

# PROGETTARE UN TUTOR POTENZIATO DA INTELLIGENZA ARTIFICIALE PER LA DIDATTICA: IL CASO STUDIO ASKLEA

## DESIGNING AN ARTIFICIAL INTELLIGENCE-ENHANCED TUTOR FOR EDUCATION: THE ASKLEA CASE STUDY

▪ Lorenzo Cesaretti<sup>1</sup> - Federica Marchesini<sup>2</sup> - Pietro Monari<sup>3</sup>

### RIASSUNTO

Il presente articolo propone un *framework* metodologico per la progettazione e validazione pedagogica di sistemi di *tutoring* potenziati da intelligenza artificiale, affrontando la questione cruciale dell'allineamento tra principi pedagogici scientificamente validati e comportamento effettivo di tali sistemi. Attraverso un'articolata metodologia in quattro fasi, che integra design pedagogico, implementazione tecnologica e valuta-

<sup>1</sup> È cofondatore e responsabile tecnico-scientifico di Talent S.r.l. È il responsabile dello sviluppo delle piattaforme *weturtle.org* (formazione docenti) e *askLea.ai* (tutor potenziato da intelligenza artificiale a supporto di studenti e docenti). Nel 2020 ha conseguito il Dottorato di Ricerca in Ingegneria dell'Informazione presso l'Università Politecnica delle Marche. Collabora con Università di Camerino, Università Europea di Roma, Università Politecnica delle Marche e Pontificia Facoltà di Scienze dell'Educazione «Auxilium» di Roma. È autore di numerose pubblicazioni scientifiche in ambito nazionale e internazionale.

<sup>2</sup> Educatrice e pedagogista specializzata in didattica digitale e metodologie innovative. Dal 2022 formatrice per Talent S.r.l., dove si occupa di progettare e realizzare percorsi didattici STEAM per studenti e docenti, con l'obiettivo di incrementare la consapevolezza nell'utilizzo multidisciplinare delle nuove tecnologie per la didattica.

<sup>3</sup> È un fisico e musicista con una lunga esperienza come insegnante di matematica e fisica nelle scuole superiori. Negli ultimi sette anni, si è dedicato all'educazione riguardante l'intelligenza artificiale e all'impiego della stessa in ambito pedagogico. Ha condotto diverse sperimentazioni nell'educazione interdisciplinare STEAM e, per *Ammagamma*, parte del gruppo *Accenture*, è stato ideatore e promotore di progetti quali "Educare a Pensare: Umanità dell'IA" e *Lucy*, il primo percorso curricolare di IA per la scuola secondaria di I grado in Italia, realizzato in collaborazione con l'I.C. 3 Mattarella di Modena. Attualmente, collabora con Talent S.r.l. nelle sperimentazioni del progetto *askLea* nelle scuole, ricoprendo il ruolo di AI Learning & Innovation Strategist.

zione sistematica, viene presentato il caso studio della piattaforma askLea, con particolare attenzione alla funzione “Ripassiamo un argomento”. L’innovativa triangolazione tra valutazione esperta e analisi automatizzata mediante *Large Language Model* dimostra la possibilità di implementare processi di verifica dell’allineamento pedagogico su larga scala. I risultati evidenziano come l’approccio proposto costituisca un avanzamento significativo rispetto ai modelli valutativi generalisti, offrendo una prospettiva granulare e contestualizzata. Si sottolinea inoltre l’importanza di considerare l’allineamento pedagogico come prerequisito fondamentale per qualsiasi sperimentazione educativa che coinvolga sistemi di intelligenza artificiale, preservando un paradigma autenticamente *human-centered*.

## PAROLE CHIAVE

Intelligenza Artificiale, allineamento pedagogico, tutor IA, *Large Language Models* a scuola.

## ABSTRACT

This article proposes a methodological framework for the design and pedagogical validation of AI-enhanced tutoring systems, addressing the crucial issue of alignment between scientifically validated pedagogical principles and the actual behavior of such systems. Through a structured four-phase methodology that integrates pedagogical design, technological implementation, and systematic evaluation, the paper presents the case study of the askLea platform, with particular attention to the “Let’s Review a Topic” function. The innovative triangulation between expert evaluation and automated analysis using Large Language Models demonstrates the possibility of implementing pedagogical alignment verification processes at scale. The results highlight how the proposed approach constitutes a significant advancement compared to generalist evaluation models, offering a granular and contextualized perspective. The article also emphasizes the importance of considering pedagogical alignment as a fundamental prerequisite for any educational experimentation involving artificial intelligence systems, preserving an authentically human-centered paradigm.

## KEYWORDS

Artificial Intelligence, pedagogical alignment, AI tutors, Large Language Models at school.

## RESUMEN

El presente artículo propone un *framework* metodológico para el diseño y validación pedagógica de sistemas de *tutoring* potenciados por inteligencia artificial, abordando la cuestión crucial del alineamiento entre principios pedagógicos científicamente validados y el comportamiento efectivo de tales sistemas. A través de una articulada metodología en cuatro fases, que integra diseño pedagógico, implementación tecnológica y evaluación sistemática, se presenta el caso de estudio de la plataforma askLea, con particular atención a la función “Repasemos un tema”. La innovadora triangulación entre evaluación experta y análisis automatizado mediante *Large Language Model* demuestra la posibilidad de implementar procesos de verificación del alineamiento pedagógico a gran escala. Los resultados evidencian cómo el enfoque propuesto constituye un avance significativo respecto a los modelos evaluativos generalistas, ofreciendo una perspectiva granular y contextualizada. Se subraya además la importancia de considerar el alineamiento pedagógico como prerrequisito fundamental para cualquier experimentación educativa que involucre sistemas de inteligencia artificial, preservando un paradigma auténticamente *human-centered*.

## PALABRAS CLAVE

Inteligencia Artificial, alineamiento pedagógico, tutor IA, Large Language Models en la escuela.

## Introduzione

I *Large Language Models* (LLMs) rappresentano potenzialmente una delle rivoluzioni più significative nel panorama educativo contemporaneo, come sostenuto da diversi esperti del settore.<sup>4</sup> Le prospettive di applicazione di questi modelli generativi in ambito didattico sono molteplici e promettenti.

In contesto educativo, i LLMs potrebbero consentire lo sviluppo di sistemi di *tutoring* intelligenti capaci di:

- fornire feedback personalizzati agli studenti, basati su attività svolte in classe o a casa;
- adattare i contenuti didattici alle specifiche esigenze dello studen-

<sup>4</sup> Cf KHAN Sal, *How AI could save (not destroy) education* [Video]. TED Conferences, 1 may 2023, in [https://www.ted.com/talks/sal\\_khan\\_how\\_ai\\_could\\_save\\_not\\_destroy\\_education](https://www.ted.com/talks/sal_khan_how_ai_could_save_not_destroy_education); MOLICK Ethan R. - MOLICK Lilach, *Instructors as Innovators: A Future-focused Approach to New AI Learning Opportunities, With Prompts*, in *The Wharton School Research Paper* (2024) April 23. Available at [https://papers.ssrn.com/sol3/papers.cfm?Abstract\\_id=4802463](https://papers.ssrn.com/sol3/papers.cfm?Abstract_id=4802463).

te, operando su diverse dimensioni (complessità concettuale, leggibilità, integrazione di elementi visivi e schematici);

- implementare tecniche di *scaffolding*, offrendo un supporto strutturato nella zona di sviluppo prossimale dello studente,<sup>5</sup> fungendo da “impalcatura” cognitiva che permette l’acquisizione progressiva di conoscenze e competenze.

Se da un lato queste potenzialità suscitano notevole interesse verso l’integrazione dell’intelligenza artificiale (IA) in ambito educativo, dall’altro emergono legittime perplessità che meritano un’attenta considerazione. La necessità di validazione scientifica, analisi rigorosa dei risultati e valutazione dell’impatto effettivo dei LLMs sugli apprendimenti diventa sempre più urgente.

In questo articolo intendiamo approfondire due questioni fondamentali: in che misura uno strumento didattico (a supporto dell’insegnamento-apprendimento) basato su LLMs è in grado di rispettare i principi pedagogici stabiliti in fase di progettazione? Come può essere misurato e valutato il grado di aderenza tra il comportamento del tutor IA e i principi selezionati che derivano dalle scienze dell’educazione e dalle scienze cognitive?

La letteratura scientifica offre numerose ricerche sull’impatto di questi strumenti sui processi di apprendimento, ma riteniamo preliminare valutare con precisione quali principi pedagogici dovrebbero guidare lo strumento e, soprattutto, quanto esso sia effettivamente in grado di implementarli. Se un tutor IA non riesce a garantire un output metodologicamente robusto, come possiamo considerare validi i risultati di sperimentazioni basate su tale soluzione? Proponiamo quindi una validazione pedagogica dello strumento come fase propedeutica alla sperimentazione in classe con gli studenti.

