e-ISSN 2280-6792

EL.LE
Vol. 14 — Num. 1 — Marzo 2025

ChatGPT nella valutazione dell'elaborato scritto
Applicazioni pratiche in un corso d'inglese lingua straniera

Salvatore Nizzolino
Sapienza Università di Roma, Italia

Abstract This case study explores ChatGPT's practical use in analysing textual content from an educational forum. It examines three texts from a sample of 26, written by first-year university students during a guided English as a foreign language activity. This highlights generative AI's potential in teaching. ChatGPT is used to create an evaluative rubric and conduct linguistic analysis. It shows logical cohesion, output consistency, and an ability to identify both global and specific textual features. With few documented case studies on AI in education, this contribution presents a practical experience to foster experimentation and results.

Keywords Artificial intelligence for education. Teaching assistance with ChatGPT. Instructional design with ChatGPT. Foreign language with ChatGPT. English language and AI.

Sommario 1 Introduzione . – 2 Il chatbot ChatGPT. – 3 Metodologia. – 4 Il caso di studio. – 5 Discussione. – 6 Prospettive di ricerca.

Edizioni Ca'Foscari

Peer review

Submitted 2024-04-02
Accepted 2025-01-09
Published 2025-05-08

Open access
© 2025 Nizzolino | 4.0

Citation Nizzolino, S. (2025). "ChatGPT nella valutazione dell'elaborato scritto. Applicazioni pratiche in un corso d'inglese lingua straniera". *EL.LE*, 14(1), 29-46.

DOI 10.30687/ELLE/2280-6792/2025/01/002

1 Introduzione

Nell'era dei *big data*, l'intelligenza artificiale (IA) trova applicazioni diffuse e sta progressivamente permeando un numero crescente di settori, rivoluzionando processi e contesti in maniera inedita (Zhai et al. 2021). L'IA utilizza una serie di passaggi logici chiamati algoritmi e tecnologie cognitive avanzate di calcolo. In quanto campo interdisciplinare, è utilizzata in informatica, scienze cognitive e dell'apprendimento, *game design* (Melhárt et al. 2023), psicologia, sociologia, filosofia, matematica, linguistica, industria della difesa, neuroscienze, medicina e istruzione (Yassin et al. 2022; Sharma et al. 2023). L'IA ha una portata diversa in termini di sviluppi nei campi sopra citati, in quanto include le capacità analogiche e statistiche durante l'elaborazione di un linguaggio naturale per sviluppare la comunicazione tra uomo e computer (Bozkurt et al. 2015). L'educazione è uno degli ambiti in cui l'IA promette di generare un impatto rivoluzionario, laddove le tecnologie intelligenti possono trasformare il modo in cui insegnanti e studenti affronteranno il processo di apprendimento (Chen, Chen, Lin 2020). Affermare in modo generico che questa trasformazione può migliorare l'accesso all'istruzione, ma anche la qualità e la personalizzazione dell'apprendimento, può risultare aleatorio (Ezzaim et al. 2023). Un primo approccio dovrebbe circoscrivere il contesto in cui definire le nuove relazioni tra IA, docente e discente. A titolo esemplificativo, se si intenda adottare soluzioni digitali sincrone o asincrone basate sull'IA, se il discente sarà un mero recettore dei contenuti generati dall'IA oppure avrà un ruolo attivo nell'interazione, se il docente rivestirà il ruolo di tramite tra la classe e l'IA oppure se coordinerà gruppi di lavoro che adottano soluzioni basate sull'IA (Ouyang, Jiao 2021). Tuttavia, l'introduzione di sistemi di IA nell'educazione, soprattutto pre-universitaria, attualmente sembra guidata da un atteggiamento votato alla prudenza, sia da parte delle istituzioni (O'Dea, O'Dea 2023) sia da parte degli stessi docenti. Tra i fattori che pongono un freno alla repentina diffusione dei nuovi strumenti, nonostante la diffusa digitalizzazione al traino delle discipline STEM, annoveriamo la carenza di esperienze e casi studio documentati (Chen et al. 2020; Papadopoulos et al. 2020) e di conseguenza la mancanza di linee guida per gli educatori (Zawacki-Richter et al. 2019).

2 Il chatbot ChatGPT

È possibile far risalire l'origine dei chatbot agli anni Sessanta del XX secolo, con l'introduzione di ELIZA, che forniva semplici risposte a seguito di parole chiave inserite dall'utente (Chuah, Kabilan 2021). Tuttavia, l'accesso aperto agli strumenti di IA è diventato disponibile per la prima volta grazie al lancio del chatbot di OpenAI, meglio conosciuto come ChatGPT (GPT), nel novembre 2022. Un chatbot è un software che consente il dialogo tra l'utente e il complesso convolvolo di algoritmi che costituiscono la rete cognitiva dell'IA. In altri termini, potremmo raffigurare il chatbot come un interprete tra l'uomo e la 'mente' dell'IA.

GPT adotta una tecnologia basta su una classe di modelli di apprendimento automatico per l'elaborazione del linguaggio naturale noti come Large Language Model (LLM). Per questo motivo, è possibile interagire semplicemente utilizzando il linguaggio naturale. I LLM, oltre ad analizzare enormi quantità di dati testuali e dedurre le relazioni tra le parole all'interno del testo, accrescono le loro capacità con l'aumentare delle dimensioni dei dataset e degli input. Le possibilità di applicazione di questa tecnologia riguardano i più svariati ambiti, inclusi la generazione di testi complessi, anche di livello accademico, fino alla scrittura di intere sezioni di coding dei linguaggi di programmazione (Aljanabi et al. 2023). Attualmente, i maggiori timori associati all'implementazione di queste tecnologie nel settore educativo ruotano attorno alle implicazioni etiche, in quanto un testo generato dall'IA è irriconoscibile da un testo scritto da un essere umano. Già numerosi studi riconoscono che il grado di complessità e le conoscenze integrate nel testo, nonché i registri linguistici soddisfano gli standard accademici più esigenti.[1] Riguardo all'interesse che GPT sta suscitando nel mondo dell'educazione, si registra una crescente attenzione da parte della ricerca accademica, per scopi diversi, mentre la sperimentazione didattica non è stata ancora coinvolta in modo significativo. La ricerca sta esplorando non solo i benefici dell'IA per l'apprendimento, ma ne sta anche documentando gli attuali limiti. Pertanto, lo scopo del presente contributo è quello di adottare GPT per eseguire alcune operazioni di routine, legate alla valutazione degli elaborati scritti. Verranno osservati gli output di GPT quando viene interrogato per analizzare i testi prodotti da un gruppo classe. Il presente lavoro non esamina le implicazioni dell'IA associate al plagio o all'integrità delle fonti associate a un testo (Adeshola, Adepoju 2023).

