

Tra sapere e saggezza: il framework MISE per la scuola dell'IA

di Alba Forni

Abstract

La diffusione della formazione all'intelligenza artificiale generativa nella scuola tende a concentrarsi sull'uso operativo degli strumenti, lasciando sullo sfondo la comprensione del meccanismo che li governa. Questo contributo sostiene che tale impostazione riposi su un equivoco epistemologico prima ancora che didattico: un modello linguistico di grandi dimensioni non recupera risposte da un archivio, ma le genera al momento dell'inferenza come completamento statisticamente plausibile del prompt, sulla base di rappresentazioni distribuite costruite in fase di addestramento. Da questa distinzione discendono conseguenze dirette per la pratica d'aula, dal prompt come co-costruttore del contenuto all'allucinazione come proprietà strutturale e non come guasto occasionale, che il senso comune insegnante, oscillante tra l'immagine dell'IA come archivio onnisciente e quella di oracolo infallibile, fraintende sistematicamente. Richiamando la definizione operativa di intelligenza di Nello Cristianini e un'attività di simulazione sul bias posizionale condotta in diversi corsi di formazione docenti, il paper mostra come la conoscenza tecnica del funzionamento degli LLM costituisca la premessa necessaria, e non un'opzione accessoria, per una progettazione didattica efficace. Vengono inoltre discussi la crisi della mediazione del sapere e dell'autorialità del testo prodotto con l'IA, e le ricadute sulla valutazione, proponendo di spostarne l'oggetto dal prodotto consegnato al processo che lo ha generato. Il lavoro si conclude presentando MISE (Meccanismo, Idea, Sperimentazione, Estensione), un framework in quattro fasi mappato sulle aree di competenza DigCompEdu, pensato per orientare la formazione dei docenti dalla competenza allo strumento, anziché dallo strumento alla competenza.

Parole chiave: *intelligenza artificiale generativa; large language model; AI literacy; formazione docenti; DigCompEdu*

1 Introduzione

C'è un momento, nei miei corsi di formazione sull'intelligenza artificiale per docenti, in cui si ripete sempre la stessa cosa. Mostro uno strumento nuovo, l'ultima app, l'ultimo assistente o l'ultima funzione appena rilasciata, e vedo negli occhi dei partecipanti accendersi l'entusiasmo per la novità. Si prova, si sperimenta, ci si diverte e si scattano screenshot da condividere nel gruppo WhatsApp dei colleghi. Poi, una volta tornati in classe, quegli stessi insegnanti fanno tendenzialmente ciò che facevano prima.

Non per pigrizia e nemmeno per resistenza al cambiamento. La maggior parte di loro è tutt'altro che refrattaria alla tecnologia. Il problema è un altro, più sottile: hanno imparato a utilizzare uno strumento, ma non hanno avuto l'occasione di chiedersi quale problema didattico quello strumento dovrebbe contribuire a risolvere. Hanno un'app a disposizione, ma nessuna idea precisa di come valorizzarla.

Questa osservazione nasce da un anno di attività formativa sul campo: una decina di corsi, alcuni dei quali giunti alla seconda edizione con contenuti aggiornati, per una media di circa cinquanta docenti a edizione, provenienti da istituti di ogni ordine e grado, dalla scuola dell'infanzia alla secondaria di secondo grado. Non si tratta di una rilevazione sistematica in senso statistico, ma di un'osservazione ripetuta con sufficiente costanza da poter essere considerata un pattern, più che un'impressione isolata: l'attenzione riservata al metodo, più che allo strumento, tende a diminuire progressivamente salendo lungo i gradi scolastici. I docenti della scuola dell'infanzia e della primaria, meno esposti alla pressione della programmazione disciplinare rigida, si sono mostrati in media più disponibili ad accogliere un'impalcatura progettuale complessiva prima di sperimentare con l'IA in classe; nella

secondaria, e in particolare nel secondo grado, la richiesta più ricorrente è stata invece quella di indicazioni operative immediate – «cosa posso fare da domani» – con una resistenza maggiore, in alcuni casi esplicita, verso l’articolazione metodologica proposta, percepita come un passaggio che sottrae tempo a un’urgenza didattica già sentita come pressante.

Non è un’osservazione originale in senso assoluto dato che chiunque si occupi di innovazione tecnologica nelle organizzazioni la riconosce, sotto altro nome, come uno degli errori più comuni e più costosi: l’attrazione verso lo strumento come sostituto della progettazione. Ma nella scuola questo errore ha una gravità particolare, perché il prodotto di cui parliamo non è un report aziendale o un processo produttivo: è la crescita cognitiva di una persona. Un’app usata senza un’idea didattica precisa non produce nella migliore delle ipotesi nulla, e nella peggiore produce l’illusione dell’innovazione con la sensazione di aver fatto qualcosa di nuovo, mentre in realtà si è solo cambiato lo strumento con cui si fa la stessa cosa di sempre.

