

INTEGRARE IL *PROMPT ENGINEERING* TRA LE ABILITÀ TRASVERSALI PER L’EDUCAZIONE LINGUISTICA: UNO STUDIO DI CASO

Marco Mezzadri, Mariapaola Paita¹

1. INTRODUZIONE

La natura interdisciplinare che caratterizza l’intelligenza artificiale (IA) richiede un approccio “implicativo” volto a indagare criticamente il suo impatto nei processi di apprendimento e insegnamento, favorendo al contempo l’elaborazione di nuove prospettive e pratiche di interazione con queste tecnologie nel campo dell’Educazione Linguistica (Mezzadri, Paita, 2025). In tal senso, l’integrazione dei *Large Language Models* (LLM) non è da considerarsi meramente di natura tecnico-applicativa, ma presuppone l’acquisizione di conoscenze e competenze comunicative, metacognitive e trasversali che permettano una collaborazione dinamica e sinergica con l’IA, inclusa l’IA generativa (IAG) (Wang *et al.*, 2024). Tra queste, la capacità di mediazione e il pensiero critico basato su un approccio di tipo *problem solving* diventano competenze fondamentali per poter comunicare efficacemente e responsabilmente con i modelli linguistici generativi. ChatGPT, con cui è stata sviluppata la ricerca oggetto di questo contributo, ne è al momento l’esempio più conosciuto.

Alla luce di queste riflessioni preliminari, l’articolo si propone di esplorare come l’implementazione dei LLM in ambito edulinguistico richieda una comprensione approfondita delle dinamiche di interazione, del loro potenziale educativo e delle loro limitazioni. Particolare attenzione è rivolta a identificare le competenze trasversali e le strategie di mediazione richieste per ottimizzare l’interazione con i LLM, con l’obiettivo, perseguito da insegnanti e apprendenti, di promuoverne un uso critico ed efficace di tali strumenti. Questi obiettivi sono stati esplorati nell’ambito di una ricerca sperimentale, che costituisce il nucleo del presente contributo proponendosi di analizzare le interazioni tra gli apprendenti e ChatGPT durante attività di potenziamento delle abilità di scrittura accademica, al fine di identificare le competenze trasversali e digitali legate all’*ingegneria del prompt* (Chen *et al.*, 2025). Sulla base dei dati raccolti, il contributo propone un modello operativo delle competenze trasversali per un uso consapevole dei LLM nell’Educazione Linguistica, adottando una prospettiva orientata allo sviluppo del pensiero critico, dell’autonomia e di una sinergia educativa tra l’umano e l’IA.

2. POTENZIALITÀ E LIMITI DELL’INTELLIGENZA ARTIFICIALE GENERATIVA IN AMBITO EDULINGUISTICO

La rapida e pervasiva evoluzione della IAG, con particolare riferimento ai LLM, solleva importanti questioni riguardo agli effetti trasformativi di tali tecnologie in ambito

¹ Università degli Studi di Parma.

L’articolo è stato realizzato con il contributo congiunto di entrambi gli autori. Tuttavia, i §§ 1 e 2 sono da attribuire a Marco Mezzadri; i §§ 3, 4, 5, 6 a Mariapaola Paita e il § 7 a entrambi.

educativo (Lim *et al.*, 2023). Le capacità innovative dei LLM nel campo del Trattamento Automatico del Linguaggio, tra cui la capacità di comprendere e generare testi coerenti e contestualmente rilevanti attraverso input testuali, o *prompt* (P. Liu *et al.*, 2021), basati sul linguaggio naturale, ha reso imprescindibile una riflessione critica sulle potenzialità e sulle criticità derivanti dalla loro integrazione nei processi di apprendimento e nelle pratiche didattiche (Kasneci *et al.* 2023). In tal senso, diversi studi hanno contribuito a delineare le opportunità e i rischi connessi al loro utilizzo per scopi educativi (Adiguzel *et al.*, 2023; Farrokhnia *et al.*, 2023).

2.1. *Potenzialità e benefici*

Tra i benefici più significativi, emerge la capacità di promuovere l'autonomia e il senso di auto-efficacia degli apprendenti, grazie alla possibilità di costruire percorsi di apprendimento personalizzati e adattivi (Chiu *et al.*, 2023). Questi modelli possono agire come tutor virtuali, capaci di adattare l'interazione alle caratteristiche e ai bisogni del discente (Farrokhnia *et al.*, 2023). Possono, ad esempio, creare contenuti e materiali didattici personalizzati e offrire *feedback* mirati e in tempo reale (Rawas, 2024). Inoltre, Lai *et al.* (2023) suggeriscono come la didattica personalizzata facilitata dai LLM stimoli gli apprendenti a riflettere attivamente sul proprio processo di apprendimento, favorendo al contempo un aumento della motivazione intrinseca, dell'interesse e della partecipazione attiva.

Diversi studi hanno anche evidenziato come la capacità di comprendere e generare testo in tempo reale favorisca lo sviluppo delle competenze linguistiche e strategiche degli apprendenti (Barrot, 2024). Innanzitutto, i LLM possono agire come assistenti virtuali durante attività di produzione scritta, offrendo opportunità di *brainstorming* durante la fase di pianificazione o suggerimenti per la revisione del testo in termini di correttezza morfosintattica, coerenza e coesione testuale (Su *et al.*, 2023). Possono inoltre assumere il ruolo di partner comunicativi, grazie alla loro capacità di simulare dialoghi interattivi che si adattano al livello di competenza linguistica degli apprendenti (Shaikh *et al.*, 2023; Young, Shishido, 2023). Autori come Steele (2023) e Kasneci *et al.* (2023) sottolineano infine come i LLM possano contribuire al potenziamento delle abilità ricettive, attraverso la sintesi e l'analisi testuale, supportando al contempo gli educatori nella progettazione di attività mirate allo sviluppo della comprensione scritta.

2.2. *Limiti e rischi*

Parallelamente, la letteratura ha evidenziato diversi rischi legati all'utilizzo dei LLM nella didattica (Farrokhnia *et al.*, 2023; Kasneci *et al.*, 2023). Un primo aspetto riguarda la possibilità che un uso acritico e strumentale conduca a una dipendenza e a una diminuzione di competenze trasversali, come il *problem solving*, il pensiero critico e la creatività (Tlili *et al.*, 2023). Per esempio, in assenza di una comprensione dei meccanismi interni che regolano il modello e delle competenze necessarie per valutare criticamente la qualità dell'interazione, gli apprendenti potrebbero accettare in modo passivo la veridicità e l'accuratezza delle risposte, le quali possono talvolta contenere informazioni errate, incomplete o fittizie (Anderson *et al.*, 2024). Questo rischio è amplificato dalla difficoltà ad accedere in modo trasparente ai processi decisionali che determinano gli *output* dei LLM, spesso descritti come *black box* proprio per l'opacità che caratterizza il loro funzionamento (Ribeiro *et al.*, 2016).

