L’INTELLIGENZA ARTIFICIALE COME GRANDE SFIDA. CHE COSA DEL RAGIONAMENTO UMANO NON PUÒ ESSERE IMPLEMENTATO DA UNA AI

Antonino Drago

Università “Federico II” di Naples, via Benvenuti 3, Calci 56011, Italy – drago@unina.it

**Sommario.** Il presente articolo caratterizza in un quadro storico e filosofico la sfida improvvisa rappresentata dall’introduzione dell’intelligenza artificiale nella vita sociale. È drammatico anche perché è molto aggravato dalla povertà culturale di questo quadro. Qui viene sfruttata una novità culturale: un modo alternativo di organizzare una teoria scientifica che già secoli fa venne applicata da molti eminenti scienziati, anche quelli più rivoluzionari nella storia della scienza. In questa organizzazione il modo di argomentare appartiene alla logica intuizionista, che si avvale in particolare delle proposizioni doppiamente negate, degli argomenti ad assurdo e del principio di ragione sufficiente. È dimostrato che l’intelligenza artificiale, gestita secondo la logica classica, non può implementare nessuna di queste tre caratteristiche di un’organizzazione teorica di questo tipo. Questo risultato implica che il test di Turing può essere superato. Si suggeriscono piuttosto una nuova definizione di “intelligenza” e un nuovo test di intelligenza sulla base della logica intuizionista. Come misura politica concreta si suggerisce di abolire le guerre, riducendo gli esseri umani ad agire secondo la logica classica come fanno i robot.

**Parole chiave: Punti di vista della riflessione sull'intelligenza artificiale, Attuale mancanza di coscienza dell'umanità, Organizzazione scientifica alternativa, Test di Turing, Impotenza di un computer nell'implementazione dell'organizzazione alternativa, Nuova definizione di intelligenza, Nuovo test di intelligenza, Niente più umani come robot assassini.**

 D R A F T

**1. L’angosciante situazione della coscienza di sé da parte del genere umano nell’epoca della nascita delle AI.**

Il dibattito sulle intelligenze artificiali è spesso complesso perché spesso il punto di vista degli interlocutori non è sufficientemente manifesto. Potrebbe essere quello di una persona individualista, un consumatore, un venditore, un giurista, un leader politico o spirituale, ecc. In realtà la dimensione sociale dei problemi posti dalle AI è di livello mondiale per le presenti e le future generazioni. Per la prima volta nel XX secolo il progresso sociale dell'umanità ha imposto il punto di vista dell'intera umanità come specie e anche come la specie migliore. Questo punto di vista richiede ad ogni essere umano un’attenzione maggiore verso la la ita dell’umanità e gli innumerevoli problemi legati ad essa; inoltre il processo di crescita della coscienza umana a questo livello è complesso e ad oggi non ancora completo. Non è quindi sorprendente che molti studiosi che dibattono sulla questione delle AI non hanno considerato questo punto di vista. In quanto segue prenderò in considerazione esclusivamente il punto di vista umano in quanto più appropriato per discutere il tema in termini più generali.

Per quanto riguarda i pericoli dell’intelligenza artificiale, a mio avviso la considerazione seguente è quella più importante. La storia ci insegna che la crescita delle conseguenze letali del gesto di una persona è impressionante. Nell'antichità ogni volta che una persona utilizzava armi letali poteva danneggiare un numero limitato di esseri umani; le conseguenze delle sue azioni furono molto limitate nello spazio e nel tempo. Oggi un essere umano può premere un pulsante (comandando il lancio di una bomba nucleare su una città) provocando la morte di un milione e più di persone e infine addirittura l'estinzione della specie umana. Attualmente gli esseri umani hanno la capacità di una distruzione globale.

Tuttavia, nel caso delle armi nucleari, la capacità e la responsabilità ultima di questo atto spetta allo Stato, perché la preparazione di questo tipo di arma è molto complessa e difficile. Invece i pericoli dell’intelligenza artificiale esaltano la responsabilità individuale.

Ad oggi infatti, l’intelligenza artificiale permette a chiunque di creare enormi danni al genere umano. Pertanto, la possibilità di commettere ad es. un genocidio è stata abbassata dagli Stati, che sono grandi e razionali gruppi sociali collettivi, a piccoli gruppi e infine a una sola persona. Ciò è impressionante: non siamo preparati a gestire collettivamente una capacità di distruzione così grande come quella offerta dall’intelligenza artificiale anche a una singola persona.

Un po’ di luce su questa nuova situazione storica viene da un mero parallelismo tra l’attuale processo di interazione degli esseri umani con i computer e il processo storico di addomesticazione degli animali. I progressi dell’intelligenza artificiale obbligano l’umanità a convivere con una moltitudine di computer capaci di acquisire un comportamento autonomo, così come nella storia passata era costretta a convivere con una moltitudine di animali buoni e cattivi. Fu necessario un lungo processo di più di diecimila anni per riuscire ad addomesticare qualcuna tra la moltitudine di specie animali. Risultò però impossibile addomesticare alcune specie, dette perciò “feroci”; gli esseri umani hanno dovuto lottare duramente per espellerli dall'ambiente umano.

Tuttavia, in questo parallelismo esiste anche una differenza nel possibile pericolo di un “processo di addomesticamento” fallito. In passato ogni persona era l'unica responsabile del processo di addomesticamento di un certo animale; il pericolo (anche di morte) conseguente al suo fallimento era individuale. Nel caso dell’intelligenza artificiale, invece, il pericolo di un “processo di addomesticamento” fallito può avere conseguenze imprevedibili anche sull’intera società e sulle prossime generazioni di esseri umani, fino alla soppressione della specie umana. La sfida attuale è quindi molto più ampia di quella animale perché oggi i pericoli anche di un solo robot possono riguardare l’intera umanità.