Questo lavoro sostiene una tesi precisa: prima di scegliere uno strumento di intelligenza artificiale, un docente dovrebbe avere già chiara un’idea derivata da un bisogno didattico reale, radicato nel proprio contesto, nella propria classe, nelle competenze che intende sviluppare. È un’idea che deve nascere prima del prompt, non dopo. E perché quell’idea possa nascere in modo consapevole, non ingenuo, non timoroso, non affidato al caso – è necessario un secondo ingrediente, che nella formazione docenti viene troppo spesso saltato: capire, almeno nei suoi tratti essenziali, come funziona davvero un sistema di intelligenza artificiale generativa per non cadere nelle trappole cognitive che questi sistemi tendono a costruire, per loro stessa natura, attorno a chi li usa senza conoscerli. Un’attività da me condotta in diverse occasioni di formazione docenti aiuta a rendere percepibile, in modo quasi fisico, un fenomeno che altrimenti resterebbe puramente astratto: il modo in cui un modello linguistico distribuisce la propria attenzione lungo un prompt, privilegiando sistematicamente le informazioni collocate all’inizio e alla fine, a scapito di quelle centrali. Ai partecipanti ho chiesto di disporsi in fila, ciascuno con in mano un foglietto recante una singola informazione una condizione, un vincolo, un dato, che nel suo insieme comporrrebbe un’istruzione articolata, del tipo che un docente potrebbe scrivere per ottenere una scheda didattica personalizzata: livello della classe, argomento, durata dell’attività, stile di linguaggio richiesto, vincoli di valutazione, formato di output atteso. Ogni partecipante riceve il compito di trasmettere al successivo, in pochi secondi e senza rileggere l’intero elenco, solo ciò che ricorda di aver ricevuto. Alla fine della fila, si chiede al gruppo di ricostruire collettivamente l’istruzione originaria.

Il risultato è pressoché costante, e ha un effetto comico che aiuta non poco a fissarlo in memoria: le informazioni di apertura e di chiusura sopravvivono quasi intatte, mentre quelle centrali, e spesso proprio le più specifiche, come il vincolo di valutazione o il livello della classe, si perdono, si confondono, vengono in alcuni casi sostituite da dettagli inventati per colmare il vuoto. I docenti, ridendo del proprio stesso malfunzionamento collettivo, arrivano da soli alla conclusione che l’esercizio intende suggerire: un sistema che elabora informazione in sequenza tende naturalmente a pesare in modo diseguale le diverse parti di un testo lungo, e questo non è un errore accidentale ma una caratteristica strutturale del meccanismo, nel caso dei modelli linguistici, dell’architettura di attenzione che ne governa il funzionamento.

Il fenomeno che l’esercizio rende tangibile è documentato empiricamente in letteratura: analizzando la capacità dei modelli linguistici di recuperare informazioni rilevanti collocate in punti diversi di un contesto lungo, uno studio ha mostrato che le prestazioni sono sistematicamente migliori quando l’informazione si trova all’inizio o alla fine dell’input, e degradano quando si trova al centro; un pattern osservato con costanza anche nei modelli esplicitamente progettati per gestire contesti estesi (Liu et al., 2024).

È solo un esempio tra un ventaglio di possibilità, ma illustra bene il senso di questo lavoro: comprendere come un sistema elabora davvero le informazioni, non se *capisce* o *pensa* nel senso in cui lo facciamo noi, ma come distribuisce concretamente il proprio calcolo tra le parole che riceve e permette di evitare errori che altrimenti vengono scambiati per stranezze imprevedibili dello strumento, quando sono in realtà comportamenti regolari e prevedibili di un sistema che elabora i dati in un modo strutturalmente diverso da come fa una mente umana. Non diverso perché più o meno intelligente in senso antropologico, ma diverso nel meccanismo stesso di elaborazione: un modello linguistico non recupera un ricordo, genera una sequenza di parole plausibile calcolando probabilità; non ha una gerarchia di importanza tra le informazioni che riceve se non quella che deriva dalla loro posizione e dalla loro ripetizione nel testo; non possiede una memoria di ciò che ha detto un minuto prima, se non quella che l’utente stesso reintroduce nella conversazione o nelle istruzioni.

Il percorso di questo lavoro procede dunque su un doppio binario che si intreccia costantemente. Da un lato, la ricostruzione tecnica necessariamente sintetica, seppur rigorosa, di alcuni meccanismi di funzionamento dei modelli linguistici di grandi dimensioni: come nascono le rappresentazioni

vettoriali delle parole, come funziona il meccanismo di attenzione, perché la generazione è un processo probabilistico e non un recupero di informazioni verificate, perché l'allucinazione non è un guasto occasionale ma una conseguenza strutturale di questo stesso meccanismo. Dall'altro lato, la traduzione pedagogica di questa comprensione tecnica: come cambia la progettazione didattica, come cambia la valutazione, come cambia il ruolo stesso del docente, quando si smette di trattare l'IA come una scatola magica onnisciente o infallibile, a seconda dei momenti, e la si tratta per quello che è: uno strumento potente ma parziale, che restituisce ciò che gli si chiede nel modo in cui gliela si chiede, e che per questo richiede, più di ogni strumento precedente, un pensiero che lo preceda.

2 Come funziona un LLM: dai vettori numerici al comportamento emergente

La fluidità con cui un modello linguistico risponde induce facilmente a proiettarvi un'intenzionalità che non possiede. Ciò che osserviamo come conversazione è, al livello sottostante, una sequenza di operazioni algebriche su vettori numerici: non c'è comprensione così come la intendiamo per un interlocutore umano, ma un processo che, proprio per la sua rigorosa regolarità matematica, produce un comportamento sorprendentemente simile a quello di un ragionamento. Per capire come questo avvenga occorre partire dal livello più elementare della rappresentazione. Ogni parola, o frammento di parola – token, nel lessico tecnico – viene tradotta in un vettore numerico, un embedding, collocato in uno spazio ad altissima dimensionalità. Parole che ricorrono in contesti simili finiscono in posizioni vicine di questo spazio: il significato, in questa architettura, non è definito ma relazionale, emerge dalla posizione reciproca dei vettori. È qui che si annida il primo equivoco su cui vale la pena insistere con i docenti: il modello *non sa* cosa significhi una parola, ne codifica la distribuzione d'uso.