Un ulteriore ambito di criticità riguarda la dimensione etica (Vaccino-Salvadore, 2023). In particolare, emergono questioni rilevanti in merito alla salvaguardia della *privacy* e alla protezione dei dati sensibili degli utenti (Ge, 2024). Gli apprendenti, infatti, potrebbero divulgare involontariamente informazioni personali, che rischiano di essere esposte a violazioni o utilizzi impropri, come la generazione di contenuti illeciti (Ray, 2023). Un esempio emblematico è rappresentato dal fenomeno del *jailbreaking*, ovvero la manipolazione intenzionale dei *prompt* da parte degli utenti per aggirare i sistemi di sicurezza e di moderazione integrati nei LLM, ottenendo così risposte non filtrate e potenzialmente dannose o illegali (Y. Liu *et al.*, 2024).

Ci sentiamo di condividere questa visione in merito alle aree di criticità menzionate e crediamo sia opportuno enfatizzare ulteriormente la centralità dello sviluppo delle abilità trasversali legate al pensiero critico in campo educativo, in particolare nei percorsi di educazione linguistica. L'avvicinamento tra due mondi apparentemente distanti non può avvenire se non con scelte di tipo “implicativo”. Come ricorda Balboni (2011: 24), «l’implicazione rimanda al meccanismo *se... allora...*, è quindi selettiva e molto più responsabilizzante per il soggetto che trae implicazioni da aree esterne». In quest’ottica, abbiamo già avuto modo di affermare che

la glottodidattica non si limita alle fascinazioni derivanti da altri ambiti scientifici, tecnici o di altro tipo, né ne è vittima, bensì riflette, prima di applicarli, sulle implicazioni che l’adozione di determinati strumenti, tecniche ecc. determina nella didattica delle lingue Mezzadri (2024: 390).

3. IL RUOLO DELL’INGEGNERIA DEL PROMPT NELL’INTERAZIONE CON I *LARGE LANGUAGE MODELS*

Alla luce delle opportunità e dei rischi discussi, appare evidente che l’integrazione dei LLM nei processi di apprendimento e insegnamento delle lingue necessita non solo di una consapevolezza del potenziale tecnico riguardo allo strumento, ma anche, e soprattutto, di una riflessione più ampia sulle competenze digitali e trasversali che educatori e apprendenti devono acquisire per interagire in modo efficace, critico e consapevole con la IAG. Tra queste competenze rientrano la conoscenza e la comprensione dei meccanismi di funzionamento dei modelli linguistici, nonché la capacità di interagire con essi tramite la progettazione di *prompt* testuali (Tseng, Warschauer, 2023; Warschauer *et al.*, 2023). In questo contesto, il campo emergente dell’ingegneria del *prompt*, o *prompt engineering*, si configura come

the systematic design and optimization of input *prompts* to guide the response of LLMs, ensuring high levels of accuracy, relevance, coherence and usability in the generated *output*. This process is crucial for harnessing the full potential of these models, making them more accessible and applicable across diverse domains (Chen *et al.*, 2025: 2).

La ricerca in questo campo si propone di identificare le strategie di *prompting* più efficaci attraverso l’analisi delle interazioni con i modelli linguistici generativi, con l’obiettivo di acquisire una comprensione più approfondita dei loro meccanismi interni di funzionamento (Linardatos *et al.*, 2021). Da questa prospettiva, il *prompt* si pone pertanto come strumento di mediazione nell’interazione tra l’umano e la IAG, permettendo di orientare in modo intenzionale e strategico il comportamento dei LLM (Mondal *et al.*, 2024). In particolare, dal momento che minime variazioni nella struttura e nel contenuto

del *prompt* possono influenzare significativamente l'*output* generato, un fenomeno noto come *prompt sensitivity* (Zhao *et al.*, 2021; Lu *et al.*, 2022), la qualità dell'interazione dipende dalla capacità di progettare *prompt* efficaci, che possano guidare il LLM durante l'esecuzione di un *task* specifico, generando risposte rilevanti e coerenti rispetto agli obiettivi e alle intenzioni degli utenti (White *et al.*, 2023; M. Wang *et al.*, 2024). Per queste ragioni, il *prompt engineering* emerge come elemento centrale, rendendo necessaria un'esplorazione più approfondita e sistematica delle competenze digitali e trasversali richieste per promuovere un uso efficace e consapevole di queste tecnologie in ambito educativo (Cain, 2024; Walter, 2024).

3.1. Il *prompt engineering* come competenza emergente della IAG

Nella riflessione sulla *Human-AI Interaction* (HAI), il *prompt engineering* si configura come una competenza emergente e una forma di mediazione tra l'umano e la IAG, strettamente connessa allo sviluppo di *AI literacy skills* necessarie per interagire efficacemente con i modelli linguistici (Jiang *et al.*, 2024). Nel contesto educativo, l'importanza di tale competenza risulta centrale negli studi che analizzano le interazioni tra apprendenti e LLM. Essi riportano, in particolare, una scarsa consapevolezza delle strategie di *prompting* da parte di utenti inesperti, spesso inclini a un uso disorganico e opportunistico dei modelli, basato su un approccio per tentativi ed errori (Dang *et al.*, 2022; Zamfirescu-Pereira *et al.*, 2023). Le difficoltà riscontrate sono attribuibili, da un lato, all'antropomorfizzazione del modello, che porta a impostare l'interazione come un dialogo naturale piuttosto che una pianificazione strategica basata sulla comprensione del funzionamento dei LLM e sulla progettazione di *prompt* efficaci (Knoth *et al.*, 2024, Sawalha *et al.*, 2024); dall'altro, alla limitata capacità di valutare criticamente l'*output* generato e di migliorarne la qualità attraverso un processo iterativo di ridefinizione del *prompt* (Zamfirescu-Pereira *et al.*, 2023). Sulla base di queste considerazioni, il *prompt engineering* è riconducibile quindi a una «non-intuitive skill that people first need to acquire before it can be applied in meaningful ways» (Oppenlaender *et al.*, 2024: 3) e che deve essere promossa all'interno dei percorsi di alfabetizzazione critica all'IA.

La crescente consapevolezza del ruolo cruciale del *prompt engineering* nell'interazione con i modelli linguistici ha spinto la ricerca a elaborare diversi *framework* teorici e linee guida per guidare l'utente nella formulazione dei *prompt*. Alcuni contributi, come il *framework* CLEAR (Lo, 2023) si soffermano sulle caratteristiche strutturali e funzionali che un *prompt* dovrebbe includere per orientare efficacemente l'interazione con i LLM. In questo contesto si colloca anche il modello AIRPROMPT (Korzyński *et al.*, 2023), che individua gli elementi chiave e propone linee guida per la progettazione di un *prompt*. A un livello più tecnico, Anderson *et al.* (2024) analizzano i meccanismi interni dei LLM, mettendo in evidenza i fattori che ne influenzano la performance e che possono essere modellati attraverso un uso consapevole delle tecniche di *prompting*. In modo analogo, White *et al.* (2023) forniscono un catalogo di *prompt pattern*, basato sulle caratteristiche dei LLM, che gli utenti possono adattare a seconda del contesto comunicativo e del task da svolgere. Altri studi, infine, descrivono il *prompt engineering* come un'attività iterativa di costruzione, sperimentazione e affinamento, utile per migliorare progressivamente la qualità dell'*output* generato dal modello (Eager, Brunton, 2023; Cain, 2024). Tuttavia, Zamfirescu-Pereira *et al.* (2023) evidenziano la necessità di una sperimentazione sistematica del *prompt engineering* in contesti autentici e con utenti inesperti, sottolineando l'importanza di esplorare le dinamiche di interazione e le difficoltà operative che emergono dall'utilizzo dei LLM.

