Virtù Intellettuali e Intelligenza Artificiale: Il Machine Learning può fornire una Comprensione Scientifica del mondo?

di Vera Matarese

The pervasive use of Artificial Intelligence (AI) in the contemporary scientific method raises a crucial question: can AI provide a scientific understanding of the world? In this paper, we will answer this question by interpreting the concept of scientific understanding as an intellectual virtue inherent to the person. This interpretation is fruitful, as it allows us to transcend the conceptual horizons of science, offering a fertile ground for discussion for all those areas in which AI is used for epistemic purposes.

Keywords: Machine Learning; Intellectual Virtues; Scientific Understanding; Interpretability; Modal relations

Introduzione

L'intelligenza artificiale (IA) rappresenta oggi una componente essenziale e pervasiva del metodo scientifico, intervenendo in ogni fase del processo di ricerca scientifica[1]. L'IA viene impiegata per raccogliere, generare e interpretare i dati, per formulare ipotesi che possano spiegarli e selezionare quella più probabile. Viene anche utilizzata per progettare gli esperimenti atti a verificare le ipotesi scientifiche indagate e per ottimizzare i processi sperimentali stessi, ad esempio trovando la combinazione di parametri sperimentali che sia ottimale per la riuscita dell'esperimento. L'IA si sta affermando come pilastro imprescindibile anche per il suo ruolo rivoluzionario nel processo di predizione e di scoperta scientifica, come dimostrano casi emblematici quali l'individuazione di nuovi esopianeti[2] o la sintesi di nuovi nanomateriali[3]. Le sue capacità, che sono trasversali a ogni ambito disciplinare scientifico, stanno rivoluzionando il *modus operandi* della scienza. I benefici sono innegabili. Predicendo quali sono le ipotesi più promettenti e affinando in modo autonomo le condizioni sperimentali rendendole ottimali, l'IA consente un'ottimizzazione significativa di tempi e costi, imprimendo un'accelerazione senza precedenti al progresso scientifico. È importante sottolineare che questa rivoluzione non si esaurisce nella dimensione operativa, ma investe profondamente anche la sfera epistemica. Anzitutto, l'IA sta emergendo come un autentico *peer* epistemico, un collaboratore scientifico a tutti gli effetti, al punto che molte scoperte iniziano a essere quasi attribuite a essa. Inoltre, si osserva che l'uso dell'IA consente alla scienza di ottenere progressi epistemici notevoli, sia in campo predittivo sia concernenti la scoperta di nuove realtà, e che però questi, non necessariamente, si accompagnano a un progresso nella comprensione dei sistemi fisici indagati[4]. A proposito di questo, la diffusione capillare dell'IA solleva una questione cruciale: può l'IA fornire una comprensione scientifica dei sistemi fisici? Se la risposta a questa domanda dovesse essere dubbia, o negativa, il rischio è che la diffusione capillare dell'IA potrebbe portare a uno slittamento dell'asse epistemica della scienza verso il suo carattere predittivo piuttosto che quello propriamente conoscitivo che riguarda la comprensione del mondo. L'interrogativo se l'IA possa fornire

1 - Per una discussione sui ruoli dell'IA nella ricerca scientifica, si veda: H. Wang et al., *Scientific discovery in the age of artificial intelligence*, in «Nature», 620-7972 (2023), pp. 47-60.

2 - A. Valizadegan, et al., *ExoMiner: A highly accurate and explainable deep learning classifier that validates 301 new exoplane*ts, in «The Astrophysical Journal», 926-2 (2022), p. 120.

3 - N.J. Szymanski., Rendy, B., Fei, Y. et al., *An autonomous laboratory for the accelerated synthesis of novel materials*, in «Nature» 624 (2023), pp. 86-91 (https://doi.org/10.1038/s41586-023-06734-w).

4 - L. Messeri - M.J. Crockett, *Artificial intelligence and illusions of understanding in scientific research* in «Nature», 627-8002 (2024), pp. 49-58.

comprensione scientifica è reso ancor più pressante dal fatto che le tecniche di Machine Learning (ML), ovvero di apprendimento automatico, spesso operano come 'scatole nere' (black-box), la cui opacità sembra ostacolare il raggiungimento di una comprensione autentica[5].

