simboli sono riferiti dei *fatti* o delle

⁶J. L. McClelland, D. E. Rumelhart e G. E. Hinton, “The Appeal of Parallel Distributed Processing,” in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Volume 1, Capitolo 1, consultabile liberamente su <https://stanford.edu/~jlmcc/papers/PDP/Chapter1.pdf>.

⁷Ibid.

regole di inferenza, proposizioni⁸ che descrivono lo *stato del mondo*, ossia i fatti che nel mondo accadono. La macchina procede per situazioni, per scenari o disposizioni di fatti nel reale. Per fare qualche esempio:

Regola Logica 1 (RL1): *Se il paziente ha la febbre e la tosse, allora è possibile abbia l'influenza.*

RL2: *Se RL1 e difficoltà a respirare, allora potrebbe essere polmonite.*

RL3: *Se RL2 (RL1+difficoltà respiratorie) e "febbre alta", allora è necessario ricoverarlo al più presto!*

Questo discorso può essere ovviamente ampliato a scenari molto più complessi e possono essere date contemporaneamente anche centinaia di inferenze logiche a cui la macchina deve attenersi.

Una seconda differenza chiave tra i due tipi di reti è il modo in cui vengono addestrate. Mentre le reti di tipo connessionista vengono sommerse di dati da tokenizzare e da trasformare in relazioni probabilistiche, queste seconde di tipo rappresentazionale-simbolico sono pilotate da esseri umani. Le inferenze logiche vengono formulate da studiosi esperti con una preparazione matematica, logica e filosofica eccellente, cosicché l'input della macchina, non sarà una semplice stringa di testo, ma la rappresentazione simbolica di una serie di fatti, di *stati di cose* da cui trarre le inferenze progettate dai ricercatori.

Il metodo di verifica umana delle proposizioni logiche rappresenta una svolta cruciale nel dibattito sulle reti neurali, poiché affronta direttamente la questione dell'interpretabilità dei loro processi interni. Nelle reti neurali connessioniste, il "ragionamento" avviene su base statistica: le associazioni tra i dati, scomposti e trasformati in vettori matematici, si basano sulle probabilità con cui tali correlazioni sono emerse durante la fase di addestramento. Di conseguenza, il singolo passaggio dell'elaborazione non è direttamente interpretabile dall'uomo, poiché i dati vengono frammentati e resi inaccessibili al programmatore. Al contrario, nelle reti rappresentazionali il processo non si basa esclusivamente su calcoli algebrici applicati a unità tokenizzate in modo statistico, ma su operazioni secondo schemi logici completi. Questo approccio rende il ragionamento della macchina più trasparente e comprensibile per l'essere umano.

Da queste osservazioni si evince quali siano i contesti di applicazione delle reti neurali rappresentazionali rispetto a quelle connessioniste: contesti discriminatori o generativi nel caso delle reti neurali connessioniste e problemi logici, anche complessi, con situazioni che richiedono un ragionamento formale per quelle neuro-simboliste.

2.2) Una rete simbolista all'opera: il lupo, la capra e i cavoli

Prima di addentrarci ulteriormente nella nostra esposizione, guardiamo una rete neuro-simbolista all'opera con un famoso rompicapo logico. Per spiegare con rigore i passaggi necessari alla risoluzione del problema, dunque comportandoci analogamente alla macchina intelligente, sono stati inclusi nei passaggi della risoluzione degli elementi di logica dei predicati. Per chi non avesse

⁸L'introduzione delle regole logiche di inferenza che strutturano lo stato di cose nel mondo, è a nostro avviso una grande innovazione nel campo dell'IA. Questo perché il sistema non si limita ad associare probabilisticamente i dati - come avveniva nella rete connessionista - ma cerca di dare loro un senso logico, come facciamo noi umani quando pensiamo, quando facciamo filosofia. E da cosa dovrebbe essere composta la filosofia se non da proposizioni? "La filosofia non dovrebbe mai dimenticare che essa lavora sempre, comunque, con proposizioni. È la sua forza, e il suo limite". S. Benvenuto, "Il setaccio di Wittgenstein. Wittgenstein e l'etica," *aut aut* 394 (2023), p.. 49.

familiarità con questa disciplina, le spiegazioni parafrasate saranno più che sufficienti alla comprensione del nostro argomento.

Il rompicapo è famoso risale almeno al IX secolo ed è trascritto nel testo medievale *Propositiones ad acuendos juvenes*, di Alcuino di York (735-804), famoso filosofo e teologo anglosassone. Il problema è così formulato: un contadino si reca al mercato. Compra una capra, un lupo e un cesto di cavoli. Sulla via di casa è costretto ad attraversare con una chiatto un fiume, ma la barca può trasportare soltanto uno tra il lupo, la capra e i cavoli, assieme a lui. Se lasciati soli il lupo e la capra, la capra viene mangiata dal lupo. Se lasciata sola la capra con i cavoli, i cavoli vengono divorati dalla capra. Se il lupo viene lasciato solo con i cavoli, i cavoli non vengono mangiati, poiché il lupo è carnivoro. Come può il contadino trasportare tutti dall'altra parte del fiume senza perdere nessuno dei suoi averi?

Rappresentiamo il problema secondo il metodo simbolista. Le posizioni degli oggetti rispetto al fiume possono essere indicate con destra e sinistra. Le regole della logica sono le seguenti:

$L(x, t)$: indica che l'oggetto x è sulla sponda sinistra al tempo t considerato.

$R(x, t)$: indica che l'oggetto x è sulla sponda destra.

$T(x, t)$: il contadino T è con l'oggetto x (il contadino sta trasportando l'oggetto x in barca sull'altra sponda).

$E(x, y, t)$: indica che x è assieme ad y sulla stessa sponda (vorremmo quindi evitare che $E(\text{Lupo}, \text{Capra}, t)$ oppure $E(\text{Capra}, \text{Cavolo}, t)$). L'aggiunta della variabile temporale ci permette di formalizzare il fatto che, ad esempio, x e y compiono l'azione contemporaneamente, al medesimo tempo t .

$M(x, y, t)$: indica che x può mangiare y , condizione che non vogliamo assolutamente si realizzi.

(RL1) Se il contadino lascia capra e lupo non supervisionate *allora* il lupo mangia la capra:

- $\forall t. E(\text{Capra}, \text{Lupo}, t) \wedge \neg E(\text{Contadino}, \text{Lupo}, t) \Rightarrow M(\text{Lupo}, \text{Capra}, t)$

(RL2) Se il contadino lascia la capra con il cesto di cavoli senza una supervisione *allora* la capra mangerà i cavoli:

- $\forall t. E(\text{Capra}, \text{Cavolo}, t) \wedge \neg E(\text{Contadino}, \text{Capra}, t) \Rightarrow M(\text{Capra}, \text{Cavolo}, t)$

(RL3) Se il contadino trasporta x , al tempo t , *allora* x cambia sponda, al tempo $t+1$: da sinistra a destra o viceversa.

