

Le competenze linguistiche e matematiche usate nel lavoro non sono sostituibili dall'IA

## L'INTELLIGENZA ARTIFICIALE SA FARE SOLO I COMPITI

di Claudio Franchi

**Le possibili implicazioni della IA nei settori dell'istruzione e del lavoro, tra limiti e prospettive. Uno studio dell'OCSE. La valutazione in base ai test e l'ideologia liberista. L'insostituibile dimensione relazionale umana.**

**L**o sviluppo continuo di sistemi digitali complessi, più comunemente circoscritti sotto il nominativo di *intelligenza artificiale*, sta avendo un impatto enorme tagliando in trasversale l'elemento sociale e culturale mentre tutte le grandi organizzazioni internazionali stanno dedicando studi, ricerche e approfondimenti a questo fenomeno. Nel tentativo, ufficialmente dichiarato, di comprenderne i limiti e le possibilità.

In particolare, ultimamente, l'attenzione è rivolta alle NLP (Natural Language Processing) ovvero quegli algoritmi che, attraverso reti neurali, sono in grado di comprendere e riprodurre il discorso umano. L'esempio più famoso è Chat GPT (Generative Pre-trained Trasformer), prodotto da OpenAI, ma con un sostegno più che ingente di Microsoft (quasi 10 miliardi di dollari), che si configura come il sistema che ha avuto la più rapida diffusione della storia delle tecnologie, ad oggi, nel mondo occidentale e comunque l'applicazione che ha suscitato più interesse in tutto il mondo, per le sue "incredibili" capacità.

Negli ultimi mesi, l'OCSE, l'Organizzazione per il Commercio e lo Sviluppo Economico, nel corso di un progetto generale sulle capacità di IA in relazione alle possibili implicazioni che potrebbe comportare nei settori dell'istruzione e del lavoro, effettuato dal CERI (il Centro per l'Innovazione e la Ricerca nell'Istruzione), ha pubblicato *Is Education Losing the Race with Technology? AI's*

*Progress in Maths and Reading*, con l'obiettivo esplicito di «sviluppare misurazioni delle capacità dell'IA che siano comprensibili, comprensive, ripetibili e rilevanti per le politiche» (OECD 2023, p. 3). L'idea centrale dello studio è cercare di comparare i risultati ottenuti sottoponendo all'IA i test contenuti nelle indagini internazionali PISA e PIAAC con quelli raggiunti da studenti e popolazione adulta rispettivamente.

La prima grande evidenza, che rappresenta poi lo snodo su cui costruire il senso dello studio in questione, riguarda il miglioramento rapido dei risultati ottenuti dall'IA nei test sul *reading*, rispetto a quelli ottenuti nel 2016, mentre nella soluzione di quelli relativi alle capacità matematiche i risultati restano più o meno comparabili.

Questa discrasia viene spiegata e si spiega principalmente dalle linee di sviluppo industriale sulle quali si sono concentrati gli investimenti maggiori negli ultimi anni, tutti riguardanti il rafforzamento degli NLP, cioè delle capacità linguistiche dell'IA. Di qui, se le risorse delle società si rivolgessero con la stessa intensità e potenza economico-industriale anche verso l'altra direzione, resta comunque semplice dedurre che, nei prossimi anni, si potrebbe giungere senza particolari problemi a progressi più che sensibili anche rispetto allo sviluppo di sistemi relativi alle capacità di calcolo e matematica, che tra l'altro per alcuni versi meglio si adatterebbero alla struttura digitale.

## PISA E PIAAC: problematiche epistemologiche e modelli neoliberisti

Per comprendere appieno il valore di questo studio e le sue possibili e conseguenze o indicazioni ai decisori politici, si deve sempre tenere conto e ricordare chiaramente che da molti anni, se non addirittura dalle prime esperienze (dal PISA 2000), l'efficacia di queste indagini internazionali è sempre stata messa in discussione da molti studiosi – solo per fare un esempio tra decine si guardi a Bonderup Dohn (2007) –, perché gran parte dell'importanza di queste indagini sembra provenire esclusivamente dall'autorevolezza delle istituzioni stesse. In un vero e proprio cortocircuito semantico, esse risultano attendibili e accreditate perché promosse dalle grandi e famose Istituzioni attendibili e accreditate che le finanziano e non perché siano effettivamente il risultato di indagini date da quadri di riferimento teorici o da metodologie scelte e applicate.

L'argomento fondamentale – e allo stesso tempo il maggior punto critico – è che in questi *test* non si riuscirebbero a valutare le competenze o le conoscenze, *skills and knowledge*, ma semplicemente le competenze e le conoscenze necessarie per risolvere *esattamente* quel test o quella tipologia di test. Questo nodo cruciale, tra l'altro, non è solo inerente al *test* stesso ma anche all'intero sistema educativo che queste indagini dovrebbero riuscire a interpretare e analizzare o, secondo le parole usate nelle pubblicazioni, a giudicare, “*to assess*”.