Dopo una rassegna della letteratura sul tema, si presenterà una metodologia specificamente progettata per realizzare questo tipo di validazione, illustrando la sua applicazione al caso studio askLea (una piattaforma italiana potenziata da IA pensata per il settore educativo). Si fornirà inoltre un esempio concreto di applicazione della metodologia progettata, per evidenziare l’affidabilità e la robustezza dell’approccio proposto.

## 1. Revisione della letteratura

L’aspirazione alla progettazione di sistemi di *tutoring* potenziati dall’IA, che costituisce il fondamento teorico-metodologico del presente contri-

<sup>5</sup> Cf WASS Rob - HARLAND Tony - MERCER Alison, *Scaffolding critical thinking in the zone of proximal development. Higher Education*, in *Research & Development* 30(2011)3, 317-328, in DOI:10.1080/07294360.2010.489237.

buto, si inserisce in una tradizione di ricerca consolidata che precede significativamente l'emergere dei recenti LLMs. La genealogia epistemologica di questa ricerca affonda le proprie radici negli studi pionieristici sugli *Intelligent Tutoring Systems* (ITS), sistemi sviluppati a partire dagli Anni '70 del secolo scorso, basati su IA e progettati per fornire insegnamento personalizzato agli studenti.

Da questi studi,<sup>6</sup> in cui si pongono a confronto tutoraggio umano e tutoraggio tramite ITS, sono emerse alcune considerazioni che hanno messo in luce criticità e punti di forza di entrambi, riportando una superiorità nei risultati, seppur molto lieve, del tutoraggio umano, per motivazioni che riguardano in modo particolare le modalità di interazione. Nello specifico si evidenzia che l'efficacia del tutoraggio umano fa forza su due punti chiave: la capacità di fornire feedback immediato e la messa in atto di tecniche di *scaffolding* più vicine alle esigenze cognitive ed emotive degli studenti. Gli ITS, soprattutto nelle loro forme meno recenti, forniscono supporto passo dopo passo o a livello di microcompiti, riescono ad adattare e a personalizzare i contenuti in base al livello e agli errori dello studente, ma sono meno accurati nel cogliere le dinamiche relative agli errori, meno efficaci nell'insegnamento di abilità concettuali o aperte e, soprattutto fino all'integrazione dell'elaborazione del linguaggio naturale, non si adattavano con flessibilità all'interazione con lo studente. Anche per questo motivo l'analisi degli ITS ha portato risultati migliori quando impiegati nell'ambito delle discipline STEM,<sup>7</sup> ossia nell'elaborazione di competenze procedurali e problemi ben definiti (matematica, ingegneria, scienza, tecnologia), dando allo studente feedback preimpostati e immediati sulle loro risposte.

Un approccio meno unidirezionale in questo senso è quello adottato dai sistemi DBTS (*Sistemi di Tutoraggio Basati su Dialogo*), che basano l'interazione con lo studente sulla conversazione in una modalità che potrebbe essere definita "socratica", cioè partendo da una domanda e proseguendo con stimoli e altri quesiti andando avanti con l'interazione. Queste analisi si sono concentrate sull'utilizzo dei sistemi IA come strumento per potenziare l'apprendimento, ma la ricerca ha preso in esame anche un altro filone: l'IA come oggetto di studio a scuola per sviluppare consapevolezza nei cittadini del futuro riguardo queste tematiche.<sup>8</sup> Portando in classe attività di *machine learning* e di studio dell'IA, gli studenti possono comprendere che il funzionamento di questi si basa su dati e

<sup>6</sup> Cf VanLEHN Kurt, *The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems*, in *Educational Psychologist* 46(2011)4, 197-221, in DOI:10.1080/00461520.2011.611369.

<sup>7</sup> Cf PALADINES José - RAMIREZ Jaime, *A systematic literature review of intelligent tutoring systems with dialogue in natural language*, in *IEEE Access* 8(2020) 164246-164267, in DOI:10.1109/ACCESS.2020.3021383.

<sup>8</sup> Cf CESARETTI Lorenzo, *Intelligenza artificiale e educazione: un incontro tra due mondi. Rischi e opportunità*, in *Rivista di Scienze dell'Educazione* 59(2021)1, 81-98.

leggi della statistica e riescono a riportarli con consapevolezza agli ambiti della matematica e dell'informatica, senza sconfinare nella "magia".<sup>9</sup>

Questo ci porta anche a considerare alcune ricerche che hanno analizzato l'impatto e gli effetti dell'utilizzo dell'IA da parte degli studenti, quando non supportata da progettazione pedagogica ed educativa. Lo studio di Michael Gerlich<sup>10</sup> analizza gli effetti dell'uso di questi strumenti sulle capacità di pensiero critico, con particolare attenzione al fenomeno del *cognitive offloading* (delegazione di compiti cognitivi a strumenti esterni per alleggerire il carico). Dai dati quantitativi e qualitativi raccolti dagli studenti implicati in questa sperimentazione emerge che esiste una correlazione negativa significativa tra l'uso frequente degli strumenti IA e le abilità di pensiero critico, che gli studenti più giovani tendono a diventare più dipendenti dallo strumento, ma anche che chi ha livelli educativi più alti tende a essere più consapevole degli effetti negativi dell'IA e ad adottare strategie per mantenere un pensiero critico attivo.

A una conclusione affine arriva la sperimentazione messa in atto da Gregor Jošt, Viktor Taneski e Sašo Karakatič<sup>11</sup> su 32 studenti universitari per dieci settimane, con l'obiettivo di esaminare l'uso non regolamentato degli LLMs e la correlazione con le prestazioni degli studenti nella programmazione, in particolare in attività di generazione di codice, *debugging* e richiesta di spiegazioni aggiuntive. I risultati ottenuti tramite le due fasi della sperimentazione (una prima in cui gli studenti avevano piena libertà nell'utilizzo degli LLMs e una seconda in cui dovevano portare a termine l'esercizio senza di essi), hanno portato i ricercatori a concludere che gli studenti che hanno fatto molto affidamento sugli LLMs nel momento iniziale hanno ottenuto poi risultati peggiori nel test senza LLMs, dunque gli LLMs possono essere utili per chiarire concetti, ma affidarsi troppo a essi per generare codice e risolvere *bug* porta a difficoltà nel *problem-solving* autonomo.

L'utilizzo senza regolamentazione o senza supporto pedagogico di un LLM come ChatGPT porta gli studenti a delegare al sistema la soluzione di problemi ed esercizi, influenzando negativamente le capacità di *problem-solving* a lungo termine. Questo è quanto emerge da un'ulteriore ricerca, focalizzata proprio sui risultati di apprendimento degli studenti della scuola superiore, quando supportati da due strumenti differenti: ChatGPT Base e ChatGPT Tutor.<sup>12</sup> Il divario tra i risultati nell'utilizzo dei

<sup>9</sup> Cf BITMARK Inc., *Technology is Not Magic: The Hacker's Point of View - Bitmark Ambassador "bunnie" Huang* in *Medium.com* (16-07-2019), in <https://medium.com/clean-titles/technology-is-not-magic-the-hackers-point-of-view-bitmark-ambassador-bunnie-huang-6750ab772f33>.

<sup>10</sup> Cf GERLICH Michael, *AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking*, in *Societies* 15(2025)1, 6, in DOI: 10.3390/soc15010006.

<sup>11</sup> Cf JOŠT Gregor - TANESKI Viktor - KARAKATIČ Sašo, *The Impact of Large Language Models on Programming Education and Student Learning Outcomes*, in *Applied Sciences* 14(2024)10, 4115, in DOI: 10.3390/app14104115.

<sup>12</sup> Cf BASTANI Hamsa et alii, *Generative AI Can Harm Learning* (July 15, 2024). University of Penn-

due supporti è notevole: gli studenti che hanno usato GPT Base hanno avuto un miglioramento del 48% nei compiti di pratica, ma in fase di esame finale (senza utilizzo di sistemi IA), le loro prestazioni sono peggiorate del 17% rispetto al gruppo di controllo; quelli che avevano lavorato con la versione Tutor hanno migliorato le prestazioni nelle fasi di pratica del 127%, senza compromettere le capacità di apprendimento e senza mostrare cali di prestazione nell'esame finale. Ciò è reso possibile, secondo lo studio, dalla particolare caratteristica di ChatGPT Tutor di non fornire risposte immediate, ma di offrire suggerimenti gradualmente e guidare lo studente nella risoluzione del problema, in modo tale da renderlo partecipe del processo che porta al completamento del compito e mantenere attive le capacità di ragionamento e *problem-solving*.

