ΣΥΖΗΤΗΣΙΣ

La trasformazione di un'ipotesi nell'Intelligenza Artificiale

di

Nicole Dalia Cilia

Abstract: The paper aims to discuss the role of hypotheses starting from the treatment that since the 40s has been given of "synthetic method" in which the simulation was seen as a test or as a control of a theory. Secondly, the well-known problem of the under-determination of the models generated by the use of the synthetic method will be highlighted, which cannot be easily solved, or neglected, starting from obtained performances. This question in fact becomes crucial if we consider that, with the current Machine Learning techniques, it is possible to obtain the same result starting from different implementation constructs. Then the problem is building the right hypothesis in favor of the explanation of the phenomenon under consideration. Thirdly, a reconstruction of the various paths that a scientific hypothesis follows from its conception to its validation in the current methods of Artificial Intelligence will be provided. In trying to evaluate the movement of scientific hypotheses from the 50s to the present, through two experimental examples that use Machine Learning techniques, we will show the passage from an explicit use of the hypothesis, which remains subject to validation or denial, to a type of hypothesis that seems emerging only *a posteriori*.

Keywords: Artificial Intelligence, Hypotheses, Machine Learning, Synthetic Method, Undertermination Problem

Abstract: Il presente saggio ha lo scopo di discutere il ruolo delle ipotesi partendo dalla trattazione che già dagli anni '40 è stata fornita del "metodo sintetico" in cui la simulazione era vista come test o come controllo di una teoria. In secondo luogo, verrà messo in evidenza il noto problema della sottodeterminazione dei modelli generato dall'impiego del metodo sintetico, il quale non può essere risolto facilmente, o trascurato, a partire dalla scelta del livello di spiegazione più adeguato quanto a prestazioni ottenute. Tale questione infatti diventa cruciale se si considera che con le tecniche di apprendimento automatico attuali è possibile ricavare lo stesso risultato a partire da costrutti implementativi differenti. E allora il problema diventa costruire la giusta ipotesi a favore della spiegazione del fenomeno preso in

esame. In terzo luogo, verrà fornita una ricostruzione dei vari percorsi che un'ipotesi scientifica segue dalla sua ideazione alla sua validazione nelle attuali metodologie dell'Intelligenza Artificiale. Nel cercare di valutare il movimento delle ipotesi scientifiche dagli anni '50 ad oggi, attraverso due esempi sperimentali che vedono l'impiego di tecniche di *Machine Learning*, verrà evidenziato il passaggio da un uso esplicito dell'ipotesi, la quale rimane comunque soggetta a convalida o smentita, ad un tipo di ipotesi che sembra, più che altro, emergente solo a posteriori.

Keywords: apprendimento automatico, intelligenza artificiale, ipotesi, metodo sintetico, sottodeterminazione dei modelli

1. *Introduzione*

All'interno del panorama attuale dell'Intelligenza Artificiale sembrano convivere due linee di ricerca differenti, caratterizzate da differenti obiettivi. Un primo obiettivo è quello di costruire manufatti tecnologicamente avanzati (ad esempio, robot, veicoli senza pilota) e macchine intelligenti, come quelli che sfidano gli esseri umani nei giochi o quelli chiamati a svolgere il ruolo di assistenti virtuali, per far progredire la nostra tecnologia e la nostra economia. La ricerca nelle scienze comportamentali e nelle scienze cognitive ha contribuito a questo obiettivo tecnologico, e gli organismi biologici hanno spesso ispirato la costruzione di robot e altri manufatti, dando luogo a diversi programmi di ricerca che sono chiaramente bioispirati o biomimetici[1]. Un secondo obiettivo è quello di costruire (o simulare) macchine intelligenti in modo da riprodurre e infine comprendere l'intelligenza biologica.

In questa prospettiva, le simulazioni al computer e soprattutto i robot possono influenzare profondamente il nostro modo di concettualizzare il comportamento e la cognizione. Il dibattito, ancora acceso, ha contribuito a dichiarare una sorta di "dualismo" negli obiettivi dei ricercatori che operano in questi campi (e una separazione parziale dei programmi di ricerca, dei convegni e delle comunità). Riconoscere questa distinzione concettuale e questa sorta di "dualismo" è utile per evitare malintesi e false dichiarazioni dei progressi nei vari campi di ricerca. Allo stesso tempo, il dualismo non è necessariamente rigido, poiché concettualmente questi due estremi hanno due possibili direzioni di influenza: dalle

[1] V. M. A. Arbib, *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge 2003.

scienze naturali alle scienze dell'artificiale, e dalle scienze dell'artificiale alle scienze naturali. È infatti a questo proposito che Webb afferma: «Come dovrebbe essere modellato il comportamento biologico? Un approccio relativamente nuovo è quello di indagare i problemi in neuroetologia attraverso la costruzione di modelli di robot fisici dei sistemi sensorimotori biologici»[2]. In modo analogo Pfeifer e Bongard sostengono che gli scienziati cognitivi e i neuroscienziati hanno molto da imparare dalla robotica[3]. Un argomento che è stato più volte sottolineato è che l'approccio simulativo e la robotica chiedono ai ricercatori di formulare le loro richieste e le loro ipotesi in maniera più precisa. Inoltre, i robot possono essere utilizzati in vari modi, ad esempio,

> come modelli operativi di confronto tra teorie specifiche, come prova di concetti, come strumenti di esplorazione concettuale per generare nuove ipotesi, o possono essere utilizzati come prove sperimentali per scoprire particolari proprietà comportamentali negli animali o negli esseri umani. [...] Addirittura possono essere utilizzati come strumenti terapeutici[4].

Le influenze tra le scienze naturali e le scienze artificiali e tecnologiche possono essere, quindi, bidirezionali, e ci sono ricercatori che non si posizionano in uno dei due ambiti, ma in mezzo o in entrambi. Ne deriva che alcuni dei metodi che sono al giorno d'oggi influenti nelle neuroscienze (ad esempio, gli approcci alla funzione della dopamina che derivano dalla ricerca nell'apprendimento per rinforzo, gli approcci statistici per i processi decisionali e l'integrazione sensomotoria) non sono stati inizialmente sviluppati come modelli di intelligenza biologica e corroborano ulteriormente l'idea che il dualismo, di cui sopra, non è rigido[5]. I robot che sono stati inizialmente costruiti per convalidare le ipotesi scientifiche hanno successivamente proposto ai ricercatori nuovi modi di concettualizzare il problema in questione.