La comprensione scientifica costituisce il fine ultimo della scienza, ben al di là della mera capacità predittiva. In questo contributo tratteremo la comprensione scientifica come una delle virtù epistemiche o intellettuali più elevate. La filosofia delle virtù intellettuali o epistemiche è un insieme eterogeneo di posizioni in cui la nozione di virtù intellettuale ricopre un ruolo centrale o nel risolvere grandi questioni epistemologiche tradizionali o nel tracciare nuove direzioni nella epistemologia, enfatizzando le abilità intellettuali nel soggetto epistemico[6]. Le virtù epistemiche più comunemente discusse sono l'accuratezza, la precisione, il rigore logico, la responsabilità intellettuale, il pensiero critico, e anche l'empatia verso i propri peer epistemici e il coraggio nel sostenere le proprie idee[7]. Grimm ha avuto il merito di interpretare la comprensione scientifica proprio come una virtù intellettuale[8]. La ragione per la quale in questo contributo si tratterà la comprensione scientifica come virtù intellettuale è che il suo interesse non si limita esclusivamente all'ambito scientifico, ma si estende a una riflessione più ampia sul valore dell'IA per la comprensione del mondo, del quale la comprensione scientifica è sicuramente un aspetto cruciale, e quindi che, se analizzato, può essere informativo sugli altri ambiti in cui la comprensione è cruciale. Adottare una concezione della comprensione così ampia come virtù intellettuale, dunque, ci permette di discutere come l'IA non solo sia strumentale per la ricerca di una verità scientifica o una comprensione scientifica, ma possa anche essere generatrice di virtù che trascendono quelle meramente proprie dell'ambito scientifico.

In particolare, nel presente contributo, si intende affrontare questa problematica in tre passaggi distinti. In primo luogo, verrà inquadrata la concezione di comprensione scientifica attraverso il prisma delle virtù intellettuali, evidenziando come queste possano offrire una chiave di lettura 'pienamente umana' per valutare il ruolo dell'IA nella pratica scientifica. Infine, si esaminerà se e in che misura le

---

5 - E. Sullivan, *Understanding from machine learning models* in «The British Journal for the Philosophy of Science» (2022) pubblicazione online first; F.J. Boge, *Two dimensions of opacity and the deep learning predicament*, in «Minds and Machines», 32 (1/2022), pp. 43-75; C. Beisbart - T. Räz, *Philosophy of science at sea: Clarifying the interpretability of machine learning* in «Philosophy Compass», 17 (2022/6), e12830.

6 - M. Croce, *Epistemologia delle virtù*, in «Aphex» 15 (2017).

7 - J. Baehr, *Educating for intellectual virtues: From theory to practice* in «Education and the growth of knowledge: Perspectives from social and virtue epistemology», (2013), pp. 106-123.

8 - S.R. Grimm, *Understanding as an intellectual virtue* in H. Battaly (ed.), *The Routledge handbook of virtue epistemology*, Routledge, New York 2018, pp. 340-351.

tecniche di Machine Learning possano contribuire alla comprensione scientifica. L'auspicio è che la prospettiva offerta in questo lavoro non si limiti all'ambito scientifico, ma offra spunti di riflessione anche in altri domini, dove l'IA sta assumendo un ruolo sempre crescente, e sempre più meritevole di attenzione filosofica.

Prima di addentrarci nella nostra analisi, è opportuno chiarire alcuni aspetti fondamentali. Innanzitutto, sebbene esistano molteplici forme di intelligenza artificiale, il presente contributo si concentrerà sul Machine Learning, ovvero alle tecniche di apprendimento automatico[9]. A tal proposito, la nostra analisi non discuterà se il machine learning può raggiungere, in quanto *artificial mind*, una reale comprensione scientifica, bensì se gli scienziati, in quanto soggetti epistemici, utilizzando tali strumenti di apprendimento automatico, possano raggiungere comprensione scientifica nonostante la loro opacità. In secondo luogo, è opportuno sottolineare che esistono varie tipologie di opacità che caratterizzano questi sistemi. La cosiddetta black box o opacità del machine learning si riferisce al fatto che, pur conoscendo gli input and osservando gli output, non sia possibile interpretare il processo che ha condotto a tali risultati. Questa opacità però può derivare o dalla difficoltà degli agenti epistemici nel comprendere i meccanismi interni del modello, o all'impossibilità di capire la funzione 'target' che il modello computa[10]. Queste distinzioni non saranno discusse ulteriormente perché, è bene ribadirlo, la questione di questo contributo non è principalmente sulla nostra comprensione dell'IA, bensì sulla possibilità che l'IA possa fornirci una comprensione del mondo, nonostante la sua parziale natura opaca e non interpretabile. Mentre la prima domanda necessiterebbe di un'analisi approfondita dei vari tipi di opacità coinvolti, il nostro quesito, pur intrecciandosi col primo, è parzialmente indipendente, riguardando, appunto, la comprensione del mondo nonostante l'opacità dell'IA. Infine, il nostro quesito può essere formulato in due modi distinti. Nel primo modo, il quesito posto è se l'IA si limiti ad aiutarci nella comprensione del mondo, fungendo da mero strumento facilitatore per il raggiungimento di tale obiettivo. Nella seconda accezione, invece, viene chiesto se l'IA fornisca una comprensione del mondo in senso proprio. In questo contributo terremo in considerazione entrambi i tipi di quesiti.

9 - In particolare, si discuteranno due esempi di Machine Learning, ovvero il GPTune, basato sull'utilizzo di modelli surrogati e ottimizzazione Bayesiana, e l'XGBoost, basato su alberi decisionali potenziati dalla tecnica del *gradient boosting*. Entrambi gli algoritmi, dunque, non si avvalgono di reti neurali.