Nonostante l’illuminante parallelismo precedente, nulla dice dell’aspetto più sconcertante dell’attuale problema dell’IA: il suo verificarsi imprevisto[[1]](#footnote-1). “Ma qualcuno non avrebbe dovuto avvisarci?” è la famosa domanda della regina del Regno Unito quando si verificò la crisi economica dei mutui subprime nel 2008. In altre parole, poiché viviamo in una società moderna, dove la vita può contare su molti grandi vantaggi e privilegi economici, che tutti insieme ci offrono una vita da semidio, come è possibile che una crisi economica così radicale si verifichi all'improvviso?

Lo stesso può essere ripetuto riguardo al verificarsi del problema dell’intelligenza artificiale. Parallelamente alla domanda della precedente regina ci si può chiedere: perché in uno stadio così avanzato di sviluppo della nostra società scientifica nessuno ci ha avvertito che una scienza relativamente giovane, l'informatica, poteva generare una crisi così dolorosa? Perché avviene proprio oggigiorno questo sconvolgimento intellettuale così grande come quello causato dall’intelligenza artificiale? Se la scienza (e ancor più l’informatica) è l’espressione della nostra razionalità, perché noi, esseri razionali, veniamo colti di sorpresa? Perché attualmente non siamo in grado di riconoscere chiaramente il problema centrale di questa crisi?

Siamo padroni quasi assoluti del mondo attraverso la scienza e la tecnologia. Tuttavia, nell’ultimo mezzo secolo l’intelligenza artificiale è diventata sempre più potente rispetto alle facoltà intellettive umane, tanto che attualmente essa appare incontrollabile nella sua capacità di provocare disordini sociali. In particolare, ci siamo improvvisamente confrontati con una domanda drammatica: può il nostro strumento di potere più avanzato, l’intelligenza artificiale, rivoltarsi contro di noi e ridurci in servitù? Possiamo evitare il dominio dell'umanità da parte dei computer?

A mio parere, il fatto che l’intelligenza artificiale sia diventata un problema sociale è stata essenzialmente una sorpresa perché attualmente l’umanità manca di autocoscienza. Tutte le passate esperienze di utile riflessione sull’impresa scientifica sono infatti fallite.

1) La lotta tra la Chiesa e Galileo Galilei staccò la scienza dalla fede e quindi anche la saggezza accumulata nei millenni precedenti. Galilei sosteneva la concezione secondo cui “l'intenzione dello Spirito Santo essere d'insegnarci come si vadia al cielo, e non come vadia il cielo” (l'intenzione dello Spirito Santo è di insegnarci come andare in cielo e non come il cielo va). (Galilei 1978, pp. 128-135) In seguito, i due mondi, quello religioso e quello scientifico, seguirono due percorsi storici distinti, senza reciproca comunicazione. Tuttavia, quest'ultimo accumulò un numero così grande di novità da portare a considerare largamente superato il mondo religioso con tutti i suoi antichi testi sacri. Che quindi la società moderna considera inutili per orientare l’evoluzione dell’umanità.

2) La Scienza è nata nell'antica Grecia in forte connessione con la filosofia. Tuttavia, la filosofia moderna non è riuscita a comprendere la scienza contemporanea (la sua origine storica, le nozioni assolute della meccanica di Newton (spazio, tempo e forza gravitazionale), il suo rapporto con la matematica e i suoi fondamenti). Alla fine Kant affermò di aver offerto una soluzione, che però presto fallì di fronte alla chimica, all’ottica fisica e alla termodinamica; in seguito fu ignorato sia da Einstein che dagli scienziati successivi che costruirono le nuove teorie fisiche. Non è quindi una sorpresa che la maggior parte degli scienziati ha concluso che la filosofia tradizionale della conoscenza rappresenta un grave fallimento della moderna filosofia occidentale; e se ne staccarono.

3) Se la scienza moderna fosse davvero la migliore rappresentazione della razionalità stessa, perché dopo la sua nascita non ha progredito in modo lineare e semplicemente cumulativo? Nei primi anni del Novecento la scienza subì una crisi talmente radicale da far pensare ad un suo completo fallimento. Perché ha subito una crisi? Tuttavia la crisi sopra menzionata è stata risolta; inoltre è stata risolta in modo così brillante che ad es. la meccanica quantistica, nonostante riguardi un nuovo mondo invisibile e microscopico, è una teoria così perfetta da essere stata la prima teoria fisica senza alcuna divergenza tra le sue previsioni e i dati sperimentali. Perché dopo questa crisi la scienza è diventata ancora più potente? Non si conosce alcuna risposta.

4) Alcuni studiosi hanno tentato di comprendere almeno la storia della scienza moderna. Fin dagli anni ’50 Alexander Koyré, poi Thomas Kuhn e Imre Lakatos hanno offerto valutazioni storiche rispettivamente sulla nascita della scienza moderna e sullo sviluppo storico della fisica classica. Ma le successive critiche radicali ai loro risultati portarono a valutarli come tentativi insufficienti di inaugurare una nuova storiografia, senza aver suggerito alcun decisivo miglioramento di questa disciplina. Resta quindi senza risposta anche la seguente domanda: perché lo studio della storia della scienza del passato non ci ha portato a prevedere nulla di simile al problema dell’IA? In particolare, la principale rivoluzione nella storia della fisica teorica fu causata dalla nascita contemporanea della relatività e della teoria quantistica all’inizio del 1900; hanno detronizzato il paradigma newtoniano di lunga data. Ma qual è una caratterizzazione accurata del paradigma newtoniano? In quale ambito la fisica teorica attuale ha cambiato questo paradigma o meglio lo ha ripristinato secondo una nuova versione? Per quanto mi risulta da storico della fisica, non esistono risposte accreditate a queste domande.

5) La ricerca scientifica si sviluppa apparentemente in tutte le direzioni possibili; perché non ci ha avvisato? Resta quindi senza risposta anche la seguente domanda: perché all'improvviso l'assoluta certezza comunicata dalla scienza si è trasformata in una dolorosa incertezza sulla sopravvivenza del genere umano stesso?