1 Cooper 2023; Crawford et al. 2023; Dwivedi et al. 2023; Halaweh 2023; Sullivan et al. 2023; Tlili et al. 2023.

3 Metodologia

3.1 Pseudonimizzazione del campione

I testi analizzati provengono da un corso di inglese lingua straniera, livello B2, erogato nel II semestre del 2023 in modalità *blended*, presso la Facoltà di Farmacia e Medicina, LM di Chimica e Tecnologie Farmaceutiche, Sapienza Università di Roma. Il campione è stato filtrato attraverso un processo di pseudonimizzazione, in modo da rispettare le norme vigenti in conformità con il regolamento generale sulla protezione dei dati (GDPR 2016). Questa pratica è particolarmente rilevante nei contesti educativi e di ricerca, soprattutto quando i dati vengono condivisi con terze parti. La pseudonimizzazione consiste nel sostituire le informazioni sensibili con un identificatore artificiale o con pseudonimi in modo che i dati non possano essere ricondotti a un individuo specifico senza l'utilizzo di informazioni aggiuntive. Il processo ha seguito due step:

1. a ogni studente del gruppo campione è stato assegnato un unico pseudonimo costituito dalla lettera 's' e da un numero casuale, utilizzato in tutte le fasi dell'analisi per sostituire le informazioni personali;
2. i dati pseudonimizzati e i contenuti testuali sono stati archiviati, cancellando dal file anche i metadati associati all'autore del salvataggio. Questo consente la condivisione sicura del file e previene qualsiasi tentativo di ricostruzione inversa delle informazioni.

3.2 GPT-4

Le interazioni che seguono sono state generate tramite la versione GPT-4, disponibile al momento della sperimentazione. Il primo passo necessario è la creazione di un profilo sulla piattaforma `https://chat.openai.com/auth/login`; successivamente, si effettua il login con le proprie credenziali. Nella schermata di benvenuto **[fig. 1]**, si viene accolti dal messaggio «How can I help you today?», oltre a quattro suggerimenti casuali che campeggiano al centro della schermata e cambiano a ogni nuovo *login*. Il campo attivo in fondo alla pagina è pronto ad accogliere le nostre richieste (*prompt*) che non richiedono altro che un comando o una domanda formulata nella nostra lingua. Attualmente, GPT supporta oltre 50 lingue e può gestire operazioni simultanee in più lingue. In altre parole, è possibile richiedere l'analisi di testi scritti in inglese mentre si forniscono le istruzioni in italiano, in finlandese o in arabo. Per quanto riguarda l'inserimento del testo da valutare, questo può essere effettuato tramite una semplice operazione di copia e incolla. Nel nostro caso di studio abbiamo

utilizzato la sottoscrizione a ChatGPT-4, per esplorare tutte le potenzialità dello strumento.

Durante l'utilizzo di GPT-4 per la valutazione degli elaborati scritti, è stato necessario monitorare le risposte dell'IA con particolare attenzione, per individuare eventuali fenomeni di *hallucination*. Questa categoria di eventi, che causa l'inserimento di dati non verificati o errati nei testi generati, sottolinea l'importanza di una supervisione continua da parte del docente per garantire che i risultati riflettano informazioni accurate. Questo fenomeno rappresenta uno dei limiti documentati più subdoli, in cui l'IA produce informazioni false o non aderenti alla realtà, pur mantenendo una coerenza linguistica e strutturale. Quindi, si tratta di discrepanze non sempre immediatamente riconoscibili. La mancanza di una definizione univoca del fenomeno delle *hallucination* dell'IA complica l'approccio a tale problema, rendendo necessario un maggiore sforzo per standardizzare i termini e migliorare l'accuratezza delle risposte nei vari campi applicativi, dalla sanità all'educazione. Questo potrebbe rappresentare una sfida significativa nel contesto educativo, dove l'attenzione alle informazioni è cruciale (Bang et al. 2023). Le *hallucination* possono portare l'IA a inventare fatti, nomi, o eventi inesistenti, rendendo essenziale la supervisione umana per verificare la correttezza dei contenuti generati (Maleki et al. 2024). Infine, oltre ad amplificare le implicazioni etiche associate a false elaborazioni, gli eventi di *hallucination* possono generare sfiducia negli utenti finali, che nel nostro scenario sono docenti e studenti (Athaluri et al. 2023; Judge, Jones, McCreery 2023). Nel nostro specifico caso di studio, che ha interrogato le funzioni deduttive di GPT, il fenomeno non si è presentato. Pertanto, abbiamo constatato la piena affidabilità di GPT in relazione alle istruzioni e al contesto forniti.