Non a caso anche un approccio sistemico, se guardato attentamente, pone delle problematiche invero insormontabili, a meno che non si decida di saltarle e ignorarle e di seguire la strada *mainstream*. Le modalità di insegnamento nazionale vengono, infatti, valutate proprio mediante questa tipologia di *test*. Alla stregua di quanto si è detto, questo potrebbe anche significare che alcune modalità porterebbero a risultati migliori *non* nelle competenze e nelle conoscenze dispensate e poi acquisite, ma *solo* in quelle particolari e “tecniche” capacità di affrontare le batterie di *test* dell'indagine. Esemplificativo, in questo senso, è che quando alcuni Paesi hanno deciso di tentare di “migliorare” le *performances* dei propri studenti, si sono trovati costretti a inserire nei programmi scolastici proprio l'apprendimento indispensabile per la preparazione a questi *test* specifici. Incredibilmente non si è pensato affatto di effettuare una parziale o completa trasformazione in senso pedagogico dell'approccio del sistema dell'istruzione nazionale, ma solo di migliorare le *performances* valutate dalle indagini internazionali. In questo modo, tra l'altro, l'ipotetico vantaggio dell'indagine PISA di essere “*curriculum free*”, cioè di non valutare direttamente gli elementi inseriti nei programmi, nella pratica si ribalta e diventa l'evidenza di una paradossale inefficacia sulla programmazione educativa. Si aggiunga a questo, come provato da Fiore e Poliandri, che molto spesso, almeno in Italia, neanche un vero dibattito pubblico si è riuscito a sollevare negli anni, lasciando inoltre sempre inascoltata la comunità scientifica di riferimento.



A questo bisogna aggiungere un dato di realtà politica, pur senza nessuna presa di posizione: la politica culturale e economica che promuove l'OCSE è da sempre di stampo marcatamente neo-liberale e scientemente anche le indagini si sono spesso concentrate a valutare competenze e conoscenze capaci di rendere il sistema educativo funzionale alle esigenze dirette del mercato del lavoro e ai suoi nuovi bisogni. Non a caso solo un vero e proprio movimento di 'opposizione' accademica ha spinto i redattori dei *test* a valutare anche capacità relazionali o legate al benessere e la crescita individuale e sociale, interpretando in modo diverso i possibili obiettivi dei sistemi educativi nazionali.

Oltre al PISA – ma con gli stessi limiti sottolineati dalla comunità scientifica – l'OCSE conduce storicamente un'ulteriore indagine, su campo internazionale, chiamata PIAAC, Programme for the International Assessment of Adult Competencies, finalizzata a raccogliere dati cercando di individuare il livello di competenze degli adulti tra 16 e 65 anni su *literacy*, *numeracy* e *problem solving*, sempre su base campionaria, e a comprendere come queste competenze vengano utilizzate sia nel lavoro che nella vita quotidiana, a casa e in società. In parole semplici: le banche dati del PIAAC, secondo gli intenti dichiarati, dovrebbero permettere di comprendere la relazione tra competenze, istruzione e lavoro e evidenziare il ruolo preciso di queste competenze rispetto alla dimensione occupazionale.

C'è da ricordare, però, che l'obiettivo ufficiale è sempre stato quello di fornire gli strumenti ai decisori politici per implementare strategie di sviluppo in subordine all'approccio neoliberista, e cioè quello di comprendere quali competenze devono essere fornite dai sistemi educativi tra quelle che in quel momento sono richieste dal mercato del lavoro e dalle imprese private.

## I "risultati" dello studio OCSE: problemi interpretativi o illusioni ideologiche?

Nonostante le perplessità e problematiche sollevate dalla comunità scientifica e da quella politica e sindacale, un centro di ricerca interno all'OCSE, il CERI, nel

quadro del progetto AIFS (*Artificial Intelligence and the Future of Skills*) decide comunque di utilizzare le stesse batterie di *test* di PISA e PIAAC, per valutare le *performances* dell'IA nella loro soluzione, esplicitando chiaramente come questa comparazione possa «aiutare i decisori politici a dare una nuova forma ai sistemi educativi in modo da preparare al meglio gli studenti per il futuro e a fornire opportunità agli adulti che stanno apprendendo a rinnovare le loro competenze».

La metodologia usata per la valutazione di questa comparazione è stata quella di sottoporre i risultati delle prove affrontate dall'IA all'analisi di esperti di Computer Science.

L'IA, anche grazie all'enorme sviluppo degli ultimi anni dell'NLP, cioè del lavoro sui Natural Language Processes, è stata in grado di rispondere a più dell'80% dei *test* di *Literacy* del PIAAC. Per quanto riguarda invece quelli di *Numeracy*, i risultati sono stati peggiori, attestandosi intorno a poco più del 60%, probabilmente proprio a causa dei mancati investimenti dei produttori di IA su questo sviluppo specifico.