6) In particolare, la riflessione filosofica sull’informatica è ancora a un livello così basso che la domanda “L’informatica è una scienza?” (Denning 2005; Eden 2007) non riceve una risposta decisiva. Inoltre, non si sa perché questa teoria si basi invece che su calcoli già eseguiti e ripetibili in un arco di tempo finito, su una parola modale (funzioni computabili) che allude a calcoli considerati semplicemente possibili lungo un arco di tempo indeterminato. Inoltre la sua teoria si basa essenzialmente su una proposizione, la tesi di Turing-Church, che falsamente è una tesi, perché non precede alcun teorema, per la semplice ragione che un teorema che equipara una nozione formale (di computabilità) con la corrispondente nozione intuitiva non può esistere .

7) La tecnologia moderna, essendo informata dalla scienza moderna, è riuscita a trasformare radicalmente l'intero ambiente terrestre e anche il modo di vivere dell'umanità al suo interno, tanto da indurre un cambiamento antropologico. Pertanto il modo in cui l’uomo affronta la realtà è così radicalmente cambiato che l’etica tradizionale viene svalutata o addirittura abbandonata, senza essere sostituita da alcuna etica ben formulata (si pensi ad esempio alla “pillola contraccettiva” e più in generale alla sessualità umana).

In sintesi, la drammatica questione dell’intelligenza artificiale si verifica in un momento caratterizzato sia da un’enorme capacità umana di conoscere i dati scientifici sia da una grande espansione della rappresentazione intellettuale del mondo. Ma questo è anche il tempo della mancanza di consapevolezza da parte dell’umanità della sua attuale situazione storica. Non avrebbe potuto esserci momento peggiore per l’umanità per affrontare una questione così cruciale come l’intelligenza artificiale.

Nel seguito del presente lavoro mostrerò essenzialmente un limite all'attività dell'IA. La sez. 2 presenta i primi due passi logici di un nuovo tipo di organizzazione sistematica di una teoria che mira a scoprire un nuovo metodo scientifico per risolvere un dato problema; il secondo riguarda le occorrenze di proposizioni doppiamente negate (DNP). La loro rilevanza per l’IA è dimostrata dalle celebri Leggi di Asimov, che risultano essere tutte DNP. La sez. 3 presenta una tipologia di questo tipo di proposizioni; ne sottolinea l'infinità di numero e anche la loro dipendenza dal contesto. Nella sez. 4 vengono presentati i restanti tre passaggi logici dell'organizzazione teorica alternativa. La sez. 5 suggerisce di chiarire una nozione preliminare alla presente ricerca: “intelligenza”; qui si parla, piuttosto che di singoli atti, della costruzione logica di teorie sistematiche. Risultano possibili otto gradi di intelligenza logica; il quarto è il più rilevante per la presente indagine: riconoscere le DNP. La sez. 6 dimostrerà che un robot è sostanzialmente incapace di riconoscere una DNP, a causa dell'incommensurabilità, segnata dal fallimento della legge della doppia negazione, tra la sua logica classica e la logica intuizionista. Nella sez. 7, applicando quanto sopra, dimostro che una DNP implica un'incapacità essenziale di un computer di implementare qualsiasi elemento logico del nuovo modo di organizzare una teoria scientifica. Pertanto suggerisco di assumere come principio teorico questa “impotenza” dell'IA. Nella conclusione avanzerò alcune considerazioni sull'etica e sulla politica nell'era dell'IA; la prima misura suggerita è l’abolizione delle guerre, riducendo gli esseri umani a robot nel seguire una logica classica di distruzione.

**2. Una novità culturale: una nuova organizzazione formale di una teoria: i primi due passi logici**

Suggerisco un nuovo modo di approcciare l'IA sfruttando un risultato recente, riguardante la storia e i fondamenti della scienza.

Dall'analisi comparativa di alcune teorie scientifiche che i rispettivi autori hanno presentato in modo non assiomatico (ad esempio la chimica classica, la meccanica di Lazare Carnot, la termodinamica di Sadi Carnot, la geometria non euclidea di Lobachevskij, la prima teoria dei quanti di Einstein, la meccanica quantistica di Dirac, la formalizzazione dei logica intuizionista), Drago (2012) ha estratto il modello di un'organizzazione teorica alternativa all'organizzazione deduttivo-assiomatica; si chiama organizzazione basata sui problemi (PO).

1) Una teoria PO dichiara un problema fondamentale.

2) Fa uso di proposizioni doppiamente negate (DNP) le cui corrispondenti affermative mancano di prove; “Due negazioni non affermano”.

 Per esempio una sentenza della Corte: “L'imputato è assolto per insufficienza delle prove negative”; cioè la Corte non dispone di prove certe.

La rilevanza delle DNP rispetto al tema dell'intelligenza artificiale è dimostrata dal fatto seguente: tutti i principi noti dell'etica suggeriti per i robot fanno uso di questo tipo di proposizione. Prendiamo ad esempio le celebri Leggi di Asimov; sono tutte DNP:

 1. Un robot non può ferire [≠ fare del bene a] un essere umano o, attraverso l'inazione, permettere che un essere umano venga danneggiato.

 2. Un robot deve obbedire a qualsiasi ordine impartitogli dagli esseri umani, tranne nei casi in cui tali ordini siano in conflitto [≠ d'accordo] con la Prima Legge.

 3. Un robot deve proteggere la propria esistenza purché tale protezione non sia in conflitto [≠ d'accordo] con la Prima o la Seconda Legge.

 0. Un robot non può danneggiare [≠ fare del bene per] l’umanità, né può permettere che l’umanità venga danneggiata [≠ il benessere dell’umanità] dalla sua inazione.