Sono stati analizzati anche i risultati di uno studio condotto in Italia<sup>13</sup> che ha preso in esame in modo più specifico l'influenza dell'utilizzo di GPT-4 nello svolgimento dei compiti a casa. L'obiettivo principale era verificare se e come questo strumento riuscisse a migliorare i risultati di apprendimento e l'*engagement* degli studenti nello studio a casa dell'inglese come lingua seconda. I membri del gruppo di sperimentazione hanno utilizzato GPT-4 per sessioni di compiti interattivi, mentre il gruppo di controllo ha seguito i compiti tradizionali. La strategia di *prompting* per GPT-4 includeva la fornitura di informazioni sullo scopo dell'esercizio, una descrizione del compito e un esempio di esercizio. I ricercatori riportano come risultati un incremento della motivazione degli studenti nel portare avanti compiti con l'utilizzo del tutor, alcuni miglioramenti nelle competenze su cui vertevano gli esercizi (soprattutto per gli studenti con più difficoltà) e un tasso di "allucinazioni" accettabile per il contesto educativo (inferiore all'1%). Tuttavia, dai questionari è emerso anche che gli studenti del gruppo GPT-4 hanno riportato un calo nella percezione delle proprie capacità e il rischio di fare troppo affidamento sullo strumento, limitando le opportunità di comprensione profonda degli argomenti.

Una risultante frequente dalle varie sperimentazioni è la maggior usabilità dei modelli LLMs per compiti e attività con risposta chiusa o univoca (esercizi di grammatica/matematica). Nell'ambito specifico della matematica, è stato elaborato un nuovo approccio<sup>14</sup> per dedurre dalla conoscenza dei modelli di linguaggio di grandi dimensioni (come GPT-3), modelli più piccoli ed efficienti, con il fine di migliorare la personalizzazione dell'apprendimento attraverso esercizi generati su misura.

---

sylvania - The Wharton School Research Paper, in <http://dx.doi.org/10.2139/ssrn.4895486>.

<sup>13</sup> Cf VANZO Alessandro - PAL CHOWDHURY Sankalan - SACHAN Mrinmaya, *GPT-4 as a homework tutor can improve student engagement and learning outcomes*, ETH Zürich 2024, in DOI:10.48550/arXiv.2409.15981.

<sup>14</sup> Cf LIANG Zhenwen et alii, *Let GPT be a math tutor: Teaching math word problem solvers with customized exercise generation*, Paper (22 May 2023). University of Notre Dame, Georgia Institute of Technology, Allen Institute for AI, in DOI: 10.48550/arXiv.2305.14386.

Il metodo CEMAL (*Customized Exercise for Match Learning*) procede valutando il livello di conoscenza dello studente e proponendo nuovi problemi personalizzati per affrontare in modo più approfondito i punti critici, allenando le competenze meno sviluppate. Anche in questo caso emerge come punto di forza la personalizzazione dell'interazione con la creazione di esercizi *ad hoc* in base alle difficoltà e agli errori dello studente, che ha l'opportunità di incrementare abilità e competenze in modo graduale e progressivo seguendo il proprio ritmo di apprendimento.

Più in generale, gli studi<sup>15</sup> condotti sugli effetti dell'utilizzo di Chat-GPT evidenziano come risultati positivi l'incremento di motivazione e interesse da parte degli studenti e alcuni miglioramenti nelle loro performance. Tuttavia, l'utilizzo del *chatbot*, sembra avere effetti negativi sulla percezione degli studenti riguardo le loro capacità e, diminuendo il carico cognitivo, non offre loro l'opportunità del ragionamento profondo e dello sviluppo del pensiero critico a lungo termine.

Un ulteriore elemento da tenere in considerazione tra le criticità è la disponibilità dei dispositivi, della rete internet o dell'accesso a questi strumenti per tutti gli studenti. In Ghana è stato testato un tutor IA per la matematica<sup>16</sup> che punta in modo particolare a essere accessibile a livello economico e di contesto culturale anche a territori con risorse limitate. *Rori* ha un costo di implementazione di soli \$5 per studente all'anno ed è compatibile con dispositivi a basso livello tecnologico (connessione 3G). Rispetto al gruppo di controllo, gli studenti che hanno avuto la possibilità di lavorare con questo strumento hanno ottenuto risultati notevolmente migliori grazie all'adattabilità del tutor alle loro esigenze personali *step-by-step*. Nonostante la robustezza del disegno sperimentale e i significativi miglioramenti nell'apprendimento, lo studio presenta alcune limitazioni metodologiche rilevanti. Gli autori evidenziano il potenziale effetto Hawthorne, per cui i comportamenti di amministratori e insegnanti nelle scuole sperimentali potrebbero essere stati influenzati dall'aumentata osservazione, benché le scuole della *Rising Academies Network* siano abituate a test e osservazioni di routine. Inoltre, nonostante le molteplici analisi condotte per verificare l'equivalenza iniziale dei due gruppi, potrebbero esistere differenze non osservate nelle capacità pregresse degli studenti o nelle loro caratteristiche socioeconomiche, potenzialmente influenti sui risultati. Un'ulteriore criticità riguarda la progettazione della valutazione: per garantire la comparabilità dei progressi tra i diversi gradi

<sup>15</sup> Cf DENG Ruiqi et alii, *Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies*, in *Computers & Education* 227(2025)105224, in DOI: 10.1016/j.compedu.2024.105224.

<sup>16</sup> Cf HENKEL Owen - HORNE-ROBINSON Hannah - KOZHAKHMETOVA Nessie - LEE Amanda, *Effective and Scalable Math Support: Experimental Evidence on the Impact of an AI-Math Tutor in Ghana*, in *Book of International Conference on Artificial Intelligence in Education* (2024, July), 373-381, in DOI: 10.1007/978-3-031-64315-6\_34.

scolastici, è stato utilizzato lo stesso strumento valutativo per tutti gli studenti. Questo ha comportato che alcuni studenti, particolarmente quelli di livello superiore, abbiano ottenuto punteggi perfetti già nella valutazione iniziale, rendendo impossibile osservare il loro miglioramento. Sebbene questo suggerisca una possibile sottostima dei punteggi di crescita, è anche plausibile che l'efficacia di *Rori* sia limitata a temi matematici relativamente più semplici, e che i vantaggi di apprendimento potrebbero ridursi se gli studenti più avanzati fossero valutati su argomenti più complessi. Mancano inoltre dati longitudinali che consentano di verificare la persistenza dei miglioramenti nel tempo.

Sono degne di sottolineatura anche sperimentazioni che guardano l'utilizzo degli strumenti IA da un altro punto di vista: non direttamente in mano agli studenti, ma come potenziamento per i tutor. In particolar modo uno studio condotto su nove scuole di un distretto statunitense ha messo a confronto gli output di studenti i cui tutor avevano accesso a *Tutor CoPilot* come strumento di supporto nel tutoraggio K-12<sup>17</sup> con quelli degli studenti i cui tutor lavoravano in maniera "tradizionale". Anche in questo caso uno dei risultati positivi emersi è rintracciabile nel basso costo dello strumento a disposizione degli insegnanti/tutor, rispetto a una tradizionale formazione. I tutor che avevano accesso a *Tutor CoPilot* hanno ottenuto risultati migliori rispetto al gruppo di controllo e il divario è particolarmente evidente per i tutor con valutazioni iniziali inferiori. Ciò che ha influito in misura maggiore nell'efficacia dei tutor del gruppo di sperimentazione sono state le seguenti strategie: l'utilizzo di numerose domande guida per stimolare il ragionamento, la richiesta frequente agli studenti di spiegare i propri ragionamenti e la mancanza di risposte dirette fornite senza coinvolgimento attivo dello studente.

L'efficacia di un tutor IA personalizzato per studenti è stata messa a paragone anche con i feedback ottenuti dai docenti di un corso universitario di fisica che utilizzano come metodologia didattica l'apprendimento attivo.<sup>18</sup> In questo caso il fine era quello di valutare il coinvolgimento e la motivazione degli studenti nelle due differenti situazioni (con tutor IA e con docente). Sebbene da una prima analisi sono emersi risultati incoraggianti dal punto di vista della motivazione e dei risultati di apprendimento degli studenti con il tutor IA (il 70% degli studenti ha dichiarato di sentirsi più coinvolto rispetto alla lezione frontale), ci sono alcune criticità che vale la pena di sottolineare. Il tutor IA procede in modo da fornire feedback immediati e dettagliati, il che può ridurre le occasioni di riflessione autonoma sugli errori e portare, nel lungo periodo, a limitare lo sviluppo del

<sup>17</sup> Cf WANG Rose E. et alii, *Tutor CoPilot: A Human-AI Approach for Scaling Real-Time Expertise*, EdWorkingPaper No. 24-1054 (2024), Annenberg Institute for School Reform at Brown University, in <https://eric.ed.gov/?q=1800&ff1=eduMiddle+Schools&id=ED661562>.