Al momento dell'analisi, per quanto riguarda il PIAAC, in *Literacy* l'IA raggiunge il livello 3 di 5, laddove il 90% della popolazione adulta è al livello 3 o inferiore, mentre solo il 10% supera questo livello, arrivando a 4 o a 5. In *Numeracy*, invece, le performance dell'IA sono di livello due per i *test* semplici e quelli intermedi, mentre per le domande più difficili arriva al livello 3. La popolazione adulta in questo caso arriva nel 57% dei casi massimo al livello 2, e nell'88% massimo al livello 3.

In ogni caso, secondo tutti gli esperti consultati, con l'attuale frequenza di avanzamento tecnologico, entro il 2026 l'IA sarà in grado senza problemi di superare l'intera batteria di *test*, lasciando definitivamente dietro la popolazione umana. Alla luce del nostro discorso, quindi, sono fondamentali le conclusioni che se ne trarrebbero da questo studio: non si registrano incrementi sensibili negli ultimi anni nelle capacità di risoluzione umana, mentre si prevede una progressione inarrestabile da parte dell'IA (si prendano in considerazione i dati citati: la *Literacy* risulterebbe usata, al livello dell'IA, dal 59% della forza lavoro,



mentre la *Numeracy*, sempre al livello comparabile con l'IA, tra il 27% e il 44%). Da questo assunto, se ne deduce una assoluta condizione futura: le macchine potranno sostituire e sostituiranno donne e uomini, almeno nella percentuale dei compiti rilevata dalle indagini.

### L'intelligenza come questione centrale

L'apparente fondatezza di questa conclusione però si scontra con un elemento chiave intrinseco che è esattamente il centro dell'intero discorso 'scientifico'. La questione infatti rimane sempre in una domanda che ha mille possibili risposte, e che deve – o dovrebbe – essere affrontata come presupposto: cos'è l'intelligenza? O quantomeno, dal punto di vista dell'analisi delle competenze linguistiche e matematiche necessarie e utilizzate nella vita e soprattutto sul lavoro, quale tipo di intelligenza viene usata? Lungi dal ripercorrere qui una genealogia di questo concetto, limitiamoci però a indicare che tipo di procedure razionali l'IA mette in campo e come queste possano avere una relazione con quanto indagato da PISA e PIAAC.

Nel 1956 John McCarthy aveva coniato la definizione di Intelligenza Artificiale, descrivendola così: «ogni aspetto dell'apprendimento od ogni altra caratteristica dell'intelligenza può in principio essere descritta così precisamente che può essere simulata da una macchina».

In realtà, una delle prove della difficoltà della risposta alla domanda è anche data dalle mille differenti forme di elaborazione e determinazione dell'uso delle tecnologie digitali per svolgere compiti differenti.

Restringiamo allora il campo alla dimensione linguistica, che nello studio dell'OCSE avrebbe rappresentato un possibile *vulnus* nella popolazione adulta, in particolare sul lavoro, dal quale sarebbe poi potuta dipendere una massiccia e vasta "sostituzione" della gestione di queste competenze da parte dell'IA.

La grande svolta "epistemologica" che ha poi definitivamente aperto le strade al *deep learning* è venuta soprattutto a fronte degli scarsi risultati ottenuti partendo da ipotetici paradigmi linguistici generali che poi le macchine avrebbero dovuto e potuto utilizzare per comprendere e produrre testi. Questo meccanismo *top down* infatti non ha mai raggiunto dei risultati in grado di rappresentare una vera e propria innovazione tecnologica capace di entrare a pieno titolo nei meccanismi di produzione e vita quotidiana.

Invece, una delle intuizioni di Frederick Jelinek, da quando nel 1972 era nel Continuous Speech Recognition Group della IBM, fu proprio quella di ribaltare la prospettiva e indirizzare la dimensione linguistica verso una ricerca probabilistica, praticamente *bottom up*. Quindi, da allora in poi fino a Chat GPT 4: la plau-

## Abbonati al Pepeverde

Trimestrale di letture e letterature giovanili,  
analizza le tendenze dell'editoria, segnala le pubblicazioni migliori.



**Il costo dell'abbonamento annuale è in offerta a 35 euro anziché 45.  
L'abbonamento alla versione digitale del Pepeverde: 25 euro**

- online sul sito <https://www.edizioniconoscenza.it>
- con bonifico bancario a favore di Valore Scuola coop.  
IBAN: IT44 Q0103003202 00000 2356139

inviando poi la ricevuta del pagamento a [commerciale@edizioniconoscenza.it](mailto:commerciale@edizioniconoscenza.it)

- scrivendo una e-mail a [commerciale@edizioniconoscenza.it](mailto:commerciale@edizioniconoscenza.it).