3.2. La ricerca sul prompt engineering nell'Educazione Linguistica

Spostandoci in ambito edulinguistico, la ricerca sul *prompting* è ancora agli inizi, ma è possibile individuare due principali aree di interesse. La prima riguarda l'uso dei LLM per la generazione di materiali e proposte didattiche volte a promuovere le competenze linguistiche degli apprendenti². A titolo esemplificativo, la ricerca condotta da L. Wang *et al.* (2024) sull'accuratezza del *feedback* di ChatGPT in risposta alle abilità di scrittura argomentativa evidenzia come specifiche caratteristiche dell'input, tra cui l'assenza di connettivi, l'elevata complessità argomentativa o la lunghezza eccessiva dei testi, possano ostacolare la capacità del modello di valutare con precisione la coerenza e la coesione testuale. Altri contributi hanno esplorato le potenzialità offerte dai LLM nelle diverse fasi di scrittura, dalla pianificazione alla revisione, suggerendo strategie per migliorare la qualità del *feedback* attraverso la formulazione più accurata dei *prompt* (Su *et al.*, 2023). Questi studi riconoscono l'importanza del *prompt engineering* nel potenziare le abilità generative dei modelli linguistici; tuttavia, si concentrano sull'analisi dei comportamenti del modello, senza esplorare le strategie di *prompting* e il loro effetto sull'*output* con una maggiore sistematicità.

La seconda area si nutre della ricerca sull'alfabetizzazione critica all'IA e si concentra sull'identificazione delle competenze trasversali necessarie per interagire con i LLM (Cain, 2024; Walter, 2024). Tra queste, la conoscenza e la capacità di elaborare strategie di *prompting* adeguate sono considerate competenze emergenti per un'interazione efficace con i LLM; ciononostante, diversi studi evidenziano come né gli apprendenti né gli educatori dispongano attualmente di un livello adeguato di tali competenze, determinando una limitazione del potenziale formativo di queste tecnologie ed evidenziando la necessità di ulteriori approfondimenti in relazione al *prompt engineering* (Ou *et al.*, 2024; M. Liu *et al.*, 2025). In questa direzione, lo studio condotto da M. Liu *et al.* (2025) riporta la difficoltà degli apprendenti nell'ottimizzare l'interazione con i modelli generativi a causa una scarsa familiarità con le strategie di *prompting*, sottolineando la necessità di fornire conoscenze mirate e linee guida operative per la costruzione di *prompt* efficaci e funzionali all'apprendimento linguistico. Analogamente, il modello ASPE per l'alfabetizzazione critica all'IA proposta da Wang e Wang (2025) insiste sull'importanza di sviluppare una consapevolezza critica delle potenzialità e dei limiti dei LLM, nonché delle strategie mirate di interazione, considerando l'introduzione al *prompt engineering* come competenza chiave per un uso efficace e riflessivo della IAG.

4. IL DISEGNO DI RICERCA

4.1. *Contesto*

Il presente studio si colloca all'interno di una sperimentazione realizzata tra ottobre e dicembre 2023 e frutto della collaborazione tra l'Università di Parma e il Liceo Scientifico Statale "G. Ulivi". L'obiettivo principale della sperimentazione consisteva nell'indagare l'impatto dell'IA, tramite l'uso di ChatGPT, sull'acquisizione e il potenziamento delle competenze di scrittura argomentativa. Il progetto ha coinvolto un totale di sei classi (quattro del quarto anno e due del secondo) suddivise in un gruppo sperimentale (58 studenti) e un gruppo di controllo (38 studenti). Tutti i partecipanti hanno dato il consenso informato per il trattamento anonimo dei dati a fini di ricerca. Per entrambi i gruppi sono

² Si vedano tra i contributi disponibili gli studi condotti da Lee *et al.* (2023) e Liu (2023).

state predisposte le stesse attività e gli stessi materiali didattici, condividendo i medesimi obiettivi formativi sulla produzione e la revisione di un testo argomentativo. La differenza principale ha riguardato l'integrazione di ChatGPT, poiché solamente gli studenti del gruppo sperimentale hanno avuto accesso al modello linguistico durante lo svolgimento delle attività didattiche.

4.2. *Partecipanti*

Lo studio prende in esame un corpus di 58 interazioni con ChatGPT realizzate dagli studenti appartenenti al gruppo sperimentale. In particolare, i partecipanti hanno redatto due testi argomentativi su tracce fornite dagli autori e senza l'ausilio della IAG e in seguito hanno interrogato individualmente il modello per ricevere un *feedback* personalizzato sul proprio elaborato. Complessivamente, sono state raccolte e analizzate 58 interazioni, suddivise come segue: 36 interazioni hanno riguardato la valutazione di un testo argomentativo sull'impatto dei *social media* nella società contemporanea, di cui 25 provenienti da studenti di classe quarta e 11 da studenti di classe seconda. Le restanti 22 interazioni si sono focalizzate sulla valutazione di un testo argomentativo sul tema della tecnologia e dell'automazione nel futuro del lavoro, con 14 interazioni da parte di studenti di classe quarta e 8 da studenti di classe seconda.

Sebbene ciascun partecipante abbia interagito con ChatGPT per ottenere un *feedback* su entrambi gli elaborati, non è stato possibile raccogliere tutte le interazioni per l'analisi. Tale limitazione è motivata dal fatto che la raccolta dei dati è avvenuta al termine della sperimentazione, quando alcune conversazioni con il modello non erano più accessibili per la condivisione con utenti esterni. Tuttavia, riteniamo il presente corpus adeguato agli scopi della ricerca, in quanto la tipologia testuale, così come gli aspetti formali e relativi alla struttura argomentativa oggetto di valutazione, sono invariati in entrambe le attività, indipendentemente dalla differenza tematica delle tracce proposte.