Su questo substrato geometrico opera il meccanismo che ha reso possibile il salto qualitativo degli ultimi anni. L'architettura Transformer ha superato il limite delle reti sequenziali precedenti – RNN, LSTM – introducendo il meccanismo di attenzione: il modello pesa simultaneamente ogni parola della frase in relazione a tutte le altre, indipendentemente dalla distanza, per stabilire cosa sia rilevante ai fini dell'interpretazione. È questo meccanismo, non una qualche forma di comprensione olistica, a permettere al modello di gestire relazioni linguistiche complesse, ambiguità e riferimenti a lungo raggio.

Il punto più contro-intuitivo, e più utile in chiave formativa, è che la risposta non preesiste alla domanda: a ogni passo il modello calcola una distribuzione di probabilità sul token successivo, condizionata da tutto ciò che precede, incluso il prompt, e ne campiona uno. La conoscenza del modello non è dunque un archivio consultabile, ma una capacità di continuazione statisticamente plausibile del testo. Da qui discendono due conseguenze dirette per la didattica: il prompt non è una semplice richiesta ma un co-costruttore del contenuto generato, e l'allucinazione non è un malfunzionamento occasionale, bensì una proprietà strutturale dello stesso meccanismo che produce le risposte corrette in quanto il modello non distingue internamente tra genero un fatto vero e genero un fatto plausibile.

Per rendere tangibile quanto descritto, si consideri un caso tipico di uso scolastico: un docente chiede al sistema un riferimento bibliografico su un argomento di nicchia.

Prompt: *Indicami uno studio del 2019 che dimostri l'efficacia della flipped classroom nella scuola primaria italiana, con titolo, autore e rivista.*

Risposta tipica del modello: *Certamente. Lo studio di riferimento è: Bianchi, M. (2019), 'La flipped classroom nella scuola primaria: uno studio quasi-sperimentale', pubblicato su Giornale Italiano della Ricerca Educativa, vol. 12, n. 2, pp. 45-61.*

Il riferimento ha tutte le caratteristiche formali di una citazione reale con: nome plausibile, rivista esistente, formato corretto, ma può non corrispondere ad alcuna pubblicazione effettiva. Non si tratta di un errore di trascrizione o di un dato recuperato male perché il modello ha generato la sequenza di token statisticamente più coerente con la struttura tipica di una citazione accademica su quel tema, senza alcun meccanismo interno che verifichi l'esistenza referenziale della fonte. Il modello non mente né inventa nel senso intenzionale del termine: semplicemente completa il prompt nel modo più probabile dato il proprio addestramento, e in assenza di un forte segnale statistico verso una fonte reale specifica, il completamento più probabile è comunque uno formalmente corretto.

Si confronti ora lo stesso prompt riformulato:

Prompt corretto: *Esistono studi peer-reviewed sull'efficacia della flipped classroom nella scuola primaria italiana? Se non sei certo dell'esistenza di una fonte specifica, dillo esplicitamente invece di generarne una.*

Questa riformulazione non sblocca un accesso a dati diversi, infatti il modello resta lo stesso, l'addestramento resta lo stesso ma modifica la distribuzione di probabilità sui token successivi, rendendo più probabile una risposta di cautela epistemica. È la dimostrazione più diretta del concetto

di prompt come co-costruttore: la qualità della risposta non dipende solo da quanto sa il modello, ma da come la domanda orienta il processo generativo. L'esempio è particolarmente utile in formazione perché smonta insieme due posizioni: quello del docente tecnico che si limita a insegnare a scrivere un prompt efficace senza spiegare perché funzioni, e quello del docente rinunciatario che scarta lo strumento bollandolo come inaffidabile in blocco. Comprendere che l'allucinazione nasce dalla stessa identica meccanica che produce le risposte utili permette di formare gli insegnanti, e a cascata gli studenti, a un uso critico e non superstizioso: si verifica sempre l'output, non perché il sistema sia difettoso, ma perché la verifica è strutturalmente esterna al suo funzionamento.

Con l'aumento di scala (parametri, dati, calcolo), questi sistemi mostrano capacità emergenti non esplicitamente addestrate come ragionamento step-by-step, transfer tra domini, abbozzi di pianificazione attraverso un fenomeno discusso ad esempio nel report *Sparks of Artificial General Intelligence* (Bubeck et al., Microsoft Research, 2023) a proposito di GPT-4. È importante restituire ai docenti l'idea che si tratta di comportamenti emergenti da un processo statistico su larga scala, non di segnali di una comprensione simile alla nostra: la distinzione è cruciale per calibrare aspettative realistiche, né mitizzate né liquidatorie.

Se la formazione docenti si concentra sull'interfaccia 'come si usa questa app' senza toccare tali meccanismi, resta esposta esattamente ai tre equivoci individuati in apertura del paper: trattare l'IA come un archivio onnisciente porta a fidarsi acriticamente delle risposte; trattarla come un motore di ricerca potenziato porta a non verificarne le fonti (che spesso non esistono come tali); trattarla come un interlocutore comprendente porta a delegarle compiti valutativi o relazionali che richiedono un giudizio che il sistema non possiede. Solo comprendendo che ogni risposta è generata e non recuperata, e che il prompt co-determina l'output, un docente può trasformare l'IA in un dispositivo didattico per esempio insegnando esplicitamente agli studenti a interrogarla, verificarla, correggerla, invece che in uno strumento da subire o da imitare acriticamente.