[2] B. Webb, *Using Robots to Understand Animal Behaviour*, «Advances in the Study of Behavior» 38 (2008), pp. 1-58.
[3] R. E. Pfeifer-J. C. Bongard, *How the Body Shapes the Way We Think*, MIT Press, Cambridge 2006.
[4] P. Y. Oudeyer, *On the Impact of Robotics in Behavioral and Cognitive Sciences: From Insect Navigation to Human Cognitive Development*, «IEEE Transactions on Autonomous Mental Development» 2 (2010), pp. 2-16, p. 1.
[5] V. G. Santucci-N. D. Cilia-G. Pezzulo, *The Status of the Simulative Method in Cognitive Science: Current Debates and Future Prospects*, «Paradigmi. Rivista di Critica Filosofica» 3 (2016), pp. 51-74.

A partire da queste assunzioni, il presente saggio ha lo scopo di discutere il ruolo delle ipotesi partendo dalla trattazione che già dagli anni '40 è stata fornita del "metodo sintetico" in cui la simulazione era vista come test o come controllo di una teoria. In secondo luogo, verrà messo in evidenza il noto problema della sottodeterminazione dei modelli generato dall'impiego del metodo sintetico, il quale non può essere risolto facilmente, o trascurato, a partire dalla scelta del livello di spiegazione più adeguato quanto a prestazioni ottenute. Tale questione infatti diventa cruciale se si considera che con le tecniche di apprendimento automatico attuali è possibile ricavare lo stesso risultato a partire da costrutti implementativi differenti. E allora il problema diventa costruire la giusta ipotesi a favore della spiegazione del fenomeno preso in esame. In terzo luogo, verrà fornita una ricostruzione dei vari percorsi che un'ipotesi scientifica segue dalla sua ideazione alla sua validazione nelle attuali metodologie dell'Intelligenza Artificiale. Nel cercare di valutare il movimento delle ipotesi scientifiche dagli anni '50 ad oggi, attraverso due esempi sperimentali che vedono l'impiego di tecniche di *Machine Learning*, verrà evidenziato il passaggio da un uso esplicito dell'ipotesi, la quale rimane comunque soggetta a convalida o smentita, ad un tipo di ipotesi che sembra, più che altro, emergente solo a posteriori.

2. *Metodo sintetico*

Dal 1940, al fine di riprodurre e studiare i meccanismi di alcune funzioni cognitive, gli scienziati hanno seguito una metodologia conosciuta come *synthetic method*[6]. L'obiettivo del metodo sintetico è quello di testare il "meccanismo" sottostante la costruzione della macchina, non quello di riprodurre un meccanismo cognitivo. Ciò è possibile comparando il comportamento della macchina con quello dell'organismo.

Il primo tentativo esplicito di applicare il metodo sintetico fu la macchina descritta da S. Bent Russell nel 1913[7], trent'anni prima dalla pubblicazione dell'articolo di Rosenblueth, Wiener e Bigelow[8], solitamente

[6] R. Cordeschi, *The Discovery of the Artificial: Behaviour, Mind and Machines Before and Beyond Cybernetics*, Kluwer, Dordrecht 2002.
[7] S. Bent Russell, *A Practical Device to Simulate the Working of Nervous Discharges*, «Journal of Animal Behaviour» 3 (1913), pp. 1535.
[8] A. Rosenblueth-N. Wiener-J. Bigelow, *Behaviour, Purpose and Teleology*, «Philosophy

considerato il manifesto della nascente cibernetica. Tale macchina era un dispositivo idraulico, che simulava alcune semplici forme di apprendimento associativo. La metodologia modellistica impiegata prevedeva i due passi che caratterizzano tutt'oggi il metodo sintetico:

i. Esposizione delle ipotesi. Nel caso della macchina idraulica le ipotesi sono due: 1) la stimolazione ripetuta e ravvicinata nel tempo di neuroni dà luogo al rafforzamento delle reciproche connessioni e a un aumento della conduzione; 2) la stimolazione non ripetuta e distanziata nel tempo di neuroni dà luogo all'indebolimento delle reciproche connessioni e a una diminuzione della conduzione nervosa.
ii. Descrizione del progetto di una macchina idraulica funzionante che "incorpora" le ipotesi e il successivo confronto dei risultati ottenuti dalla macchina con quelli delle connessioni nervose organiche per verificare se la macchina simula effettivamente le caratteristiche essenziali delle connessioni nervose.

Questa macchina rappresenta una svolta sorprendente perché per l'epoca l'idea di un dispositivo in grado di modificare il proprio comportamento in relazione all'ambiente, cioè in grado di apprendere, richiedeva un ampliamento del concetto stesso di macchina, considerata invece un mero automatismo. È proprio in questa macchina che, secondo Cordeschi, si ritrovano gli ingredienti fondamentali del metodo sintetico:

> [la macchina] si comportava come previsto dalla teoria che essa incorpora, ed era un dispositivo (idraulico) funzionante, e non una delle tante generiche analogie (idrauliche) con il sistema nervoso. In questo senso, essa costituiva un test o un controllo di quella teoria, giacché «organismo meccanico (o macchina) e «organismo biologico» (o organismo propriamente detto) condividevano alcune «caratteristiche essenziali» del fenomeno indagato (l'apprendimento), rivelando una comune organizzazione funzionale al di là delle differenti strutture fisiche[9].

La speranza era di riuscire ad ottenere su questa stessa base un test per i

of Science» 10 (1943), pp. 18-24.
[9] R. Cordeschi, *Il Metodo Sintetico: Problemi Epistemologici nella Scienza Cognitiva*, «Sistemi intelligenti» 20/2 (2008), pp. 167-191, p. 170.

tipi di apprendimento più complessi, che al momento la macchina non riusciva a manifestare. In ogni caso, la stessa esistenza della macchina «era una prova a sostegno della sufficienza delle ipotesi neurologiche invocate nella spiegazione del fenomeno indagato»[10]. Già nel 1935, infatti, Thomas Ross scriveva: «La speranza è che diventi possibile controllare le diverse ipotesi psicologiche sulla natura del pensiero costruendo macchine ispirate ai principi che implicano tali ipotesi e confrontando il comportamento delle macchine con quello delle creature intelligenti»[11]. È chiaro che questo metodo non si propone di dare alcuna indicazione sulla natura delle strutture meccaniche o le funzioni fisiche del cervello stesso, ma solo di determinare, nel modo più adeguato possibile, i tipi di funzione che possono aver luogo tra «stimolo» e «risposta». Infatti, un atteggiamento erroneo è l'idea che il costruttore di una macchina che apprende debba pensare di costruire un meccanismo fisicamente simile a quello che sta alla base dell'apprendimento umano o animale. In questa occasione Ross sembrava già essere consapevole delle caratteristiche più controverse del metodo sintetico: parlando di artefatti fisicamente diversi ma con uguali funzioni, egli formula infatti sia i problemi relativi al test di sufficienza, sia quelli relativi alla realizzabilità multipla e a quelli del funzionalismo.