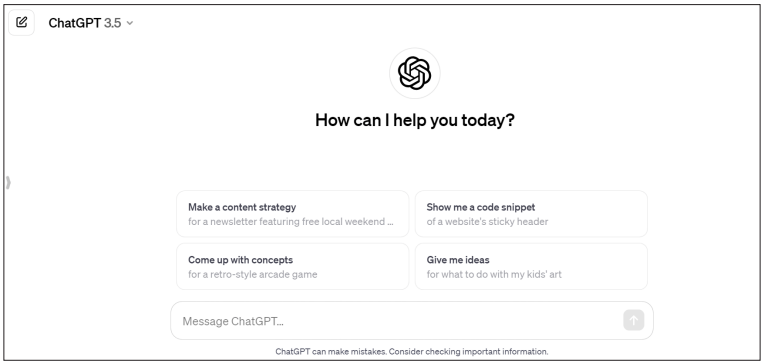

Figura 1 Schermata di benvenuto di ChatGPT

4 Il caso di studio

Seguendo un criterio basato su diversi tentativi, abbiamo fissato una sequenza riproducibile, che è possibile proporre come un kit per il sillabo. L'attività si sviluppa attraverso tre *prompt* per GPT-4, che guideranno tre fasi: (1) creazione di una rubrica valutativa; (2) contestualizzazione dell'attività da valutare; (3) valutazione dei singoli testi tramite i criteri della rubrica.

4.1 Creazione di una rubrica valutativa

Recenti studi suggeriscono che GPT, con la sua capacità di comprendere aspetti connotativi e denotativi dei testi, potrebbe potenzialmente aiutare a creare rubriche di valutazione semplificandole e personalizzandole per specifici contesti educativi (Bhattacharyya et al. 2022; Kondo, Nishigori, Vleuten 2023). Secondo Joseph et al. (2019), l'uso di rubriche può agevolare maggiormente gli educatori nella valutazione rispetto agli studenti, in quanto gli educatori comprendono i requisiti meglio degli studenti. Ciò suggerisce che, sebbene GPT sia in grado di creare rubriche autonomamente, è essenziale garantire che queste siano progettate con criteri chiari e comprensibili per supportare l'apprendimento e la valutazione degli studenti.

Nel nostro caso, testeremo la capacità dell'IA di generare una rubrica tramite i processi predefiniti degli algoritmi, quindi forniremo pochi elementi fondamentali per configurare il contesto. L'output di GPT spesso impiega gli asterischi assieme ai numeri per formattare gli elenchi; nel nostro caso abbiamo cancellato gli asterischi e mantenuto solo la numerazione, oltre a normalizzare l'uso della maiuscola secondo le norme stilistiche della lingua italiana.

Prompt che tipo di criteri suggerisci per valutare un'attività di produzione scritta in lingua inglese, assegnata a un gruppo di studenti di inglese L2, ai quali ho chiesto di commentare un video che tratta l'argomento della neuroplasticità?
Output per valutare un'attività di produzione scritta in inglese L2 in cui gli studenti sono chiamati a commentare un video sulla neuroplasticità, si possono considerare i seguenti criteri.

1. Comprensione del contenuto: gli studenti hanno dimostrato di aver compreso il video? Hanno identificato i concetti principali e li hanno spiegati correttamente nei loro commenti?
2. Correttezza linguistica:

– Grammatica: utilizzo corretto della grammatica inglese, incluso tempo verbale, accordo soggetto-verbo, articoli e preposizioni.
– Ortografia: correttezza nell'ortografia delle parole.

- Punteggiatura: uso appropriato dei segni di punteggiatura per articolare e strutturare le frasi.

3. Ricchezza lessicale: hanno usato vocabolario vario ed espressioni appropriate al contesto? Si sono allontanati dal ricorso a terminologia troppo basilare o ripetitiva?
4. Coerenza e coesione:
- Coerenza: i commenti presentano idee ordinate logicamente che portano a una conclusione chiara?
- Coesione: hanno utilizzato connettivi logici per legare frasi e paragrafi tra di loro?

5. Capacità critica e analisi:
- Gli studenti hanno mostrato capacità di pensiero critico nel loro approccio ai video?
- Hanno fornito analisi personalizzate piuttosto che limitarsi a riassumere il contenuto dei video?

6. Riflessione personale: il testo riflette una comprensione personale dell'apprendimento delle lingue? Gli studenti hanno collegato il materiale del video con le proprie esperienze o opinioni personali?
7. Adesione alle istruzioni: il testo rispetta la lunghezza richiesta, il formato (ad esempio: saggio breve, riflessioni puntuali) e tutte le altre direttive specificate nell'assegnazione?
8. Originalità: gli studenti hanno mostrato originalità nei loro commenti? Hanno integrato le proprie idee e prospettive uniche in modo efficace?
   Ricorda che ognuno di questi criteri dovrebbe essere adattato in base all'età e al livello di competenza linguistica degli studenti.

La rubrica generata include aspetti per valutare elaborati di un livello intermedio-avanzato, e altri integrativi da calibrare in contesti specifici. Ovvero, la comprensione del contenuto (punto 1), che nel nostro caso coinvolge *listening* (video-audio) e *reading* (sottotitoli); due criteri di *writing* (punti 2 e 3); tre criteri associati al pensiero logico-analitico (punti 4, 5 e 7); un criterio legato al pensiero riflessivo (punto 6) e un criterio per l'originalità (punto 8). Senz'altro, questa rubrica, sebbene generata sulla base di un vasto database di esempi standardizzati, si presta egregiamente alla nostra sperimentazione, in quanto include numerosi aspetti che, in un contesto di valutazione operato dal docente, richiedono un monte ore significativo.

4.2 Contestualizzazione dell'attività da valutare

Le richieste assegnate a GPT in un contesto accademico, dimostrano la sua validità per l'applicazione pratica in compiti e contesti complessi. La capacità di generare risposte valide, coerenti e pertinenti rappresenta un fattore di inesplorate possibilità (Sudheesh et al. 2023), tra cui quella di delegare all'IA le attività ripetitive e monotone, quali la valutazione degli elaborati scritti, liberando così tempo da impiegare in attività intellettualmente più stimolanti o creative (Iskender 2023).

Nel nostro caso, la risorsa da visionare è un video presente su YouTube. Tuttavia, sebbene GPT sia in grado di esaminare i contenuti di un video su questa piattaforma, può solo coglierne i temi chiave, mentre non è in grado di generare la trascrizione completa dei contenuti. Pertanto, abbiamo scaricato la trascrizione del video direttamente da YouTube, quindi copiato e incollato tutto come *prompt*, inclusi tutti i *timestamp*, che indicano il segmento temporale in cui il testo viene pronunciato nel video.