10 - T. Räz, *ML interpretability: Simple isn't easy*, in «Studies in history and philosophy of science», 103 (2024), pp. 159-167.

La comprensione scientifica come virtù intellettuale

È senza dubbio un'opinione ampiamente condivisa che uno dei fini ultimi della scienza sia quello di offrire una comprensione scientifica del mondo. Indiscutibile è anche il fatto che, fino ad oggi, tale obiettivo sia stato ampiamente raggiunto. Grazie alla fisica, comprendiamo i processi di reazione nucleari; grazie alla genetica possiamo capire perché alcune patologie vengono tramandate da generazione in generazione; grazie alla geologia, invece, conosciamo le ragioni per cui alcune zone del pianeta Terra sono più suscettibili a terremoti o a eruzioni vulcaniche. Non si può dunque negare che la nostra comprensione del mondo sia stata profondamente ampliata dalla scienza.

Ma che cosa intendiamo esattamente con l'espressione 'comprensione scientifica'? Tale quesito ha ricoperto un ruolo centrale nel dibattito filosofico contemporaneo a partire dal celebre contributo di Michael Friedman "Explanation and Scientific Understanding" (1974), e dall'importante saggio di Salmon "What is Scientific Understanding?". Tale rilevanza stride con la scarsissima attenzione che era stata dedicata alla comprensione scientifica nel periodo precedente. Per Hempel (1965), la comprensione scientifica appariva come una nozione dal carattere pragmatico, soggettivo e psicologico, e dunque irrilevante o fuorviante per la scienza. La scienza, invece, si fonda su connessioni oggettive tra evidenza e teorie, e deve dunque trattare nozioni più oggettive come la spiegazione scientifica, basata, per Hempel, su nessi inferenziali tra le leggi di natura e fenomeni particolari[11]. In questa prospettiva, si riteneva che non vi fosse dunque alcuna dimensione della comprensione scientifica meritevole di attenzione, che non fosse già compresa nel concetto di spiegazione scientifica: fornire spiegazioni equivaleva a garantire comprensione scientifica. Tale prospettiva riduzionista ha trovato eco e sostegno anche in autori più recenti come Trout, il quale ha nuovamente sminuito la comprensione scientifica, interpretandola come un mero sentimento soggettivo, un momento di 'eureka', o una sensazione di 'soddisfazione intellettuale' provata dagli scienziati di fronte a una spiegazione[12].

Sebbene il dibattito sulla comprensione scientifica ricopra ora un ruolo prominente, e senz'altro vi sono filosofi che ammettono la natura poliedrica e contestua-

11 - C. Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, Free Press, New York 1965; C. Hempel, *Philosophy of Natural Science*, Englewood Cliffs, Prentice Hall New York 1966.

12 - J.D. Trout, *Scientific Explanation And The Sense Of Understanding*, in «Philosophy of Science», 69 (2/2002), pp. 212-233 (doi:10.1086/341050).

le della comprensione scientifica[13], la nozione di comprensione scientifica che viene discussa è prevalentemente di tipo esplicativo, ovvero fondata sulla capacità di rispondere a domande sul 'perché' di determinati fenomeni. In particolare, secondo la formulazione di comprensione scientifica articolata nel contributo di de Regt, la nozione di comprensione scientifica si articola in due piani diversi ma strettamente legati, ed entrambi di tipo esplicativo[14]. Da una parte, vi è la comprensione esplicativa di un fenomeno, dall'altra parte, vi è la comprensione esplicativa di una teoria. Una comprensione esplicativa di un fenomeno, infatti, richiede la comprensione di una teoria, e affinché una teoria possa essere compresa, essa deve essere intellegibile. Il concetto chiave della comprensione scientifica diventa dunque quello di intelligibilità delle teorie scientifiche. Senza intelligibilità delle teorie scientifiche non vi può essere comprensione scientifica.

Questo modo di approcciare la comprensione scientifica è certamente meritevole perché enfatizza il fatto che si stia parlando di comprensione scientifica, basata, tipicamente, su teorie scientifiche costituite da leggi della natura che rilevano la vera descrizione del mondo fenomenico. Il rischio però è che tale concetto di comprensione scientifica sia riduttivo. Si pensi a campi o discipline scientifiche dove ancora manca un impianto teorico valido e corroborato, come la fisica delle altissime energie, dove abbiamo solo modelli approssimativi, o in campo climatico o economico. In tali ambiti scientifici, sebbene gli scienziati lavorino con modelli che possono svolgere un ruolo teorico, in realtà essi non costituiscono un impianto epistemico solido come le teorie scientifiche vere e proprie. Sono piuttosto modelli ipotetici e provvisori. Inoltre, tale concezione di comprensione scientifica è limitante perché pone l'enfasi sulla teoria stessa, che deve essere intellegibile, piuttosto che sull'agente epistemico stesso. Ovviamente il vantaggio di tale definizione è di fondarsi su un aspetto 'oggettivo' di teorie scientifiche e dunque di evitare elegantemente ogni possibile critica riguardo il carattere soggettivo e psicologico della comprensione scientifica. Ma il rischio viene annullato, a mio avviso, con una manovra infelice, che sopprime la dimensione più virtuosa del soggetto epistemico che comprende. Qual è la condizione epistemica del soggetto che comprende e il valore epistemico della comprensione? Il mio obiettivo, in questo contributo, è di distogliere lo sguardo da una concezione di comprensione scientifica come il frutto dell'applicazione di una teoria intellegibile a un caso particolare, e di invece interpretarla come una manifestazione di una virtù propria del pensatore-scienziato, che indipendentemente da una teoria scientifica, riesce a comprendere il fenomeno indagato.