<sup>18</sup> Cf KESTIN Gregory et alii, *AI tutoring outperforms active learning*, in *Research Square* (14 May 2024), in <https://www.researchsquare.com/article/rs-4243877/v1>.

pensiero critico e della capacità di risolvere problemi senza assistenza. Inoltre il tutor può comunque generare risposte errate o fuorvianti (fenomeno delle “allucinazioni” dell’IA) e senza un controllo umano, gli studenti potrebbero interiorizzare concetti errati o sviluppare una fiducia eccessiva nelle risposte dell’IA.

Uno step ulteriore in questo senso è stato fatto da *Google DeepMind* con *LearnLM*, un modello di IA specifico per la didattica.<sup>19</sup> Attraverso una sperimentazione su 49 scenari educativi realistici, l’efficacia di *LearnLM*, addestrato su *System Instructions* pedagogiche e con l’integrazione di *Reinforcement Learning from Human Feedback* (RLHF), è stata messa a confronto con quella di GPT-4o (OpenAI), *Claude 3.5 Sonnet* (Anthropic) e *Gemini 1.5 Pro* (base di *LearnLM*).

*LearnLM* è stato preferito agli altri con delle rilevanti percentuali (31%, 11% e 13%) a motivo della maggior vicinanza a un tutor umano nelle interazioni, con una migliore adattabilità alle esigenze dello studente e stimolazione dell’apprendimento attivo, evitando gli output con risposte dirette. Ciò che aiuta lo studente nella metacognizione e la riflessione sui suoi ragionamenti e sulle sue risposte sono le tecniche di “socratic questioning” e di *scaffolding* integrate nel modello, che sostengono l’interazione facendo sì che lo studente sviluppi autonomia e venga frequentemente interrogato su ciò che sta apprendendo. Le istruzioni pedagogiche migliorano il modello da questo punto di vista, ma dai risultati emerge anche che l’aderenza troppo stretta a tali strutture potrebbe limitare la creatività nelle interazioni e far apparire il tono di *LearnLM* troppo formale rispetto a modelli come GPT-4o. Per le stesse caratteristiche, il modello è ottimizzato per il contesto educativo, ma potrebbe essere meno efficace rispetto agli altri LLMs in conversazioni meno didattiche e più generaliste.

## 2. Il caso studio: la progettazione del sistema askLea

La presente sezione illustra la metodologia sviluppata dal team interdisciplinare di Talent S.r.l.,<sup>20</sup> organizzazione italiana specializzata nella progettazione e implementazione di soluzioni formative e piattaforme educative destinate sia a discenti che a personale docente. Tale approccio metodologico ha trovato applicazione nella progettazione e sviluppo di *askLea*,<sup>21</sup> un sistema conversazionale basato su IA specificamente concepito per il contesto educativo, con funzionalità differenziate per docenti e studenti.

<sup>19</sup> Cf LearnLM Team Google: MODI Abhinit et alii, *LearnLM: Improving Gemini for Learning* (25 Dec 2024), in <https://arxiv.org/html/2412.16429v2>.

<sup>20</sup> Cf Talent, in <https://www.talenteducation.it/>.

<sup>21</sup> Cf *askLea*, in <https://www.asklea.ai/>

L'architettura di *askLea* comprende un articolato insieme di funzionalità, organizzate secondo una tassonomia che distingue tra supporto alla professionalità docente e sostegno ai processi di apprendimento degli studenti. Di seguito si presenta l'elenco delle funzioni implementate.

### Funzioni docenti

- Progettiamo una rubrica di valutazione
- Progettiamo una lezione o un percorso didattico
- Progettiamo materiali didattici
- Progettiamo un quiz o una prova di verifica
- Progettiamo un gioco didattico

### Funzioni studenti

#### *Esercizi:*

- Risolviamo un esercizio
- Crea un esempio di esercizio svolto
- Inventa 5 esercizi per me

#### *Ripasso/Studio:*

- Ripassiamo un argomento
- Comprensione del testo

#### *Approfondimento:*

- Un viaggio nella conoscenza
- Rispondi come se fossi
- Dal libro al mondo

#### *Progettazione:*

- Scriviamo un testo
- Facciamo un project work
- Segui la tua passione

A differenza di sistemi conversazionali generalisti, privi di una specifica impostazione pedagogica, le funzioni studenti di *askLea* sono state progettate secondo un paradigma *student-centered*, che preserva la centralità del discente nel processo di apprendimento. Questa impostazione si traduce nell'evitare risposte preconfezionate o soluzioni che potrebbero sostituirsi all'attività cognitiva dello studente. Analogamente, nell'interazione con i docenti, il sistema mantiene una posizione di supporto alla progettualità didattica, preservando la centralità dell'insegnante nei processi decisionali e progettuali inerenti alla pratica educativa.

Per implementare questo paradigma autenticamente *human-centered* nei tutor basati su IA, il team di ricerca e sviluppo ha elaborato una metodologia strutturata in quattro fasi sequenziali e iterative. Tale approccio metodologico si caratterizza per il rigore scientifico e l'integrazione sinergica di competenze pedagogiche e tecnologiche.

### *Fase 1: Design pedagogico fondato su evidenze scientifiche*

La prima fase consiste nell'identificazione sistematica dei principi pedagogici che informeranno la funzionalità del sistema conversazionale. Questa selezione si basa su un'approfondita analisi della letteratura scientifica *evidence-based*, con particolare attenzione agli studi che presentano solide evidenze empiriche in relazione agli specifici obiettivi didattici della funzionalità in oggetto. La deliberata scelta di fondare lo sviluppo su principi pedagogici validati dalla ricerca costituisce un elemento distintivo rispetto ad approcci meramente tecnologici.

### *Fase 2: Integrazione dei principi pedagogici nell'architettura del sistema*

Questa fase prevede la traduzione dei principi pedagogici selezionati in istruzioni efficaci all'interno del *system message*, componente fondamentale dell'architettura di interazione con il modello linguistico. Il *system message* può essere definito come «un insieme di istruzioni o contesti specifici per una determinata funzione, forniti a un modello di IA generativa (ad esempio, GPT-4o, Claude Sonnet 3.7, ecc.) per orientare e migliorare la qualità e la sicurezza delle risposte del modello. Questo è utile in situazioni che richiedono determinati livelli di formalità, linguaggio tecnico o termini specifici di settore».<sup>22</sup> La progettazione di tale componente richiede l'applicazione di tecniche avanzate di *prompt engineering*<sup>23</sup> finalizzate all'allineamento del comportamento del sistema con i principi pedagogici identificati.

### *Fase 3: Validazione interna tramite simulazioni controllate*

Prima dell'implementazione in contesti autentici, il sistema viene sottoposto a un rigoroso protocollo di test da parte di esperti in pedagogia. Questi test simulano un'ampia gamma di potenziali interazioni, considerando diversi profili di utenza e vari scenari didattici. Le evidenze raccolte durante questa fase possono determinare iterazioni successive di affinamento del *system message*, in un processo ciclico di continuo miglioramento fino al raggiungimento di standard qualitativi predefiniti. Solo quando si raggiunge un livello soddisfacente di comportamento del sistema in ambiente controllato, si procede alla sperimentazione con utenti reali.

<sup>22</sup> Safety system messages, in <https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/system-message>. Traduzione nostra.

<sup>23</sup> Cf MARVIN Ggaliwango - NAKAYIZA Hellen Raudha - JJINGO Daudi - NAKATUMBA-NABENDE Joyce, *Prompt Engineering in Large Language Models* (January 2024), in JACOB I.J. - PIRAMUTHU S. - FALKOWSKI-GILSKI P. (eds), *Data Intelligence and Cognitive Informatics. ICDICI 2023. Algorithms for Intelligent Systems*, Singapore, Springer Nature Singapore 2024, 387-402, in DOI: 10.1007/978-981-99-7962-2\_30.