> **Prompt** il testo che sto per caricare è la trascrizione di un video che gli studenti sono stati invitati a guardare. Dopo averlo guardato devono rispondere a 7 domande. Devi valutare le loro risposte e applicare la rubrica per valutare l'intervento di ogni studente. Gli studenti sono identificati da una 's' seguita da un numero.
> **Output** va bene, ho capito. Per iniziare, per favore incolla la trascrizione del video, le 7 domande a cui gli studenti dovevano rispondere. Dopo di che, potrai condividere i testi degli studenti uno alla volta e io ti aiuterò a valutarli utilizzando la rubrica.
> **Prompt** questa è la trascrizione del video (segue copia e incolla dell'intera sequenza della trascrizione).
> **Output** grazie per aver condiviso la trascrizione del video. Ora, potrebbe fornire le 7 domande a cui gli studenti dovevano rispondere e la rubrica che devo usare per la valutazione?
> **Prompt** (copia e incolla delle 7 domande in lingua inglese – vedi allegato)[2]
> **Output** grazie per aver condiviso le domande. Ora, potrebbe fornire la rubrica che devo usare per valutare le risposte degli studenti?
> **Prompt** (segue il copia e incolla dell'intera rubrica prima generata da GPT)
> **Output** perfetto, grazie per aver fornito la rubrica. Ora, potresti condividere i testi degli studenti uno alla volta e io ti aiuterò a valutarli utilizzando la rubrica fornita.

2 Link all'allegato: `https://drive.google.com/file/d/16GgkP8jOy5lsQX1zAHI9HN7pJEciD0cN/view?usp=sharing`.

4.3 Valutazione dei singoli testi

Procediamo incollando in GPT il primo testo dello studente da valutare secondo i criteri della rubrica. Quindi, una volta ottenuto l'output, procediamo nello stesso modo con i testi successivi. Notiamo come GPT concluda la prima valutazione con una frase conativa; questo commento compare sistematicamente a chiusura di ogni valutazione, ma lo abbiamo eliminato dalle successive. Ricordiamo che GPT è progettato per imitare lo stile e le connotazioni del linguaggio umano in base ai dati a cui è stato esposto. Questo include fattori come il vocabolario, la sintassi ma anche il registro. Pertanto, può imitare stili specifici, come quello formale, informale, accademico o colloquiale, a seconda del contesto. Il corsivo per distinguere le frasi in inglese è nostro.

Prompt (segue il copia e incolla del testo dello studente s2, visionabile in allegato)[3]
Output grazie per aver condiviso il primo testo degli studenti. Ecco la mia valutazione basata sulla rubrica fornita:

1. Comprensione del contenuto: il testo dimostra una buona comprensione dei concetti principali presentati nel video.
2. Correttezza linguistica:
– Grammatica: l'uso della grammatica inglese è generalmente corretto, ma ci sono alcune frasi che potrebbero essere migliorate, come "*It has been reported that neuroplasticity is negatively related to new habits and knowledge because the brain forgets information, details, over time*".
– Ortografia: non ci sono errori di ortografia evidenti.
– Punteggiatura: l'uso della punteggiatura è appropriato.
3. Ricchezza lessicale: il vocabolario usato è abbastanza vario e appropriato al contesto.
4. Coerenza e coesione:
– Coerenza: le idee sono logiche e portano a conclusioni chiare.
– Coesione: la coesione del testo potrebbe essere migliorata con un uso più efficace dei connettivi.
5. Capacità critica e analisi: lo studente ha principalmente riassunto il contenuto del video senza mostrare molto pensiero critico o analisi personale.
6. Riflessione personale: mancano collegamenti significativi con le esperienze personali dello studente o riflessioni personali sull'apprendimento delle lingue.
7. Adesione alle istruzioni: il testo sembra rispettare le istruzioni fornite in termini di lunghezza e formato, ma manca una

3 Vedi nota 2.

risposta alla domanda *"Can you apply the content of this talk to improve your learning habits? How?"*.
8. Originalità: lo studente ha rielaborato i contenuti del video piuttosto che integrare idee o prospettive personali originali.

Spero che questo ti sia d'aiuto! Se hai altri testi da valutare, non esitare a condividerli!

Prompt (segue il copia e incolla del testo dello studente s29, visionabile in allegato)[4]
Output grazie per aver condiviso il secondo testo degli studenti. Ecco la mia valutazione basata sulla rubrica fornita:
1. Comprensione del contenuto: il testo dimostra una buona comprensione dei concetti principali presentati nel video.
2. Correttezza linguistica:
– Grammatica: l'uso della grammatica inglese è corretto.
– Ortografia: non ci sono errori di ortografia evidenti.
– Punteggiatura: l'uso della punteggiatura è appropriato.
3. Ricchezza lessicale: il vocabolario usato è vario e appropriato al contesto.
4. Coerenza e coesione:
– Coerenza: le idee sono logiche e portano a conclusioni chiare.
– Coesione: i connettivi sono utilizzati in modo efficace per legare frasi e paragrafi tra di loro.
5. Capacità critica e analisi: lo studente ha riassunto il contenuto del video, ma ha anche mostrato un certo livello di pensiero critico, specialmente nella risposta alla domanda 6.
6. Riflessione personale: lo studente ha collegato le informazioni presentate nel video ai propri interessi personali, mostrando un livello di riflessione personale.
7. Adesione alle istruzioni: il testo rispetta le istruzioni fornite in termini di lunghezza e formato.
8. Originalità: anche se gran parte del saggio riguarda la rielaborazione delle informazioni presentate nel video, lo studente include anche alcune riflessioni personali originali (ad esempio, l'interesse per le neuroscienze).