**3. Una tipologia di DNP**

Consideriamo il seguente problema: *decidere se l'autore di un testo letterario argomenta secondo una logica non classica oppure no.* Se si prende come prova l’uso da parte di un autore della legge del terzo escluso, questa decisione è possibile solo quando l’autore avverte esplicitamente il lettore del suo uso – o del suo rifiuto – di un’alternativa esclusiva; sicuramente questo caso è raro in letteratura. Studi recenti invece (Prawitz e Melmnaas 1968); Grigio 1970; Prawitz 1976); Dummett 1977, pag. 24) suggeriscono che il modo migliore per discriminare tra i due tipi di logica è fare riferimento al fallimento della legge della doppia negazione; questo fallimento introduce ad argomentare entro una logica non classica; quella più rilevante è la logica intuizionista. Inoltre, la differenza tra i due tipi di logica non è superabile né approssimabile (Gödel 1933).

Pertanto, siamo portati a prendere come prova l’uso da parte di un autore della legge della doppia negazione. Di fatto il riconoscimento di questa figura logica all'interno di un testo non è un compito difficile; bisogna riconoscere quattro aspetti del testo:

*i*) le parole negative, che spesso sono evidenti;

*ii*) le proposizioni comprendenti due parole negative;

*iii*) le proposizioni del genere che non sono equivalenti alle corrispondenti proposizioni affermative perché prive di prove a supporto (DNP);

*iv*) il ruolo essenziale svolto all'interno del testo da ciascuna delle proposizioni risultanti.

Il passo cruciale di questo metodo è quindi decidere se una proposizione doppiamente negata è davvero una DNP. Questo compito richiede un esame accurato perché esiste una varietà di forme linguistiche di proposizioni doppiamente negate.

Come risultato della mia lunga esperienza nell'investigazione delle occorrenze di DNP all'interno di diversi tipi di testi, suggerisco un metodo accurato e generale per ispezionare le DNP all'interno di un testo. Innanzitutto bisogna riconoscere tre casi di proposizioni doppiamente negate che non sono DNP e quindi devono essere scartate:

*a*) Una proposizione doppiamente negata, che è meramente retorica, perché la verifica della sua corrispondenza con la realtà rende evidente che si tratta di un fatto oggettivo; in tal caso le due negazioni si affermano, secondo la legge classica della doppia negazione; quindi la proposizione appartiene alla logica classica; per esempio. “Non ho altri cinque euro”; “Questa mossa non ti porta fuori dalla stanza”.

 *b*) Una proposizione doppiamente negata che equivale ad una proposizione meramente negata perché una negazione è rinforzata psicologicamente dall'altra; per esempio. “Non posso andare oltre”; "Non riesco a ottenere alcuna soddisfazione".

 *c*) Una proposizione doppiamente negata in cui una negazione spiega l'altra, ad es. “Non ho risposta da quest’uomo sordo” (in realtà questa proposizione è una versione abbreviata di due proposizioni: “Non ho risposta da questo uomo perché è sordo”). Quindi questa proposizione comporta una sola negazione, la prima.

Successivamente, suggerisco la seguente tipologia di tutte le forme di DNP; essa ha considerato anche le proposizioni doppiamente negate, formulate in modo oscuro o ambiguo; le traduco in precise DNP. Poiché il lavoro di decisione sulla natura negativa di una parola è il primo passo di questa ispezione di un testo, la tipologia seguente è ordinata in base al crescente grado di attenzione richiesto per definire una parola negativa.

 Una proposizione doppiamente negata è una vera DNP quando include:

*i*) Due negazioni indipendenti; per esempio. “Non è vero che non è…”, oppure due parole negative indipendenti, ad esempio “Le proposizioni irreali vengono respinte”; “senza contraddizione”.

*ii*) Una singola parola composta da due parole negative; per esempio. "im-mobile" (≠ fisso) e "in-variante" (≠ costante).

*iii*) Una parola che sintetizza una DNP (sottolineerò puntualmente questo genere di parole); per esempio. "solo", che significa "nient'altro che...".

*iv*) Una parola negativa più una parola modale, che equivale sempre a una DNP; per esempio. “possibile = non è vero che non è…” Si noti che quando un autore stabilisce “l’uguaglianza salvo i seguenti vincoli…” ciò equivale ad affermare che l’uguaglianza vale secondo una modalità e quindi non abbiamo un’uguaglianza ma una DNP di Si noti inoltre che l’uso comune delle parole “equivalente” e “simile” copre una moltitudine di modalità di quasi-uguaglianza, senza spiegare nel caso in questione quale esatta modalità di uguaglianza venga scelta..[[2]](#footnote-2)

*v*) Una parola negativa più una disuguaglianza – solitamente rappresentata da un aggettivo comparativo; quest'ultima è una parola negativa perché è la negazione del caso di uguaglianza desiderato. Per esempio. “più di…” e “meno di…”. Si noti che "non è minore di..." ≠ "è maggiore o uguale a..." che è solo una negazione.

*vi*) Una parola negativa più un punto interrogativo apparentemente in attesa di una risposta negativa; per esempio. “Sono stupido? [Certo che no, non sono stupido]”; "Perché no?" [= Non vedo motivo contrario]”.

*vii*)Una parola negativa più una parola idealistica che ha cancellato (spesso in modo eufemistico) un'altra parola negativa; quindi quest'ultimo deve essere restaurato; per esempio. “creare = produrre dal nulla”, “chimerico = non reale”, “amore platonico = amore in separazione”, “perpetuo = senza fine”, ecc..

*viii*) Una parola negativa più una negativa implicita; quindi quest'ultimo deve essere scoperto; ad esempio, "la portata [eccessiva] di quest'opera ha forse ostacolato i miei connazionali..." (Lobachevskij 1840, Prefazione).

*ix*) Una parola concepita dall'autore come negativa all'interno di un dato contesto specifico; per esempio. nella fisica teorica le parole “cambiamento”, “variazione”, ecc. sono negative poiché richiedono spiegazioni teoriche; in un contesto matematico anche l'aggettivo “positivo” può avere un significato negativo, ad es. quando indica un valore “diverso da zero”. Questi esempi sottolineano che dobbiamo superare un'imprecisione principale riguardante la natura (affermativa o negativa) di una parola candidata a essere riconosciuta come parola negativa.