#### *Fase 4: Valutazione sistematica in contesti autentici*

L'ultima fase prevede un articolato processo di valutazione dell'efficacia pedagogica del sistema in contesti di utilizzo reale, strutturato nelle seguenti sotto-fasi:

- a) *co-progettazione di una rubrica valutativa*: sviluppo collaborativo, con il supporto di LLMs, di una rubrica strutturata per valutare l'aderenza del comportamento del sistema ai principi pedagogici selezionati;
- b) *validazione della rubrica*: sottomissione dello strumento valutativo a un panel di esperti in pedagogia per verificarne validità, affidabilità e applicabilità;
- c) *analisi automatizzata*: applicazione della rubrica validata a un campione significativo di N conversazioni reali tra studenti e sistema, mediante un LLM opportunamente configurato;
- d) *analisi esperta*: valutazione parallela di un sottocampione di M (con M molto minore di N) conversazioni da parte di esperti in pedagogia, utilizzando la medesima rubrica;
- e) *triangolazione metodologica*: confronto sistematico tra le valutazioni generate dal sistema automatizzato e quelle prodotte dagli esperti, finalizzato alla validazione del processo valutativo automatizzato;
- f) *scaling dell'analisi*: in caso di concordanza significativa tra le valutazioni automatizzate e quelle esperte, estensione dell'analisi all'intero corpus di interazioni raccolte, consentendo una valutazione complessiva delle performance del sistema di *tutoring*.

Questo approccio metodologico in quattro fasi costituisce un avanzamento nella progettazione di sistemi di *tutoring* basati su IA, integrando rigorosamente fondamenti pedagogici, validazione empirica e valutazione sistematica in un *framework* coerente e replicabile.

La quarta fase del *framework* metodologico proposto rappresenta un'evoluzione rispetto all'approccio delineato in "Towards Responsible Development of Generative AI for Education: An Evaluation-Driven Approach" da Irina Jurenka et al.<sup>24</sup> Il contributo innovativo della metodologia qui proposta risiede nella granularità e nella specificità contestuale degli strumenti valutativi impiegati per l'analisi dell'allineamento pedagogico del *chatbot*.

Mentre il modello proposto da Jurenka et al. adotta un approccio valutativo monolitico, basato sull'applicazione di una singola rubrica universale a tutte le funzionalità del sistema di *tutoring* indipendentemente

<sup>24</sup> Cf JURENKA Irina et alii, *Towards responsible development of generative AI for education: An evaluation-driven approach* (May 2024), in <https://arxiv.org/abs/2407.12687>.

dai contesti d'uso e dagli obiettivi pedagogici specifici, l'approccio sopra descritto propone una strategia valutativa differenziata. Tale strategia prevede la progettazione e validazione di rubriche valutative distinte e specifiche per ciascuna funzione del sistema, calibrate rispetto ai principi pedagogici selezionati durante la fase iniziale di design.

Questa differenziazione degli strumenti valutativi consente di superare le limitazioni intrinseche di un approccio generalista, in particolare favorisce:

1. *il superamento della generalizzazione inappropriata*: l'utilizzo di dimensioni educative generiche rischia di non cogliere le specificità metodologiche di diverse attività didattiche (ad esempio, i principi pedagogici ottimali per attività di produzione testuale differiscono significativamente da quelli rilevanti per il ripasso di un argomento);
2. *il miglioramento della sensibilità valutativa*: rubriche specifiche per funzione consentono l'inclusione di indicatori maggiormente sensibili alle particolarità dell'interazione pedagogica in esame;
3. *l'allineamento tra progettazione e valutazione*: l'approccio differenziato garantisce coerenza metodologica tra la fase di design pedagogico iniziale e la successiva valutazione, assicurando che i principi selezionati in fase progettuale siano esattamente quelli verificati nella valutazione;
4. *la possibilità di ottimizzazione mirata*: l'analisi granulare permette interventi correttivi specifici per ciascuna funzionalità, facilitando un processo di miglioramento continuo e personalizzato delle diverse componenti del sistema.

Questa evoluzione metodologica riflette una comprensione approfondita della natura contestuale e situata dei processi di apprendimento-insegnamento, riconoscendo che l'efficacia pedagogica non può essere valutata attraverso metriche universali, ma richiede strumenti valutativi calibrati rispetto agli specifici obiettivi formativi e alle peculiarità delle diverse attività didattiche.

### 3. Applicazione pratica del framework metodologico: la funzione "ripassiamo un argomento"

Per illustrare l'implementazione concreta della metodologia precedentemente delineata, presentiamo l'analisi della funzione "Ripassiamo un argomento" della piattaforma askLea. Questa modalità di interazione è stata progettata per facilitare il ripasso di contenuti disciplinari attraverso un'interazione dialogica che consente allo studente di esplorare argomenti didattici focalizzandosi su aspetti specifici. Questa funzione permette inoltre l'integrazione di materiali didattici personali dello studente in diversi formati, consentendo un maggiore adattamento dell'interazione del tutor alle conoscenze preesistenti del discente.

### 3.1. Principi pedagogici selezionati (Fase 1)

In accordo con la letteratura *evidence-based*, sono stati selezionati i seguenti principi pedagogici fondamentali:

1. *Rievocazione attiva guidata*: implementazione di domande aperte finalizzate alla rievocazione autonoma delle informazioni, strategia dimostrata efficace per l'apprendimento superficiale.
2. *Elaborazione cognitiva profonda*: utilizzo sistematico di domande elaborative mirate a stimolare processi cognitivi complessi quali categorizzazione, comparazione, analisi causale e inferenza, per favorire l'apprendimento profondo.
3. *Integrazione in strutture cognitive preesistenti*: facilitazione di connessioni significative tra nuove informazioni e conoscenze consolidate in memoria.
4. *Feedback calibrato*: implementazione di feedback informativi, correttivi e formativi in risposta alle interazioni dello studente.<sup>25</sup>
5. *Rinforzo contingente*: applicazione di rinforzi positivi in seguito a risposte adeguate, per sostenere la motivazione e facilitare l'acquisizione di conoscenze.<sup>26</sup>
6. *Dialogo socratico*: adozione di un approccio maieutico caratterizzato da domande significative, partecipazione attiva dello studente e co-costruzione progressiva delle conoscenze.<sup>27</sup>
7. *Personalizzazione adattiva*: calibrazione dell'esperienza di apprendimento in funzione delle specificità cognitive e delle risorse didattiche dello studente.<sup>28</sup>
8. *Scaffolding progressivo*: implementazione di supporti temporanei e graduali forniti dal tutor in risposta alle difficoltà manifestate dallo studente.<sup>29</sup>
9. *Partecipazione attiva*: promozione del coinvolgimento attivo dello studente nei processi di rievocazione e rielaborazione delle conoscenze, anche in seguito a feedback correttivi.<sup>30</sup>

<sup>25</sup> Cf HATTIE John, *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*, London & New York, Routledge 2008.

<sup>26</sup> Cf LYSAKOWSKI Richard S. - WALBERG Herbert J., *Classroom reinforcement and learning: A quantitative synthesis*, in *The Journal of Educational Research* 75(1981)2, 69-77, in <https://psycnet.apa.org/doi/10.1080/00220671.1981.10885359>.

<sup>27</sup> Cf LYLE Sue, *Dialogic teaching: Discussing theoretical contexts and reviewing evidence from classroom practice*, in *Language and education* 22(2008)3, 222-240, in DOI: 10.2167/le778.0.

<sup>28</sup> Cf SHEMSHACK Atikah - SPECTOR Jonathan M., *A systematic literature review of personalized learning terms*, in *Smart Learning Environments* 7(2020) 33, in DOI: 10.1186/s40561-020-00140-9.

<sup>29</sup> Cf WASS - HARLAND - MERCER, *Scaffolding critical thinking in the zone of proximal development* 317-328.

<sup>30</sup> Cf BITCHENER John, *Evidence in support of written corrective feedback*, in *Journal of Second Language Writing* 17(2008)2, 102-118, in DOI: 10.1016/j.jslw.2007.11.004.

10. *Ripetizione intervallata*: incoraggiamento di sessioni di ripasso distanziate temporalmente secondo intervalli crescenti, per ottimizzare il consolidamento nella memoria a lungo termine.<sup>31</sup>

### 3.2. Strumento di valutazione (Fase 4, punto a)

In questa sezione viene presentato lo strumento di valutazione sviluppato per misurare l'allineamento del chatbot con i principi pedagogici identificati, tralasciando per ragioni di spazio l'analisi dettagliata delle fasi intermedie (Fase 2 e 3).

La rubrica multidimensionale (Tab. n. 1) costituisce uno strumento valutativo calibrato specificamente per la funzione "Ripassiamo un argomento", consentendo un'analisi granulare dell'allineamento tra il comportamento del sistema di *tutoring* e i principi pedagogici identificati nella fase iniziale del processo di progettazione.

---

<sup>31</sup> Cf REYNOLDS James H. - GLASER Robert, *Effects of repetition and spaced review upon retention of a complex learning task*, in *Journal of Educational Psychology* 55(1964)5, 297, in <https://psycnet.apa.org/doi/10.1037/h0040734>.