Prompt (segue il copia e incolla del testo dello studente s10, visionabile in allegato)[5]
Output grazie per aver condiviso il terzo testo degli studenti. Ecco la mia valutazione basata sulla rubrica fornita:

4 Vedi nota 2.
5 Vedi nota 2.

1. Comprensione del contenuto: il testo dimostra una buona comprensione dei concetti principali presentati nel video, anche se alcune risposte potrebbero essere più dettagliate.
2. Correttezza linguistica:
– Grammatica: ci sono alcuni errori grammaticali e di sintassi, come "*the our brain changes*" o "*that once time we knew*".
– Ortografia: non ci sono errori di ortografia evidenti.
– Punteggiatura: l'uso della punteggiatura è appropriato.
3. Ricchezza lessicale: il vocabolario usato è adeguato, ma non particolarmente vario o avanzato.
4. Coerenza e coesione:
– Coerenza: le idee sono in genere logiche, ma alcune risposte potrebbero essere più chiare.
– Coesione: i connettivi sono utilizzati in modo abbastanza efficace per legare frasi e paragrafi tra loro.
5. Capacità critica e analisi: lo studente ha principalmente riassunto il contenuto del video senza mostrare molto pensiero critico o analisi personale.
6. Riflessione personale: c'è qualche riflessione personale nell'ultima risposta, ma potrebbe essere sviluppata ulteriormente per mostrare una maggiore connessione con l'esperienza dello studente.
7. Adesione alle istruzioni: li testo sembra rispettare le istruzioni fornite in termini di lunghezza e formato, anche se alcune risposte (come la numero 2) potrebbero non aver affrontato pienamente la domanda posta.
8. Originalità: non vi è molta originalità nelle risposte dello studente; le sue osservazioni si basano principalmente sul contenuto del video invece che su idee o prospettive uniche dell'allievo.

4.4 Attribuzione di una valutazione globale e una votazione numerica

Secondo uno studio di Dai et al. (2023), GPT è in grado di fornire un feedback dettagliato e coerente sulle prestazioni degli studenti, nonché di raggiungere un elevato grado di corrispondenza nei criteri di valutazione proposti dal docente. Inoltre, i risultati di Watts et al. (2023) rivelano che i chatbot generativi rispondono in modo diverso ai vari compiti assegnati e, in genere, gli output non sono il risultato di ragionamenti meccanici e univoci. In altri termini, l'IA generativa adegua le dinamiche di analisi alle particolarità del compito richiesto.

In questa fase, chiediamo a GPT di formulare una valutazione secondo il criterio dei 'voti'. Notiamo che nel *prompt* iniziale non abbiamo richiesto una valutazione descrittiva, che invece GPT fornisce

motu proprio per giustificare la valutazione numerica finale; così come la scelta di optare per la valutazione in decimi.

> **Prompt** sulla base della rubrica e della tua analisi, che votazione attribuisci a questi tre studenti?
> **Output** basandomi sulla rubrica e sulla mia analisi, suggerirei le seguenti valutazioni:
> Studente s2: questo studente ha dimostrato una buona comprensione del contenuto e ha fornito risposte grammaticalmente corrette, ma ha mancato in parte nell'analisi critica e nella riflessione personale. Potrebbe beneficiare di un maggiore impegno nel collegare il materiale con le proprie esperienze o riflessioni personali. Inoltre, non ha risposto pienamente a tutte le domande fornite. Valutazione: 6/10.
> Studente s29: questo studente ha dimostrato una forte comprensione del contenuto e un'ottima correzione linguistica. Ha anche offerto alcune riflessioni personali, ma avrebbe potuto fare un uso più efficace dell'analisi critica nel suo approccio ai video. Valutazione: 8/10.
> Studente s10: questo studente ha mostrato una comprensione adeguata del contenuto, ma la sua valutazione è limitata da vari errori grammaticali e da mancanza di profondità sia nelle sue analisi che nelle sue riflessioni personali. Valutazione: 5/10.
> Ricorda che queste sono solo linee guida basate sulla mia interpretazione della rubrica; come insegnante hai la discrezione finale sull'attribuzione delle valutazioni!

5 Discussione

Il chatbot GPT ha superato le funzionalità tradizionali orientate al linguaggio umano offerte da strumenti come DeepL, InstaText o Grammarly. Questo avanzato sistema va oltre la semplice correzione grammaticale, lessicale e stilistica, introducendo un'analisi semantica sofisticata che consente di valutare sia il significato connotativo sia denotativo di un testo. Questa caratteristica rappresenta una svolta epocale, poiché permette all'IA di elaborare interpretazioni complesse, prima ritenute prerogativa esclusiva dell'intelligenza umana e dell'esperienza degli educatori.

Un aspetto rivoluzionario del chatbot GPT è la sua capacità di interagire con gli utenti attraverso il linguaggio naturale, eliminando la necessità di dominare i linguaggi di programmazione complessi e aprendo le porte dell'IA a settori al di fuori delle tradizionali aree specializzate. Questo livello di interattività democratizza l'accesso all'IA e rende possibile utilizzare la tecnologia anche per coloro che non sono esperti di informatica.

Basandoci sulla nostra esperienza pratica, abbiamo formulato una serie di osservazioni utili per i docenti che desiderano utilizzare GPT in ambito educativo. Attualmente, il sistema si dimostra particolarmente efficace nel supportare una serie di compiti tipici della professione docente, tra cui la progettazione didattica, la valutazione dei testi scritti, l'analisi di grandi quantità di dati testuali e la profilazione degli studenti basata su tali dati. Grazie alla sua capacità di inferire significati dai contenuti forniti, GPT offre risultati obiettivi e coerenti, operando deduttivamente ed esclusivamente attraverso l'analisi del testo inserito.