- $\forall x, t. L(x, t) \wedge T(x, t) \Rightarrow R(x, t+1) \wedge \neg L(x, t+1)$
- $\forall x, t. R(x, t) \wedge T(x, t) \Rightarrow L(x, t+1) \wedge \neg R(x, t+1)$

(RL4) Se all'istante t , x e y sono assieme sulla stessa sponda ($E(x,y,t)$) e il contadino trasporta x sull'altra sponda ($T(x,t)$), allora, dopo il trasporto, x e y non saranno più sulla stessa sponda.

$$\bullet \forall x, y, t. (E(x, y, t) \wedge T(x, t)) \Rightarrow \neg E(x, y, t+1)$$

Problema: ritrovarsi in uno *stato di cose* in cui lupo, capra e cavoli siano integri sulla sponda di destra contemporaneamente (al tempo t):

$$\forall t. (R(\text{Lupo}, t) \wedge R(\text{Capra}, t) \wedge R(\text{Cavolo}, t) \wedge R(\text{Contadino}, t))$$

Input iniziali, stati di cose del mondo: (1) lupo, capra, cavolo e contadino si trovano sulla sponda sinistra del fiume. (2) Gli elementi sono integri sulla sponda di sinistra.

A questo punto il motore logico-inferenziale della nostra rete rappresentazionale procede ad analizzare gli stati di cose del mondo in relazione alle regole opportunamente stabilite, meditando sulle condizioni logiche di risoluzione del problema. L'obiettivo è di trovare uno stato finale in cui tutti gli input iniziali siano soddisfatti; questo deve avvenire tramite una rigida consequenzialità logica.

La sequenza di azioni che porta alla soluzione può essere formalizzata come segue utilizzando la logica dei predicati⁹:

1. Passo primo: il contadino trasporta la capra sulla sponda destra.

$$T(\text{Capra}, 0) \wedge L(\text{Lupo}, 0) \wedge L(\text{Cavolo}, 0) \Rightarrow R(\text{Capra}, 1)$$

2. Il contadino torna da solo sulla sponda sinistra (la capra è sulla destra, il lupo col cavolo sulla sinistra).

$$E(\text{Lupo}, \text{Cavolo}, 1) \wedge R(\text{Capra}, 1)$$

3. Il contadino trasporta il lupo sulla sponda destra.

$$T(\text{Lupo}, 1) \wedge L(\text{Cavolo}, 1) \wedge R(\text{Capra}, 1) \Rightarrow R(\text{Lupo}, 2)$$

4. Il contadino torna con la capra sulla sponda sinistra.

$$T(\text{Capra}, 2) \wedge R(\text{Lupo}, 2) \wedge L(\text{Cavolo}, 2) \Rightarrow L(\text{Capra}, 3)$$

5. Il contadino trasporta il cavolo sulla sponda destra, lasciando la capra a sinistra.

$$L(\text{Capra}, 3) \wedge T(\text{Cavolo}, 3) \wedge R(\text{Lupo}, 3) \Rightarrow R(\text{Cavolo}, 4)$$

6. Il contadino torna da solo sulla sponda sinistra.

⁹ In questa sede non si vuole affrontare con rigore una dimostrazione logica del tutto esauriente, poiché la costruzione degli alberi semantici della logica dei predicati impegnerebbe il lettore più di quanto questa sezione dello scritto vuole fare. E' a sua discrezione decidere se affidarsi alle nostre inferenze o verificarle personalmente.

$$E(\text{Cavolo}, \text{Lupo}, 4) \wedge L(\text{Capra}, 4)$$

7. Il contadino trasporta la capra sulla sponda destra.

$$T(\text{Capra}, 5) \wedge E(\text{Cavolo}, \text{Lupo}, 5) \Rightarrow R(\text{Capra}, 6)$$

La rete rappresentazionale è dunque giunta al risultato che speravamo, facendo sì che ogni stato di cose fosse verificato nelle regole logiche stabilite dai suoi programmatori, ossia che si evitasse quella conformazione degli eventi in cui Capra e Cavolo condividessero la stessa sponda ($\forall t. E(\text{Capra}, \text{Cavolo}, t)$) oppure lo facessero Lupo e Capra ($\forall t. E(\text{Lupo}, \text{Capra}, t)$). La situazione a cui siamo giunti è:

$$R(\text{Capra}, 6) \wedge R(\text{Lupo}, 6) \wedge R(\text{Cavolo}, 6).$$

2.3) La tokenizzazione

Chiarito il modo in cui la rete neuro-simbolista opera logicamente, trattiamo in modo più approfondito le differenze nel metodo di tokenizzazione all'interno di una rete connessionista e neuro-simbolica. In generale, la tokenizzazione riguarda la traduzione delle espressioni di input in vettori matematici, ai fini dell'analisi. Nell'informatica delle reti artificiali, è un passaggio preliminare fondamentale per permettere ai modelli linguistici di elaborare il linguaggio naturale che viene dettato dall'utente, trasformandolo in sequenze numeriche interpretabili dalla macchina.

La tokenizzazione funziona nel seguente modo: l'input testuale viene suddiviso in unità di testo più piccole e frammentate, dette *token*. Questi token vengono poi trasformati in numeri, nello specifico in valori vettoriali, che poi la rete neurale può apprendere ed elaborare per fare delle *previsioni* o *analisi del testo*. In questo modo la rete ipotizza le risposte future. In pratica, questi token sono parole, sillabe, o persino singoli caratteri, a scelta del programmatore. La frase "adoro scoprire come funziona l'Intelligenza Artificiale" potrebbe ad esempio essere descritta dapprima come ["adoro", "scoprire", "come", "funziona", "l'", "Intelligenza", "Artificiale"], poi in ["a", "do", "ro", "sco", "pri", "re", ecc...], riducendo le unità semantiche in parti sempre più piccole. Ognuna di queste unità va poi descritta matematicamente, attraverso una rappresentazione vettoriale, ad esempio il nostro ["a", "do", "ro"] si trasforma in ["[1,2,3]", "[1,3,3]", "[1,4,3]"]¹⁰, facendo sì che la macchina possa elaborare algebricamente queste informazioni. Le espressioni linguistiche sono così diventate dei valori vettoriali.