4.3. *Obiettivi e domande di ricerca*

Nel contesto della sperimentazione, la ricerca si propone di esplorare le variabili che contribuiscono all'interdipendenza tra la formulazione del *prompt* e l'*output* generato da ChatGPT durante la valutazione dei testi argomentativi prodotti dai partecipanti. L'indagine adotta un approccio di tipo *evidence-based* per analizzare l'impatto che il processo iterativo di costruzione e di adattamento del *prompt* da parte degli studenti può esercitare sul comportamento del LLM in termini di qualità e pertinenza del feedback sulla struttura argomentativa e sugli aspetti formali del testo. I risultati emersi dall'analisi delle interazioni tra gli studenti e ChatGPT costituiscono la base per una riflessione critica circa le competenze trasversali rese necessarie per una collaborazione efficace e consapevole con la IAG nell'ambito dell'Educazione Linguistica. In particolare, questo approccio "implicativo", basato sull'osservazione delle strategie di *prompting* adottate dai partecipanti e arricchito dalle riflessioni teoriche presenti nella letteratura scientifica di riferimento, ha permesso di delineare un modello operativo delle competenze di *prompt engineering*. Il modello intende offrire una cornice di riferimento per mediare l'interazione con i LLM secondo una prospettiva critica a supporto dei processi di apprendimento e insegnamento delle lingue. Pertanto, la ricerca si basa sulle seguenti domande:

1. Quali variabili influenzano l'interdipendenza tra la formulazione del *prompt* e la qualità delle risposte di ChatGPT?

2. Quali criticità emergono nel processo iterativo di costruzione e adattamento del *prompt* e in che modo influenzano la qualità dell'interazione con ChatGPT?
3. Quali competenze trasversali emergono dall'osservazione delle interazioni tra i partecipanti e ChatGPT?

4.2. Metodologia

La ricerca si articola in due fasi distinte ma interconnesse. La prima è finalizzata ad analizzare l'interdipendenza tra la formulazione del *prompt* e la qualità delle risposte generate da ChatGPT nella valutazione dei testi argomentativi. La seconda si concentra sull'osservazione del processo iterativo di costruzione e adattamento dei *prompt*, con l'obiettivo di individuare potenziali criticità riscontrate dai partecipanti nel corso dell'interazione. Entrambe le fasi concorrono a far emergere, attraverso un approccio *evidence-based*, le competenze trasversali necessarie per progettare, valutare e ottimizzare l'interazione con i LLM, delineando così un quadro operativo a supporto dell'integrazione efficace e consapevole di queste tecnologie in ambito edulinguistico.

La prima fase consiste nell'analisi integrata delle risposte generate da ChatGPT e dei *prompt* elaborati dai partecipanti. Inizialmente, è stata esaminata la qualità del *feedback* del modello per determinare quali aspetti della scrittura argomentativa venissero esplicitamente menzionati nell'*output* e con quale frequenza. A tale scopo, gli autori hanno predisposto una griglia di osservazione costruita sulla base degli obiettivi didattici del percorso formativo erogato durante la sperimentazione, con l'obiettivo di valutare sistematicamente due aree complementari:

- aspetti legati alla struttura argomentativa, tra cui l'introduzione, il *thesis statement*, lo sviluppo di argomentazioni e controargomentazioni, gli esempi a sostegno della tesi e la conclusione;
- aspetti formali, quali la chiarezza espositiva e la coesione testuale, l'adeguatezza del registro linguistico, l'accuratezza grammaticale e ortografica, l'uso appropriato della punteggiatura e la struttura sintattica delle frasi.

Per ogni aspetto sono state elaborate categorie specifiche, corredate da descrittori (Tabelle 1 e 2). La griglia è stata quindi utilizzata per valutare il grado di aderenza del *feedback* di ChatGPT alle caratteristiche della tipologia testuale, consentendo di rilevare la frequenza con cui il modello faceva riferimento esplicito a ciascuna categoria e di osservare eventuali variazioni nella distribuzione degli aspetti chiave all'interno delle risposte in relazione alle strategie di *prompting* adottate dai partecipanti.

Parallelamente, i *prompt* sono stati analizzati qualitativamente per determinare l'influenza delle variazioni linguistiche e strutturali sulla qualità del *feedback*. A tal fine, sono stati presi in considerazione tre ambiti principali, individuati sulla base della letteratura scientifica di riferimento (O'Connor, Andreas, 2021; Eager, Brunton, 2023; Korzynski *et al.*, 2023; Lo, 2023; Anderson *et al.*, 2024). Il primo fattore riguarda le componenti del *prompt*, che includono un'istruzione chiara e dettagliata relativa al compito da svolgere, le informazioni su cui si fonda il *task* (dati di *input*), e il contesto, che orienta il modello rispetto alle modalità di esecuzione del *task*. Il linguaggio rappresenta il secondo elemento chiave e richiede precisione, esaustività e chiarezza nella definizione degli obiettivi e della tipologia di attività desiderata. Infine, si è osservata la struttura del *prompt* in termini di organizzazione testuale e complessità del compito assegnato. L'analisi delle interazioni sulla base di questi fattori ha permesso di osservare eventuali cambiamenti nella qualità del *feedback* rispetto alle diverse strategie di *prompting* adottate dai partecipanti.

La seconda fase si è concentrata sull'esplorazione del processo iterativo di costruzione e adattamento del *prompt* in relazione al *feedback* di ChatGPT. In particolare, sono state analizzate qualitativamente le variazioni tra *input* e *output* durante il corso delle interazioni per identificare eventuali criticità riscontrate dai partecipanti nel raggiungimento degli obiettivi prestabiliti. L'osservazione ha riguardato sia la capacità dei partecipanti di valutare criticamente la pertinenza e la qualità delle risposte ricevute, sia le strategie adottate per affinare progressivamente la formulazione del *prompt* in risposta al comportamento del modello.

5. ANALISI DEI RISULTATI E DISCUSSIONE

5.1. *Interdipendenza tra la qualità del feedback e la formulazione del prompt*

L'analisi delle interazioni rivela uno sbilanciamento marcato nella distribuzione percentuale degli aspetti relativi alla struttura argomentativa inclusi nel *feedback* di ChatGPT, con un'attenzione maggiore posta sulle controargomentazioni e gli esempi a supporto della tesi (Tabella 1). Osservando la frequenza con cui il modello menziona ciascun elemento, è possibile notare valori superiori al 50% per la presenza di controargomentazioni (69%), l'equilibrio tra argomentazioni e controargomentazioni (56%), la credibilità e la validità degli esempi forniti (75%), nonché la loro capacità di sostenere in modo efficace la tesi (52%). Al contrario, aspetti legati all'introduzione (23% per la chiarezza della tesi e 17% per il contesto), alle argomentazioni (13% per lo sviluppo logico e 21% per la pertinenza rispetto alla tesi) e alla conclusione (27% per il riassunto delle argomentazioni e il *thesis re-statement*, 29% per l'invito all'azione o alla riflessione) risultano inclusi nel *feedback* con una frequenza minore del 30%. In particolare, il *thesis statement* registra valori molto bassi (10% per chiarezza e specificità e 4% per la pertinenza rispetto alla tesi), suggerendo una maggiore difficoltà del modello nell'identificare e nel valutare gli aspetti che richiedono una distinzione esplicita tra la tesi e l'antitesi.