3 Le differenze IA/mente umana come chiave didattica

Se la macchina non conserva risposte ma le produce nell'istante dell'interrogazione, come argomentato nel paragrafo precedente, il passo successivo consiste nel chiedersi che cosa significhi, in questo contesto, intelligenza. È una domanda che il senso comune tende a eludere, dando per scontato un solo metro di paragone possibile: quello umano. Gran parte della retorica giornalistica e divulgativa sull'IA e, di riflesso, molte rappresentazioni che i docenti si portano in aula, resta ancorata a un'idea implicita di intelligenza come somiglianza: una macchina sarebbe intelligente nella misura in cui riesce a pensare come noi, a possedere coscienza, intenzionalità, comprensione nel senso in cui la intendiamo per un essere umano. Questo criterio, per quanto intuitivo, produce sistematicamente due equivoci speculari, entrambi ricorrenti nell'esperienza diretta di formazione con gli insegnanti: da un lato l'idea dell'IA come cervello elettronico, un'entità onnisciente che custodisce da qualche parte tutto lo scibile e lo restituisce su richiesta; dall'altro, per reazione, la sua immagine di oracolo infallibile, la cui risposta viene accettata con la stessa fiducia acritica che si riserverebbe a un'autorità superiore. Sono due errori opposti che nascono dalla stessa radice: misurare la macchina sul metro dell'uomo, invece che sul proprio.

Un contributo utile per sciogliere questo nodo viene da Nello Cristianini, che propone di abbandonare il criterio dell'imitazione a favore di una definizione operativa: intelligenza come capacità di comportarsi in modo efficace in situazioni nuove (Cristianini 2023). Lo spostamento non è terminologico ma di prospettiva. Se l'intelligenza si misura sulla prestazione in contesti non previsti, e non sulla somiglianza a un processo cognitivo umano, allora un sistema può risultare efficacissimo in un compito circoscritto come riconoscere pattern linguistici, generare testo coerente, individuare correlazioni in enormi quantità di dati senza che questo implichi alcuna forma di comprensione, coscienza o intenzionalità paragonabile alla nostra. È esattamente ciò che serve a un insegnante per orientarsi: non uno strumento onnisciente, né un oracolo, ma un sistema capace di prestazioni molto elevate in domini specifici e altrettanto fragile fuori da essi, tanto più utile quanto meglio se ne conoscono i confini operativi. Portare questa definizione in classe, o meglio nella formazione dei docenti, significa offrire un antidoto concreto a entrambi gli equivoci: non è necessario temere una macchina *che pensa come noi* perché non è questo il criterio con cui va giudicata, né affidarsi ciecamente a una macchina *che sa tutto* perché la sua efficacia è sempre situata, dipendente dai dati e dai contesti di addestramento. Il docente che comprende l'IA come sistema funzionale, e non come simulacro di mente, è nella posizione migliore per insegnare agli studenti a usarla con la stessa lucidità: non per delega del pensiero, ma come strumento le cui prestazioni vanno verificate, non date per fede.

4 Perché serve un metodo, non solo lo strumento

Il ragionamento sviluppato finora resta però incompleto se non si affronta un secondo equivoco, altrettanto diffuso quanto quello descritto nel paragrafo precedente: l'idea che l'adozione dello strumento coincida, di per sé, con la trasformazione della pratica didattica. È una convinzione che ha un preciso corrispettivo nel mondo delle organizzazioni aziendali, dove da tempo si osserva che dotarsi di un modello linguistico avanzato, o distribuire una licenza a tutto il personale, non produce automaticamente valore se non si ripensano i processi che quello strumento dovrebbe attraversare. Nella mia esperienza di formazione, il caso più ricorrente di *foundation first* applicato alla scuola ha spesso lo stesso copione: un istituto acquista una licenza collettiva per una piattaforma di IA, organizza un webinar introduttivo di due ore aperto a tutto il collegio, distribuisce le credenziali di accesso e poi, tre mesi dopo, quella piattaforma risulta utilizzata da una minoranza di docenti, quasi sempre i più propensi alla sperimentazione digitale prima ancora dell'acquisto. Non è mancata la tecnologia, né la buona volontà dell'istituto: è mancata, a monte, la definizione di quale problema didattico specifico quella piattaforma avrebbe dovuto risolvere.