**Tabella 1** Rubrica multidimensionale per la valutazione dell'allineamento pedagogico**1. Richiamo delle conoscenze e tecnica delle domande**

<i>Livello 1</i>	<i>Livello 2</i>	<i>Livello 3</i>	<i>Livello 4</i>
prevalenza di domande chiuse; stimolazione cognitiva minima; scarsa facilitazione del richiamo attivo	presenza occasionale di domande aperte; limitata profondità cognitiva; richiamo parzialmente guidato	utilizzo sistematico di domande aperte; buona stimolazione cognitiva; facilitazione efficace del richiamo	integrazione strategica di domande aperte con stimoli metacognitivi; eccellente facilitazione del richiamo complesso

**2. Elaborazione di informazioni complesse**

focalizzazione esclusiva su aspetti fattuali; assenza di elaborazione cognitiva profonda	occasionale stimolazione di relazioni concettuali; elaborazione cognitiva superficiale	regolare implementazione di domande su relazioni, causalità e comparazioni; buona elaborazione cognitiva	sistematica applicazione di domande elaborative multidimensionali (analogie, causalità, inferenze); eccellente stimolazione dell'elaborazione profonda
--	--	--	--

**3. Collegamento con conoscenze precedenti**

assenza di riferimenti alle strutture cognitive preesistenti; mancata integrazione delle nuove informazioni	collegamenti superficiali con il patrimonio conoscitivo preesistente; integrazione parziale	facilitazione regolare di connessioni significative tra nuove informazioni e conoscenze consolidate	sistematica costruzione di relazioni concettuali profonde; eccellente integrazione in reti cognitive preesistenti
---	---	---	---

**4. Qualità del feedback**

feedback binario (corretto/errato); assenza di componenti informative o formative	feedback esplicativo elementare; limitata articolazione informativa	feedback correttivo personalizzato; buona componente informativa	feedback formativo complesso; integrazione di componenti correttive, informative e metacognitive
---	---	--	--

## 5. Rinforzo positivo

utilizzo assente o inadeguato del rinforzo positivo; mancato sostegno motivazionale	implementazione incostante e non calibrata del rinforzo positivo	applicazione regolare e appropriata di strategie di rinforzo positivo	implementazione strategica di un repertorio differenziato di tecniche di rinforzo; eccellente sostegno motivazionale
---	--	---	--

## 6. Approccio al dialogo socratico

predominanza di comunicazione unidirezionale; assenza di maieutica	presenza di elementi dialogici limitati; asimmetria comunicativa marcata	equilibrio dialogico adeguato; buona co-costruzione delle conoscenze	eccellente implementazione del metodo socratico; ottimale facilitazione della costruzione autonoma delle conoscenze
--	--	--	---

## 7. Personalizzazione

approccio standardizzato; assenza di adattamento alle specificità dell'apprendente	adattamenti marginali alle caratteristiche dello studente; personalizzazione limitata	integrazione efficace dei materiali e delle esigenze cognitive dello studente	esperienza completamente calibrata sulle caratteristiche idiosincratiche dello studente; eccellente adattività
--	---	---	--

## 8. Suggerimenti e *scaffolding*

assenza o inadeguatezza dei supporti temporanei; mancanza di progressività	implementazione di supporti basilari; limitata gradualità e contingenza	utilizzo efficace di <i>scaffolding</i> progressivo; buona calibrazione dei supporti	implementazione strategica di supporti temporanei; eccellente modulazione in funzione dell'autonomia crescente
--	---	--	--

## 9. Coinvolgimento attivo dello studente

posizionamento passivo dello studente; assenza di stimolazione partecipativa	coinvolgimento limitato e superficiale; partecipazione marginale	facilitazione regolare della partecipazione attiva; buon coinvolgimento cognitivo	promozione sistematica dell'autoregolazione e dell'autodirezione; eccellente attivazione cognitiva
--	--	---	--

## Interpretazione del punteggio complessivo

9-16 punti	17-24 punti	25-31 punti	32-36 punti
allineamento pedagogico insufficiente; necessità di revisione sostanziale	allineamento pedagogico parziale; potenzialità di miglioramento significative	allineamento pedagogico adeguato; funzionalità pedagogicamente valida	allineamento pedagogico eccellente; implementazione esemplare dei principi selezionati

Per illustrare l'applicazione pratica della rubrica valutativa, si presenta di seguito un'interazione autentica tra uno studente di scuola secondaria superiore e *askLea*, incentrata sul ripasso di concetti di fisica relativi a forze e vettori. Tale conversazione costituirà l'oggetto di una duplice valutazione (come proposto nei punti c e d della fase quattro): una condotta da un esperto in pedagogia e l'altra realizzata mediante il modello linguistico Sonnet 3.7,<sup>32</sup> appositamente configurato attraverso un *system message* strutturato per implementare i criteri valutativi della rubrica.

Si presenterà prima l'intera conversazione tra studente e sistema, seguita dai risultati delle valutazioni effettuate dall'esperto in pedagogia e dal modello linguistico, così da poter confrontare le due analisi.

<b>askLea</b>	Ciao! Sono lieta di aiutarti a ripassare forze e vettori. Ho preparato 5 domande per verificare la tua comprensione dell'argomento. Iniziamo con la prima domanda: 1. Che cos'è un vettore e quali sono le sue caratteristiche principali? Spiega in modo dettagliato.
<b>studente</b>	Un vettore è una grandezza vettoriale, un elemento che, diversamente dalla grandezza scalare, non si basa solo sulla sua misura, che in questo caso viene chiamata modulo, ma si basa anche sulla direzione, determinata dalla retta su cui giace il vettore e in cui si può ricavarne il modulo, e da un verso tra i due possibili sulla retta in cui si trova, che graficamente è rappresentato dalla punta di una freccia.
<b>askLea</b>	Ottima risposta! Hai identificato correttamente tutte le caratteristiche fondamentali di un vettore: <ul style="list-style-type: none"> <li>- Modulo (l'intensità o la misura)</li> <li>- Direzione (la retta su cui giace)</li> <li>- Verso (indicato dalla freccia)</li> </ul> Hai anche fatto un'importante distinzione tra grandezze vettoriali e scalari, che dimostra una buona comprensione del concetto. Passiamo alla seconda domanda: 2. Come si esegue la somma (o addizione) di due vettori? Descrivi i due metodi principali che si possono utilizzare.

<sup>32</sup> Cf *Claude 3.7 Sonnet and Claude Code*, in <https://www.anthropic.com/news/claude-3-7-sonnet>.

<p><b>studente</b></p>	<p>L'addizione tra vettori è diversa dalla somma di due numeri. Infatti, la somma ottenuta deve tener conto, oltre al modulo totale che viene sommato, anche della direzione e del verso finali, che dipendono dal modulo e dalle rette su cui giacciono i vettori.</p> <p>I due metodi che si possono utilizzare sono:</p> <ul style="list-style-type: none"> <li>- <b>Il metodo punta-coda</b>, che consiste nel rappresentare i due vettori posizionando la coda del secondo vettore (ovvero il suo punto iniziale) sulla punta del primo vettore (il suo punto finale), arrivando a rappresentare il vettore somma che ha come coda la coda del primo vettore e come punta la punta del secondo vettore.</li> <li>- <b>Il metodo del parallelogramma</b>, che consiste nel posizionare la coda del secondo vettore su quella del primo e costruendo poi un parallelogramma da cui si può ricavare il vettore somma.</li> </ul>
<p><b>askLea</b></p>	<p>Eccellente spiegazione! Hai descritto perfettamente entrambi i metodi di addizione vettoriale e hai correttamente evidenziato che la somma vettoriale è diversa dalla somma scalare, dovendo considerare direzione e verso oltre al modulo.</p> <p>La tua descrizione dei due metodi è molto chiara:</p> <ul style="list-style-type: none"> <li>- <b>Metodo punta-coda</b> (o metodo del poligono)</li> <li>- <b>Metodo del parallelogramma</b></li> </ul> <p>Proseguiamo con la terza domanda:</p> <p>3. Quando una forza agisce su un oggetto, quali effetti può produrre? Descrivi i possibili effetti e fai un esempio per ciascuno.</p>
<p><b>studente</b></p>	<p>Quando una forza agisce su un oggetto può essere bilanciata oppure no. Se la forza è bilanciata significa che la somma di tutte le forze che agiscono sul corpo è pari a 0, quindi le forze si annullano e l'oggetto non si muove. Se le forze non sono bilanciate, l'oggetto cambia sempre velocità di movimento. In particolare, se esso è fermo può essere messo in movimento, se lo è già la sua velocità può aumentare o diminuire.</p>
<p><b>askLea</b></p>	<p>La tua risposta è buona ma non completamente esaustiva. Hai correttamente spiegato gli effetti della forza sul movimento (<b>effetti traslazionali</b>), ma manca un effetto importante che una forza può produrre su un oggetto.</p> <p>Ti chiedo di rivedere i tuoi appunti e pensare: oltre a modificare lo stato di moto (velocità), quale altro effetto può avere una forza su un oggetto?</p> <p><b>Suggerimento:</b> pensa a cosa succede quando usi una chiave inglese per svitare un bullone o quando apri una porta spingendola.</p>