In quegli stessi anni anche Kenneth Craik aveva già chiara l'importanza di tale metodo. Egli infatti scriveva che vi è differenza tra un metodo "analitico", interessato alla struttura anatomica e neurofisiologica degli organismi, e un metodo "sintetico"[12]. Quest'ultimo ingloba «i principi generali» che valgono sia per gli organismi viventi sia per le macchine, considerando entrambi come complessi sistemi adattativi. Craik ne indicava tuttavia anche i rischi, sottolineando come fosse possibile che

> i modelli finissero per ridursi a pure imitazioni del fenomeno studiato, dunque a esperimenti privi di ogni interesse scientifico per quanto riguarda la spiegazione del comportamento degli organismi, [...] proprio perché non condividevano con gli organismi, che si limitavano ad imitare, nessun principio funzionale comune[13].

[10] *Ibidem*.
[11] *Ibidem*.
[12] K. J. W. Craik, *The Nature of Explanation*, Cambridge University Press, Cambridge 1943.
[13] R. Cordeschi, *Il Metodo Sintetico*, cit., p. 171.

In effetti già Rosemblueth *et al.* (1943) avevano sottolineato come gli artefatti potevano essere di grande interesse per la spiegazione all'interno della scienza cognitiva, solo se «erano instanziazioni di un "modello teorico", il quale garantiva la base per il confronto tra il sistema naturale e il sistema artificiale»[14]. Di nuovo, dunque, compare l'idea del modello come test di una teoria, ma soprattutto viene formulato esplicitamente quel ciclo metodologico teoria-modello che sarebbe diventato pervasivo nella successiva evoluzione del metodo sintetico fino ai nostri giorni. Si tratta di elementi del metodo sintetico che hanno caratterizzato l'approccio simulativo della *Information Processing Psychology* di Newell e Simon[15], poi confluita nella scienza cognitiva. Per fare un esempio ormai storico, la messa a punto del *Logic Theorist* aveva mostrato la necessità che ne venisse elaborata una "versione modificata", come si esprimevano Newell e Simon, che prenderà la forma del *General Problem Solver* (GPS)[16]. Si tratta del processo "elicoidale" teoria-modello, verso modelli sempre più realistici del fenomeno indagato, i quali includessero dunque restrizioni sempre più esigenti.

2.1. *Il ciclo metodologico*

Per comprendere meglio il ciclo metodologico di cui si è parlato si consideri la Fig. 1, in cui sono riportatati i principali agenti presenti in uno studio cognitivo.

[14] A. Rosemblueth *et al.*, *Behaviour*, cit. Si veda anche G. Tamburrini-E. Datteri, *Machine Experiments and Theoretical Modelling: From Cybernetic Methodology to NeuroRobotics*, «Minds and Machines» 15/3 (2005), pp. 335-358.
[15] A Newell-H. Simon, *Human Problem Solving*, Prentice-Hall, Englewood Cliffs (NJ) 1972. Oppure A. Newell-H. Simon, *Computer Simulation of Human Thinking*, «Science» 134 (1961), pp. 2011-2017.
[16] Il GPS, programma realmente implementato, rappresentava una teoria del *problem solving* umano, poiché tentava di spiegare tutto il comportamento in funzione delle operazioni di memoria, dei processi di controllo e delle regole. Il GPS aveva lo scopo di fornire un insieme di processi utilizzati per risolvere una varietà di diversi tipi di problemi, attraverso la definizione dello spazio del problema in termini di obiettivi da raggiungere e regole di trasformazione impiegabili per passare da uno stato all'altro all'interno del problema.

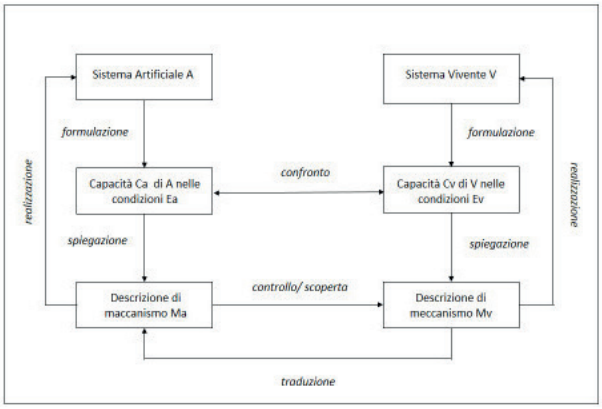

Fig. 1 – Schema metodologico per l'analisi degli studi simulativi. Tratto da E. Datteri, *Filosofia delle scienze cognitive: spiegazione, previsione, simulazione*, Carocci, Roma 2012.

Ci serviremo di un esempio per indagare come questo ciclo sia effettivamente impiegato nel processo della modellizzazione. Alcuni ricercatori statunitensi si sono proposti di scoprire il meccanismo che guida i movimenti degli astici verso le fonti di cibo attraverso scie chimiche dissolte nell'acqua, probabili indizi di fonti di nutrimento[17]. Per far ciò hanno costruito un piccolo robot chiamato RoboLobster, capace di muoversi sott'acqua. L'idea di fondo è che gli astici riescano a raggiungere le fonti di cibo perché sono in grado di seguire le scie chimiche emesse dal cibo e disperse dalle turbolenze marine. L'ipotesi esplicativa soggiacente dei costruttori della macchina era che l'intensità dello stimolo percepito da ogni chemiorecettore stimolasse, in proporzionalità diretta, la velocità degli organi motori del lato opposto, portando dunque l'animale a sterzare nella direzione corrispondente allo stesso lato del sensore. Attraverso il comportamento manifesto del robot – della forma per nulla somigliante a quella di un astice – hanno poi tratto la conclusione che l'ipotesi da loro formulata per spiegare il comportamento dell'animale non fosse adeguata.