La maggiore preoccupazione risiede nel fatto che, una volta mappati come vettori, i token perdono in un certo senso il valore logico/semantico che avevano all'interno della frase, "scadendo" a rappresentazioni informatico-algoritmiche. Nella rete neuro-simbolista i token non sono solo vettori numerici, ma sono anche associati a simboli e regole esplicite: prima che la frase sia tokenizzata come nei sistemi neurali connessionisti, dei modelli di analisi del sistema linguistico associano all'enunciato di partenza una serie di altri

¹⁰ I numeri sono arbitrariamente scelti e non hanno una valenza matematica reale in questo contesto.

enunciati collegato a esso da relazioni inferenziali, catturando in questo modo parte del suo contenuto semantico. Queste inferenze vengono poi a loro volta tradotte nei token del tradizionale processo connessionista, ma questa volta è il simbolo a venire mappato in un oggetto matematico. In questa fase, le reti neurali neuro-simboliste sono implementate con sistemi come Word2Vec¹¹ e GloVe¹², algoritmi che preservano le relazioni semantiche tra i token, anche avendo matematizzato le informazioni.

Avviene poi il processo che riguarda il vero e proprio *ragionamento* della rete neuro-simbolica. Ogni strato della rete cattura i valori trasmessi dalle relazioni complesse tra i simboli logici tokenizzati e valutati con attribuzione di pesi sinaptici, permettendo alla rete di inferire nuove proposizioni con significato semantico autentico.

¹¹Si tratta di un algoritmo implementato da un team di ricercatori di Google che funziona rappresentando le parole come vettori numerici densi (dove allora ogni componente del vettore contiene un valore numerico diverso da zero), facendo sì che le parole con significati simili siano logicamente posizionate all'interno dello spazio vettoriale multidimensionale in modo ravvicinato, permettendo un'associazione semantica di significato, non su base statistica completa. Una completa illustrazione matematica dei parametri di apprendimento di questo algoritmo è reperibile in Xin Rong, *word2vec Parameter Learning Explained*, disponibile su <https://formacion.actuarios.org/wp-content/uploads/2024/05/1411.2738-word2vec-Parameter-Learning-Explained.pdf>.

¹²Un ulteriore modello di *word embedding* con capacità di predizione semantica. Per un approfondimento, cfr. J. Pennington, R. Socher, C. D. Manning, *GloVe: Global Vectors for Word Representation*, disponibile su <https://aclanthology.org/D14-1162.pdf>.

3) Il Tractatus e lo spazio logico

Il metodo di rappresentazione sopra descritto apre una prospettiva estremamente innovativa: invece di rappresentare i dati come semplici vettori su base statistica, questi vengono espressi attraverso simboli logici che ne catturano almeno in parte il significato. Ciò consente di associare vettori vicini nello spazio vettoriale anche quando si riferiscono a oggetti apparentemente distanti nella realtà. L'affinità tra questi oggetti non dipende solo dalla loro distanza matematica, ma anche dalla struttura logica che condividono, ovvero dal loro modo di esistere e di relazionarsi nel mondo. Discutiamone attraverso le proposizioni¹³ del celeberrimo *Tractatus Logico-Philosophicus* di Ludwig Wittgenstein, testo imprescindibile se in filosofia, matematica o logica, si vuol trattare la rappresentazione dei dati. L'autore nel *Tractatus* enuncia [2.18]:

*“Ciò che ogni immagine, di qualunque forma, deve avere in comune con la realtà per poterla raffigurare – in modo corretto o errato – è la forma logica, cioè la forma della realtà”*¹⁴.

Sostanzialmente, Wittgenstein afferma che un'immagine, oppure un simbolo, può rappresentare la realtà nella misura in cui ne condivide la struttura logica interna. La suddetta struttura logica è ciò che permette all'immagine o al simbolo di corrispondere a una situazione del mondo reale, ad uno stato di cose veritiero, rappresentando il mondo attraverso un simbolo o un'immagine. Anche se l'immagine potrebbe differire dall'oggetto che rappresenta, non viene meno la possibilità di rappresentazione secondo una struttura logica identica: La forma logica è dunque quell'isomorfismo tra l'immagine e la realtà che consente all'immagine, o al simbolo, di rappresentare la realtà fedelmente. Ma nelle reti connessioniste, poiché statistiche e prive di vero fondamento linguistico, la nozione di forma logica veniva smarrita tra gli strati neurali nascosti della rete. Per banalizzare ai fini della comprensione, una proposizione come “La capra mangia il cavolo”, per restare nel nostro esempio, avrebbe le sue difficoltà se dovessimo sostituirla con “La pecora mangia il cavolo” in una rete connessionista, poiché sarebbe tokenizzata secondo delle regole algoritmiche che la scomporrebbero in sillabe e la forma logica verrebbe persa in una sequenza statistica di simboli. Nella rete neuro-simbolica, invece, condividendo questi due elementi la stessa struttura logica - che potremmo esplicitare come una struttura logica soggetto-verbo-oggetto - sarebbero caratterizzati da una vicinanza matematica in uno spazio vettoriale, perché logicamente simili.

Approfondiamo un poco quanto si intende qui per struttura logica e come Wittgenstein affronta nelle proposizioni a seguire questo concetto.

*“Se la forma della raffigurazione è la forma logica, allora l'immagine è chiamata l'immagine logica.”*¹⁵.

Wittgenstein sottolinea come l'immagine possa asserire qualcosa sul mondo. La forma della raffigurazione, ossia il modo in cui gli elementi di un'immagine si pongono per essere una raffigurazione, sono la forma logica, ovvero quanto prima si discuteva: la struttura che

¹³Per specificare cosa si intenda con il termine “proposizione” all'interno del contesto wittgensteiniano ci rifacciamo all'analisi di Massimo de Carolis, che nota come questo concetto muti nel pensiero dell'autore. “Nel *Tractatus* [...] una proposizione era considerata un'unità puramente *logica*, indipendente dalla forma materiale in cui è espressa”, e diverrà in una seconda fase poi “un'unità *linguistica*, che ha senso solo se concretamente *enunciata* all'interno di un gioco [...]”. Massimo de Carolis, “Prassi, forme di vita e dinamica delle convenzioni in Wittgenstein,” *aut aut*, 394 (2023), pp. 49-67.

¹⁴Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, prop. 2.18.

¹⁵Ibid. prop. 2.181

permette all'immagine di corrispondere alla realtà. Se il modo in cui gli elementi dell'immagine si pongono per rappresentare aderisce alla struttura logica di fondo, si può allora dire che l'immagine è immagine logica. Questa immagine può infatti “[...] raffigurare il mondo.”¹⁶ ed “ha in comune con il raffigurato la forma logica della raffigurazione.”¹⁷. Ma queste sono solo le prime assonanze della “struttura” che il pensiero wittgensteiniano condivide con la rappresentazione dei dati nell’Intelligenza Artificiale. La proposizione 2.202 ci dice qualcosa di ancora più profondo, ossia che:

*“L’immagine presenta uno stato di cose possibile nello spazio logico”*¹⁸.