Un rapporto recente sullo stato dell'arte nell'adozione dell'IA nelle organizzazioni descrive la tecnologia come un amplificatore: essa potenzia le realtà già solide nei propri processi e, simmetricamente, mette a nudo le inefficienze di quelle più fragili. La tentazione, di fronte a questa constatazione, è di rovesciare l'ordine delle priorità: prima mettere ordine, costruire l'infrastruttura, definire le procedure, formare capillarmente il personale, e solo dopo introdurre l'IA nella pratica quotidiana. Anche in ambito scolastico questa tentazione è riconoscibile, ed è alla base di molti piani di formazione che puntano a un'alfabetizzazione digitale generale prima di qualunque sperimentazione concreta in classe. Si tratta di un impulso comprensibile, ma che rischia di trasformarsi in un rinvio permanente, perché le fondamenta pensate in astratto senza un problema didattico reale da risolvere sono quasi sempre sbagliate: costruite su ipotesi su cosa potrebbe servire, anziché su ciò che effettivamente serve. Un framework pedagogico condiviso serve esattamente a impedire questa deriva: significa definire, prima ancora di introdurre lo strumento, quali traguardi di apprendimento si intendono perseguire, partendo da un problema didattico specifico anziché dalla piattaforma o dal corso di formazione più esaustivo possibile e costruire modalità di verifica capaci di intercettare il percorso con cui lo studente è arrivato a un risultato, non solo il risultato stesso. Le competenze e le procedure che si rivelano necessarie durante questo percorso vanno consolidate passo dopo passo, e diventano a loro volta il punto di partenza per l'iniziativa successiva.

Per rendere operativo questo capovolgimento è utile riprendere, riformulandoli in chiave scolastica, quattro criteri emersi nella riflessione sull'adozione dell'IA nelle organizzazioni.

Criterio	Traduzione pedagogica	Rischio che previene
Focalizzati, non esaustivi	Partire da un obiettivo di apprendimento circoscritto, una competenza precisa da consolidare, invece di tentare di coprire l'IA nella sua interezza in un solo intervento didattico	Un intervento formativo che vuole dire tutto sull'IA finisce per non insegnare nulla in profondità
Efficacia, non efficienza	Chiedersi non se l'IA renda il lavoro più rapido, ma se renda lo studente davvero più competente	Un compito svolto in metà tempo grazie a un chatbot, che non lascia alcuna traccia di apprendimento
Essenziale, non speculativo	Introdurre solo le funzionalità realmente necessarie all'obiettivo dato	La tentazione, diffusa nella formazione docenti, di mostrare tutto ciò che uno strumento sa fare, indipendentemente dalla pertinenza rispetto al traguardo formativo del momento
Centrato sullo studente, non sullo strumento	Il punto di riferimento è lo studente e il suo processo di apprendimento	Lo spostamento del focus sulle specifiche tecniche dello strumento o sull'entusiasmo per la sua novità

Tabella 1: i quattro criteri pedagogici derivati dai principi aziendali sull'adozione dell'IA

Applicati insieme, questi quattro criteri delineano un principio unico, che è anche la tesi centrale di questo lavoro: l'IA non entra in classe come sostituto di una progettazione didattica, ma può diventarne parte integrante solo a valle di una scelta pedagogica già definita.

A questo si accompagna un secondo elemento, altrettanto rilevante: nessun processo che coinvolga l'IA può fare a meno della supervisione umana. Quanto più uno strumento automatizza compiti, dalla generazione di esercizi alla correzione automatica, tanto più cresce, e non diminuisce, il valore della responsabilità del docente nel validare, correggere e, quando necessario, scartare ciò che il sistema produce. È un principio che vale anche per lo studente, e anticipa direttamente il nodo trattato nel paragrafo seguente: la supervisione umana è ciò che rende affidabile un processo automatizzato nella didattica e la stessa funzione spetta al criterio con cui l'allievo, guidato dal docente, impara a vagliare, e non semplicemente ad accogliere, ciò che l'IA gli restituisce.

5 Il rischio della 'conoscenza senza saggezza'

Nella progressione della conoscenza umana, il sapiente – e con lui l'insegnante – ha svolto una funzione precisa: non semplicemente possedere il sapere, ma saperlo interrogare, verificarne l'attendibilità, condurre l'allievo lungo il sentiero che porta dall'indizio alla conoscenza verificata. È il ruolo che gli umanisti quattrocenteschi incarnarono nel riscoprire i classici greci e latini nelle biblioteche monastiche: non lettori qualsiasi, ma mediatori capaci di decifrare, contestualizzare, distinguere l'autentico dal corrotto.

Questa funzione di mediazione è rimasta sostanzialmente intatta anche nell'epoca di internet, che pure aveva già reso la ricerca della conoscenza più rapida ed economica: restava comunque necessario un metodo per distinguere una fonte affidabile da una fallace, e quel metodo continuava a essere insegnato a scuola. Con l'intelligenza artificiale generativa, invece, il salto è di natura diversa. Non si tratta più di rendere più veloce l'accesso a un sapere che il sapiente doveva comunque vagliare: è la funzione stessa della mediazione a essere messa in discussione, perché il modello restituisce direttamente una sintesi già confezionata, senza che l'utente debba più inerparsi lungo il sentiero che un tempo era il segno distintivo della competenza acquisita.

Un secondo effetto, meno immediato ma altrettanto rilevante per la scuola, riguarda la crisi dell'autorialità. Nella tradizione occidentale un testo resta inseparabile da chi lo firma: la stessa etimologia della parola *autore* rimanda a chi si assume la responsabilità di ciò che ha prodotto, non a chi si limita a metterlo per iscritto (Dal Co, 2025). Un testo generato da un modello linguistico non ha invece nessun auctor in questo senso: non risponde di sé, non può essere chiamato a rendere conto delle proprie affermazioni. È un problema che, in ambito scolastico, ha una ricaduta didattica diretta e spesso trascurata: quando uno studente consegna un elaborato prodotto o assistito dall'IA, viene meno proprio quel legame tra testo e responsabilità che tradizionalmente permetteva al docente di attribuire un giudizio non solo al contenuto, ma a chi lo aveva pensato. Questo mette in crisi, prima ancora degli strumenti valutativi, la domanda che ogni docente si porta dietro da sempre: come si accerta che un allievo abbia davvero imparato qualcosa, se la risposta che consegna può essere stata generata, anziché elaborata?