Nonostante l'interfaccia intuitiva del chatbot possa far pensare che non sia necessaria una formazione specifica per i docenti, è fondamentale condurre studi pilota per identificare le integrazioni didattiche più efficaci in diverse discipline e per comprendere come formulare i *prompt* in modo ottimale. È anche essenziale sviluppare un quadro normativo chiaro che definisca i limiti dell'integrazione dell'IA nelle pratiche curriculari e didattiche, ponendo un'enfasi particolare sull'aspetto etico. Poiché le operazioni eseguite con l'IA non sono soggette a verifica indipendente o validazione esterna, esiste il rischio che siano percepite come frutto del lavoro umano anziché dell'assistenza digitale. Pertanto, è urgente avviare un dibattito aperto sull'uso responsabile e sulla regolamentazione dell'IA nel campo dell'istruzione, al fine di garantire che questa tecnologia sia utilizzata in modo etico e responsabile.

Per gli studenti, l'introduzione di strumenti avanzati come GPT potrebbe generare preoccupazioni. Da un lato, potrebbero accogliere positivamente l'idea di avere a disposizione un aiuto tecnologico che semplifica e velocizza il processo di apprendimento e di produzione di contenuti scritti. Tuttavia, potrebbero anche sorgere dubbi riguardo alla parità di trattamento e all'equità nell'istruzione, specialmente se alcuni studenti hanno accesso a risorse tecnologiche più avanzate rispetto ad altri. Inoltre, potrebbero sorgere domande sul ruolo dell'insegnante laddove un'intelligenza artificiale può svolgere compiti educativi in modo autonomo. Dall'altro lato, i docenti potrebbero sentirsi sfidati o addirittura minacciati dall'introduzione di strumenti di IA nell'ambiente educativo. Potrebbero temere che il loro ruolo venga ridimensionato o che l'autorità professionale venga compromessa dall'uso di tecnologie che sembrano sostituire l'esperienza e la competenza umane. Allo stesso tempo, potrebbero percepire l'IA come un valido strumento di supporto, in grado di alleggerire il carico di lavoro e fornire indicazioni preziose per migliorare le prestazioni degli studenti. Inoltre, entrambi gli attori potrebbero condividere preoccupazioni etiche riguardo alla privacy e alla sicurezza dei dati.

In definitiva, l'integrazione dell'IA nell'ambito educativo solleva questioni complesse che vanno oltre la semplice efficacia didattica. È essenziale coinvolgere gli studenti, i docenti e altre parti interessate

in un dialogo aperto e inclusivo, per affrontare queste preoccupazioni etiche in modo costruttivo e garantire che l'uso dell'IA avvenga nel rispetto dei valori e degli obiettivi dell'istruzione.

6 Prospettive di ricerca

La recente ricerca di Hashem et al. (2023) esplora l'efficacia di ChatGPT come assistente didattico per ridurre il carico di lavoro e prevenire il *burnout* negli insegnanti delle scuole secondarie. L'articolo rileva che la pianificazione delle lezioni e lo sviluppo dei contenuti rappresentano le principali cause di *burnout*, e valorizza l'integrazione dei chatbot nella pianificazione personalizzata e nella collaborazione tra IA e intelligenza umana. Le future linee di ricerca dovrebbero focalizzarsi sull'implementazione di esperienze formative personalizzate, atte a modulare i contenuti didattici, le tempistiche e le metodologie d'insegnamento in base alle necessità individuali degli studenti. Questo potrebbe coinvolgere l'integrazione di sistemi in grado di monitorare costantemente il livello di competenza degli studenti e modulare dinamicamente il loro percorso formativo. La competenza 'imparare a imparare' può beneficiare dell'interazione con l'IA, così come avviene con un tutor personale. Inoltre, la gamma delle modalità per supportare l'insegnamento può includere l'aiuto nella progettazione delle lezioni, la valutazione delle performance dei discenti e la gestione della classe. È fondamentale che tali tecnologie siano guidate da appropriate teorie pedagogiche ed empiricamente testate (Luan et al. 2020). Ulteriormente, un ambito promettente riguarda la valutazione automatizzata: strumenti in grado di esaminare aspetti quali la comprensione testuale, la creatività e il pensiero critico possono contribuire a rendere più agevole ed equa la valutazione globale del lavoro dello studente. Le tecnologie *proctoring* hanno già trovato applicazione nel settore universitario, al fine di mantenere elevata l'integrità accademica; queste potrebbero essere estese anche all'istruzione secondaria attraverso linee guida appropriate. Le strategie basate sul *proctoring* continuativo governato dall'IA possono fornire indicazioni significative ai docenti sul momento più opportuno per intervenire nel processo formativo. Queste direzioni di ricerca, nonostante non siano esaurienti, possono fornire importanti spunti per plasmare un futuro in cui l'IA assuma un ruolo sempre più integrato e trasparente nel contesto dell'evoluzione dei sistemi educativi.

Per quanto riguarda il fenomeno delle *hallucination*, l'attendibilità di GPT varia notevolmente a seconda che i dati siano ricercati all'interno di un corpus caricato direttamente nel chatbot o recuperati dall'esterno, come sul web. Nel primo caso, l'IA lavora su un insieme definito di dati che l'utente ha caricato (ad esempio, un PDF,