Da questo esempio, come da altri, è possibile trarre uno schema metodologico per l'analisi dei modelli simulativi (si veda la Fig. 1): si

[17] F. Grasso-T. Consi-D. Mountain-J. Atema, *Biomimetic Robot Lobster Performs Chemo-Orientation in Turbulence Using a Pair of Spatially Separated Sensors: Progress and Challenges*, «Robotics and Autonomous Systems» 30 (2000), pp. 115-131.

osserva un comportamento biologico e si individua un particolare meccanismo cognitivo che possa generare qual comportamento (parte destra della figura). Allora si costruisce una simulazione informatica o, come in questo caso, robotica dell'ipotesi di meccanismo da valutare; si confronta il comportamento della simulazione con quello che costituisce l'oggetto della spiegazione; concordanze e discrepanze comportamentali vengono considerate basi empiriche per rafforzare o indebolire la valutazione della plausibilità dell'ipotesi, sotto l'assunzione che il sistema abbia simulato accuratamente tale ipotesi[18]. In alcuni casi questa rappresentazione della teoria ha avuto una forma "virtuale", in alcuni casi è divenuta una rete neurale o un ambiente simulato su calcolatore, in altri casi la rappresentazione era data da un artefatto *embodied*, di norma un robot mobile. In ogni caso, ogni artefatto, virtuale o *embodied* che sia, in questo contesto, può essere un esempio del metodo sintetico, ovvero può risultare importante per la spiegazione del fenomeno studiato.

3. *Machine Learning*

Al discorso più generico riguardo le metodologie impiegate in intelligenza artificiale si affiancano oggi le implementazioni utilizzate. Partendo dal problema teorico della sottodeterminazione dei modelli – come abbiamo visto, messo in evidenza già da Ross (1935) – per cui uno stesso risultato sperimentale è ottenibile utilizzando costrutti implementativi differenti, cercheremo allora di indagare le recenti tecniche impiegate nell'intelligenza artificiale per far luce sul ruolo che le ipotesi oggi rivestono. Alla nascita delle teorizzazioni sul metodo sintetico, come detto, la speranza era che fosse possibile ottenere su questa stessa base un test per i tipi di apprendimento più complessi, che al momento la macchina non riusciva a manifestare. Focalizzeremo pertanto la nostra attenzione sulle maggiori tecniche di *machine learning* o, in italiano, "apprendimento automatico", utilizzate oggi negli studi di Intelligenza Artificiale.

Con *apprendimento automatico* si intende la tecnica che fornisce ai computer l'abilità di apprendere senza che questi ultimi siano stati esplicitamente programmati per farlo. Il *machine learning* tuttavia non è una tecnica ben definita e univoca ma rappresenta piuttosto

[18] Per maggiori dettagli si consulti E. Datteri, *Filosofia delle scienze cognitive: spiegazione, previsione, simulazione*, Carocci, Roma 2012.

un insieme di metodi di elaborazione automatica dell'informazione sviluppati a partire dagli ultimi decenni del '900 in varie comunità scientifiche[19]. All'interno del *machine learning*, lo stesso Arthur Samuel[20], che ne coniò il termine nel 1959, identificò due approcci distinti. Il primo, indicato come rete neurale, ha lo scopo di sviluppare macchine ad apprendimento automatico (cioè in grado di imparare) in cui, grazie ad una routine di apprendimento che propaga il segnale avanti e indietro per mezzo di meccanismi di rinforzo o inibizione, la rete apprende un comportamento. Il secondo metodo, più specifico, utilizza invece reti altamente organizzate, progettate per imparare solo alcune attività specifiche. Questo approccio necessita di supervisione e richiede la riprogrammazione per ogni nuova applicazione. Tuttavia, risulta essere molto più efficiente dal punto di vista computazionale poiché è problema-specifico. Le varie direzioni di ricerca perseguite all'interno del *machine learning* sono: la statistica computazionale (*statistical computing*), il riconoscimento di pattern (*pattern recognition*), le reti neurali artificiali (*artificial neural network*), il filtraggio adattivo (*adaptive filtering technique*), la teoria dei sistemi dinamici (*dynamical systems theory*), il *data mining*, gli algoritmi adattivi (*adaptive algorithm*), ecc. L'apprendimento automatico, in particolare, esplora lo studio e la costruzione di algoritmi che possano apprendere strutture da un insieme di dati e impiegare queste strutture per fare predizioni. In altri termini, costruiscono in maniera induttiva un modello basato sui campioni presi in esame e impiegano tali modelli per fare previsioni. È facile notare che l'apprendimento automatico è strettamente legato al *pattern recognition*, cioè all'abilità di riconoscere pattern. Tale attività, evidente e cruciale per la stessa sopravvivenza degli esseri umani, ha portato a sviluppare sistemi neurali e cognitivi altamente sofisticati per progettare e costruire macchine in grado di riconoscere *pattern* forniti dal mondo esterno. Nel risolvere la miriade di problemi necessari per costruire tali sistemi, come abbiamo osservato, otteniamo una comprensione più profonda dei sistemi di riconoscimento nel mondo naturale, in particolare nell'essere umano e negli esseri animali. Per alcune applicazioni, come il riconoscimento vocale e visivo, la nostra progettazione può infatti essere influenzata dalla conoscenza di come

[19] R. O. Duda-P. E. Hart-D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York 2000.
[20] A. L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, «IBM Journal of Research and Development» 3/3 (1959), pp. 210-229.

questi compiti sono risolti in natura, sia per quanto riguarda le strutture hardware che per gli algoritmi impiegati. L'apprendimento automatico viene utilizzato in quei campi dell'informatica nei quali progettare e programmare algoritmi espliciti è impraticabile; tra le possibili applicazioni citiamo il filtraggio delle email per evitare spam, l'individuazione di intrusioni in una rete, il riconoscimento ottico dei caratteri, il riconoscimento vocale e il riconoscimento di impronte digitali, i motori di ricerca, le identificazioni di sequenze nel DNA e molto altro ancora.

3.1. *Tipologie di compiti*

Una rete neurale artificiale (ANN – *Artificial Neural Network* in inglese), normalmente chiamata solo "rete neurale" (NN – *Neural Network* in inglese), è un modello matematico-informatico di calcolo basato sulle reti neurali biologiche. Tale modello è costituito da un gruppo di interconnessioni di informazioni costituite da neuroni artificiali e processi che utilizzano un approccio di connessionismo di calcolo[21].