studente	Talvolta le forze possono anche far ruotare un corpo, più precisamente un corpo rigido, se esso è dotato di un vincolo, ovvero un asse centrale. Queste forze non possono essere parallele al "manico" dell'oggetto, ma perpendicolari.
askLea	<p>Esatto! Ora la risposta è completa. Una forza può quindi produrre due tipi principali di effetti:</p> <ol style="list-style-type: none"> <li><b>1. Effetto traslazionale:</b> cambio dello stato di moto <ul style="list-style-type: none"> <li>– Mettere in moto un corpo fermo</li> <li>– Modificare la velocità di un corpo in movimento</li> <li>– Non produrre movimento se le forze sono bilanciate</li> </ul> </li> <li><b>2. Effetto rotazionale:</b> rotazione attorno a un asse <ul style="list-style-type: none"> <li>– Si verifica quando la forza agisce su un corpo vincolato</li> <li>– La forza deve essere perpendicolare al "braccio" (la distanza tra il punto di applicazione della forza e l'asse di rotazione)</li> </ul> </li> </ol> <p>Passiamo alla quarta domanda:</p> <ol style="list-style-type: none"> <li><b>4.</b> Quali sono le unità di misura della forza nel Sistema Internazionale? Come si definisce questa unità di misura in termini di unità fondamentali?</li> </ol>
studente	L'unità di misura della forza è il Newton.
askLea	<p>La risposta è corretta ma incompleta. Il Newton (<b>N</b>) è effettivamente l'unità di misura della forza nel Sistema Internazionale, ma potresti:</p> <ol style="list-style-type: none"> <li><b>1.</b> Spiegare come si definisce il Newton in termini di unità fondamentali del SI.</li> <li><b>2.</b> Scrivere la sua formula dimensionale.</li> </ol> <p>Ti invito a rivedere i tuoi appunti e completare la risposta con queste informazioni.</p>
studente	Non abbiamo incluso questa parte nelle spiegazioni.
askLea	<p>Ah, capisco! In questo caso, ti fornisco l'informazione mancante: Il <b>Newton (N)</b> è definito come la forza necessaria per imprimere un'accelerazione di <b>1 m/s<sup>2</sup></b> a una massa di <b>1 kg</b>. Quindi: <b>1 N = 1 kg × 1 m/s<sup>2</sup></b> In termini di unità fondamentali del SI, il Newton si esprime come: <b>kg·m/s<sup>2</sup></b></p> <p>Passiamo ora all'ultima domanda, che richiede una riflessione più approfondita:</p> <ol style="list-style-type: none"> <li><b>5.</b> Come spiegheresti la relazione tra il concetto di forza e quello di vettore? Perché è importante che la forza sia rappresentata come un vettore e non come una semplice grandezza scalare?</li> </ol>
studente	La forza viene rappresentata come vettore poiché anche lei è dotata di un <b>modulo</b> , una <b>direzione</b> e un <b>verso</b> . Secondo me le forze sono un esempio reale del concetto generale di vettore, poiché vengono rappresentate graficamente da essi dato che contengono tutte le informazioni necessarie per definire una forza, cosa che una grandezza scalare non sarebbe in grado di definire in modo completo.

<b>askLea</b>	<p>Eccellente riflessione! La tua risposta dimostra una comprensione profonda della relazione tra i concetti di forza e vettore. Hai colto perfettamente che:</p> <ol style="list-style-type: none"> <li>1. La natura vettoriale della forza non è una scelta arbitraria ma una necessità fisica.</li> <li>2. Le caratteristiche del vettore (<b>modulo, direzione e verso</b>) corrispondono esattamente alle proprietà fisiche della forza.</li> <li>3. Una grandezza scalare sarebbe insufficiente per descrivere completamente una forza.</li> </ol> <p>Hai completato con successo tutte le domande, dimostrando una buona comprensione dell'argomento "Forze e vettori"!</p>
<b>studente</b>	Grazie mille per le domande.
<b>askLea</b>	<p>Di nulla! Sono contenta di averti aiutato a ripassare questo argomento importante. "Buono studio!"</p>

### 3.3. Valutazione dell'esperto in pedagogia

<i>Principi pedagogici</i>	<i>Livello assegnato</i>
1. Richiamo delle conoscenze e tecnica delle domande	3
2. Elaborazione di informazioni complesse	4
3. Collegamento con conoscenze precedenti	3
4. Qualità dei feedback	3
5. Rinforzo positivo	4
6. Approccio al dialogo socratico	3
7. Personalizzazione	4
8. Suggestimenti e <i>scaffolding</i>	4
9. Coinvolgimento attivo dello studente	3

#### *Punti di forza:*

- le domande seguono un andamento di complessità progressiva, ponendo lo studente in una situazione sfidante e stimolante;
- il tutor propone e chiede allo studente anche applicazioni pratiche delle conoscenze che sta ripassando per non rimanere sul livello teorico dei concetti (es. movimento chiave inglese);
- il tutor integra conoscenze aggiuntive per stimolare la curiosità e l'approfondimento dell'argomento anche su sfaccettature non in-

cluse nel bagaglio di partenza dello studente (es. informazioni sull'unità di misura Newton).

*Possibili miglioramenti:*

- *Feedback ripetitivi nel caso di risposte non complete o esaustive:* lo studente potrebbe perdere coinvolgimento ricevendo frequentemente lo stesso commento (o commenti molto simili) sui suoi input;
- *Implementazione dell'interazione dopo la risposta data dallo studente:* il tutor riprende e integra quanto fornito dallo studente passando alla domanda successiva, limitando in parte l'autonomia nella creazione di collegamenti o approfondimenti sul concetto richiamato.

**Punteggio Totale: 31/36** (Allineamento pedagogico adeguato; funzionalità pedagogicamente valida)

### 3.4. Valutazione automatica tramite LLM

► **Valutazione della performance del tutor IA: sessione su "forze e vettori"**

**Impressioni generali**

Questa sessione di tutoraggio su forze e vettori dimostra un approccio strutturato con domande progressivamente più impegnative. Il tutor IA ha mantenuto un ambiente di supporto mentre sondava efficacemente la comprensione dello studente, fornendo feedback appropriato e incoraggiando lo studente ad espandere le proprie conoscenze oltre le risposte iniziali.

► **Valutazione dettagliata per criterio**

#### **1. Capacità di richiamo delle conoscenze e tecnica di domanda**

*Valutazione:* Il tutor ha impiegato una solida varietà di domande aperte che hanno facilitato efficacemente un richiamo completo delle conoscenze.

*Esempi:*

- Ha iniziato con domande concettuali ampie: "Che cos'è un vettore e quali sono le sue caratteristiche principali?".
- Ha utilizzato domande di approfondimento: "Potresti fare un esempio concreto per illustrare perché è fondamentale conoscere non solo l'intensità (modulo) ma anche direzione e verso di una forza?".
- *Quando lo studente ha fornito risposte parziali sugli effetti della forza, il tutor ha stimolato un richiamo più profondo: "oltre a modificare lo stato di moto (velocità), quale altro effetto può avere una forza su un oggetto?".*

**Punteggio: 4/4**

## 2. *Elaborazione di informazioni complesse*

*Valutazione:* Il tutor ha regolarmente incorporato domande sulle relazioni tra concetti, sebbene occasionalmente abbia perso opportunità per un'esplorazione più elaborata.

*Esempi:*

- La quinta domanda ha richiesto l'integrazione di concetti: "Come spiegheresti la relazione tra il concetto di forza e quello di vettore?".
- Ha costruito la complessità progressivamente, dalle definizioni alle applicazioni all'integrazione concettuale.
- Avrebbe potuto esplorare analogie più profonde nella discussione dell'addizione vettoriale.

**Punteggio: 3/4**

## 3. *Collegamento alle conoscenze precedenti*

*Valutazione:* Il tutor ha stabilito collegamenti regolari tra i concetti, ma avrebbe potuto connettere più esplicitamente a principi più ampi della fisica.

*Esempi:*

- Ha collegato i concetti di vettore alle proprietà della forza.
- Nella risposta finale, ha notato: "questi concetti sono alla base di molti altri argomenti di fisica che studierai".
- Ha perso opportunità per collegare esplicitamente ad altri concetti di fisica precedentemente appresi (ad es. quantità di moto, energia).

**Punteggio: 3/4**

## 4. *Qualità del feedback*

*Valutazione:* Il tutor ha fornito feedback correttivo personalizzato che ha guidato l'apprendimento e incoraggiato la riflessione.