L’immagine -forma logica- presenta un modo di disporsi delle cose che è possibile in uno *spazio logico*¹⁹. Ma cosa si intende per questo spazio? Ebbene, il filosofo allude ad un contesto logico e spaziale in cui risiedono le possibilità per gli oggetti di combinarsi tra loro in un qualche modo, risultando in uno stato di cose. Lo spazio logico è il dominio delle possibilità logiche, l’insieme logico-spaziale che contiene tutte le possibili combinazioni degli oggetti affinché formino degli stati di cose, delle disposizioni delle cose nel mondo. “Con questa espressione, spazio logico, egli evoca, in effetti, lo spazio del possibile, ossia, come l’aggettivo logico suggerisce, lo spazio di tutto il possibile (vedi T, 2.0121c). Con ciò Wittgenstein ci spinge a ricordare qualcosa che, per così dire, già sappiamo: che ciò che accade avrebbe anche potuto non accadere o avrebbe potuto accadere in maniera più o meno diversa da come è accaduto”²⁰. Approfondiamo meglio questo concetto, poiché a breve lo accosteremo all’Intelligenza Artificiale in modo non del tutto scontato.

Per un’approfondita comprensione dello spazio logico, G. Piana illustra inizialmente il concetto di *spazio* come un luogo che può essere occupato da delle cose, e “l’occupazione di un luogo è la *realizzazione* di una *possibilità* di localizzazione”²¹. Questi luoghi saranno determinati relazionalmente: “determinare un luogo è niente altro che determinare una relazione”²² e ciò avviene perché “lo spazio come insieme di luoghi sarà nello stesso tempo l’insieme delle relazioni che li determinano”²³. Gli oggetti in questo determinato spazio sono dunque relazionati, ma anche inseparabili dallo spazio stesso; come fa notare Wittgenstein “Ogni cosa è in uno spazio di possibili stati di cose. Questo spazio non posso pensarlo vuoto, ma non posso certo pensare la cosa senza lo spazio”²⁴. Anche uno spazio vuoto contiene degli oggetti, cose espresse come possibilità non realizzate di essere cose. In definitiva, uno spazio è sempre e comunque *ricolmo* di possibilità di esistenza, presenza, delle cose. “Infine: ogni spazio giace nello *spazio logico*. Tutte le configurazioni di oggetti, [...] debbono essere anche configurazioni logiche. Perciò

¹⁶Ibid. prop. 2.19

¹⁷Ibid. prop. 2.2

¹⁸Ibid. prop. 2.202

¹⁹ Tratta in modo più analitico la nozione di spazio logico, anche dal punto di vista matematico e non solo filosofico L. Guidetti (2024). Lo spazio logico in Wittgenstein a confronto con l'assiomatica moderna. Algebra ipercubica e fenomenologia del significato, in V. Costa, S. Galanti Grollo, L. Guidetti, E. Mariani (eds.), Fenomenologia e realtà. Studi in onore di Stefano Besoli, Mimesis, Udine/Milano.

²⁰Luigi Perissinotto, *Introduzione a Wittgenstein*, Il Mulino 2018, pg. 65.

²¹Giovanni Piana, *Interpretazione del Tractatus di Wittgenstein*, Guerini e Associati 1994, pg. 20.

²²Ibid.

²³Ibid.

²⁴Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, prop 2.013.

possiamo affermare che nulla di illogico può accadere, che tutti i fatti sono nello spazio logico”²⁵.

Per riassumere, lo spazio logico è concepito come un intreccio di relazioni che determinano i luoghi e le possibilità di esistenza degli oggetti al loro interno. Ogni luogo è definito dalla rete di relazioni che lo costituisce, e gli oggetti stessi non possono essere pensati al di fuori dello spazio che li accoglie. Anche il vuoto si configura come un insieme di possibilità latenti, di stati di cose che possono realizzarsi. Tutto ciò che accade è inscritto nello spazio logico, dove ogni configurazione di oggetti è necessariamente anche una configurazione logica: nulla di illogico può accadere, tutto esiste nello spazio logico.

3.1) Spazio logico e spazio vettoriale multidimensionale.

Proponiamo ora un accostamento concettuale. Lo spazio logico wittgensteiniano trova una sorprendente corrispondenza con gli spazi vettoriali utilizzati nelle reti neuro-simboliche. In queste, ogni vettore rappresenta una possibilità, ovvero una specifica configurazione di dati all'interno di un sistema logico-linguistico. Lo spazio vettoriale, in cui le relazioni logico-semantiche sono descritte tramite vettori multidimensionali, condivide la stessa multidimensionalità dello spazio logico delineato nel *Tractatus*, che si configura come una struttura coerente. Così come i vettori appartengono a uno spazio matematico multidimensionale, gli stati di cose sembrano appartenere allo spazio logico wittgensteiniano. In questo senso, le proposizioni e gli stati di cose possono essere accostati nello stesso modo in cui la macchina stabilisce connessioni tra i dati.

Ma questo non è tutto. Prendiamo in considerazione quanto G. Piana afferma rispetto alla *compatibilità* di due proposizioni. “Diciamo che due proposizioni sono compatibili se la verità dell'una non esclude la verità dell'altra. [...] tale rapporto sussiste tra due proposizioni quando almeno una possibilità di verità delle proposizioni elementari è una condizione di verità per entrambe, cioè quando esse hanno un fondamento di verità comune”²⁶.

Due proposizioni si dicono tra loro compatibili quando condividono un fondamento di verità, ossia una base logica che permette loro di essere vere contemporaneamente senza contraddirsi, coesistendo nello spazio logico medesimo. Quando almeno una condizione di verità è soddisfatta per entrambe, due proposizioni possono sussistere nello spazio logico senza annullarsi o contraddirsi. Ad esempio, due proposizioni elementari come (A) “Il violino è sulla scrivania nel momento T” e (B) “La candela è accesa nel momento T”, coesistono nello spazio logico senza contraddizione e sono perciò due proposizioni atomiche compatibili. Assumere invece (C) “La penna tocca il calamaio nel momento T” e (D) “La penna è nella tasca nel momento T” rende impossibile la compatibilità tra le due proposizioni elementari: queste non potranno coesistere nello stesso spazio logico. Questo concetto di compatibilità è collegato alla visione della logica come un sistema di regole che determinano le possibilità di combinazione e di coesistenza delle proposizioni, ergo degli stati di cose nel mondo. Un fondamento di verità comune, una coesistenza all'interno dello spazio logico, è quindi detta come compatibilità tra le

²⁵Giovanni Piana, pg. 21

²⁶Piana pg. 84.

proposizioni, e permette loro di condividere alcune caratteristiche essenziali della realtà senza entrare in conflitto.