È una distinzione che riporta la valutazione al centro del processo di apprendimento invece che del suo prodotto finale coerentemente, del resto, con il capovolgimento metodologico proposto nel paragrafo precedente: se il criterio guida non è l'output ma l'efficacia dell'apprendimento reale, allora anche la valutazione deve spostarsi dal testo consegnato al ragionamento che lo sottende. Se si continua a giudicare solo l'elaborato consegnato, si esercita sullo studente la stessa pressione che spinge verso la soluzione più prevedibile e nel caso dell'IA, la risposta statisticamente più probabile anziché verso un percorso di ricerca autenticamente personale.

La stessa riflessione propone una distinzione utile a inquadrare il rischio didattico da un'altra angolazione: l'intelligenza artificiale può funzionare come strumento, il cui uso resta interamente governato da chi lo impiega, oppure come linguaggio, la cui struttura interna condiziona silenziosamente le scelte di chi la utilizza. Il rischio maggiore per l'apprendimento non risiede nell'uso dello strumento in sé, ma nel momento in cui esso smette di essere governato e comincia a orientare il pensiero dello studente. Un elemento che favorisce questo scivolamento è ciò che nella stessa analisi viene definito la seduzione dell'algoritmo: a differenza di un interlocutore umano, un sistema generativo risponde in modo sistematicamente accondiscendente, senza opporre resistenza critica. Questa caratteristica ha un lato produttivo, perché offre un punto di partenza imperfetto invece del foglio bianco (Gargiulo, 2026), e un lato rischioso quando l'uso dello strumento sostituisce il ragionamento anziché avviarlo: lo studente smette di affrontare il problema e si limita a trasferirlo al sistema, senza più rileggere né mettere in discussione ciò che riceve. Uno studente posto di fronte a una risposta già formulata dall'IA non parte da una pagina bianca da riempire, ma da un suggerimento da confermare o correggere, e la seconda condizione richiede uno sforzo critico strutturalmente

maggiore della prima. Se la scuola non insegna esplicitamente a resistere a questa scorciatoia cognitiva, il rischio non è che lo studente smetta di pensare, ma che cominci a pensare sempre a partire da un'ipotesi già data, il che è l'esatto opposto di ciò che la ricerca autonoma dovrebbe allenare. Un'osservazione recente, che ho maturato non in ambito scolastico ma nel mondo del design digitale, aiuta a mettere a fuoco la posta in gioco con particolare nitidezza. Chi progetta strumenti di intelligenza artificiale generativa ha notato che, lasciati senza indicazioni precise, questi sistemi tendono a convergere sempre sulle stesse soluzioni: uno studio citato in questa riflessione ha misurato un effetto di omogeneizzazione contenuto ma statisticamente robusto nei processi di co-creazione tra persone e IA, paragonabile agli effetti di fissazione già noti in psicologia cognitiva, per cui l'esposizione a un esempio riduce la capacità di immaginarne altri. La spiegazione tecnica è la stessa che regge l'intero impianto di questo lavoro: un modello restituisce, con grande efficienza, la versione statisticamente più plausibile di ciò che gli viene richiesto, e diventa proporzionalmente meno efficiente quando gli si chiede di scostarsene. Ciò che nel campo del design si traduce in un appiattimento estetico, nel campo dell'apprendimento si traduce in un appiattimento cognitivo: uno studente che consegna la prima risposta plausibile ottenuta dal sistema, senza avere già maturato un'idea propria da verificare o da contrapporre, non sta delegando un'esecuzione sta rinunciando a formulare il pensiero che avrebbe dovuto precederla. Il punto più rilevante, ai fini didattici, è che questo tipo di giudizio critico con la capacità di riconoscere quando una risposta non è ancora quella giusta, di scartarla e di chiederne un'altra non è innato, ma si costruisce con l'esercizio diretto e, non meno importante, per contagio: osservando come ragionano gli altri, ricevendo obiezioni su un lavoro ancora imperfetto, discutendo un'alternativa prima di scioglierla. Se l'allievo interpone sistematicamente l'IA tra sé e la fatica del primo tentativo sbagliato, rischia di non accumulare mai quel repertorio di errori attraverso cui il giudizio normalmente si allena. Questo mette in crisi, prima ancora degli strumenti valutativi, la domanda che ogni docente si porta dietro da sempre: come si accerta che un allievo abbia davvero imparato qualcosa, se la risposta che consegna può essere stata generata, anziché elaborata? È una domanda che molti insegnanti vivono con smarrimento, ed è comprensibile: gli strumenti di verifica tradizionali quali il compito scritto, l'elaborato a casa, la relazione, presuppongono tutti che il testo prodotto sia un indizio affidabile del processo cognitivo che lo ha generato. Con l'IA generativa questa presunzione salta, e non è recuperabile vietando lo strumento, perché il divieto è aggirabile e, soprattutto, rinuncia a insegnare l'unica competenza che davvero conterà nei prossimi anni: usare l'IA con criterio.