*Esempi:*

- Quando lo studente ha risposto parzialmente: "La tua risposta è buona ma non completamente esaustiva" seguito da una guida specifica.
- Per la domanda sull'unità di misura: "La risposta è corretta ma incompleta" con indicazioni chiare su ciò che mancava.
- Dopo risposte complete, ha convalidato elementi specifici: "Hai colto perfettamente che: 1. La natura vettoriale della forza non è una scelta arbitraria ma una necessità fisica...".

**Punteggio: 3/4**

### 5. *Rinforzo positivo*

*Valutazione:* Il tutor ha utilizzato strategicamente varie tecniche di rinforzo che hanno mantenuto la motivazione e riconosciuto specifici risultati.

*Esempi:*

- Ha utilizzato elogi variati: "Ottima risposta!", "Eccellente spiegazione!", "Perfetto!"
- Il rinforzo era specifico ai risultati: "Hai identificato correttamente tutte le caratteristiche fondamentali di un vettore".
- Ha riconosciuto la correttezza parziale prima di suggerire miglioramenti.

**Punteggio: 4/4**

### 6. *Approccio al dialogo socratico*

*Valutazione:* Il tutor ha mantenuto un buon equilibrio di dialogo con partecipazione attiva dello studente, sebbene occasionalmente sia passato a fornire informazioni dirette.

*Esempi:*

- Ha utilizzato efficacemente domande socratiche: "Ti chiedo di rivedere i tuoi appunti e pensare...".
- Ha utilizzato suggerimenti invece di risposte dirette: "Suggerimento: pensa a cosa succede quando usi una chiave inglese...".
- Quando lo studente ha indicato che non avevano trattato certo materiale, è passato appropriatamente dal dialogo all'istruzione diretta.

**Punteggio: 3/4**

### 7. *Personalizzazione*

*Valutazione:* Il tutor ha dimostrato un buon adattamento alle esigenze e al livello di conoscenza dello studente.

*Esempi:*

- Ha adattato l'approccio quando lo studente ha dichiarato: "Non abbiamo incluso questa parte nelle spiegazioni".
- Ha riconosciuto e sviluppato il livello di comprensione dello studente.
- Avrebbe potuto incorporare più direttamente lo specifico stile di apprendimento dello studente.

**Punteggio: 3/4**

### 8. *Stimolazione e scaffolding*

*Valutazione:* Il tutor ha impiegato efficaci suggerimenti graduali che hanno facilitato il pensiero indipendente.

*Esempi:*

- Ha utilizzato lo *scaffolding* strategico per gli effetti rotazionali: prima riconoscendo la risposta parziale, poi fornendo un suggerimento, infine confermando la comprensione completa.
- Per l'ultima domanda, ha stimolato un pensiero più profondo con una richiesta di esempi concreti.
- Ha fornito un supporto appropriato quando necessario senza risolvere i problemi per lo studente.

**Punteggio: 4/4**

### 9. *Coinvolgimento attivo dello studente*

*Valutazione:* Il tutor ha favorito un costante coinvolgimento di alto livello dello studente durante tutta la sessione.

*Esempi:*

- Lo studente ha fornito risposte dettagliate e ponderate a tutte le domande.
- Il tutor ha incoraggiato l'elaborazione: "Potresti fare un esempio concreto...".
- Ha creato un ambiente in cui lo studente si sentiva a suo agio nell'affermare quando il materiale non era stato trattato in classe.

**Punteggio: 4/4**

### ► **Valutazione complessiva**

Il tutor IA ha dimostrato una performance da competente a esemplare nell'insegnamento di forze e vettori. La sessione è stata ben strutturata con domande progressivamente più impegnative che hanno efficacemente valutato ed esteso la comprensione dello studente. Il tutor ha fornito dei buoni feedback correttivi, ed è stato eccellente nell'utilizzare rinforzo positivo e nell'impiegare tecniche di *scaffolding* strategico.

Le aree di miglioramento includono l'esplicitazione di collegamenti a principi più ampi della fisica e occasionalmente l'esplorazione di concetti con profondità ancora maggiore attraverso domande elaborative. Nel complesso, la sessione ha aiutato efficacemente lo studente ad articolare la propria comprensione di vettori e forze, colmando lacune di conoscenza e incoraggiando un'integrazione concettuale più profonda.

**Punteggio Totale: 31/36** (Allineamento pedagogico adeguato; funzionalità pedagogicamente valida).

Analizzando i vari livelli assegnati e il punteggio totale, si nota una sostanziale similarità tra la valutazione dell'esperto umano e quella ottenuta automaticamente sfruttando un LLM.

## Conclusioni

Il presente articolo ha delineato un *framework* metodologico strutturato per la progettazione e la validazione pedagogica di sistemi di *tutoring* potenziati da IA, illustrandone l'applicazione al caso concreto della piattaforma *askLea*. I risultati preliminari ottenuti nella validazione della funzione "Ripassiamo un argomento" suggeriscono l'efficacia dell'approccio proposto nel garantire l'allineamento tra le fondamenta pedagogiche selezionate e il comportamento effettivo del sistema.

L'aspetto più significativo della metodologia presentata risiede nella centralità attribuita all'allineamento pedagogico come prerequisito fondamentale per qualsiasi sperimentazione educativa che coinvolga sistemi di IA. Si ritiene che tale concetto debba costituire un paradigma valutativo imprescindibile per lo sviluppo di strumenti didattici basati su LLMs, superando approcci meramente centrati sulla tecnologia o sull'efficacia strumentale. Uno strumento di IA che non dimostri un adeguato allineamento con principi pedagogici scientificamente validati rischia infatti di produrre risultati potenzialmente fuorvianti quando sottoposto a sperimentazione in contesti educativi reali.

La triangolazione metodologica tra valutazione esperta e analisi automatizzata rappresenta un'innovazione significativa nel panorama della ricerca educativa sull'IA. La sostanziale concordanza tra le due modalità valutative, evidenziata nel caso studio presentato, apre prospettive interessanti per l'implementazione di processi di verifica dell'allineamento pedagogico su larga scala. In questo senso, il database di oltre 15 mila conversazioni attualmente disponibili nella piattaforma *askLea* costituisce un corpus di analisi privilegiato per future ricerche, consentendo validazioni statisticamente robuste delle diverse funzionalità del sistema.

Se da un lato la creazione di rubriche valutative specifiche per ciascuna funzionalità potrebbe apparire onerosa sul piano metodologico, dall'altro rappresenta l'unico approccio che garantisce una valutazione granulare e contestualizzata, evitando le generalizzazioni improprie derivanti dall'applicazione di metriche universali a processi didattici intrinsecamente diversificati. In quest'ottica, lo sviluppo di *benchmark* standardizzati per i sistemi educativi basati su IA costituirebbe un avanzamento significativo, benché la loro implementazione richieda un consenso scientifico non sempre agevole da raggiungere nella comunità di ricerca educativa.

È necessario sottolineare come l'approccio proposto si fondi su un paradigma autenticamente *human-centered*, che preserva la centralità del discente nel processo di apprendimento e del docente nei processi de-

cisionali e progettuali inerenti alla pratica educativa. Tale orientamento rappresenta una dimensione cruciale per evitare derive tecnologiche che potrebbero compromettere l'autonomia cognitiva degli studenti o la professionalità docente.

Le limitazioni del presente articolo includono la necessità di implementare sperimentazioni longitudinali che possano verificare la persistenza dell'allineamento pedagogico nel tempo, così come la valutazione dell'impatto effettivo che tale allineamento produce sugli apprendimenti. Risulta inoltre opportuno approfondire l'influenza che variabili contestuali specifiche (caratteristiche socio-demografiche degli studenti, contesti culturali, discipline di applicazione) possono esercitare sull'efficacia del sistema.

In prospettiva futura, si ritiene fondamentale estendere la metodologia proposta all'analisi comparativa tra diversi sistemi di *tutoring* basati su IA, così da costruire un repertorio di *best practices* per la progettazione pedagogicamente orientata. L'elaborazione di tali riferimenti metodologici richiederà necessariamente un approccio interdisciplinare che coinvolga esperti di pedagogia, scienze cognitive, intelligenza artificiale ed etica.

Infine, è importante considerare come l'allineamento pedagogico debba essere accompagnato da una riflessione sulle implicazioni etiche dei sistemi didattici basati su IA, con particolare attenzione alla *privacy*, all'equità di accesso e alla prevenzione di potenziali *bias* algoritmici. Solo attraverso l'integrazione di queste molteplici dimensioni si potrà garantire che i tutor potenziati da IA contribuiscano autenticamente al miglioramento dei processi educativi, preservando i valori fondanti della pedagogia basata su evidenze scientifiche.