Word o contenuti testuali inseriti direttamente, come nel nostro caso di studio). In questi contesti, l'IA può estrarre, cercare e analizzare solo le informazioni fornite, e i limiti contestuali si traducono in un vantaggio a favore dell'accuratezza, in quanto se l'informazione desiderata non è nell'input, l'IA non può rispondere. Per questo, attualmente GPT si presenta più affidabile nel ragionamento deduttivo, rispetto a quello induttivo (Bang et al. 2023). In questo scenario, occorre specificare che GPT offre due configurazioni di base, quella standard (utilizzata da molti utenti) che non ha il browsing abilitato e quella avanzata, che richiede la sottoscrizione. La versione standard può rispondere solo basandosi sui dati che ha visto fino alla data del suo ultimo addestramento (che, in molti ambiti, è limitato a dati raccolti fino a settembre 2021). Le risposte in questo caso sono limitate a ciò che il modello ha 'imparato' durante l'addestramento e non può accedere a nuove informazioni o eventi recenti. Invece, le versioni di GPT che prevedono la sottoscrizione includono la funzionalità di browsing, quindi a ogni immissione di un prompt si avvia una ricerca in tempo reale anche nelle fonti web. Nei casi maggiormente documentati nei quali l'IA cerca informazioni all'esterno, ovvero su internet,[6] la molteplicità delle risorse, inclusi siti web, banche dati pubbliche, notizie e altre risorse disponibili online, copre una gamma più ampia di riferimenti rispetto a un singolo documento. In questi scenari, l'IA può interfacciarsi con motori di ricerca o applicazioni per ottenere informazioni pertinenti, ma il risultato può dipendere dalla qualità e dall'attualità delle informazioni disponibili al momento della ricerca. Inoltre, l'IA potrebbe imbattersi in risorse non affidabili o in contenuti di bassa qualità. Altresì, il ruolo del rumore informativo non deve essere sottovalutato, in quanto l'IA potrebbe ottenere risultati meno pertinenti a causa di contenuto eccessivo o ambiguo. Pertanto, un ambito di ricerca confacente al settore educativo riguarda proprio gli aspetti associati all'affidabilità delle risposte e alle condizioni atte a favorirne l'attendibilità.

6 Athaluri et al. 2023; Bang et al. 2023; Deng et al. 2024; Judge et al. 2023; Maleki et al. 2024.

Bibliografia

Adeshola, I.; Adepoju, A.P. (2023). «The Opportunities and Challenges of ChatGPT in Education». *Interactive Learning Environments*, 1-14.
https://doi.org/10.1080/10494820.2023.2253858

Aljanabi, M.; Ghazi, M.; Hussein Ali, A.; Abas Abed, S. (2023). «ChatGpt: Open Possibilities». *Iraqi Journal for Computer Science and Mathematics*, 4(1), 62-4.
https://doi.org/10.52866/ijcsm.2023.01.01.0018

Athaluri, S.A.; Manthena, S.V.; Kesapragada, V.S.R.K.M.; Yarlagadda, V.; Dave, T.; Duddumpudi, R.T.S. (2023). «Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References». *Cureus*, 15(4).
https://doi.org/10.7759/cureus.37432

Bang, Y.; Cahyawijaya, S.; Lee, N.; Dai, W.; Su, D.; Wilie, B.; Lovenia, H.; Ji, Z.; Yu, T.; Chung, W.; Do, V.Q.; Xu, Y.; Fung, P. (2023). "A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity". *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*. Nusa Dua: Association for Computational Linguistics, 675-718.
https://doi.org/10.18653/v1/2023.ijcnlp-main.45

Bhattacharyya, S.; De, S.; Mršić, L.; Pan, I.; Muhammad, K.; Mukherjee, A.; Konar, D. (2022). «A Quantitative Assessment of Rubrics Using a Soft Computing Approach». *Preprints*.
https://doi.org/10.20944/preprints202210.0197.v1

Bozkurt, A.; Akgun-Ozbek, E.; Yilmazel, S.; Erdogdu, E.; Ucar, H.; Guler, E.; Sezgin, S.; Karadeniz, A.; Sen-Ersoy, N.; Goksel-Canbek, N.; Dincer, G.D.; Ari, S.; Aydin, C.H. (2015). «Trends in Distance Education Research: A Content Analysis of Journals 2009-2013». *The International Review of Research in Open and Distributed Learning*, 16(1), 330-63.
https://doi.org/10.19173/IRRODL.V16I1.1953

Chen, L.; Chen, P.; Lin, Z. (2020). «Artificial Intelligence in Education: A Review». *IEEE Access*, 8, 75264-78.
https://doi.org/10.1109/ACCESS.2020.2988510

Chen, X.; Xie, H.; Zou, D.; Hwang, G.-J. (2020). «Application and Theory Gaps During the Rise of Artificial Intelligence in Education». *Computers and Education: Artificial Intelligence*, 1.
https://doi.org/10.1016/j.caeai.2020.100002

Chuah, K.M.; Kabilan, M.K. (2021). «Teachers' Views on The Use of Chatbots to Support English Language Teaching in a Mobile Environment». *International Journal of Emerging Technologies in Learning (IJET)*, 16(20), 223-37.
https://doi.org/10.3991/IJET.V16I20.24917

Cooper, G. (2023). «Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence». *Journal of Science Education and Technology*, 32(3), 444-52.
https://doi.org/10.1007/s10956-023-10039-y

Crawford, J.; Cowling, M.; Allen, K.-A. (2023). «Leadership is Needed for Ethical ChatGPT: Character, Assessment, and Learning Using Artificial Intelligence (AI)». *Journal of University Teaching and Learning Practice*, 20(3).
https://doi.org/10.53761/1.20.3.02

Dai, W.; Lin, J.; Jin, F.; Li, T.; Tsai, Y.; Gašević, D.; Chen, G. (2023). «Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT».
https://doi.org/10.35542/osf.io/hcgzj

Deng, K.; Huang, Z.; Li, C.; Lin, C.; Gao, M.; Rong, W. (2024). «PFME: A Modular Approach for Fine-grained Hallucination Detection and Editing of Large Language Models». *ArXiv, abs/2407.00488*.
https://doi.org/10.48550/arXiv.2407.00488

Dwivedi, Y.K.; Kshetri, N.; Hughes, L.; Slade, E.L.; Jeyaraj, A.; Kar, A.K.; Baabdullah, A.M.; Koohang, A.; Raghavan, V.; Ahuja, M.; Wirtz, J.; Wright, R. (2023). «"So What If ChatGPT Wrote it?". Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy». *International Journal of Information Management*, 71.
https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ezzaim, A.; Dahbi, A.; Haidine, A.; Aqqal, A. (2023). «AI-Based Adaptive Learning: A Systematic Mapping of the Literature». *JUCS – Journal of Universal Computer Science*, 29(10), 1161-97.
https://doi.org/10.3897/JUCS.90528