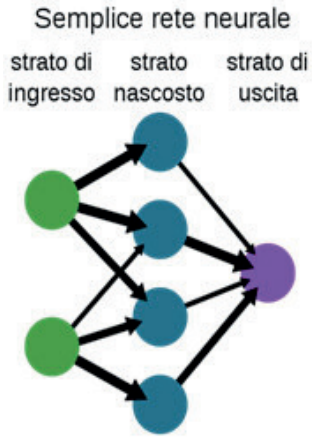

Fig. 2 – Esempio di rete neurale elementare. Tratto da: www.intelligenzaartificiale.it/reti-neurali/ (23.05.2019).

[21] Per una sintetica ricostruzione storica, si consulti N. D. Cilia-L. Tonetti, *Introduction*, in N. D. Cilia-L. Tonetti (eds.), *Wired Bodies. New Perspectives on the Machine-Organism Analogy*, CNR Edizioni, Roma 2017, pp. 13-25.

Nella maggior parte dei casi una rete neurale artificiale è un sistema adattivo che cambia la sua struttura sulla base di informazioni esterne o interne che scorrono attraverso la rete durante la fase di apprendimento. In termini pratici le reti neurali sono strutture non-lineari di dati statistici organizzate come strumenti di modellazione. Una rete neurale artificiale riceve segnali esterni su uno strato di nodi (unità di input) d'ingresso, ciascuno dei quali è collegato con numerosi nodi interni, organizzati in più livelli. Ogni nodo elabora i segnali ricevuti e trasmette il risultato ai nodi successivi.

Tipicamente, vi sono tre grandi tecniche impiegate nel campo dell'apprendimento automatico. L'impiego di queste tecniche varia in relazione alla natura del *segnale*, o dell'input, utilizzato per l'apprendimento e dal *feedback* disponibile al sistema. Queste generiche categorie sono:

i. Nell'apprendimento supervisionato, al programma vengono forniti degli esempi di possibili input e i rispettivi output desiderati. L'obiettivo è quello di estrarre una regola generale che associ l'input all'output corretto.
ii. Nell'apprendimento non supervisionato, non viene fornita invece alcuna descrizione dell'output desiderato. Il programma ha quindi lo scopo di trovare una struttura negli input forniti, senza che questi siano stati etichettati in alcun modo.
iii. Infine, nell'apprendimento per rinforzo il programma interagisce costantemente con un ambiente dinamico, cercando di raggiungere un obiettivo specifico. In questo caso al programma viene fornito solo un suggerimento riguardo il raggiungimento dell'obiettivo desiderato. Un esempio dell'impiego dell'apprendimento per rinforzo è quello di imparare le regole di un qualsiasi gioco, dall'esercizio costante con un avversario.

A metà strada tra l'apprendimento supervisionato e quello non supervisionato si pone infine l'apprendimento semi-supervisionato. Nell'utilizzo di questa tecnica si ha a disposizione un dataset incompleto per la fase di addestramento della rete (il *training*), cioè un insieme di dati, per alcuni tra i quali non è fornito il rispettivo output.

Considerando invece l'output del sistema, si potrebbe avanzare un altro tipo di tripartizione:

i. Nella classificazione, gli input sono divisi in due o più classi e il sistema di apprendimento deve produrre un modello che permetta di assegnare ad un nuovo input una o più classi tra quelle definite. Questo compito viene solitamente affrontato

in maniera supervisionata e i possibili output sono definiti a priori. Un esempio di classificazione è il filtraggio email anti spam: le email, cioè gli input, vengono classificate nelle due cartelle "spam" e "non spam".

ii. Nella regressione, l'output desiderato è, invece, solitamente, un valore continuo. Non ci sono classi entro cui far ricadere il valore di output. Un esempio di regressione è la predizione dell'andamento del valore di un immobile, in un paese, avendo come input i suoi valori nel passato.

iii. Il *clustering*, infine, così come la classificazione, permette di dividere gli input in gruppi, classi o *cluster* appunto. Tuttavia, diversamente da quanto accade per la classificazione, gli output non sono definiti a priori e quindi la rete cerca delle "somiglianze" tra i dati autonomamente per poi spartire gli output in diversi gruppi. Tale problema impiega tipicamente, quindi, un approccio non supervisionato.

3.2. *L'applicazione del machine learning attraverso due studi sperimentali*

Presenteremo adesso due studi sperimentali per analizzare il ruolo che le ipotesi hanno assunto con la nascita del *machine learning*.

Nel primo studio sperimentale che presenteremo, lo scopo è stato quello di riprodurre il riconoscimento di un'analogia percettiva, basata sulla similarità o differenza interna ad ogni coppia di stimolo. In altre parole, il modello costruito doveva riprodurre la capacità umana di riconoscere un'analogia percettiva. Il compito da eseguire era dunque quello di visualizzare due immagini, composte a loro volta da due figure tra loro uguali o differenti; riconoscere la relazione appartenente a queste due figure e infine scegliere l'immagine (tra le altre due presentate) che godesse della stessa relazione rispetto l'immagine target (la prima presentata). Come mostrato in Fig. 3, parte destra, in cui l'immagine target è quella riportata in basso, il compito prevedeva di scegliere l'immagine, tra le due mostrate in alto, che avesse la stessa relazione (di similarità o differenza) dell'immagine target. In questo caso, poichè l'immagine target riporta una relazione di differenza tra le figure componenti, l'immagine corretta da scegliere sarebbe dovuta essere la C. In linea generale, l'ipotesi del modello prevede una porzione semplificata del sistema visivo, in cui le aree LGN[22], V1 e

[22] Il nucleo genicolato laterale (NGL) del talamo è una parte del cervello preposta al

V2[23] vengono campionate due volte in corrispondenza della presentazione di due oggetti, che possono essere diversi o uguali tra loro (Fig. 3, parte destra)[24]. Le due campionature vengono poi valutate in modo topografico da un'area superiore, rappresentata in modo generico come parte della corteccia prefrontale PFC, entro cui viene identificata una distribuzione di neuroni che apprendono la relazione di eguaglianza e diversità (Fig. 3, parte sinistra). Per l'implementazione