Immaginando, come prima si era fatto, lo spazio logico come un insieme matematico complesso dove le proposizioni sono descritte da vettori multidimensionali, a cosa si potrebbe riferire la compatibilità di due proposizioni? Ebbene, il grado di compatibilità di due proposizioni è dato dal numero di realtà condivise in cui queste possono essere entrambe vere. Dunque maggiore sarà il loro grado di compatibilità, maggiore sarà la loro vicinanza nello spazio logico. Se la compatibilità tra due proposizioni dipende dal numero di possibilità di verità condivise, allora maggiore è questa condivisione, minore sarà la distanza tra le due proposizioni nello spazio logico. Questo suggerisce che lo spazio logico wittgensteiniano può essere concepito come una struttura in cui la vicinanza tra proposizioni riflette il grado di coesistenza tra le loro condizioni di verità. Similmente, nello spazio matematico, due vettori vengono assunti come ravvicinati quando condividono molte componenti riferite a stesse dimensioni. Si potrebbe allora immaginare lo spazio logico come un insieme in cui giacciono le proposizioni accadute e possibili, rappresentate sotto forma di vettori. Questi vettori, definiti dalle loro dimensioni, risultano tanto più vicini tra loro quanto maggiore è il numero di componenti dimensionali che condividono, ovvero quanto più alto è il loro grado di compatibilità e il numero di condizioni di verità comuni.

Riassumendo, lo spazio logico wittgensteiniano si trova in analogia con gli spazi vettoriali delle reti neuro-simboliche, dove ogni vettore rappresenta una possibilità logico-linguistica. La compatibilità tra proposizioni, intesa come condivisione di un fondamento di verità comune, si riflette nella capacità delle Intelligenze Artificiali di accostare contenuti logici, avvicinandole alla cognizione umana. Per generare una risposta, l'IA si affida allo spazio logico e, una volta attivati determinati vettori, raccoglie informazioni dai vettori ad essi più vicini. Questo processo garantisce coerenza di senso, poiché seleziona proposizioni con un alto grado di compatibilità rispetto a quella considerata.

Riconoscere questa analogia concettuale non è un semplice intellettualismo, ma un modello estremamente efficace per affrontare la rappresentazione dei dati logici nel contesto dell'Intelligenza Artificiale. Trattare lo spazio matematico multidimensionale come lo spazio logico wittgensteiniano significa avere già un'ampia letteratura sull'argomento, poiché Wittgenstein, descrivendo la natura della logica e della realtà, ha contemporaneamente implementato i modelli logico-matematici utili allo sviluppo della rappresentazione dei dati per le reti neurali e le IA. Il neuro-simbolismo, sotto la guida indiretta del filosofo, sembrerebbe essere un sistema già largamente implementato nella teoria che, una volta riconosciuta come valida, aspetta di essere applicata. Le nuove tecnologie sembrano ripercorrere, a vari livelli, i passi operati dal filosofo del *Tractatus*, evidenziando una crescente attenzione accademica verso il contributo implicito del testo logico-filosofico alle teorie della rappresentazione computazionale²⁷.

²⁷ Oltre alla sezione quarta del testo, che esplora un range più ampio di studi, rimandiamo, per quanto riguarda il pensiero di Wittgenstein e l'Intelligenza Artificiale, a Klaus K. Obermeier, Wittgenstein on Language and Artificial Intelligence: The Chinese-Room Thought Experiment Revisited, in *Synthese*, 56/3, 1983; Stuart S. Shanker, Wittgenstein's Remarks on the Foundations of AI, Routledge, 1998; Marko Kardum, Wittgenstein's Tractatus, the Vienna Circle, and Artificial Neural Networks as the First Computational Theory of Mind: Does Development of AI Owe Anything to Wittgenstein?, in Skelac, Ines, Belić, Ante (eds.), What Can Be Shown Cannot Be Said. Proceedings of the International Ludwig Wittgenstein Symposium, Zagreb, Croatia, 2021, Münster: LIT Verlag, 2023; Alice C.

Ecco che riconosciamo la potenza di questo sistema: è possibile per il neuro-simbolismo, come per la mente umana, generalizzare e astrarre i contenuti logici degli oggetti riscontrati nel reale. È forse possibile che questo modo di rappresentazione della conoscenza avvicini ancor più le macchine a noi? Secondo un primo Wittgenstein no. Come fa notare C. Penco²⁸, “Wittgenstein considered attributing thinking to machines a category mistake, like attributing colour to numbers²⁹ or speaking of “artificial pain”³⁰. Eppure, specifica poi l'autore, “his later remarks are more ambivalent, and interpretations are divided between antagonists, who interpret Wittgenstein’s work as a means to contrast AI, and compatibilists, who see Wittgenstein as an inspiration for AI”. Ciò che conta nel dibattito rispetto l’IA pensante è che l’IA rappresentazionale pone, come fa notare Marco Brigaglia, un’immagine simbolica della mente, la cosiddetta “teoria rappresentazionale della mente”. L’idea centrale è che la cognizione naturale possa essere interamente spiegata in termini di manipolazione di simboli governata da regole sintattiche. Il pensiero, in particolare, avrebbe struttura linguistica: pensare significa concepire frasi in ‘linguaggio del pensiero’, stringhe di simboli discreti composti e trasformati sulla base di regole sintattiche. Per quello che qui più conta, in questo modello la mente è raffigurata come strettamente analoga ad una IA classica: una macchina simbolica. Proprio, in virtù di questa corrispondenza, si ritiene, una IA classica può emulare con successo le caratteristiche della cognizione naturale”³¹. Sebbene la teoria rappresentazionale della mente, proposta da Fodor³² e qui chiarita da Brigaglia, sia ora stata ampiamente superata³³, pone comunque delle interessanti questioni non solo rispetto ad un futuro in cui l’Intelligenza Artificiale possa modellare forme complesse di pensiero simbolico, ma anche in relazione alla natura stessa del pensiero umano, inteso come manipolazione strutturata di rappresentazioni. Essa costringe infatti a confrontarsi con il problema della sintassi e della semantica, con le domande più profonde rispetto il linguaggio.

Wittgenstein, con il suo *Tractatus*, non solo getta le basi della logica tradizionale così come la conosciamo oggi, ma spiana anche la strada all’architettura dei modelli di rappresentazione che le macchine potrebbero sfruttare per essere sempre più *efficacemente umane*. Prendere consapevolezza di questo significa osservare i modelli matematici rappresentazionali con uno sguardo filosofico, strutturandoli architettonicamente per renderli non solo più efficienti, ma anche più aderenti alla complessità del linguaggio, della cognizione umana e della logica immanente nel mondo.

Helliwell, Alessandro Rossi, Brian Ball, Wittgenstein and Artificial Intelligence. Volume I e Volume II, Anthem, 2024.

²⁸Carlo Penco, *Wittgenstein, Contexts, and Artificial Intelligence. An Engineer Among Philosophers, a Philosopher Among Engineers*.

²⁹Ludwig Wittgenstein, *Blue Book*, p. 47.