Tabella 1. *Frequenza degli aspetti relativi alla struttura argomentativa nel feedback di ChatGPT*

Criteri di valutazione	Descrittori	Frequenza (%)
Introduzione	Chiarezza della tesi	23
	Contesto	17
Thesis statement	Chiarezza e specificità	10
	Spunti per la riflessione	0
	Pertinenza rispetto alla tesi	4
Argomentazioni	Forza delle argomentazioni	46
	Sviluppo logico	13
	Pertinenza rispetto alla tesi	21
Controargomentazioni	Presenza di controargomentazioni	69
	Efficacia delle confutazioni	37
	Approccio bilanciato	56
Esempi	Pertinenza rispetto alla tesi	40
	Credibilità e validità	75
	Sufficiente supporto alla tesi	52
Conclusione	Riassunto delle argomentazioni	27
	Thesis re-statement	27
	Invito all'azione o riflessione	29

La frequenza degli aspetti formali inclusi nel *feedback* di ChatGPT mostra una distribuzione percentuale fortemente sbilanciata a favore della chiarezza e della coesione testuale (Tabella 2). In particolare, possiamo osservare una maggiore frequenza nel menzionare gli elementi relativi all'organizzazione testuale (92%), seguita dalla transizione tra paragrafi (48%) e dalla progressione logica delle idee (37%). Inoltre, è possibile notare una riduzione della frequenza con cui vengono menzionati gli aspetti relativi alla chiarezza sintattica (23%) e alla complessità sintattica (17%), mentre l'accuratezza grammaticale e nell'uso della punteggiatura compaiono nel *feedback* con una frequenza compresa tra il 19% e il 15%. Un ulteriore aspetto interessante emerge in relazione agli elementi del registro linguistico, che risultano sottorappresentati; in particolare, la frequenza con cui vengono inclusi il linguaggio formale e accademico, l'assenza di colloquialismi e l'uso di un lessico accademico assume un valore pari o inferiore al 10%.

Tabella 2. *Frequenza degli aspetti formali nel feedback di ChatGPT*

Criteri di valutazione	Descrittori	Frequenza (%)
Chiarezza e coesione	Organizzazione testuale	92
	Progressione logica delle idee	37
	Transizione tra paragrafi	48
Registro	Linguaggio formale	10
	Linguaggio accademico	10
	Assenza di linguaggio colloquiale	6
	Lessico accademico	10
Grammatica	Accuratezza grammaticale	19
Punteggiatura	Accuratezza nell'uso della punteggiatura	15
Ortografia	Accuratezza ortografica	2
Sintassi	Chiarezza sintattica	23
	Complessità sintattica	17

La distribuzione delle frequenze relative agli aspetti formali e di contenuto presenti nel *feedback* di ChatGPT sembra riflettere, almeno in parte, alcune caratteristiche legate al funzionamento e ai meccanismi interni dei LLM. La maggiore attenzione posta sulla coesione testuale, l'accuratezza grammaticale e la sintassi potrebbero essere attribuite alla natura probabilistica dei modelli linguistici generativi, i quali producono *output* testuali calcolando statisticamente la sequenza di parole più probabile sulla base della distribuzione statistica delle parole nei corpora testuali utilizzati durante l'addestramento e alla loro architettura interna, senza tuttavia disporre di una reale comprensione del significato semantico del testo (Bender, Koller, 2020; P. Liu *et al.*, 2021). Ciò potrebbe indurre il modello a focalizzarsi maggiormente sugli aspetti formali che non richiedono una contestualizzazione ulteriore della tipologia testuale, come nel caso degli aspetti relativi al registro.

Questa caratteristica può inoltre contribuire a interpretare la distribuzione delle frequenze osservata nel *feedback* in relazione agli aspetti strettamente connessi alla tesi e all'antitesi, soprattutto se la distinzione tra le due non è esplicitamente formulata nel *prompt*. L'incapacità dei LLM di comprendere le sfumature semantiche e di rappresentare sul piano concettuale le relazioni logiche tra le diverse parti del discorso (L. Wang *et al.*, 2024), unita al principio di funzionamento basato sulla predizione della sequenza linguistica più plausibile, sembra favorire una tendenza del modello a concentrarsi

maggiormente nel fornire un *feedback* sulla qualità degli esempi forniti e sull'approccio bilanciato tra gli argomenti a favore e contrari, piuttosto che valutarne la pertinenza o la chiarezza rispetto alla tesi.

Un ulteriore fattore che ha influito negativamente sulla capacità di ChatGPT di fornire risposte rilevanti e pertinenti rispetto al *task* è attribuibile alla formulazione del *prompt* da parte di utenti inesperti. Poiché i partecipanti non hanno ricevuto una formazione preliminare né sulle strategie di *prompting* né sulle potenzialità e sulle limitazioni dei LLM, i *prompt* utilizzati per completare l'attività sono stati oggetto di variazioni linguistiche e strutturali. In particolare, l'analisi delle interazioni ha permesso di identificare una serie di caratteristiche dei *prompt*, tra cui la mancanza di componenti principali, l'uso di un linguaggio ambiguo o impreciso e una struttura errata delle richieste rivolte al modello, nonché una difficoltà da parte dei partecipanti nell'ottimizzare iterativamente l'interazione stessa.

5.1.1. *Variazioni nelle componenti del prompt*

La mancanza di componenti essenziali nel *prompt* (cfr. Korzynski *et al.*, 2023; Lo, 2023) può contribuire alle variazioni nella qualità del *feedback* rispetto agli aspetti formali e relativi alla struttura argomentativa. Questo fenomeno si riscontra in particolar modo quando nella formulazione del *prompt* non vengono riportate informazioni contestuali, istruzioni e dati di *input* necessari per orientare il adeguatamente il LLM rispetto al *task* da svolgere. Nello specifico, la mancanza di istruzioni chiare, dettagliate e precise sul *task* influenza negativamente sulla capacità di ChatGPT di valutare in modo efficace ed esauritivo il testo argomentativo. Di seguito riportiamo tre esempi di *output* generati dal modello in risposta a *prompt* privi di istruzioni³:

Esempio 1. ChatGPT	<p>Capisco le preoccupazioni riguardo al futuro del lavoro e l'impatto dell'IA sull'occupazione. È vero che il mondo del lavoro è in continua evoluzione e alcuni settori stanno sperimentando una riduzione dei posti di lavoro a causa dell'automazione e dell'introduzione di tecnologie avanzate.</p> <p>Tuttavia, c'è anche un lato positivo: l'IA e l'automazione possono liberare le persone da compiti ripetitivi e consentire loro di concentrarsi su mansioni più creative e complesse. È essenziale che i ragazzi vengano preparati per questo cambiamento. La scuola può svolgere un ruolo cruciale nell'insegnare loro non solo le competenze digitali, ma anche la flessibilità, la creatività e la capacità di adattamento, che sono competenze fondamentali nel mondo del lavoro moderno.</p> <p>Sì, l'IA potrebbe comportare una riduzione di alcuni tipi di lavoro, ma allo stesso tempo crea nuove opportunità. Ad esempio, emergono nuove professioni legate all'IA e alla gestione dei dati che richiedono competenze specifiche. Invece di competere con le macchine, i giovani potrebbero imparare a lavorare con esse, sviluppando le competenze necessarie per gestire e migliorare le tecnologie intelligenti.</p> <p>È importante trovare un equilibrio tra l'utilizzo dell'IA per migliorare l'efficienza e la produttività e garantire che le persone non vengano sostituite completamente dai robot. Ciò potrebbe significare investire in</p>
---------------------------	--

³ Tutti gli esempi relativi alla formulazione dei *prompt*, o estratti di essi, sono stati riportati mantenendo la punteggiatura e la formattazione originali.