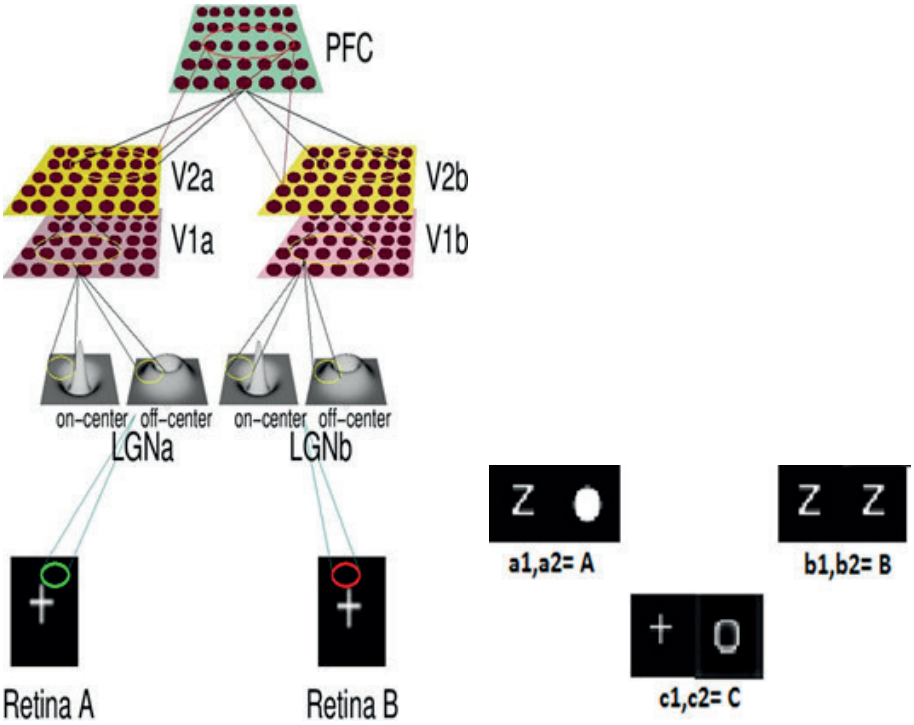

Fig. 3 – A sinistra la struttura del modello, a destra un esempio di stimolo utilizzato.

trattamento dell'informazione visiva proveniente dalla retina.

[23] V1 si riferisce alla corteccia visiva primaria e V2 alle aree aree visive corticali secondarie extra-striate.

[24] A. Plebe-N. D. Cilia, *La difficoltà nel simulare la semplicità*, in G. Airenti-M. Cruciani-M. Tirassa (eds.), *Mind the Gap: Brain, Cognition and Society*, 13th Annual Conference of the Italian Association for Cognitive Science, Università degli Studi di Torino, Torino 2016, pp. 17-24.

è stato utilizzato Topographica[25], il quale integra reti biologicamente realistiche di decine o centinaia di migliaia di neuroni, che formano mappe topografiche che contengono decine o centinaia di milioni di connessioni e consente di simulare qualsiasi regione corticale o sottocorticale bidimensionale, come quella visiva, uditiva, somatosensoriale. Tipicamente, i modelli comprendono più regioni cerebrali, come una parte di un percorso di elaborazione.

Per rendere più pratica la modellizzazione l'unità neurale fondamentale nella simulazione è un foglio bidimensionale di neuroni, piuttosto che un neurone o una parte di un neurone[26]. Concettualmente, un foglio è una porzione continua bidimensionale che è in genere approssimata da una serie finita di singoli neuroni. I modelli sono costituiti da un insieme interconnesso di tali fogli, in cui ciascuna regione cerebrale è rappresentata da uno o più fogli.

L'ipotesi di apprendimento si basa sulla LISSOM (*Laterally Interconnected Synergetically Self-Organizing Map*), la quale implementa connessioni laterali modificabili e flessibili di tipo inibitorio ed eccitatorio, connessioni afferenti, il rafforzamento della coefficienza sinaptica per l'apprendimento di Hebb ecc. L'equazione è la seguente:

$$x_i^{(k)} = f\Big(\gamma_A g_A\left(\mathbf{a}_{r_A,i} \cdot \mathbf{v}_{r_A,i}\right) + \gamma_B g_B\left(\mathbf{b}_{r_B,i} \cdot \mathbf{u}_{r_B,i}\right) + \gamma_E \mathbf{e}_{r_E,i} \cdot \mathbf{x}_{r_E,i}^{(k-1)} - \gamma_I \mathbf{i}_{r_I,i} \cdot \mathbf{x}_{r_I,i}^{(k-1)}\Big).$$

L'architettura LISSOM descrive l'attivazione x_ι di ogni neurone ι ad un certo time step k. Possiamo dire che accanto all'ipotesi fondante il modello, che riguarda la scelta dell'implementazine specifica utilizzata (rete neurale), e all'ipotesi riguardante il modello di apprendimento, basato sulla LISSOM, le sotto ipotesi che andranno ad influenzare la bontà del modello (Fig. 5) sono strettamente legate alla modifica dei parametri mostrati in Fig. 4.

[25] A. Plebe, *Neurocomputational Model of Moral Behavior*, «Biological Cybernetics» 109/6 (2015), pp. 685-699.
[26] J. Sirosh-R. Miikkulainen-Y. Choe (eds.), *Lateral Interactions in the Cortex: Structure and Function*, The UTCS Neural Networks Research Group, Austin 1996.

layer	r_A	r_B	r_{BCK}	r_E	r_H	γ_A	γ_B	γ_{BCK}	γ_E	γ_H
LGN	0.2	-	-	-	-	-	-	-	-	-
V1	0.1	-	-	0.5	0.2	2.0	-	-	1.3	−1.4
V2	0.5	-	0.5	0.1	0.9	1.	0.0	-	1.2	−1.1
PFC	0.6	0.7	-	0.1	0.8	1.2	1.5	-	2.0	−1.9

Fig. 4 – Tabella dei parametri implementati dal modello.

Questi parametri rappresentano varie caratteristiche del sistema visivo, come l'ampiezza del raggio di visione sul campo osservato o l'influenza di un neurone sul suo neurone prossimo. La modifica di tali parametri è manuale e guidata dalla letteratura sul campo. Si potrebbe dunque sostenere che le ipotesi sottostanti l'assunzione del modello vengono riformulate in funzione della prestazione finale del modello stesso.