13 - H.W. de Regt - D. Dieks, *A Contextual Approach to Scientific Understanding*, in «Synthese», 144 (1/2005), pp. 137-170 (doi: 10.1007/s11229-005-5000-4).

14 - H.W. de Regt, *Understanding scientific understanding*, Oxford University Press, Oxford 2017.

Questo 'atto epistemico', tra l'altro, potrebbe essere cruciale e prioritario rispetto alla creazione di una teoria scientifica. In particolare, in questo contributo tratteremo la comprensione scientifica come una virtù intellettuale/epistemica.

Fin dalla pubblicazione di "Virtues of the Mind" di Zagzebski (1996), l'epistemologia delle virtù ha ampliato gli orizzonti dell'epistemologia, spostando l'attenzione dal valore epistemico della conoscenza ai cosiddetti beni epistemici superiori, quali la comprensione e la saggezza[15]. In questo senso, dunque, l'epistemologia abbraccia l'intera gamma delle 'eccellenze intellettuali' che ci permettono di cogliere le strutture della realtà e sviluppare un sapere saggio. Il soggetto che comprende il mondo (anche) a livello scientifico manifesta senza dubbio una eccellenza intellettuale. Se dunque l'IA fornisce o contribuisce a questa comprensione, il suo ruolo può essere visto come un potenziamento di tale virtù epistemica dell'uomo. Questo è un primo punto da indagare: fino a che punto l'IA può rafforzare tale virtù? Come abbiamo visto nella introduzione, una delle preoccupazioni principali è che l'IA eccella nel suo carattere predittivo, ma non necessariamente nel suo ruolo di fornire una comprensione scientifica. In ambiti in cui l'IA non si dimostra neppure particolarmente abile nelle predizioni scientifiche, come per esempio in ambito finanziario, si tende infatti a privilegiare un'analisi 'umana' e manuale dei dati, mentre l'uso dell'IA viene proprio abbandonato. Ciò è segno del fatto che l'IA non abbia potere di comprensione. In secondo luogo, se adottiamo la prospettiva del 'responsabilismo epistemico', la vera virtù intellettuale implica 'un'agenzia' (agency) attiva da parte del soggetto che quindi è responsabile della propria comprensione[16]. L'agente virtuoso non è colui che riceve passivamente la comprensione, ma colui il quale partecipa attivamente alla sua acquisizione. Sorge dunque un altro interrogativo: che ruolo svolge l'IA nella comprensione scientifica, sempre che svolga un ruolo? Rischia di sostituire l'agenzia intellettuale dell'uomo? Questa inquietudine si riflette anche nella vita quotidiana: si pensi, ad esempio, al bambino che svolge i compiti con l'ausilio di un algoritmo di machine learning. Il timore è che, affidandosi a uno strumento esterno, egli non sviluppi appieno la propria virtù intellettuale del comprendere. Infine, trattare la comprensione scientifica come una virtù intellettuale consente di oltrepassare i confini della filosofia della scienza e di esplorare il ruolo dell'IA in altri domini della conoscenza.

A questo punto, è necessario chiarire come possiamo concettualizzare la comprensione scientifica in termini di virtù epistemica.

---

15 - L.T. Zagzebski, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge University Press, New York 1999 (doi:10.1017/CBO9781139174763).

16 - Per una discussione sulla differenza tra responsabilismo e affidabilismo epistemico, si veda L. Code, *Epistemic Responsibility*, University Press of New England, Hanover, NH 1987.

Secondo Grimm, la comprensione è la capacità di cogliere le connessioni tra i fenomeni indagati. Sebbene non esista una definizione di comprensione universalmente accettata in letteratura, essa è tradizionalmente intesa come un atto di cogliere o afferrare le relazioni strutturali, modali e di dipendenza che intercorrono tra gli elementi chiave di un dato dominio[17]. Ad esempio, per comprendere il cambiamento climatico, è necessario afferrare le relazioni tra effetto serra, emissioni di gas serra, aumento delle temperature, eventi meteorologici estremi, acidificazione degli oceani e altri fattori. Analogamente, comprendere un fenomeno come il fuoco significa cogliere la relazione tra combustibile, ossigeno e calore.