	<p>nuove forme di istruzione e formazione continua per adattarsi a questo cambiamento.</p> <p>La comprensione delle implicazioni dell'IA e dell'automazione sul lavoro è essenziale per affrontare le sfide future. È un tema complesso e richiede un approccio olistico che coinvolga l'istruzione, le politiche pubbliche e l'innovazione per trovare soluzioni sostenibili e garantire opportunità per i giovani nel mondo del lavoro.</p>
Esempio 2. ChatGPT	<p>Hai toccato diversi punti importanti riguardo all'innovazione tecnologica e all'uso dei robot nel mondo del lavoro. Ecco alcune riflessioni e considerazioni:</p> <p>1. Efficienza e precisione: I robot sono noti per la loro efficienza e precisione nelle attività assegnate. Possono svolgere compiti ripetitivi in modo costante senza stancarsi, riducendo gli errori umani e aumentando la produttività.</p> <p>2. Svantaggi e sfide: Come hai sottolineato, i robot richiedono tempo per essere costruiti e programmati, possono rompersi e rappresentano una spesa significativa iniziale. Inoltre, la dipendenza eccessiva dai robot potrebbe portare a problemi quando si verificano guasti o malfunzionamenti.</p> <p>[...]</p> <p>In conclusione, l'equilibrio tra l'uso intelligente dei robot e la preservazione delle opportunità di lavoro umano è una sfida chiave per la società. Una gestione attenta e ponderata dell'innovazione tecnologica può massimizzare i benefici senza compromettere il benessere umano e l'occupazione.</p>
Esempio 3. ChatGPT	<p>La tua tesi affronta un dibattito importante sull'automazione del lavoro e il suo impatto sugli adolescenti. L'antitesi mette in luce il potenziale sviluppo tecnologico e la semplificazione delle mansioni più complesse grazie all'automazione, ma evidenzia anche i rischi associati all'affidare responsabilità significative alle macchine.</p> <p>La conclusione sottolinea il possibile impatto negativo sull'esperienza lavorativa degli adolescenti a causa della perdita di opportunità di lavoro e della privazione di esperienze lavorative. Inoltre, si evidenziano i rischi legati alla affidabilità delle macchine e ai potenziali malfunzionamenti.</p> <p>Per rendere la tua argomentazione più completa, potresti considerare l'inclusione di soluzioni o suggerimenti per mitigare gli impatti negativi dell'automazione, come ad esempio l'implementazione di programmi di formazione per le competenze necessarie nel nuovo contesto lavorativo o politiche che promuovano la responsabilità aziendale nel garantire la sicurezza delle macchine.</p> <p>Infine, potrebbe essere utile citare esempi specifici o studi di caso che supportino le tue affermazioni e forniscano concretezza alla discussione. In questo modo, la tua tesi sarà più convincente e fornirà una visione più approfondita della complessità dell'argomento.</p>

Da questi esempi è possibile constatare come la presenza di istruzioni influisca marcatamente sulla capacità di ChatGPT di fornire una valutazione efficace del testo argomentativo. I *prompt* che riportano esclusivamente il testo da analizzare, senza

specificare la tipologia di attività attesa dal modello, tendono a generare *output* caratterizzati da considerazioni personali o riflessioni generali sul contenuto, come negli esempi (1) e (2), senza fornire una valutazione della struttura argomentativa o degli aspetti linguistici specifici della tipologia testuale. Interessante è l'esempio (3), che menziona aspetti relativi alla struttura argomentativa, quali la tesi, la conclusione e le argomentazioni. Nonostante vi sia una maggiore attenzione nel commentare i contenuti espressi nel testo, emerge un focus specifico sugli esempi a sostegno delle argomentazioni. Seppur non esaustivo, questo rimando ad aspetti della struttura argomentativa potrebbe essere attribuito all'organizzazione testuale adottata dallo studente nel *prompt*, che suddivide il testo in paragrafi etichettati rispettivamente come “tesi”, “antitesi” e “conclusione”. È possibile quindi ipotizzare che tali riferimenti abbiano orientato, almeno parzialmente, la risposta di ChatGPT nonostante a mancanza di istruzioni esplicite sul *task*.

Un fenomeno simile è stato riscontrato anche nei *prompt* le cui istruzioni non includono i dati di *input*. In questo caso specifico, facciamo riferimento alle interazioni in cui i partecipanti forniscono al modello il proprio testo argomentativo, senza istruzioni, e successivamente formulano un secondo *prompt* sottoforma di domanda “È scritto bene?” senza tuttavia specificare la tipologia testuale da sottoporre a valutazione. Si veda la seguente interazione tra uno studente e ChatGPT:

Esempio 4. Prompt	L'impatto dei social media sulla società contemporanea. Secondo me i social hanno un impatto positivo sulla società contemporanea, promuovendo la connessione e la condivisione di idee. I social hanno un impatto positivo perché possono aiutare le persone più timide a socializzare con più sicurezza. Inoltre aiutano a comunicare nonostante le distanze, che separano le persone. I social sono un luogo di svago dove si possono vedere video divertenti, ovunque si voglia ed in qualunque momento. Oltre a questo possono essere anche un posto di scambio di idee, infatti nei social si possono trovare educatori e divulgatori di idee e notizie [...]
ChatGPT	Hai toccato diversi aspetti importanti riguardo all'impatto dei social media sulla società contemporanea, evidenziando sia gli aspetti positivi che quelli negativi. La tua analisi riflette la complessità di questo argomento e mostra come l'uso dei social media possa variare in base alle modalità di utilizzo [...]
Prompt	è scritto bene ?
ChatGPT	Sì, il tuo testo è scritto in modo chiaro e coerente. Hai presentato argomenti sia a favore che contro l'impatto dei social media sulla società contemporanea e hai articolato le tue idee in modo comprensibile. Inoltre, hai fornito esempi pertinenti e hai concluso con un'analisi equilibrata, sottolineando l'importanza dell'uso consapevole dei social media. Nel complesso, il testo è ben strutturato e esprime efficacemente le tue opinioni sull'argomento.