Una possibile via d'uscita, che non è una soluzione definitiva, ma un primo spostamento del problema nella giusta direzione, consiste nel rovesciare l'oggetto della valutazione: non più (solo) il testo consegnato, ma il percorso che lo ha prodotto. Chiedere a uno studente che cosa ha domandato al sistema, perché ha formulato la domanda in quel modo, come ha giudicato la risposta ottenuta, quali parti ha corretto o scartato e sulla base di quale criterio, permette di distinguere con relativa chiarezza chi ha usato l'IA come amplificatore del proprio ragionamento da chi l'ha usata come scorciatoia per evitarlo. È una distinzione che riporta la valutazione al centro del processo di apprendimento invece che del suo prodotto finale coerentemente, del resto, con il capovolgimento metodologico proposto nel paragrafo precedente: se il criterio guida non è l'output ma l'efficacia dell'apprendimento reale, allora anche la valutazione deve spostarsi dal testo consegnato al ragionamento che lo sottende.

Resta, tuttavia, un interrogativo più radicale, che la sola tecnica valutativa non esaurisce e che vale la pena lasciare aperto in questa sede, perché tocca il senso stesso dell'operazione educativa: se la conoscenza diventa sempre più disponibile, istantanea e a basso costo cognitivo, che cosa garantisce che cresca insieme a essa anche la capacità di usarla bene, quella che da sempre chiamiamo saggezza? È una domanda che, nella riflessione più ampia sull'IA e l'educazione, resta senza risposta rassicurante: l'accumulo di sapere disponibile non implica automaticamente un progresso nella capacità di giudizio, e anzi rischia di allargare il divario tra chi possiede gli strumenti critici per interrogare l'IA con discernimento e chi, non possedendoli, ne diventa semplicemente dipendente. È precisamente in questo spazio tra l'abbondanza di risposte e la scarsità di criteri per valutarle che si gioca oggi il compito della scuola, ed è per questo che la sezione propositiva di questo lavoro, nel paragrafo successivo, prova a tradurre l'urgenza descritta fin qui in indicazioni operative per la formazione dei docenti.

6 Il framework MISE per la formazione docenti

- (1) L'intreccio tra competenza tecnica, competenza disciplinare e competenza pedagogica non è una novità introdotta dall'IA, ma è precisamente ciò che il modello TPACK descrive da quasi vent'anni come nucleo della professionalità docente in contesti tecnologicamente mediati: la conoscenza tecnologica da sola, priva di un'integrazione con la conoscenza pedagogica e disciplinare, non produce alcun miglioramento didattico automatico (Mishra e Koehler,

2006). Il *competence-first* qui proposto può essere letto come un'applicazione di questo principio al caso specifico dell'IA generativa. Le sezioni precedenti hanno mostrato perché una formazione centrata sull'interfaccia sia strutturalmente insufficiente: non avendo compreso il meccanismo generativo, il docente non dispone degli strumenti concettuali per calibrare fiducia, verifica e uso didattico dello strumento.

Si propone qui un framework a quattro fasi, pensato per invertire l'ordine con cui solitamente si progetta la formazione – dallo strumento alla didattica – riportandolo all'ordine inverso: dalla competenza da sviluppare allo strumento più coerente.

Fase	Denominazione	Domanda guida	Contenuto della fase	Area DigCompEdu corrispondente
1	Meccanismo	Come funziona davvero lo strumento?	Alfabetizzazione tecnico-epistemologica minima: generazione vs. recupero, ruolo del prompt, natura strutturale dell'allucinazione. Non formazione all'uso dell'app, ma al suo funzionamento. La <i>letteratura di base</i> della disciplina digitale.	Propedeutica, trasversale a tutte le aree.
2	Idea	Quale competenza voglio far emergere?	Il punto di partenza è la competenza disciplinare o trasversale da sviluppare (argomentazione, sintesi, pensiero critico, scrittura creativa), non lo strumento. Solo dopo si valuta se e quale IA sia coerente con l'obiettivo.	Insegnamento e apprendimento (preceduta intenzionalmente dalla scelta pedagogica)
3	Sperimentazione	Come faccio esperire il meccanismo agli studenti?	L'attività non usa l'IA solo per produrre un output, ma è progettata perché gli studenti osservino il meccanismo in azione: variazione del prompt, verifica delle fonti, confronto tra output.	Facilitare la competenza digitale degli studenti
4	Estensione	Chi resta responsabile del risultato?	Chiusura del ciclo con riflessione metacognitiva guidata: il modello	Valutazione ed empowerment degli studenti

			genera, la validazione resta un atto umano non delegabile. Principio di controllo umano già formalizzato in altri contesti di governance dell'IA, qui trasposto come garanzia che lo strumento rafforzi, non eroda, il giudizio critico.	
--	--	--	--	--

Tabella 2: il framework MISE (Meccanismo, Idea, Sperimentazione, Estensione) e la sua corrispondenza con le aree di competenza DigCompEdu

Un'indicazione che la letteratura più recente conferma con chiarezza è che l'alfabetizzazione sull'intelligenza artificiale non può ridursi a un intervento episodico: uno studio dedicato allo sviluppo di questa competenza nelle organizzazioni sottolinea che va costruita su più livelli – dai vertici decisionali al personale operativo – e integrata stabilmente nella programmazione, non trattata come un corso una tantum (Benlian, Pinski). Trasposta alla scuola, l'indicazione conferma l'impianto proposto in questo lavoro: un framework realmente *competence-first* non può esaurirsi in un singolo momento formativo rivolto ai docenti, ma richiede un percorso che coinvolga progressivamente la dirigenza, il collegio docenti e, infine, la classe stessa ciascun livello con un proprio grado di approfondimento tecnico e una propria responsabilità nell'adozione.