³⁰Ludwig Wittgenstein, *Philosophical Grammar*, p. 64.

³¹Marco Brigaglia, *L’IA e la rappresentazione di noi stessi, come tristi macchine allo specchio*.

³²Fodor, Jerry A., *The Language of Thought*, Harvard University Press, 1975.

³³Cfr. ad esempio dibattito tra Searle vs Dennett (Chua, John Moses. *Formulating Consciousness: A Comparative Analysis of Searle’s and Dennett’s Theory of Consciousness*. Talisik: An Undergraduate Journal of Philosophy, vol. 4, n. 1, 2017, pp. 43–58. Disponibile su: <https://philarchive.org/archive/CHUFCA>.) o ll’emergentismo cognitivo di Hofstadter (Hofstadter, Douglas R. & Dennett, Daniel C., *The Mind’s I: Fantasies and Reflections on Self and Soul*, Basic Books, 1981.)

4) Perché Wittgenstein. Un inevitabile confronto

Ma in cosa risiede il motivo preciso per cui è opportuno rispolverare i testi, o il testo, wittgensteiniani nel campo dei nuovi modelli di Intelligenza Artificiale? In fondo, si potrebbe obiettare che tra la logica filosofica dell'autore e le prime forme di Intelligenza Artificiale vi siano diversi anni di scarto; questo scarto temporale renderebbe il pensiero di Wittgenstein inattuale, anacronistico, per i nostri argomenti. Come fa giustamente notare G. Casey, "Wittgenstein died in 1951 and the term artificial intelligence didn't come into use until 1956 so that it seems unlikely that one could have anything to do with the other. However, establishing a connection between Wittgenstein and artificial intelligence is not as insuperable a problem as it might appear at first glance"³⁴. Nello specifico, "While it is true that artificial intelligence as a quasi-distinct discipline is of recent vintage, some of its concerns, especially those of a philosophical nature, have been around for quite some time"³⁵; in effetti "The term 'Artificial Intelligence' was introduced to the world by John McCarthy and Marvin Minsky at a conference in Dartmouth, New Hampshire, in 1956. Whether or not it had a clear and unambiguous meaning at that time, since then it has acquired a range of interpretations"³⁶. A partire da Descartes, che si chiedeva se fosse possibile creare una macchina che fosse fenomenologicamente indistinguibile dall'uomo³⁷, passando per l'esperimento mentale della "macchina mentale" di Leibniz, fino ad arrivare a Turing, tra i padri fondatori dell'Intelligenza Artificiale moderna e risaputo uditor delle lezioni di Wittgenstein sulle fondamenta della matematica³⁸, la filosofia si è spesso interrogata su come e se le macchine possano imitare la mente umana. Wittgenstein non è indifferente a questa discussione; per quanto indirettamente e forse senza nessuna volontà di farlo, i suoi contributi sono stati - come a breve vedremo - e sono - come dalla nostra ricerca si evince - di vitale importanza per lo sviluppo di queste macchine.

È necessario allora spiegare ordinatamente i motivi per cui ci sembra che il pensiero dell'autore si situi al centro del dibattito sulle nuove tecnologie.

4.1) Psicologia della *Gestalt* e filosofia della percezione

Primariamente, osserviamo che i suoi lavori nel contesto dello studio della psicologia della percezione gestaltista siano di vitale importanza: ci aiutano a considerare le dimensioni essenziali dell'esperienza e della percezione umana, mai scadendo nel pensiero riduzionista. Come si evince dal saggio "*Osservazioni sulla Filosofia della Psicologia*"³⁹, una serie di appunti raccolti e pubblicati postumi in un unico volume, il pensiero

³⁴Gerard Casey, *Artificial Intelligence and Wittgenstein.*, pg. 2.

³⁵Ibid.

³⁶ Ibid, pg 4.

³⁷Cfr. René Descartes, *Discorso sul metodo* (1637), in *Opere filosofiche*, Laterza, 2005 e anche René Descartes, *Meditazioni metafisiche* (1641), in *Opere filosofiche*, Laterza, 2005

³⁸Per una più ampia illustrazione di come Wittgenstein contribuì indirettamente, tramite gli studi di Turing, allo sviluppo delle macchine di Intelligenza Artificiale, rimandiamo a Antonio Pennisi e Alessandra Falzone, "Wittgenstein, Turing e l'IA. Un'intervista a ChatGPT." *Università di Messina*. Un altro testo che ci sembra di fondamentale rilevanza nel campo di IA e Machine Learning -sempre confrontandoci con Turing- è Fabio Grigenti, "Intelligenza artificiale. Dalla macchina di Turing al machine learning", in *Filosofia del digitale*, a cura di Andrea Taddio e Carlo Giacomini (Milano: Mimesis, 2021).

³⁹Ludwig Wittgenstein, *Osservazioni sulla filosofia della psicologia*.

wittgensteiniano comprende le nuove teorie della percezione gestaltista, vitali nello scorso secolo. La psicologia della *Gestalt* si concentra sul tema della percezione, attribuendo all'esperienza umana un valore ontologico profondo. Questa teoria sostiene che la mente non si limita a registrare passivamente gli stimoli, ma li organizza attivamente in strutture unitarie e significative. Segue principi come quello della *prossimità* – per cui gli elementi vicini tendono a essere percepiti come un'unica entità – e quello della *somiglianza*, secondo cui elementi simili vengono naturalmente raggruppati. Il punto centrale della *Gestalt* è che il tutto non è la somma delle parti. La percezione non è una semplice aggregazione di dati sensoriali, ma il risultato di un'organizzazione intrinseca della mente.

Come questo influisce nel contesto dello studio dell'AI? Ebbene, se la tecnologia vuole davvero avvicinarsi alla perfetta riproduzione delle qualità umane, deve tener conto del funzionamento profondo di queste e delle loro proprietà di strutturazione psicologica delle informazioni. Ciò significa che un modello di Intelligenza Artificiale che vuole percorrere il sentiero della “macchina umana” non può essere costruito senza delle nozioni sulla percezione ed organizzazione cognitiva dei dati che immagazina. Ricostruire il valore umano dell'esperienza ci permette di non abbandonare la nozione di percezione nel senso fenomenologico del termine, e quindi di costruire un'IA che tenga conto delle considerazioni fatte dai campi di ricerca psicologici e anche fenomenologici⁴⁰. In una realtà sempre più “matematizzata”, la riflessione wittgensteiniana ci ricorda cosa significhi “percepire” e comprendere nel senso più strettamente umano del termine.

Il lettore potrebbe ora intendere che qui si stia discutendo la nozione di “contesto” che la filosofia di Wittgenstein e della *Gestalt* condividono. Piuttosto che soffermarci su questo confronto, già ampiamente discusso in svariati articoli accademici⁴¹, preferiamo mantenere la nostra attenzione su quanto già introdotto a proposito della proposizione 2.18.