Nell'esempio (4), è possibile osservare come l'assenza di riferimenti esplicativi alla tipologia testuale (“testo argomentativo”) nei dati di *input* abbia inficiato non solo sulla capacità di ChatGPT di identificare con precisione l'attività richiesta, ma abbia anche inciso sulla frequenza con cui il modello menziona gli aspetti strettamente connessi alla distinzione tra tesi e antitesi. Il *feedback*, infatti, tende a focalizzarsi principalmente sulla credibilità e validità degli esempi forniti e sull'importanza di mantenere un approccio

bilanciato prospettive favorevoli e contrarie alla tematica trattata, nonché sugli aspetti formali legati alla chiarezza e alla coesione testuale.

Un'ulteriore componente chiave riguarda la presenza (o assenza) di informazioni contestuali, che guidano il LLM con maggiore precisione nella valutazione del testo argomentativo. In particolare, le istruzioni relative al testo argomentativo sull'impatto dei social media nella società contemporanea prevedevano una tesi e un'antitesi prestabilite, che i partecipanti dovevano rispettivamente sostenere e confutare. Analogamente, la consegna relativa al testo argomentativo sulla tecnologia e l'automazione nel futuro del lavoro lasciava agli studenti la possibilità di scegliere la tesi da sostenere e al contempo, indicando esplicitamente di utilizzare correttamente i connettivi. Nonostante tali informazioni siano cruciali per orientare ChatGPT nella valutazione più mirata della tipologia testuale in esame, dall'analisi delle interazioni si rileva che soltanto due partecipanti hanno fornito dettagli parziali sulla consegna ricevuta per lo svolgimento dell'attività di scrittura. Si osservino gli esempi (5) e (6):

Esempio 5. <i>Prompt</i>	<p>Scrivi un testo argomentativo su questo tema: L'impatto dei social media sulla società contemporanea. Sviluppa la tesi e l'antitesi riportate sotto, esprimendo accordo con la posizione espressa nella tesi. Dovrai presentare argomentazioni valide ed esempi concreti a sostegno della tesi e contro l'antitesi.</p> <p>Il crescente sviluppo dei social media ha determinato diversi cambiamenti nello stile di vita delle persone e nel modo in cui vivono le relazioni sociali. Questo fenomeno culturale ha aperto un dibattito sugli effetti della interconnessione tecnologica nella vita quotidiana [...]</p>
Esempio 6. <i>Prompt</i>	<p>Io ho scritto un testo argomentativo dove ho esposto una tesi e un'antitesi sul seguente tema:</p> <p>La tecnologia e l'automazione nel futuro del lavoro</p> <p>La tesi a favore dell'uso della tecnologia e dell'automazione del lavoro, così come l'antitesi con le opinioni di chi è contrario a queste innovazioni le ho create partendo da questi tre spunti:</p> <ul style="list-style-type: none">- Quali sono le opportunità e le sfide per gli adolescenti nel mondo del lavoro del futuro?- Come si può preparare la prossima generazione per affrontare la crescente automazione?- Quali competenze sono cruciali per il successo professionale? <p>Ecco il mio testo: [...]</p>

In entrambi gli esempi possiamo osservare come gli studenti abbiano riportato le istruzioni ricevute per la stesura del proprio elaborato. Mentre nell'esempio (5) il partecipante si limita a riportare la consegna originale nel *prompt*, nell'esempio (6) vi è una rielaborazione, seppur minima, delle istruzioni ricevute. Tuttavia, in entrambi gli esempi, il contesto fornito a ChatGPT risulta poco esaustivo e funzionale per una valutazione efficace del testo argomentativo, ma sembra piuttosto concentrarsi sul processo di scrittura anziché sugli aspetti strutturali e linguistici da analizzare. Inoltre, la modalità con cui viene riportata la consegna nell'esempio (5), ovvero la frase "Scrivi un testo argomentativo sul seguente tema", potrebbe limitare la capacità del modello di comprendere correttamente il *task* richiesto. Infatti, tale formulazione potrebbe essere interpretata da ChatGPT come un'istruzione operativa, piuttosto che come

un'informazione contestuale fornita per facilitare la valutazione del testo già prodotto dallo studente⁴. Inoltre, l'assenza di informazioni contestuali nel *prompt* incide negativamente sulla qualità del *feedback* relativo agli aspetti della struttura argomentativa, specialmente quelli che implicano una distinzione esplicita tra la tesi e l'antitesi. In questi casi, ChatGPT tende a enfatizzare maggiormente aspetti quali le controargomentazioni, gli esempi utilizzati e l'importanza di adottare una prospettiva bilanciata nel trattare posizioni favorevoli e contrari alla tesi proposta.

5.1.2. *Variazioni strutturali nel prompt*

Anche le variazioni nella struttura del *prompt* contribuiscono a ridurre la qualità del *feedback* in relazione alla valutazione del testo argomentativo. Queste variazioni sono state riscontrate principalmente nelle interazioni relative al testo argomentativo sull'impatto dei social media, strutturata in tre fasi consecutive: dapprima, i partecipanti richiedevano al LLM una valutazione del proprio elaborato; successivamente domandavano una riscrittura finalizzata a migliorarne la struttura argomentativa; infine chiedevano al modello di spiegare i miglioramenti effettuati. Dall'analisi delle interazioni emerge che la presenza di richieste multiple nello stesso input tende a limitare la capacità di ChatGPT di interpretare e gestire contemporaneamente ciascuna richiesta. Riportiamo due estratti di *prompt* esemplificativi:

Esempio 7. Prompt	<p>[...]</p> <p>In conclusione, i social media sono una realtà con cui l'essere umano deve convivere e che offre molte opportunità. Tuttavia, è necessario essere consapevoli dei loro effetti negativi e adottare un approccio equilibrato e responsabile nel loro utilizzo. Solo così l'uomo è in grado di sfruttarne appieno i benefici.</p> <p>Domanda 1: Puoi valutare il seguente testo argomentativo? È scritto bene?</p> <p>Domanda 2: Puoi riscrivere il testo apportando dei miglioramenti alla struttura argomentativa?</p> <p>Domanda 3: Puoi spiegare i tuoi interventi migliorativi?</p>
Esempio 8. Prompt	<p>Puoi valutare il seguente testo argomentativo? è scritto bene? I SOCIAL E LA SOCIETÀ</p> <p>È da molti anni che si discute sull'impatto dei social media sulla società contemporanea.</p> <p>Secondo il mio parere, condiviso da una buona parte di coloro che hanno partecipato alla discussione, L'impatto dei social media è positivo sulla società contemporanea, poiché promuove la connessione e la condivisione di idee.</p> <p>[...]</p>

⁴ Nonostante le informazioni contestuali nell'esempio (6) possano influenzare negativamente la qualità dell'*output* di ChatGPT rispetto agli obiettivi prefissati, il partecipante ha aggiunto ulteriori istruzioni al *prompt*, inclusa la richiesta di valutare il testo argomentativo. Ciò nonostante, la qualità del *feedback* risulta comunque meno accurata soprattutto a causa di variazioni strutturali nella formulazione del *prompt* (cfr. § 5.1.2).