In ambito linguistico si è soliti tenere conto di più distinzioni rispetto a quelle sopra indicate. Ad esempio, è noto che esiste una differenza semantica tra ad esempio “Non è una persona disonesta” e “È una persona non disonesta”; “Non un matrimonio felice” e “Un matrimonio infelice”. Viene inoltre sottolineata la differenza semantica tra una negazione dell'intero predicato e una negazione del termine predicativo; per esempio. “Non tutti i maschi non sono sposati” e “Tutti i maschi non sono sposati”.

Sebbene semanticamente diverse, ciascuna coppia di proposizioni include due parole negative; questo è il fatto logico che viene notato in modo oggettivo da un lettore; i restanti significati delle proposizioni appartengono ad altri mondi rispetto alla caratteristica logica che ci interessa. Presumo quindi che le ulteriori distinzioni linguistiche siano troppo sofisticate rispetto alla distinzione fondamentale tra il caso di una negazione e quello di due negazioni; cioè una proposizione che include due negazioni deve essere considerata come una DNP indipendentemente dalla loro posizione all'interno della proposizione.

Secondo la mia esperienza, i casi sopra indicati esauriscono l'insieme di tutti i casi incontrati quando si riconoscono DNP all'interno di un testo.

**4. Una nuova organizzazione formale di una teoria scientifica: gli ultimi tre passaggi logici**

3) In ciascuna delle teorie scientifiche sopra menzionate i DPN compongono prove per assurdo, la cui conclusione finale è ancora una DNP.

4) Alla conclusione finale l'autore applica sotto due vincoli (Markov 1961, p. 5) il principio di ragione sufficiente (PSR, Leibniz 1686) per tradurlo in una proposizione affermativa, l'unica che possa essere verificata con la realtà.

5) Questa proposizione e le sue conseguenze vengono poi verificate con la realtà e quindi possono convalidare la teoria. Questo è l’ultimo passo logico di tale organizzazione della teoria.

La differenza tra i due tipi di logica non è superabile né approssimabile (Goedel 1933f). Lo stesso vale per la differenza tra i due tipi di organizzazione teorica. In effetti, presentano molte variazioni radicali nei significati delle loro nozioni di base:

Proposizioni semplici: assiomi/problemi

Principi: a priori/metodologici

Passi logigi: deduzioni/induzioni

Teoremi: diretti/per assurdo

Conclusioni: teoremi infiniti/nuovo metodo scientifico

Insomma, per diverse ragioni la distanza dell’IA dalla teoria PO appare incolmabile.

**5. Una nuova definizione di intelligenza**

Affrontiamo ora un grande problema, la definizione di “significato”. Una possibile definizione è la seguente: “Il significato è la pertinenza all'azione e ai pensieri che gli esseri umani attribuiscono agli stimoli che incontrano nelle sensazioni”. Questa definizione attribuisce la massima importanza all'ambiente, che però può essere sia quello esterno che quello interno ad una persona. Quindi il problema principale per dare significato ad una proposizione è contestualizzarla. Mentre per contestualizzare una proposizione l'uomo si avvale di conoscenze pregresse, un computer deve apprendere quale può essere di volta in volta il contesto di riferimento. A prima vista sembra impossibile enumerare tutti questi contesti e sottoporli all’apprendimento automatico[[3]](#footnote-3).

Non meno difficile è il problema di definire la nozione cruciale di “intelligenza”. Ad esempio, suggerisco una vecchia definizione: “L’intelligenza è una capacità aggregata o globale dell’individuo di agire in modo mirato, di pensare razionalmente e di affrontare in modo efficace il suo ambiente”. (Wechsler 1944).

Osservo innanzitutto che esiste una distinzione tra il singolo atto intellettuale di una persona, che può risultare anche da un comportamento casuale, e la costruzione di una teoria scientifica sistematica, che dia senso pieno alle diverse attività a cui fa riferimento Wechsler. In corrispondenza di questa distinzione si possono avanzare due tesi principali sull’IA: la tesi debole, secondo la quale l’IA può possedere un po’ dell’intelligenza umana attraverso alcuni singoli atti, e la tesi forte, secondo la quale l’IA ingloba l’intera intelligenza umana.

Alla luce dei due tipi di organizzazione teorica suggerisco una nuova definizione di “intelligenza”, una definizione che è preliminare a qualsiasi conclusione sull’IA.

 Secondo i due tipi di organizzazione teorica, **l'intelligenza è definita come la capacità di utilizzare principalmente sia la logica classica che la logica intuizionista**, più le traduzioni (dirette e inverse) tra di loro. Si riferisce piuttosto ai singoli atti umani e alle conversazioni in linguaggio naturale, il ragionamento per la costruzione di teorie sistematiche.

In questa luce avanzo la proposta dei seguenti otto gradi dell'intelligenza umana che vanno dalla mera intelligenza dei singoli atti alla capacità della mente umana di costruire teorie sistematiche.

1. Pronunciare una proposizione sensata.
2. Rispondere alle domande in generale.
3. Ragionare collegando proposizioni in logica classica (deduzioni, dimostrazioni di teoremi).
4. Formulare teorie assiomatiche.
5. Distinguere i due diversi tipi di logica, logica classica e logica intuizionistica (a cui equivale la logica modale attraverso il suo modello S4 (Hughes e Cresswell 1996, pp. 224 ss.)).
6. Ragionare in almeno due diversi tipi di logica, logica classica e logica intuizionista, in accordo con i due modi di ragionare suggeriti da Platone (η e ) e Niccolò Cusano (*ratio* e *intellectus*).
7. Formulare un'intera teoria secondo una logica induttiva, quella intuizionista.
8. Tradurre da un tipo di logica all'altro.
9. Attualmente l’intelligenza artificiale è capace dei primi tre gradi di intelligenza. È discutibile se l’IA possa formulare una teoria assiomatica (Robinson e Voronkov 2001).