Un elemento centrale nella concezione della comprensione di Grimm dunque è il concetto di relazioni di dipendenza, che collegano variabili e sono suscettibili di cambiamento. Poiché l'oggetto della comprensione può essere rappresentato come una mappa causale o modale, una componente chiave del comprendere consiste nel rappresentare accuratamente queste relazioni di dipendenze tra variabili diverse. Queste mappe sono dinamiche piuttosto che statiche: le variabili possono assumere diversi valori, ciascuno dei quali influisce sul sistema in modo differente. Chi possiede una comprensione genuina è in grado di anticipare e prevedere il comportamento del sistema al variare di tali valori. Comprendere significa dunque afferrare o cogliere lo spazio modale del sistema e rispondere a interrogativi controfattuali, come "E se le cosse fossero andate diversamente?"[18]. Ciò implica anche la capacità di discernere quali variabili siano realmente rilevanti e quali invece irrilevanti nel contesto di analisi. Un individuo dotato di comprensione è quindi in grado di prevedere, inferire conseguenze modali e persino esercitare un certo controllo sul sistema, cogliendo, nella sua unitarietà, coerenza strutturale e interconnessione. Un'altra caratteristica della comprensione in quando virtù intellettuale è quella di essere un concetto 'graduale', in quanto ammette di essere 'posseduto' in gradi diversi. Questo ovviamente consente di argomentare che anche i bambini o gli animali possono acquisire gradi di comprensione del mondo, seppure non siano capaci di afferrare tutta la struttura di mappa modale.

La comprensione, dunque, si fonda su due elementi centrali. Da una parte, vi è il suo oggetto, che consiste in una mappa di relazioni modali che strutturano un

---

17 - Per una discussione puntuale della concezione di comprensione come virtù intellettuale si veda: S.R. Grimm, (2017). *Understanding and transparency*, in C. Baumberger - S.R. Grimm - S. Ammon (eds.), *Explaining understanding. New perspectives from epistemology and philosophy of science*, Routledge, London 2017, pp. 212-229. S.R. Grimm, *Understanding as an intellectual virtue*, cit., pp. 340-351; S.R. Grimm, *Varieties of Understanding: New Perspectives from Philosophy, Psychology, and Theology*, Oxford University Press, New York 2019 (doi:10.1093/oso/9780190860974.001.0001).

18 - J. Woodward, *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press, New York 2003. (doi:10.1093/0195155270.001.0001).

sistema coerente e coeso, le cui varie componenti, dunque, sono integrate tramite nessi di dipendenza. Dall'altra parte, vi è l'atto epistemico, e quindi, inevitabilmente cognitivo, mediante il quale il soggetto afferra tale struttura relazionale. La comprensione è, in questo senso, fattiva, poiché presuppone che le relazioni modali che vengono colte rappresentino effettivamente, anche se magari solo approssimativamente, il mondo in modo veritiero. Al contempo, la compresione comporta un aspetto psicologico, dato che richiede che l'agente cognitivo colga i nessi modali che strutturano il fenomeno.

Il Machine Learning e la comprensione scientifica

L'impiego degli algoritmi di apprendimento automatico, detto anche machine learning sta diventando sempre più pervasivo, assumendo funzioni di primaria importanza, tra i quali spiccano i suoi ruoli di ottimizzazione sperimentale, predizione e analisi dati. Piuttosto che soffermarsi su un'analisi tecnica della loro natura e del loro funzionamento, il presente contributo offrirà un esempio concreto della loro applicazione, soffermandosi sul loro ruolo epistemico con una particolare attenzione a ciò che questi algoritmi offrono in termine di comprensione scientifica. Come detto precedentemente, questa scelta è motivata dal fatto che l'obiettivo della nostra indagine è interrogarsi non tanto sulla nostra capacità di comprendere tali algoritmi, bensì sulla misura in cui essi stessi contribuiscano alla comprensione scientifica. In particolare, il presente lavoro esaminerà l'utilizzo da parte dei fisici del Brookhaven National Laboratory (BNL) di due algoritmi specifici di machine learning, da un lato il GPTune per ottimizzare i risultati sperimentali, e dall'altro XGBoost per analizzare e interpretare i dati ottenuti dal processo di ottimizzazione. Questi due algoritmi possono, in qualche modo, fornire o quanto meno facilitare la comprensione scientifica del mondo?