- Puoi abbonarti anche utilizzando la Carta Docente. In questo caso telefona prima per informazioni allo 065813173 oppure scrivici una e-mail.

[www.edizioniconoscenza.it](http://www.edizioniconoscenza.it)

sibilità delle frasi non viene giudicata dall'applicazione corretta di un sistema di regole ma dalla probabilità statistica, applicando la potenza di calcolo alla gestione di migliaia, milioni e oggi addirittura miliardi di parametri. In pratica, senza approfondire il ragionamento, di ogni parola si descrivono le proprietà (*features*), che poi vengono annotate in etichette (*labels*), per poi dividerle in dati di addestramento e dati di valutazione (*training e testing data*).

Il tutto con una potenza di calcolo tanto grande da assicurare risultati eccellenti. A questo punto si deve accettare che la strada che si è deciso di percorrere ha portato a esiti incredibili e forse insperati, tanto da sfiorare addirittura in un possibile nuovo paradigma epistemologico.

## L'intelligenza e l'efficacia

Però identifichiamo bene quale sia l'efficacia alla quale si è giunti e quindi quale particolare intelligenza sia operativa negli NLP.

Infatti, rispetto all'interpretazione e alla produzione dei testi, l'IA è al tempo stesso un paio di occhiali e una penna: l'occhiale è dato dal fatto che se non si sa leggere e capire esattamente un testo ci saranno solo vaghi segni grafici, non utilizzabili realmente nella società umana; la penna è, invece, legata al fatto che se non si sa cosa scrivere e come scriverlo, si avranno solo tratti di penna inutilizzabili in qualsiasi contesto.

Nei contesti ai quali fa riferimento lo studio dell'OCSE, quei compiti che richiedono competenze di *Literacy* e di *Numeracy*, potranno essere svolti molto probabilmente con l'IA, attraverso l'IA, ma solo attraverso un'IA a gestione umana, che conosca il senso, il significato e gli obiettivi multiformi di realtà lavorative che nella loro complessità umana non riescono a essere interpretate correttamente in maniera autonoma da un'IA. L'esempio chiave è quello della guida autonoma degli autoveicoli: l'ultimo stadio, quello della guida totalmente autonoma, trova proprio nell'incredibile complessità della realtà quotidiana l'ultimo, insormontabile ostacolo.

Come è chiaro alla comunità scientifica il vero bivio è quello rispetto all'ambiente in cui opera l'Intelligenza Artifi-

ziale: un ambiente circoscritto e chiuso, come una scacchiera o come la lettura di miliardi di dati, rappresenta il *locus* ideale per l'IA. Una prevedibilità assoluta, un risultato certo e garantito.

Nel momento in cui l'ambiente è quello della realtà quotidiana vissuta, nella vita sociale e in quella lavorativa, i parametri esplodono, diventano cangianti, mobili, contraddittori, diventano la vita stessa. E quando ci si sposta sul linguaggio nel suo utilizzo, che è la forma per eccellenza di relazione nella vita, tutto diventa anche qui un continuo aggiustamento a partire dai dati, esattamente come teoricamente avviene nel *deep learning*, ma con una distinzione insuperabile: gli uomini hanno desideri e bisogni, paure e poteri in gioco, e in funzione di questo prendono delle decisioni.

Le decisioni di un'IA sono su base probabilistica, ma non hanno il senso umano, della cura, della seduzione, della risposta all'ingiustizia. Sono apparentemente neutre. Le decisioni umane invece, come secoli di psicologia, di filosofia, di sociologia e di altre discipline ci hanno insegnato, si muovono in una rete di relazioni di potere, di atti desiderativi e di asimmetria sociale e economica.

Uno degli argomenti contro l'affidabilità dell'IA è quello dei possibili *bias*, cioè la dipendenza dalle banche dati utilizzate. Infatti qualora questi testi riportino opinioni razziste, autoritarie, omofobe, il risultato sarà basato anche su queste, ovviamente inficiandone l'utilizzabilità. E allo stesso tempo, proprio la modalità del *deep learning* se messa in correlazione con la provenienza privata e commerciale dei sistemi NLP, non può che essere una vera e propria dimostrazione a priori dell'inaffidabilità umana, politica e sindacale dei risultati possibili.

La questione qui invece è che se anche questi problemi non sussistessero, la dimensione relazionale umana non potrebbe mai essere sostituita da un'IA, né nell'amore, né nel lavoro.

Come si salvaguarda allora la popolazione adulta sul lavoro? Con i diritti, con i salari giusti, con una transizione ecologica e giusta, con un bilanciamento tra vita e lavoro, con i servizi sociali, educando donne e uomini a diventare donne e uomini.

Facile, no?