Consideriamo però il quinto grado di intelligenza.

**6. Un nuovo principio per l’informatica riguardante il quinto grado di intelligenza: l’AI non può riconoscere le DNP**

Consideriamo ora i tre tipi di proposizioni doppiamente negate sopra elencati che non sono DNP. Le ragioni per scartarle come DNP sono di natura semantica. Nel primo caso il rapporto con la realtà gioca un ruolo cruciale. L'intelligenza artificiale non può avere cognizione della realtà (umana), ma solo con i racconti della realtà, con tutti i limiti e le deformazioni che i racconti comportano; inoltre, la realtà è troppo grande per essere interamente assunta dalla memoria di un computer[[4]](#footnote-4).

Nel secondo caso la proposizione in questione include un'enfasi psicologica mentre AI non può intendere questa enfasi; vede le due negazioni solo come figure linguistiche recise, senza collegarle in un nesso logico estraneo a quello dato dalla logica classica.

Applichiamo queste osservazioni all'AI: non può riconoscere le DNP false tra i tre casi della sezione precedente. In particolare, nel terzo caso l'ultima negazione della proposizione doppiamente negata non conta come negazione perché si limita a spiegare la natura della prima; riconoscere il loro nesso come spiegazione richiede di comprendere il significato della precedente negazione; questo è impossibile per l'intelligenza artificiale[[5]](#footnote-5).

Come primo risultato otteniamo che l’AI non è in grado di distinguere le DNP genuine da quelle false. Di conseguenza qualsiasi discorso umano basato su proposizioni doppiamente negate può essere reiterato da un computer, ma senza condividerne la semantica.

Si noti inoltre che nei casi *i*)-*v*) della tipologia precedente un essere umano riconosce le DNP applicando regole sintattiche. Invece nei casi *vi*)-*ix*) un essere umano riconosce le DNP comprendendo i significati delle proposizioni. Tra questi ultimi gli ultimi casi *viii*)-*ix*) sono i più difficili perché il riconoscimento della natura di una DNP dipende dalla semantica del contesto della proposizione in esame; bisogna quindi prestare attenzione a un numero a priori indefinito di altre proposizioni del testo. Di conseguenza, un'analisi accurata di tutte le occorrenze di DNP all'interno di un testo richiede che un essere umano reiteri più volte la sua lettura per essere eventualmente sicuro di aver colto tutti i possibili significati sia delle parole nelle proposizioni che nel contesto di loro. Tuttavia, questo lavoro non è nelle capacità di un computer.

Suggerisco che *nessun computer possa differenziare una DNP del linguaggio umano dalla corrispondente affermazione per tre ragioni*:

1) esiste un numero infinito di DNP;

2) una DNP nel linguaggio umano può dipendere dal suo contesto;

3**)** dopo un secolo di attività in ambito informatico sappiamo che nessun programma ha mai implementato una DPN.

Quindi, è giunto il momento di riconoscere che nessun computer può implementare DNP (come dopo secoli di tentativi falliti di costruire movimenti senza fine, la fisica teorica ha affermato il principio dell’impossibilità del moto perpetuo)[[6]](#footnote-6);

Queste proposizioni includono tutte le consuete considerazioni sulla capacità della mente umana di funzionare in un mondo complesso e in continua evoluzione e in un mondo aperto, permeato di ambiguità, complessità e contingenze (Marcus e Davis 2019, 9, 11, 120- 111).

Si può decidere di ignorare questa carenza di un computer e utilizzarlo per gestire solo le DNP i) – v). In tal caso abbiamo un computer il cui comportamento è chiuso in un *isolamento schizofrenico*. Può essere utile per alcuni compiti ristretti, ma se collocato in un contesto aperto (umano) i suoi risultati sono irrealizzabili.

Di conseguenza il test di Turing è superato. È sufficiente sottoporre a un robot una proposizione doppiamente negata come la domanda “I nemici dei tuoi nemici sono nemici o amici?” Il robot può produrre una disquisizione sulla doppia negazione e addirittura suggerire che secondo i due tipi di logica, classica e intuizionista, esistono due risposte opposte; tuttavia, non può decidere sulla dicotomia perché, a seconda della proposta avanzata dal contesto (essenzialmente guerra/pace), la questione non può essere risolta; può solo applicare la logica classica che porta a rispondere “Amici”, che non sempre è la risposta corretta.

 **7. L'inaccessibilità da parte dell'IA della nuova organizzazione formale di una teoria scientifica**

Consideriamo ora le dimostrazioni ad assurdo (di un PO). Discutono per mezzo di implicazioni tra DNP. Pertanto, l’argomentazione non è riconosciuta con certezza dall’AI, a causa dell’insufficienza dell’AI nel riconoscere le DNP. Inoltre, la conclusione di questo tipo di prova è essenzialmente una DNP; altrimenti potrebbe essere riconvertito in una dimostrazione diretta (Grize 1991). Questa componente di questa organizzazione teorica è un’altra caratteristica logica non riconosciuta dall’intelligenza artificiale.

Tuttavia, ammesso e non concesso che le DNP siano ben riconosciuti dall’AI, la conclusione di una dimostrazione per assurdo è una DNP su cui il PSR deve essere applicato sotto i due vincoli di Markov, per ottenere una proposizione affermativa. Il secondo vincolo di Markov sulla conclusione della dimostrazione finale ad assurdo, cioè essere decidibile, non è facilmente soddisfatto, perché bisogna riconoscere se la conclusione deve essere logicamente decidibile, o matematicamente decidibile, o operativamente decidibile. I primi due casi sono problemi anche per l’uomo; in particolare, per decidere che una proposizione è indecidibile è necessaria una dimostrazione che deve essere inventata. Il terzo caso è generalmente irrisolvibile da parte dell’IA a causa della sua conoscenza essenzialmente parziale del mondo fisico.