Kondo, T.; Nishigori, H.; Vleuten, C.V.D. (2022). «Locally Adapting Generic Rubrics for the Implementation of Outcome-based Medical Education: A Mixed-Methods Approach». *BMC Medical Education*, 22(1).
https://doi.org/10.1186/s12909-022-03352-4

Halaweh, M. (2023). «ChatGPT In Education: Strategies for Responsible Implementation». *Contemporary Educational Technology*, 15(2).
https://doi.org/10.30935/cedtech/13036

Hashem, R.; Ali, N.; El Zein, F.; Fidalgo, P.; Abu Khurma, O. (2023). «AI to the Rescue: Exploring the Potential of ChatGPT as a Teacher Ally for Workload Relief and Burnout Prevention». *Research and Practice in Technology Enhanced Learning*, 19, 23.
https://doi.org/10.58459/rptel.2024.19023

İskender, A. (2023). «Holy or Unholy? Interview with OpenAI's ChatGPT». *European Journal of Tourism Research*, 34, 3414.
https://doi.org/10.54055/ejtr.v34i.3169

Joseph, S.; Rickett, C.; Northcote, M.T.; Christian, B. (2019). «'Who are You to Judge My Writing?': Student Collaboration in the Co-Construction of Assessment Rubrics». *New Writing*, 17(1), 31-49.
https://doi.org/10.1080/14790726.2019.1566368

Judge, B.; Jones, P.; McCreery, E. (2023). «Hallucination in AI and Its Impact on Academic Reliability». *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 15(4).
https://doi.org/10.21659/rupkatha.v15n4.17

Luan, H.; Geczy, P.; Lai, H.; Gobert, J.; Yang, S.J.H.; Ogata, H.; Baltes, J.; Guerra, R.; Li, P.; Tsai, C.C. (2020). «Challenges and Future Directions of Big Data and Artificial Intelligence in Education». *Frontiers in Psychology*, 11, 580820.
https://doi.org/10.3389/FPSYG.2020.580820/BIBTEX

Maleki, N.; Padmanabhan, B.; Dutta, K. (2024). «AI Hallucinations: A Misnomer Worth Clarifying». *2024 IEEE Conference on Artificial Intelligence (CAI)* (Singapore, Singapore), 133-8.
https://doi.org/10.1109/CAI59869.2024.00033

Melhárt, D.; Togelius, J.; Mikkelsen, B.; Holmgård, C.; Yannakakis, G. (2023). «The Ethics of AI in Games». *Ieee Transactions on Affective Computing*, 1-14.
https://doi.org/10.1109/taffc.2023.3276425

O'Dea, X.; O'Dea, M. (2023). «Is Artificial Intelligence Really the Next Big Thing in Learning and Teaching in Higher Education? A Conceptual Paper». *Journal of University Teaching and Learning Practice*, 20(5).
https://doi.org/10.53761/1.20.5.05

Ouyang, F.; Jiao, P. (2021). «Artificial Intelligence in Education: The Three Paradigms». *Computers and Education: Artificial Intelligence*, 2, 100020.
https://doi.org/10.1016/J.CAEAI.2021.100020

Papadopoulos, I.; Lazzarino, R.; Miah, S.; Weaver, T.; Thomas, B.; Koulouglioti, C. (2020). «A Systematic Review of the Literature Regarding Socially Assistive Robots in Pre-tertiary Education». *Computers and Education*, 155.
https://doi.org/10.1016/j.compedu.2020.103924

Sharma, D.K.; Patel, P.K.; Shah, M. (2023). «A Comprehensive Study on Industry 4.0 In the Pharmaceutical Industry for Sustainable Development». *Environmental Science and Pollution Research*, 30(39), 90088-98.
https://doi.org/10.1007/s11356-023-26856-y

Sudheesh, R.; Mujahid, M.; Rustam, F.; Shafique, R.; Chunduri, V.; Villar, M.G.; Ballester, J.B.; Diez, I.D.l.T.; Ashraf, I. (2023). «Analyzing Sentiments Regarding ChatGPT Using Novel BERT: A Machine Learning Approach». *Information*, 14(9), 474.
https://doi.org/10.3390/info14090474

Sullivan, M.; Kelly, A.; McLaughlan, P. (2023). «ChatGPT In Higher Education: Considerations for Academic Integrity and Student Learning». *Journal of Applied Learning and Teaching*, 6(1), 31-40.
https://doi.org/10.37074/jalt.2023.6.1.17

Tlili, A.; Shehata, B.; Adarkwah, M.A.; Bozkurt, A.; Hickey, D.T.; Huang, R.; Agyemang, B. (2023). «What If the Devil is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education». *Smart Learning Environments*, 10(1).
https://doi.org/10.1186/s40561-023-00237-x

Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. (2019). «Systematic Review of Research on Artificial Intelligence Applications in Higher Education – Where are the Educators?». *International Journal of Educational Technology in Higher Education*, 16(1).
https://doi.org/10.1186/s41239-019-0171-0

Zhai, X.; Chu, X.; Chai, C.S.; Jong, M.S.Y.; Istenic, A.; Spector, M.; Liu, J.B.; Yuan, J.; Li, Y. (2021). «A Review of Artificial Intelligence (AI) in Education from 2010 to 2020». *Complexity*.
https://doi.org/10.1155/2021/8812542

Watts, F.M.; Dood, A.J.; Shultz, G.V.; Rodriguez, J.G. (2023). «Comparing Student and Generative Artificial Intelligence Chatbot Responses to Organic Chemistry Writing-to-learn Assignments». *Journal of Chemical Education*, 100(10), 3806-17.
https://doi.org/10.1021/acs.jchemed.3c00664

Yassin, E.; Gharieb, A.; Saad, H.; Qura, O. (2022). «Robots, Artificial Intelligence, and Service Automation (Raisa) Technologies in the Egyptian Hotel Sector: A Current Situation Assessment». *International Journal of Heritage, Tourism and Hospitality*, 16(1), 51-60.
https://doi.org/10.21608/ijhth.2022.287420