7 Conclusione

Se c'è una convinzione che attraversa l'intero percorso di questo lavoro, è che la vera competenza digitale richiesta oggi a un docente non consista nel padroneggiare un numero crescente di strumenti, ma nel possedere un metodo capace di sopravvivere al ricambio continuo di quegli strumenti. Le app cambiano nome ogni sei mesi, i modelli linguistici vengono aggiornati con una frequenza che rende obsoleto qualunque manuale d'uso nel giro di una stagione scolastica; ciò che invece resta stabile, e che vale la pena insegnare, è la comprensione di come questi sistemi elaborano l'informazione, e la disciplina di partire sempre da un'idea didattica precisa prima di aprire una finestra di dialogo con l'IA.

Questo tipo di consapevolezza tecnica non è un lusso per addetti ai lavori: è la condizione minima perché un docente possa insegnare ai propri studenti a usare l'IA con criterio, invece che a subirla con fiducia cieca o a rifiutarla per timore. Non si può insegnare a riconoscere una trappola che non si conosce. Le trappole, del resto, sono più d'una, e vale la pena richiamarle in chiusura come una sorta di piccolo inventario di consapevolezza minima.

- **Trappola della posizione**, di cui si è detto: l'informazione che si perde nel mezzo di un prompt affollato.
- **Trappola della fiducia con la tendenza dei modelli a rispondere sempre con lo stesso tono sicuro**, sia che la risposta sia corretta, sia che sia completamente inventata: un'allucinazione pronunciata con la stessa autorevolezza di un fatto verificato, perché il sistema non possiede alcun meccanismo interno per distinguere ciò che sa da ciò che sta, semplicemente, generando come plausibile.
- **Trappola della compiacenza**, quella per cui un modello tende ad assecondare il tono e le premesse di chi lo interroga, restituendo conferme più che obiezioni – un rischio non da poco quando si chiede a uno studente di usare l'IA come *verifica* di un proprio ragionamento, e l'IA si limita a rimandargli indietro un'eco rassicurante.
- **Trappola della prima risposta**, la tentazione di accontentarsi del primo output plausibile, che è statisticamente la risposta più probabile e prevedibile, non la più originale: la scorciatoia più comoda è, quasi per definizione, anche la più banale.
- **Trappola dell'antropomorfizzazione**, non ultima, forse la più insidiosa perché la più naturale: la fluidità del linguaggio con cui questi sistemi rispondono induce quasi automaticamente ad attribuire loro intenzioni, stati d'animo, comprensione – mentre ciò che

sta accadendo, a livello matematico, è il calcolo della parola successiva più probabile data la sequenza precedente.

Nessuna di queste insidie si disinnescava vietando l'uso dello strumento, né si disinnescava da sola con l'entusiasmo dello smanettone di turno. È per colmare questo spazio che questo lavoro propone uno studio profondo e un metodo, un percorso che parte dalla comprensione tecnica, passa per la costruzione di un'idea didattica autenticamente radicata nei bisogni della classe, e approda a una pratica valutativa capace di guardare al processo di apprendimento e non solo al prodotto consegnato. È un metodo pensato per restare valido anche quando lo strumento con cui lo si applica sarà, inevitabilmente, sostituito da un altro.

Bibliografia

Arte, M. (2026, 6 agosto). *L'AI da sola non trasforma un'azienda: serve un metodo*, Agenda Digitale.

Benlian, A., & Pinski, M. (2025, ottobre). *The AI literacy development canvas: Assessing and building AI literacy in organizations*. ScienceDirect.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4*.

Cristianini, N. (2023). *La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna, Il Mulino.

Dal Co, M. *La conoscenza senza saggezza: l'educazione al tempo dell'AI*, Agenda Digitale.

Dangol, A., Wolfe, R., Zhao, R., Kim, J., Ramanan, T., Davis, K., & Kientz, J. A. (2025). *Children's Mental Models of AI Reasoning: Implications for AI Literacy Education*, Proceedings of the Interaction Design and Children Conference (IDC '25), 106–123.

Fullan, M. (2015). *The New Meaning of Educational Change (5th ed.)*, New York, Routledge.

Gargiulo, M. (2026, 6 agosto). *Creatività aziendale e AI: processi, strumenti e linguaggio*, Agenda Digitale.

Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). *Lost in the Middle: How Language Models Use Long Contexts*, Transactions of the Association for Computational Linguistics, 12, 157–173.

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*, Parigi, UNESCO.
<https://doi.org/10.54675/ZJTE2084>.

Mishra, P., & Koehler, M. J. (2006). *Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge*, Teachers College Record, 108(6), 1017-1054.

Newman-Griffis, D. (2025). *AI Thinking: a framework for rethinking artificial intelligence in practice*, Royal Society Open Science, 12(1), 241482.

Parrott, N. (2026, 24 luglio). *Claude design: tutti creativi grazie all'AI, ma tutti con la stessa estetica*, Blog Anthropic.

Redecker, C. (2017). *European Framework for the Digital Competence of Educators: DigCompEdu*, Lussemburgo, Publications Office of the European Union. <https://doi.org/10.2760/178382>.

Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio (AI Act), 13 giugno 2024.

Selwyn, N. (2019). *Should Robots Replace Teachers? AI and the Future of Education*, Cambridge, Polity Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*, Advances in Neural Information Processing Systems, 30.