**8. Conclusioni**

In quanto sopra abbiamo ottenuto risposte pertinenti ad alcune domande introduttive perché abbiamo posto rimedio alle insufficienze della nostra consapevolezza sui temi presentati dalle osservazioni 3) – 5), cioè la storia e i fondamenti della scienza più la riflessione filosofica sui limiti della un computer.

In termini generali, due millenni di teorie scientifiche hanno dimostrato che esistono solo due tipi di organizzazioni teoriche e inoltre le ricerche del secolo scorso hanno dimostrato che esse sono separate dalla differenza insormontabile tra la logica classica e la logica intuizionista di una teoria PO, cioè la validità o meno della legge della doppia negazione. Dobbiamo concludere che i computer non hanno la capacità di rappresentare sia un PO sia il ragionamento all'interno della logica intuizionista. Nella storia delle teorie scientifiche questa metà dell’intelligenza umana basata sulla PO e sulla logica intuizionista è stata molto produttiva (termodinamica, geometria non euclidea, rivoluzione della fisica teorica agli inizi del ‘900, ecc.); in realtà, i suoi risultati sorprendenti mostrano che rappresenta il potere più alto dell’intelligenza umana.

 Poiché la mente umana può scegliere il tipo di logica e in particolare può seguire la logica intuizionista secondo un’organizzazione teorica basata sui problemi, le sue capacità non possono essere superate nemmeno da un computer e dall’intelligenza artificiale. Questo fatto smentisce sia la tesi dell’IA debole che quella dell’IA forte. Un computer, per quanto potente sia anche negli atti di tipo intelligenza, non costituisce altro che un pappagallo estremamente avanzato.

In conclusione, tutto quanto sopra dimostra che il modo comune di concepire come valida solo la logica classica, semplicemente perché appare come quella più produttiva, spiega in gran parte il problema dell’IA. L’intelligenza artificiale sembra equiparare le prestazioni dell’intelligenza umana perché forse equipara il ragionamento umano nella logica classica, ma la mente umana include un mondo logico alternativo. Pertanto, l’introduzione dell’intelligenza artificiale porta gli esseri umani a svalutare l’argomentazione nella logica classica che l’intelligenza artificiale può gestire e implementare, per sviluppare piuttosto principalmente l’argomentazione nella logica intuizionista. Inoltre suggerisce un nuovo principio, quello dell'impotenza dell'IA.

Aggiungo una considerazione di carattere storico-filosofico. L'atteggiamento tradizionale per un'assoluta certezza della logica classica e la sua indispensabilità hanno oscurato il riconoscimento di modi alternativi sia di ragionare che di organizzare una teoria. In passato solo pochi studiosi hanno esplorato le nuove macchine da calcolo. Uno di questi fu Charles Peirce che sottolineò “l’impotenza informatica” delle “macchine intelligenti”. Al momento vediamo che il vero contenuto della valutazione di Peirce è che l’IA non può rappresentare un PO e anche le sue componenti logiche. Questo risultato rende esatto ciò che Charles Peirce affermava a proposito di una “macchina pensante”: essa ha “impotenza intrinseca”. In altri termini, non può implementare una teoria non assiomatica.

[In primo luogo] … è privo di ogni originalità, di ogni iniziativa [= comportamento non prestabilito]. Non riesce a trovare i propri problemi... Non può orientarsi tra diverse procedure possibili. [In secondo luogo] … la capacità della macchina ha limiti assoluti; è stato congegnato per fare una certa cosa e non può fare nient'altro. (Peirce 1887)

Infine aggiungo considerazioni etico-politiche sulla questione del nostro convegno: *Le macchine governeranno il mondo*?

Ho suggerito che poiché la mente umana sceglie il tipo di logica, un computer non può superare le sue capacità. Ma gli esseri umani possono interagire con i computer secondo volontà negative. Quindi i peggiori timori sollevati dall’IA sono realistici e la questione è molto attuale. A decidere la questione non sarà però la sfida tra robot e esseri umani su quale sia l’intelligenza migliore, bensì la **politica** dell’umanità sull’interazione uomo-computer.

Ciò che occorre per conquistare una gestione pacifica dei robot è un'etica che 1) richieda a tutti gli esseri umani l'adempimento dello stesso comportamento di base rispetto ai robot; quindi deve essere universale, senza eccezione dei gruppi umani; e 2) suggerisce un comportamento collettivo di tipo separativo, cioè tabù etici. Sebbene adeguata a presentare l’età matura dell’umanità, questa etica è molto lontana da noi. Tuttavia, una prima misura politica è quella di abolire tutte le attività letali che rappresentano una logica classica del sì/no, tutto/niente, bene/male, innanzitutto abolire le guerre perseguendo per eccellenza questo tipo di logica riducendo gli esseri umani a robot assassini, incapaci di ragionare secondo la logica intuizionista. Inoltre, abolire lo jus ad bellum dello Stato (diritto di fare la guerra).

 Infine, ricordiamo i saggi consigli di Einstein:

# “È diventato spaventosamente ovvio che la nostra tecnologia ha superato la nostra umanità.”

# “La scienza ha fornito agli esseri umani la possibilità di liberarsi dal duro lavoro, ma la scienza stessa non è un liberatore. Crea mezzi, non obiettivi. L’uomo dovrebbe usare [la scienza] per obiettivi ragionevoli. Quando gli ideali dell’umanità sono la guerra e la conquista, quegli strumenti diventano pericolosi quanto un rasoio nelle mani di un bambino di tre anni. Non dobbiamo condannare l’inventiva dell’uomo e la paziente conquista delle forze della natura perché ora vengono utilizzate in modo sbagliato e disobbediente. Il destino dell'umanità dipende interamente dal suo sviluppo morale.”

# “Credo che un progresso serio (nell’abolizione della guerra) possa essere raggiunto solo quando gli uomini si organizzeranno su scala internazionale e si rifiuteranno, come corpo, di entrare nel servizio militare o di guerra..” (Einstein 2010, p. 249)

**References**

Denning P.J. (2005), “Is Computer Science Science?”, *Comm. ACM*, 48, no. 4, pp. 27-31.