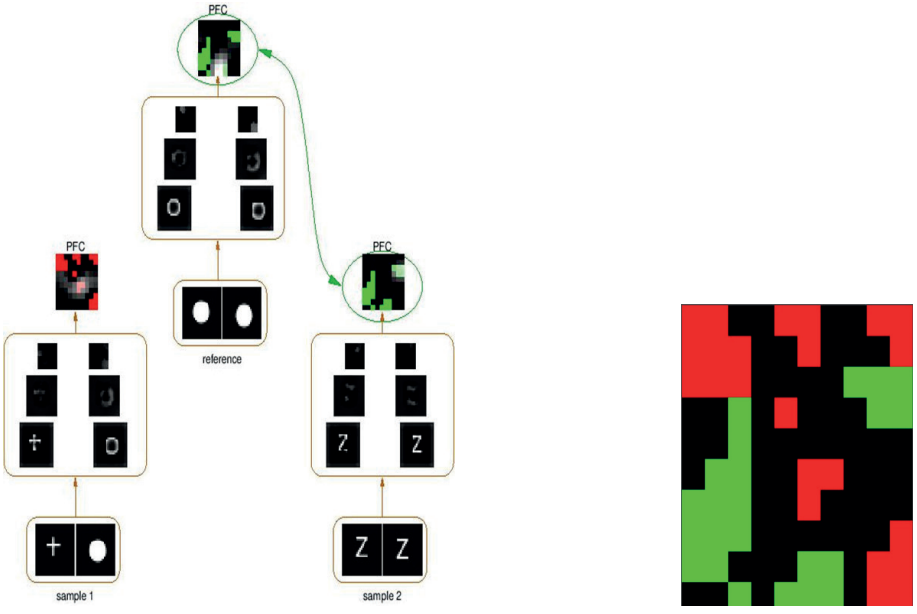

Fig. 5 – Esempio di attivazione della rete. A sinistra l'intero modello, a destra il risultato dell'addestramento.

Il secondo studio che presenteremo è stato scelto come esempio per mostrare il ruolo che le ipotesi rivestono in uno studio che utilizza le tecniche di classificazione sopra presentate. Lo scopo di questa sperimentazione è quello di utilizzare la scrittura come strumento di supporto alla diagnosi precoce di malattie neurodegenerative[27]. La fase sperimentale si è articolata nei seguenti punti:

a. Realizzazione di un protocollo sperimentale. Il protocollo sperimentale è l'insieme dei *task* utilizzabili per testare la compromissione del tratto grafico[28].
b. Acquisizione dei dati per pazienti e gruppo di controllo (coordinate x, y, z con la tavoletta grafica Bamboo Folio della Wacom).
c. Elaborazione dei dati con estrazione di *feature* (cioè le caretteristiche statiche e dinamiche estratte dal tratto). A partire dalle coordinate, si sono calcolate 24 *feature* rappresentanti il tratto grafico di ogni *task* per tutti i soggetti in termini di numero di *stroke*, pressione, accelerazione, velocità, *slant* (inclinazione della penna ripetto il foglio), *jerk* (o tremolio del tratto), dimensione verticale e orizzontale del tratto, durata, ecc.[29].
d. Analisi dei dati con tecniche di *machine learning*: classificazione e regressione. Utilizzando algoritmi di classificazione si è cercato di addestrare il sistema a riconoscere quali *feature* fossero indici prodromici della malattia.

Mantenendo in questa sede un livello di descrizione qualitativa, possiamo dire che l'applicazione del *machine learning*, attraverso algoritmi di classificazione, permette di indentificare la classe di un nuovo obiettivo sulla base di conoscenza estratta da un *training set* (insieme di addestramento). Nel nostro caso specifico, forniti dunque i dati grafici (le coordinate trasformate in *feature*), dopo opportuno

[27] N. D. Cilia-C. De Stefano-F. Fontanella-A. Scotto Di Freca, *Handwriting Analysis to Support Alzheimer Disease Diagnosis: A Preliminary Study*, «Proceedings of CAIP», Salerno (di prossima pubblicazione). Si consulti anche J. Neils Strunjas-K. Groves Wright-P. Maschima-S. Harnish, *Dysgraphia in Alzheimer's Disease: A Review for Clinical and Research Purposes*, «Journal of Speech Language Hearing Research» 49 (2006), pp. 1313-1330.

[28] N. D. Cilia-C. De Stefano-F. Fontanella-A. Scotto Di Freca, *An Experimental Protocol to Support Cognitive Impairment Diagnosis by Using Handwriting Analysis*, «Procedia Computer Science» (2018), pp. 141, 466.

[29] Per ulteriori dettagli, cfr. C. De Stefano-F. Fontanella-D. Impedovo-G. Pirlo-A. Scotto Di Freca, *Handwriting Analysis to Support Neurodegenerative Diseases Diagnosis: A Review*, «Pattern Recognition Letters» 21 (2018), pp. 23-46.

addestramento, il classificatore è in grado di restituire un grado di accuratezza nella classificazione del soggetto in uno dei due gruppi (controllo o malato). Se ci chiediamo qual è stato l'obiettivo della sperimentazione in termini teorici e non puramente implementativi, è possibile rispondere che stiamo simulando un processo di decisione. In particolar modo, si è cercato di simulare il processo decisionale di un medico, il quale, in funzione di alcuni parametri, stabilisce se un paziente è o meno affetto da una patologia neurodegenerativa[30].

La simulazione non è altro che l'attuazione della classificazione biologica o artificiale in relazione a dei parametri (dati) ottenuti da una sperimentazione. Come abbiamo largamente discusso, nel ciclo metodologico precedentemente presentato abbiamo due grandi componenti sperimentali: l'artefatto e l'organismo biologico. Abbiamo anche osservato che ciò che differenzia i due approcci simulativi è in particolar modo la possibilità di confrontare le prestazioni della macchina a quelle dell'organismo, e ciò costituisce proprio il test della teoria che essa incorpora. Nel nostro caso, allora, il problema può essere espresso nella seguente forma: è possibile confrontare la scelta o decisione del medico con quella che opera la macchina attraverso la classificazione? Rispondendo di sì a questa domanda si sostiene che la modellizzazione ricalca il metodo sintetico e che l'ipotesi sottostante è quella generale di riprodurre un meccanismo cognitivo attraverso una simulazione e comparare i risultati ottenuti dal decisore umano e dalla macchina. Tuttavia, intenzione di tale paragrafo era quella di scendere maggiormente nel dettaglio dello strumento implementativo e chiedersi quali sottoipotesi venissero formulate, come fatto per la rete neurale.