Il caso specifico che discuteremo è l'uso di questi due algoritmi per massimizzare l'intensità del fascio ionico durante le fasi di iniezione e di estrazione dalla sorgente EBIS, una sorgente che produce ioni altamente carichi sfruttando il principio della ionizzazione per impatto elettronico, ovvero attraverso la collisione di elettroni con ioni[19]. Questo processo di produzione di ioni altamente carichi richiede delle condizioni sperimentali molto precise, che coinvolgono numerosi parametri. La sfida degli scienziati, dunque, è quella di trovare la combinazione ottimale di parametri, regolandoli manualmente. Essendo il numero di parametri molto elevato, e dato che i parametri presentano interazioni molto complesse tra di loro, il pro-

19 - G. Zschornacka - M. Schmidt - A. Thorn, *Electron Beam Ion Sources,* 2014, in «*arXiv*» (arXiv:1410.8014).

cesso è estremamente complicato e dispendioso di energie e di tempo. È proprio in questa fase, dunque, che l'apprendimento automatico fornisce un valore inestimabile. In particolare, il GPTune, che è un framework di autotuning basato sull'ottimizzazione di tipo bayesiana, è estremamente efficiente per identificare i valori ottimali dei parametri e ottenere una ottimizzazione della configurazione complessiva di essi[20]. Esso cerca in modo intelligente la combinazione ottimale di parametri esplorando il loro spazio configurazionale, identificando la regione dello spazio configurazionale che presenta maggiori probabilità di contenere la migliore combinazione di parametri per l'ottimizzazione del fascio, e regolando iterativamente i parametri del sistema in modo autonomo. Questo processo si sviluppa e in tre fasi diverse. Nella prima fase, detta di campionamento, una volta impostato l'obiettivo di ottimizzazione del fascio, il GPTune campiona casualmente alcune configurazioni iniziali di parametri e li valuta; questi campioni 'pilota' forniscono una prima conoscenza di come i parametri influenzino l'obiettivo. Nella seconda fase, detta di modellizzazione, il GPTune utilizza la valutazione iniziale della fase di campionamento per addestrare il un modello di probabilità per la cosiddetta funzione obiettivo in modo non supervisionato. In terzo luogo, il GPTune aggiorna le sue probabilità e propone nuove configurazioni che hanno maggiori probabilità di ottimizzare il fascio[21]. Questo processo avviene in tempo reale ed è completamente automatizzato. È un meccanismo di feedback, nel quale l'apprendimento avviene in modo iterativo, funzionando con la raccolta continua di feedback, aggiornandosi, regolando e modificando i parametri per migliorare le prestazioni del sistema. Il GPTune si è rivelato uno strumento di grande successo per la ottimizzazione dell'intensità del fascio sia per la fase di iniezione sia per quella di estrazione del fascio, raggiungendo addirittura un notevole incremento dell'intensità del fascio pari al 70% in un certo punto di misurazione.

Sebbene il GPTune sia uno strumento altamente efficace nell'ottimizzazione in tempo reale della intensità del fascio ionico, esso da solo non è sufficiente a comprendere e interpretare i dati prodotti dal processo di ottimizzazione. GPTune si concentra solo sulla ricerca della configurazione di parametri ottimale, ma non fornisce una analisi delle correlazioni tra le variabili sperimentali né una spiegazione dettagliata dei fenomeni fisici. Per poter capire il meccanismo di ottimizzazione e comprendere il sistema fisico, i fisici del Brookhaven National Laboratory affiancano al GPTune un altro potente algoritmo di apprendimento automatico, che è l'XGBoost. Questo modello, diventato celebre per essere stato il vincitore dell'Higgs Boson Machine Learning Challenge nel 2014, si basa sulla tecnica dell'*extreme*

20 - Cho et al., *GPTune user guide,* 2022, in gptune.lbl.gov/documentation/gptune-user-guide/

21 - *Ibidem.*

gradient boosting e costruisce modelli predittivi basati su alberi di decisione. Esso funge da modello black-box, e svolge il ruolo di analizzare la grande quantità di dati sperimentali raccolta da GPTune, interpretarli individuandone pattern nascosti e relazioni complesse tra i parametri e l'intensità del fascio, e infine predire l'intensità del fascio in base ai parametri sperimentali, permettendo dunque di valutare in anticipo gli effetti di possibili regolazioni dei parametri. Infatti, addestrandosi sui dati ricevuti dal GPTune, l'XGBoost è in grado di predire con una accuratezza del 79% per l'iniezione e dell'80% per l'estrazione, mostrando dunque una piena convergenza tra il modello e il sistema fisico[22]. XGBoost però rimane un modello black-box, non riuscendo a fornire una spiegazione dei processi fisici sottostanti. Sebbene sia utile per individuare correlazioni significative tra i parametri e l'intensità del fascio, nonché pattern dei dati sperimentali, non può sopperire a una mancanza di comprensione scientifica del sistema fisico. I fisici, dunque, non dispongono di una conoscenza esplicita e dettagliata del perché e di come determinate combinazioni di parametri conducano a una ottimizzazione del fascio rispetto ad altre. Rimane loro ignota la relazione specifica tra tali configurazioni e l'intensità del fascio, così come manca una spiegazione trasparente dei processi fisici che si verificano durante le fasi di iniezione e estrazione da EBIS. Ne consegue che non possono neppure sfruttare l'analisi di dati per formulare autonomamente predizioni future, senza l'ausilio di XGBoost. Per ottenere un livello di comprensione maggiore, seppur parziale, è imprescindibile l'utilizzo di strumenti di *explainable-AI*, che sono capaci di caratterizzare e quantificare l'influenza di ciascuno parametro e quella della configurazione totale di parametri sulla intensità del fascio. Dunque, per interpretare le previsioni dell'XGBoost, i fisici lo integrano con strumenti come i grafici di dipendenza parziale, i *partial independence plot* (PDP), e le Shapley Additive exPlanations, (SHAP).