4.2) Le applicazioni nel contesto semantico.

Il secondo motivo per cui riteniamo attuale l'analisi di Wittgenstein nell'ambito delle Intelligenze Artificiali è che il suo pensiero ha influenzato direttamente studiosi come Margaret Masterman (1910-1986), che ha utilizzato le sue idee per sviluppare tecniche di elaborazione del linguaggio naturale. Le intuizioni del filosofo austriaco, in particolare quelle relative all'importanza del contesto, sono state fondamentali per costruire sistemi che tengono conto della semantica contestuale. Queste idee sono alla base di sviluppi tecnologici come i motori di ricerca, gli assistenti vocali e le reti neurosimboliche, che combinano simbolismo e apprendimento profondo. Nel suo volume *Language, Cohesion and Form: Selected Papers*, Masterman presenta e approfondisce il concetto di *thesaurus*,

⁴⁰Cfr. Luca Taddio, *Maurice Merleau-Ponty* rispetto il ruolo di Wittgenstein e della psicologia della Gestalt.

⁴¹Solo per citarne alcuni, abbiamo *Wittgenstein's Philosophy of Language The Philosophical Origins of Modern NLP Thinking e Wittgenstein, Contexts, and Artificial Intelligence. An Engineer Among Philosophers, a Philosopher Among Engineers*. Carlo Penco.... Entrambi discutono, in modo più o meno approfondito, un aspetto diverso da quello che abbiamo trattato in questa sede. Per dare una rapida idea, questi discutono come in Wittgenstein il significato di un termine simbolico dipenda dal contesto e dall'uso che viene fatto di questo nella pratica quotidiana. Così come nella psicologia gestaltista si enfatizza il modo in cui la mente umana percepisce globalmente le informazioni, alla base nei modelli di reti neurali e del linguaggio che vanno recentemente sviluppandosi e che si ispirano a queste nozioni teoriche per apprendere e rappresentare le informazioni in modo contestuale.

inteso come un sistema linguistico classificato in insiemi di contesti. Questo approccio ha avuto un ruolo centrale nella moderna IA, in particolare nello sviluppo di tecniche di clustering semantico⁴² e di classificazione concettuale.

Un esempio pratico dell'applicazione di queste idee è rintracciabile nel lavoro di K. S. Jones, *Synonymy and Semantic Classification*, dove il *thesaurus* viene esteso e utilizzato per sviluppare metodi statistici per il trattamento del linguaggio naturale. Inoltre, l'idea di costruire rappresentazioni del linguaggio come mappe concettuali o reti di significati trova ulteriore conferma nei lavori di Peter Gärdenfors, autore di *Conceptual Spaces: The Geometry of Thought* (2000). Questo approccio è oggi rilevante per comprendere le basi teoriche dei modelli di linguaggio neurali, come GPT e BERT, che elaborano il linguaggio in uno spazio multidimensionale matematico in cui i concetti sono organizzati secondo relazioni contestuali.

L'impatto delle idee wittgensteiniane tocca, infine, anche il dibattito contemporaneo sulla spiegabilità dell'IA. Come sostenuto da Tim Miller in *Explanation in Artificial Intelligence: Insights from the Social Sciences* (2019), contesto e significato d'uso sono elementi cruciali per costruire sistemi di intelligenza artificiale interpretabili.

Se quanto esposto ha dimostrato la rilevanza della filosofia e della logica di Wittgenstein nel contesto dell'Intelligenza Artificiale, allora possiamo affermare che le conclusioni tratte dai suoi testi più noti giustificano pienamente il suo inserimento nel dibattito filosofico-tecnologico contemporaneo. Non vi è alcun motivo per escludere Wittgenstein da questa discussione, poiché il suo pensiero continua a offrire strumenti essenziali per comprendere le questioni teoriche legate all'IA.

⁴²Il cluster semantico è un raggruppamento di parole o frasi che condividono significati simili o che sono strettamente correlate rispetto al loro contesto. Questo concetto è utilizzato nell'elaborazione del linguaggio naturale (NLP) per organizzare gli elementi in insiemi che risultino coerenti, rendendo più accessibili ricerche di parole e composizione di testi. I cluster si basano su tecniche statistiche e/o su modelli neurali.

5) Conclusioni

Il testo appena letto ha dimostrato l'importanza di integrare le prospettive filosofiche nello studio e nello sviluppo della tecnologia, offrendo una chiave metodologica fondamentale per la progettazione dell'IA. In particolare, nelle reti neuro-simboliche, la rappresentazione è un nodo cruciale per comprendere e guidare il rapporto tra cognizione e organizzazione spaziale delle informazioni, sia umane che artificiali.

Si evince dal testo come, quando la tecnologia tenti di avvicinarsi al funzionamento dell'essere umano, il bisogno di riflessione filosofica si faccia ancora più pressante. La soluzione probabilistica delle decisioni algoritmiche, la relazione tra percezione e rappresentazione simbolica, il modo in cui strutturiamo il nostro pensiero logico: tutto questo non può essere affrontato con il solo ausilio della matematica o dell'ingegneria. Sono necessarie delle analisi che permettano di imprimere un senso ai dati, di trasformare il calcolo in conoscenza e di distinguere tra ciò che un sistema è programmato per fare e ciò che un sistema comprende. Non è un caso che, nel corso della storia, filosofi come Wittgenstein abbiano influenzato profondamente non solo le scienze psicologiche e cognitive, ma anche l'architettura di modelli computazionali. I lavori di Margaret Masterman, K. S. Jones, Peter Gärdenfors e Tim Miller dimostrano come le intuizioni wittgensteiniane sulla natura del funzionamento del linguaggio abbiano plasmato l'analisi semantica e la rappresentazione della conoscenza nell'IA.

Se la tecnologia vuole realmente comprendere e riprodurre aspetti dell'intelligenza umana, non può evitare la filosofia, perché solo la filosofia fornisce gli strumenti utili alla concettualizzazione della conoscenza, all'indirizzamento verso una meta precisa, all'architettura del sapere.

Questo nostro intervento lascia tuttavia aperta una domanda di fondo, ossia se la convergenza tra cognizione naturale e artificiale sia davvero possibile, o se siamo destinati a riconoscere tra queste un confine invalicabile. Più nello specifico, dobbiamo comprendere se per raggiungere un risultato più vicino all'umano basterà integrare questi spazi matematici con sempre maggiori dimensioni, creando dei vettori così complessi da poter catturare matematicamente la complessità della realtà che ci circonda. E se la disamina ingegneristico/filosofica riguarda solo questi aspetti quantitativi, può una riduzione dimensionale della realtà in termini matematici soddisfare le nostre esigenze di enormi quantità di dati? La risposta dipenderà sia dalla capacità di sviluppare modelli in grado di integrare un maggior numero di elementi nelle loro formulazioni sia dall'uso di strumenti concettuali e filosofici che permettano di comprendere e guidare le nuove tecnologie. Resta il compito, filosofico e tecnico, di esplorare questa possibilità senza pregiudizi, con lo sguardo rivolto non solo al progresso tecnologico-utilitaristico, ma anche alla comprensione più profonda della natura della conoscenza e del pensiero umano.