In questo caso è in discussione il processo di diagnosi, il quale rimanda alla metodologia di *machine learning* adottata o, con ancora più precisione, allo specifico algoritmo di classificazione adottato. Come abbiamo osservato in precedenza, il sistema di classificazione, come ogni altro sistema di *machine learning* è diviso in due fasi: l'addestramento (*training*), in cui la rete riceve dei campioni e impara ad associarli all'output desiderato, cercando dei criteri per minimizzare l'errore; il *test*, in cui la rete riceve nuovi input, non osservati nella fase

[30] Per maggiori dettagli sugli aspetti metodologici di questo studio, cfr. N. D. Cilia-C. De Stefano-F. Fontanella-A. Scotto Di Freca, *La spiegazione nel Machine Learning: un caso neuroscientifico*, in F. Gagliardi-M. Cruciani (eds.), *Medicina, Filosofia e Cognizione*, Aracne Editrice, Roma 2019.

di addestramento, e li classifica secondo i criteri appresi. Potremmo allora sostenere che le ipotesi si inseriscono anche a questo livello, spingendo la rete a riconoscere il criterio o i criteri migliori per la classificazione dello stimolo. Questi criteri non sono altro che le *feature* immesse dallo sperimentatore in fase di creazione del *dataset*.

5. *Conclusioni*

Come è stato largamente presentato in 2.1., in una simulazione *model oriented* il sistema artificiale deve incorporare l'ipotesi di funzionamento, presupposta essere comune al sistema naturale. È proprio tale presupposto ad essere testato attraverso la comparazione delle prestazioni. Poiché negli studi presentati (3.2) è possibile tale confronto, si potrebbe essere tentati di concludere che il lavoro proposto sia ascrivibile completamente al metodo sintetico. In effetti, è possibile comparare le prestazioni del medico a quelle del sistema artificiale e questo ci permette di capire se l'ipotesi sottostante la categorizzazione, nel primo caso, e la classificazione nel secondo è corretta. Questo processo non è altro che l'abilità, nel nostro caso il modello teorico, comune all'uomo e alla macchina e cioè la facoltà di categorizzare o classificare.

La facoltà indagata è la categorizzazione, corrisponde a compiti di elaborazione del mero dato sensoriale e può avere come risultato finale l'individuazione di un oggetto attraverso la sua astrazione categoriale, cioè la sua inclusione in una determinata classe o categoria[31]. Nelle sperimentazioni presentate le classi sono "Stimoli Analoghi" o "Stimoli non Analoghi" e, nel secondo caso, "Paziente" o "Controllo Sano". Tuttavia, come abbiamo argomentato, le sempre più efficienti tecniche impiegate in Intelligenza Artificiale ci spingono a guardare oltre le ipotesi di costruzione del modello più alte. La forma dell'ipotesi attraverso l'uso di queste tecniche infatti sembra cambiare, prendendo posto all'interno del processo indagato solo a posteriori. Rivolgendoci alle ipotesi più prossime al fenomeno indagato, abbiamo infatti visto che nel primo studio presentato le ipotesi divengono i parametri implementati. Al variare di questi ottengo un risultato, di accuratezza nella categorizzazione dello stimolo percettivo, differente. È l'accuratezza finale che guida la bontà della mia ipotesi, che altro non è che la

[31] F. Gagliardi, *Un'analisi cognitiva delle teorie della diagnosi*, in F. Gagliardi-M. Cruciani (eds.), *Medicina, Filosofia e Cognizione*, Aracne Editrice, Roma 2019.

modifica randomizzata o quasi (parzialmente guidata dalla letteratura o dalla plausibilità biologica a riguardo) del parametro iniziale. Nel secondo studio presentato invece, si potrebbe sostenere che oltre ad ipotizzare le classi che rappresentano l'output e i classificatori dalle migliori prestazioni, le ipotesi sono l'insieme delle *feature* utilizzate per rappresentare il tratto grafico.

Tale saggio non ha pretese di esaustività rispetto alle tecniche di intelligenza artificiale adottate nel panorama attuale e una grande parte di queste è stata, al momento, trascurata. Si pensi ad esempio al *Deep Learning – apprendimento profondo* o *approfondito* – il cui uso diventa sempre più dirompente e pervasivo. Il *deep learning* ha compiuto passi da gigante, ottenendo risultati che, fino a qualche decennio fa, erano pura utopia, grazie alle conquiste soprattutto in campo hardware e alla maggiore disponibilità di dati. Solo una nota conclusiva verrà spesa in tal senso per sottolineare il costante cambiamento, tuttora in corso, che le ipotesi hanno subito dagli anni '40 ad oggi. Il *deep learning* funziona creando modelli di apprendimento su più livelli. A livello teorico, il suo funzionamento è molto semplice e ci porta, ancora una volta, ad uno stringente parallelismo con il funzionamento dell'apprendimento umano e animale. Immaginiamo di elaborare una nozione. La apprendiamo e subito dopo ne elaboriamo un'altra. Il nostro cervello raccoglie gli input della prima e la elabora insieme alla seconda, trasformandola ed astraendola sempre di più. Se iteriamo questo processo, l'apprendimento così realizzato ha la forma di una piramide: i concetti più alti sono appresi a partire dai livelli più bassi. Abbiamo visto come sia importante portare il calcolatore a fare esperienza su un quantitativo sempre maggiore di dati per addestrare la rete, tuttavia nel caso del *deep learning* non vengono fornite esplicite *feature* di riferimento sulle quali basare l'addestramento. Queste sono invece implicite e pervasive nei dati grezzi che osserva la rete. È compito degli stessi algoritmi riconoscerle. Scientificamente, è corretto definire l'azione del *deep learning* come l'apprendimento di *feature* che non sono fornite dall'uomo, ma sono apprese grazie all'utilizzo di algoritmi di calcolo statistico. Ciò significa che le ipotesi, che avevamo assunto essere le *feature* nei compiti di *machine learning*, svaniscono in principio e emergono solo a posteriori dalla rete.

Possiamo concludere sostenendo che il ruolo delle ipotesi all'interno del panorama dell'Intelligenza Artificiale è sempre più articolato e difficilmente indagabile sia a causa della complessità degli studi che nel corso della sua evoluzione l'IA si è trovata ad affrontare sul

piano cognitivo o neuroscientifico, spingendosi a livelli di accuratezza sempre maggiori, sia per quanto riguarda la complessità crescente di tecniche algoritmiche di indagine impiegate. Nel cercare di valutare il movimento delle ipotesi scientifiche dagli anni '50 ad oggi, attraverso i due esempi sperimentali presentati, ciò che risulta evidente è il passaggio da un uso esplicito dell'ipotesi, la quale rimane comunque soggetta a convalida o smentita, ad un tipo di ipotesi che sembra, più che altro, emergente solo a posteriori.

Università degli Studi di Cassino
nicoledalia.cilia@unicas.it