I grafici di dipendenza parziale (PDP) visualizzano gli effetti marginali di uno o due parametri di input sul risultato previsto. Aiutano a illustrare in che modo un singolo parametro sperimentale influisce sull'output stimato mantenendo costanti gli altri parametri. I valori SHAP *quantificano* il contributo marginale di ciascun parametro di input alla previsione del modello in modo più robusto, in quanto prendono in considerazione il fatto che i diversi parametri possono essere correlati. SHAP fornisce informazioni preziose sulle relazioni tra i diversi parametri modellati da XGBoost in quanto identifica l'importanza relativa di varie caratteristiche correlate.

22 - Gu et al., *Enhancing beam intensity in RHIC EBIS beamline via GPTune machine learning-driven optimization,* No. BNL-226095-2024-COPA, Brookhaven National Laboratory (BNL), Upton, NY - United States, 2024.

Conclusioni

Dall'analisi fin qui svolta, emergono considerazioni di notevole rilevanza per l'oggetto della nostra indagine, ovvero il contributo del machine learning per la comprensione scientifica intesa come virtù intellettuale. Il machine learning svolge un ruolo sperimentale straordinario per l'ottimizzazione delle condizioni sperimentali, ma anche per l'analisi dei dati e la formulazione di predizioni. Tuttavia, la capacità di analisi, l'interpretazione di dati, così come quella predittiva rimangono preliminari e insufficienti rispetto al raggiungimento di una comprensione scientifica autentica e piena del sistema scientifico. Se però ammettiamo, come lo abbiamo fatto, che la comprensione scientifica possa manifestarsi per gradi, lo scenario non è completamente negativo.

Innanzitutto, dal caso esaminato, si può evincere che la combinazione GPTune + XGBoost + *explainable AI* contribuisca indubbiamente a una acquisizione abbastanza profonda della struttura modale del fenomeno fisico sotto investigazione, ovvero quello del fascio ionico. In particolare, gli strumenti di *explainable AI* come SHAP o i PDP sono imprescindibili per poter afferrare le relazioni modali tra le configurazioni dei parametri e l'intensità del fascio, così come tra i parametri singoli e l'intensità del fascio. Le informazioni che gli scienziati ricavano sono molto importanti dal punto di vista modale, perché rivelano quali parametri, con quali valori e in quale misura influenzano maggiormente il fascio e contribuiscano alla sua ottimizzazione. Tali strumenti dunque, attenuano l'opacità del black box del XGBoost e rivelano connessioni modali altrimenti inaccessibili. Essendo tali relazioni modali l'oggetto proprio della comprensione scientifica, è dunque possibile concludere che l'uso degli *explainable AI* fornisce un certo grado di comprensione scientifica. Ovviamente è bene precisare che neppure gli strumenti di *explainable AI* forniscono una mappa modale esaustiva, che è ciò che costituirebbe una comprensione scientifica piena. Due sono i motivi principali di questa limitazione. In primo luogo, la selezione dei parametri per il funzionamento di GPTune e l'analisi di XGBoost è limitata, essendo il numero di parametri coinvolti nel processo fisico di ionizzazione estremamente elevato. Secondariamente, strumenti come SHAP o PDP non offrono indicazioni definitive sulla natura modale delle relazioni di dipendenza. Tuttavia, l'*explainable AI* garantisce una comprensione parziale, consentendo la costruzione – seppur incompleta e provvisoria – di relazioni modali.

A questo punto, possiamo porci un ulteriore interrogativo: se il machine learning rimanesse totalmente inintelligibile e non-interpretabile, potremmo raggiungere comunque almeno un livello minimo di comprensione scientifica? Seguendo il modello di de Regt, la comprensione scientifica è possibile solo in presenza di una teoria intellegibile. Se dunque applicassimo tale principio di intelligibilità all'IA, dovremmo concludere che solo un AI trasparente può realmente fornire