6) Bibliografia

BENVENUTO, S., *Il setaccio di Wittgenstein. Wittgenstein e l'etica*, in “aut aut”, 394 (2023).

BRIGAGLIA, M., *L'IA e la rappresentazione di noi stessi, come tristi macchine allo specchio*, in “Agenda Digitale”, disponibile al sito: <https://www.agendadigitale.eu/cultura-digitale/lia-e-la-rappresentazione-di-noi-stessi-come-tristi-macchine-allo-specchio>.

CASEY, G., *Artificial Intelligence and Wittgenstein*, s.l., s.e.

CHUA, J. M., *Formulating Consciousness: A Comparative Analysis of Searle's and Dennett's Theory of Consciousness*, in “Talisik: An Undergraduate Journal of Philosophy”, vol. 4, n. 1, 2017, pp. 43–58. Disponibile su: <https://philarchive.org/archive/CHUFCA>.

DESCARTES, R., *Discorso sul metodo*, in *Opere filosofiche*, a cura di G. Reale, Laterza, Roma-Bari, 2005.

DESCARTES, R., *Meditazioni metafisiche*, in *Opere filosofiche*, a cura di G. Reale, Laterza, Roma-Bari, 2005.

FERRARIS, M., *Documanità. Filosofia del mondo nuovo*, Laterza, Roma-Bari, 2021.

FODOR, J. A., *The Language of Thought*, Harvard University Press, Cambridge (MA), 1975.

GÄRDENFORS, P., *Conceptual Spaces: The Geometry of Thought*, MIT Press, Cambridge (MA), 2000.

GUIDETTI, L. (2024). *Lo spazio logico in Wittgenstein a confronto con l'assiomatica moderna. Algebra ipercubica e fenomenologia del significato*. In V. Costa, S. Galanti Grollo, L. Guidetti, E. Mariani (a cura di), *Fenomenologia e realtà. Studi in onore di Stefano Besoli* (pp. 667–690). Milano-Udine: Mimesis.

GRIGENTI, F., *Intelligenza artificiale. Dalla macchina di Turing al machine learning*, in *Filosofia del digitale*, a cura di A. Taddio e C. Giacomini, Mimesis, Milano, 2021.

HELLIWELL, A. C., ROSSI, A., & BALL, B. (a cura di). (2024). *Wittgenstein and Artificial Intelligence. Volume I: Mind and Language*. Londra: Anthem Press.

HELLIWELL, A. C., ROSSI, A., & BALL, B. (a cura di). (2024). *Wittgenstein and Artificial Intelligence. Volume II: Values and Governance*. Londra: Anthem Press.

HOFSTADTER, D. R., DENNETT, D. C., *The Mind's I: Fantasies and Reflections on Self and Soul*, Basic Books, New York, 1981.

JONES, K. SPÄRCK, *Synonymy and Semantic Classification*, Edinburgh University Press, Edinburgh, 1986.

KARDUM, M. (2023). *Wittgenstein's Tractatus, the Vienna Circle, and Artificial Neural Networks as the First Computational Theory of Mind: Does Development of AI Owe Anything to Wittgenstein?* In I. Skelac & A. Belić (a cura di), *What Can Be Shown Cannot Be Said. Proceedings of the International Ludwig Wittgenstein Symposium, Zagreb, Croatia, 2021* (pp. 121–132). Münster: LIT Verlag.

MASTERMAN, M., *Language, Cohesion and Form: Selected Papers*, Cambridge University Press, Cambridge, 2005.

McCLELLAND, J.L., RUMELHART, D.E., HINTON, G.E., *The Appeal of Parallel Distributed Processing*, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, cap. 1, disponibile al sito: <https://stanford.edu/~jlmcc/papers/PDP/Chapter1.pdf>.

MILLER, T., *Explanation in Artificial Intelligence: Insights from the Social Sciences*, in “Artificial Intelligence”, 267 (2019), 1–38.

OBERMEIER, K. K. (1983). *Wittgenstein on Language and Artificial Intelligence: The Chinese-Room Thought Experiment Revisited*. *Synthese*, 56(3), 339–349.

PENCO, C., *Wittgenstein, Contexts, and Artificial Intelligence: An Engineer Among Philosophers, a Philosopher Among Engineers*, s.l., s.e.

PENCO, C., *Wittgenstein e le macchine pensanti*, in M. Brigaglia (a cura di), *Logica e Intelligenza Artificiale*, Il Mulino, Bologna, 2010.

PENNINGTON, J., SOCHER, R., MANNING, C.D., *GloVe: Global Vectors for Word Representation*, disponibile al sito: <https://aclanthology.org/D14-1162.pdf>.

PENNISI, A., FALZONE, A., *Wittgenstein, Turing e l'IA. Un'intervista a ChatGPT*, Università di Messina, disponibile al sito: [inserire link se disponibile].

PERISSINOTTO, L., *Introduzione a Wittgenstein*, Il Mulino, Bologna, 2018.

PEZZANO, G., *Pensare la realtà nell'era digitale: una prospettiva filosofica*, Carocci, Roma, 2023.

PIANA, G., *Interpretazione del Tractatus di Wittgenstein*, Guerini e Associati, Milano, 1994.

RONG, X., *word2vec Parameter Learning Explained*, disponibile al sito: <https://formacion.actuarios.org/wp-content/uploads/2024/05/1411.2738-word2vec-Parameter-Learning-Explained.pdf>.

SHANKER, S. G. (1998). *Wittgenstein's Remarks on the Foundations of AI*. Londra: Routledge.

TADDIO, L., *Maurice Merleau-Ponty*, Mimesis, Sesto San Giovanni, 2017.

TADDIO, L., GIACOMINI, G. (a cura di), *Filosofia del digitale*, Mimesis, Milano, 2020.

WITTGENSTEIN, L., *Blue and Brown Books*, Basil Blackwell, Oxford, 1958.

WITTGENSTEIN, L., *Osservazioni sulla filosofia della psicologia*, ed. orig. a cura di R. Rhees, trad. it., Einaudi, Torino, 1994.

WITTGENSTEIN, L., *Philosophical Grammar*, ed. by R. Rhees, transl. by A. Kenny, Blackwell Publishers, Oxford, 1974.

WITTGENSTEIN, L., *Tractatus Logico-Philosophicus*, Feltrinelli, Milano, 2022 (Universale Economica I classici).