Drago A. (2012), “Pluralism in Logic:The Square of Opposition, Leibniz’ Principle of Sufficient Reason and Markov’s principle”, in J.-Y. Béziau and D. Jacquette (eds): *Around and Beyond the Square of Opposition*, Birkhaueser, Basel, 175-189.

Dummett M. (1977), *Principles of Intuitionism,* Clarendon Press, Oxford.

Eden A.H. (2007) “Three paradigms of Computer Science”, *Minds and Machines*, 17, pp. 135-167.

Einstein Albert (2010). “The Ultimate Quotable Einstein”, Princeton: Princeton University Press

Galilei G., Lettere, Einaudi, Torino 1978.

Gardiès J.-L (1991), *Le raisonnement par l’absurde*, Paris : PUF.

Goedel K., *Collected Works.* Oxford U.P., Oxford, 1986, papers 1932 and 1933f.

Grize J.B. (1970), « Logique »,*.* in Piaget J. (ed.), *Logique et connaissance scientifique.* Éncyclopédie de la Pléiade, Gallimard, Paris, pp. 135-288,.

Heidgger M. (1976), “Nur noch ein Gott kann uns retten”, *Der Spiegel* , May 31st.

Hughes G.E. and Cresswell M.J. (1996). *A New Introduction to Modal Logic*, London: Routledge.

Koyré A. (1957), *From the Closed Cosmos to the Infinite World*, Baltimore: Wisconsin Univ. Press..

Kuhn T.S. (1969), *The Structure of the Scientific Revolutions*, Chicago U.P., Chicago.

Landgrebe J. and Smith B. (2021), “Making AI meaningful again”, *Synthese*, 198, pp. 2061-2081.

Landgrebe J. and Smith B. (2022), *Why Machines Will Never Rule the World*, London: Routledge

Leibniz G. W. (1786), *Letter to Arnauld*, 14-7-1686, Gerh. II, Q., Opusc. 402, 513.

Lobachevsky N.I. (1955), *Geometrische Untersuchungen zur der Theorien der Parallellineen*, Finkl, Berlin, 1840 (English transl. as an Appendix to Bonola R., *Non-Euclidean Geometry.* Dover, New York).

Marcus G. and Davis E: (2019), *Rebooting AI: Building Artificial Intelligence We Can Trust*, London: Vintage.

Markov A.A. (1961), “On Constructive Mathematics”*.* *Trudy Math. Inst. Steklov*, 67 (1962) 8-14; also in *Am. Math. Soc. Translations,* 98 (2), pp. 1-9.

Peirce S.C. (1931-1935), *Collected Papers of Charles Sanders* Peirce [1931 — 1935], Cambridge U.P., Cambridge.

Peirce C.S. (1887), “Logical Machines”, *Am. J. Psychology*, 1, 165-170, pp. 168-169.

Prawitz D. (1976) *, “*Meaning and Proof. The Conflict between Classical and Intuitionist Logic. *Theoria*, 43, 6-39.

Prawitz D. and Melmnaas P.-E. (1968), “A survey of some connections between classical intuitionistic and minimal logic”, in A. Schmidt and H. Schuette (eds.), *Contributions to Mathematical Logic.* North-Holland, Amsterdam, pp. 215-229.

Robinson A. and Voronkov A. (2001), *Handbook of Automated Reasoning*, Cambridge MA: MIT Press.

Russell, B. (1903), *The Principles of Mathematics*, Cambridge U.P., Cambridge.

Weber M. (1930), *The Protestant Ethic and the Spirit of Capitalism*, New York: Routledge.

Wechsler *D. (*1944*),* *The measurement of adult intelligence*, New York: Williams & Wilkins.

1. Molti autori di teorie scientifiche hanno scritto una proposizione comprendente una parola modale più due negazioni: es. “È impossibile un movimento senza fine.” Questa associazione di una modalità con una doppia negazione fornisce prova della percezione intuitiva da parte dell’autore del legame tra logica modale e logica intuizionista; questo legame è stato formalmente stabilito in tempi recenti (Hughes e Cresswell 1996, pp. 224 ss.). [↑](#footnote-ref-1)
2. Molti autori di teorie scientifiche hanno scritto una proposizione comprendente una parola modale più due negazioni: es. “È impossibile un movimento senza fine.” Questa associazione di una modalità con una doppia negazione fornisce prova della percezione intuitiva da parte dell’autore del legame tra logica modale e logica intuizionista; questo legame è stato formalmente stabilito in tempi recenti (Hughes e Cresswell 1996, pp. 224 ss.). [↑](#footnote-ref-2)
3. Il libro (Landgrebe e Smith 2022) suggerisce argomentazioni convincenti su questo punto. Ma non è sufficiente riguardo alle potenzialità dell’IA perché 1) si basa su argomenti matematici, tuttavia la matematica può evolversi in teorie sempre nuove; 2) si basa sui modelli matematici di un sistema complesso, la cui teoria però non è completa e stabile; 3) più in generale, non considera questioni logiche. Pertanto, non può suggerire altro che risultati plausibili. [↑](#footnote-ref-3)
4. Ricordiamo che un computer che traduce inglese-francese necessita di 36 milioni di coppie di proposizioni. [↑](#footnote-ref-4)
5. A questo proposito (Landgrebe e Smith 2021, p. 2066) hanno scritto sagge parole: “Sembra che l’acquisizione del linguaggio presupponga l’attuazione di un insieme comune di distinzioni ontologiche da parte degli apprendenti di lingua, inclusa la distinzione tra oggetto e processo, tra individui e categorie, tra proprietà naturali e accidentali degli oggetti, e così via”. [↑](#footnote-ref-5)
6. In particolare un computer non può implementare tutti i DNP che esprimono la rottura della simmetria di Curie e in generale la stessa parola “simmetria”, che per la sua prima parte “syn” è una parola modale, quindi attraverso il modello S4 equivale ad una DNP. [↑](#footnote-ref-6)