comprensione scientifica. Ciò non significherebbe che un sistema di machine learning opaco sia privo di valore euristico, ma comunque non fornirebbe comprensione scientifica. In questa ottica, l'assenza di strumenti esplicativi impedirebbe infatti di chiarire le relazioni dinamiche sottostanti, limitandoci a individuare mere correlazioni empiriche tra input e output, prive di una interpretazione teorica anche solo minima. L'IA in questo caso procura una ottimizzazione dei parametri, fornendo indicazioni su quali combinazioni di parametri risultino più efficaci per raggiungere un'alta intensità del fascio ionico, ma senza fornire una vera e propria intelligibilità del processo fisico. Tuttavia, forse, una tale conclusione è troppo pessimista, e nasconde il fatto che le predizioni stesse condotte dall' XGBoost consentono di delineare delle relazioni (empiriche) modali tra configurazioni di parametri e intensità del fascio. Si potrebbe dunque argomentare che anche l'XGBoost stesso possa fornire una intuizione o conoscenza debole di alcune relazioni modali, e quindi un primo livello di comprensione scientifica. Oppure, forse, più solidamente, si potrebbe argomentare che, seppure non fornendo comprensione scientifica, anche un modello di machine learning black-box possa *aiutare* o *facilitare* il raggiungimento di comprensione scientifica.

Date queste conclusioni ottimiste sul ruolo del machine learning nella scienza, in quanto il machine learning riuscirebbe, se interpretabile, a fornire una buona comprensione scientifica, e, se non interpretabile, comunque a facilitare tale processo di comprensione scientifica, occorre comunque ricordare un importante limite. Le condizioni sperimentali che si ottengono in un laboratorio di EBIS, e in un qualunque laboratorio di fisica delle particelle, sono spesso estremamente 'fragili' e variabili a seconda del contesto specifico nel quale viene effettuato l'esperimento. Ciò implica che le informazioni estratte da un sistema potrebbero non essere replicabili in un sistema analogo, costruito in un momento successivo, o in un luogo diverso. Pertanto, le relazioni modali che si potrebbero ricavare dall'utilizzo di strumenti di IA potrebbero non essere 'robuste', ovvero potrebbero non rivelarsi veritiere in altri contesti. Questo però, è bene enfatizzare, non riguarda un problema dell'IA, quanto piuttosto la non-replicabilità dei sistemi fisici creati in laboratorio.

Infine, si pone la questione della responsabilità epistemica dello scienziato e della scienziata nella costruzione della propria comprensione scientifica. Se l'IA fornisce risultati autonomamente, senza richiedere una interazione attiva da parte degli scienziati, possiamo ancora parlare di un processo di comprensione genuina? Confrontiamo questo scenario con quello che si avrebbe senza un uso dell'IA. In quest'ultimo caso, gli scienziati e le scienziate dovrebbero manualmente cercare di cambiare le configurazioni dei parametri sperimentali fino a trovare quella ottimale. In questo processo, questi si sforzerebbero di capire delle relazioni modali tra i parametri che regolano manualmente e l'intensità del fascio. Sarebbero dunque

pienamente attivi a livello non solo pratico ma anche epistemico. L'IA, invece, offre informazioni senza che il soggetto epistemico intervenga direttamente nella loro elaborazione. Si tratta di un processo largamente, se non completamente, automatizzato di cui il ricercatore-scienziato sembra usufruire a livello passivo. A questo punto, però, la risposta parzialmente pessimista data precedentemente, ovvero che l'IA o ci fornisce solo una comprensione molto parziale, o semplicemente ci faciliti il suo raggiungimento, diventa un aspetto positivo nella questione sulla passività degli scienziati e delle scienziate nell'uso dell'IA. Se davvero aspiriamo a una scienza che fornisca comprensione scientifica piena, l'IA può solo essere concepita come un primo passo per la costruzione di comprensione scientifica, in quanto non può sostituire in alcun modo il processo di comprensione scientifica. L'IA dunque non esime e non solleva lo scienziato e la scienziata dal compito di comprendere attivamente. Se da un lato questa constatazione potrebbe apparire poco rassicurante perché implica il rischio che, qualora ci si accontenti di solo IA, la scienza diventerebbe un mero strumento predittivo, dall'altro è consolante, perché ribadisce che l'IA non annulla né sovverte l'*agency* epistemica umana. Se la scienza vuole rimanere la disciplina che fornisce una comprensione scientifica del mondo, allora preserverà il suo carattere intrinsecamente umano.

Adottare un concetto di comprensione scientifica come virtù intellettuale, intesa come capacità di afferrare relazioni modali in un sistema, si rivela, a mio avviso, fecondo, in quanto permette di applicare le conclusioni di questo articolo in altri ambiti in cui l'IA potrebbe essere utilizzato per contribuire o facilitare il processo di comprensione del mondo. L'auspicio è dunque che le tesi offerte come frutto di questa riflessione aprano la strada a considerazioni sul ruolo epistemico che l'IA svolge nella crescita delle virtù intellettuali dell'agente epistemico, in tutti i campi nei quali queste operano.

VERA MATARESE
Ricercatrice a tempo determinato in Logica e filosofia della scienza presso il Dipartimento di Filosofia, Scienze sociali, Umane e della Formazione dell'Università degli studi di Perugia
vera.matarese@unipg.it