	<p>L'utilizzo dei social perciò ha secondo me un impatto più che positivo sulla società attuale, in quanto promuove un accrescimento culturale e lo sviluppo di un proprio modo di pensare indipendente e favorisce l'instaurazione di nuove amicizie. Puoi riscrivere il testo apportando dei miglioramenti alla struttura argomentativa? Puoi spiegare i tuoi interventi migliorativi?</p>
--	--

Negli esempi (7) e (8) è possibile notare come la formulazione di un *prompt*, se non adeguatamente strutturato, possa ridurre l'accuratezza del *feedback* di ChatGPT. In entrambi i casi, infatti, l'inserimento di tre *task* distinti in un unico *prompt* ha un impatto negativo sulla capacità del modello di fornire un *feedback* esaustivo e mirato, portandolo spesso a riassumere i contenuti principali del testo o a riscrivere direttamente il testo con gli interventi migliorativi, piuttosto che valutarne gli aspetti formali e di contenuto. Ciò potrebbe essere ricondotto alle limitazioni intrinseche dei LLM nella gestione di *task* complessi e/o più *task* nello stesso *prompt* (Wu *et al.*, 2022), che richiedono processi di ragionamento strutturati e articolati su più fasi intermedie.

5.1.3. Variazioni linguistiche nel *prompt*

Una tendenza a generare *feedback* generici e incompleti viene riscontrata anche in presenza di *prompt* formulati con un linguaggio impreciso o ambiguo. In particolare, facciamo riferimento, da un lato, alle istruzioni che utilizzano frasi quali “è scritto bene” oppure “cosa ne pensi di”, che non forniscono indicazioni dettagliate e precise sull’attività da svolgere e dall’altro, ai dati di input ambigui, come nel caso di “questo testo”, che non specificano la tipologia testuale da valutare. Nell’esempio (9) il *feedback* di ChatGPT è caratterizzato dalla sintesi degli aspetti positivi e negativi e non fornisce indicazioni mirate riguardo agli elementi connessi alla struttura argomentativa. Questa risposta può essere attribuita al *prompt* formulato dallo studente (“Cosa ne pensi di questo testo”), che non esplicita né la natura argomentativa del testo, né l’obiettivo preciso della valutazione. Inoltre, la mancanza di un contesto dettagliato sulla consegna dell’attività potrebbe avere contribuito al commento finale del LLM, da cui non emerge una valutazione del testo sulla base delle istruzioni ricevute e rispetto alla tipologia testuale.

Esempio 9. ChatGPT	<p>Il testo evidenzia sia gli aspetti positivi che quelli negativi legati all’impatto dei social media sulla società contemporanea. L’autore sottolinea come i social media favoriscano la connessione, la condivisione di idee e costituiscano una fonte di cultura. Allo stesso tempo, riconosce gli svantaggi legati alla dipendenza dai dispositivi digitali, all’isolamento digitale e ai rischi associati, specialmente tra i giovani.</p> <p>La gestione dei genitori è indicata come un possibile fattore influente nella prevenzione di casi di dipendenza e isolamento. Tuttavia, la conclusione che le conseguenze positive superino quelle negative potrebbe suscitare dibattiti, poiché la valutazione dipende da prospettive e contesti variabili.</p>
-------------------------------	--

L’imprecisione e l’ambiguità linguistica nella formulazione dei *prompt* possono anche compromettere la capacità di ChatGPT di svolgere efficacemente il compito richiesto. Negli esempi (10) e (11) sono riportati due *feedback* del modello in risposta a *prompt* che

contengono istruzioni e dati di *input* generici (“Questo testo è scritto bene?” e “Cosa ne pensi di questo testo?”). Nel primo caso, la mancanza di istruzioni precise sugli obiettivi del *task* non solo contribuisce alla genericità del *feedback*, come nel caso di “linguaggio chiaro e comprensibile”, ma rende anche difficile per il modello identificare con precisione l’obiettivo del *task*. Analogamente, nel secondo esempio, ChatGPT confonde la tipologia testuale da valutare, interpretando l’elaborato come un saggio piuttosto che come un testo argomentativo. L’importanza di utilizzare un linguaggio preciso e diretto, con un riferimento esplicito al concetto di valutazione e alla tipologia di testo, è evidenziata anche da Anderson *et al.* (2024), il cui studio suggerisce come la presenza di parole chiave che descrivono dettagliatamente il *task*, come “valuta” e “testo argomentativo”, possa aumentare l’accuratezza e la pertinenza dell’*output* generato dai LLM.

Esempio 10. ChatGPT	Sì, il tuo testo è scritto bene. Hai esposto chiaramente i tuoi argomenti, evidenziando sia gli aspetti positivi che quelli negativi legati all’uso dei social media. La struttura del testo è organizzata in paragrafi coerenti, ciascuno trattando un aspetto specifico dell’argomento. Inoltre, hai utilizzato un linguaggio chiaro e comprensibile. Se il tuo obiettivo era comunicare in modo efficace la complessità del tema, direi che hai avuto successo.
Esempio 11. ChatGPT	[...] La tua conclusione, che suggerisce che la tua opinione è la più corretta, aggiunge una nota personale alla tua esposizione. È un modo comune di concludere un saggio, ma potresti voler espandere ulteriormente su come potrebbero essere implementati tali limiti sull’uso dei social media per massimizzare i benefici e ridurre gli svantaggi. [...]

5.1.4. Criticità nell’ottimizzare iterativamente l’interazione

Come precedentemente discusso, diversi studi nel campo del *prompt engineering* in ambito educativo sottolineano l’importanza di sviluppare competenze trasversali in grado di contrastare un uso acritico e inefficace dei LLM (Lo, 2023; Cain, 2024). Tuttavia, l’analisi rivela che in 31 interazioni con ChatGPT, i partecipanti hanno riformulato più di una volta il *prompt* per ottenere un *feedback* sul proprio testo argomentativo, incontrando difficoltà nel valutare l’efficacia delle risposte ricevute e nel migliorare progressivamente la qualità delle richieste. Uno degli ostacoli riscontrati riguarda la formulazione iniziale del *prompt*, che condiziona significativamente la qualità dell’interazione con il LLM (Mondal *et al.*, 2024). I seguenti esempi illustrano due interazioni con ChatGPT:

Fasi dell’interazione	Esempio 12.	Esempio 13.
Prompt	Tre istruzioni contemporaneamente	Mancanza di istruzioni
ChatGPT	Riscrittura del testo con interventi migliorativi e <i>feedback</i> sulla coesione testuale	Riassunto dei contenuti

Prompt	Puoi valutare il seguente testo argomentativo? È scritto bene?	Cosa ne pensi di questo testo?
ChatGPT	Riscrittura del testo con interventi migliorativi e commento sulla coesione testuale e sugli esempi	Riassunto dei contenuti
Prompt		È scritto bene questo testo?
ChatGPT		Riassunto dei contenuti e <i>feedback</i> sulla coesione testuale, sulle argomentazioni e sugli esempi

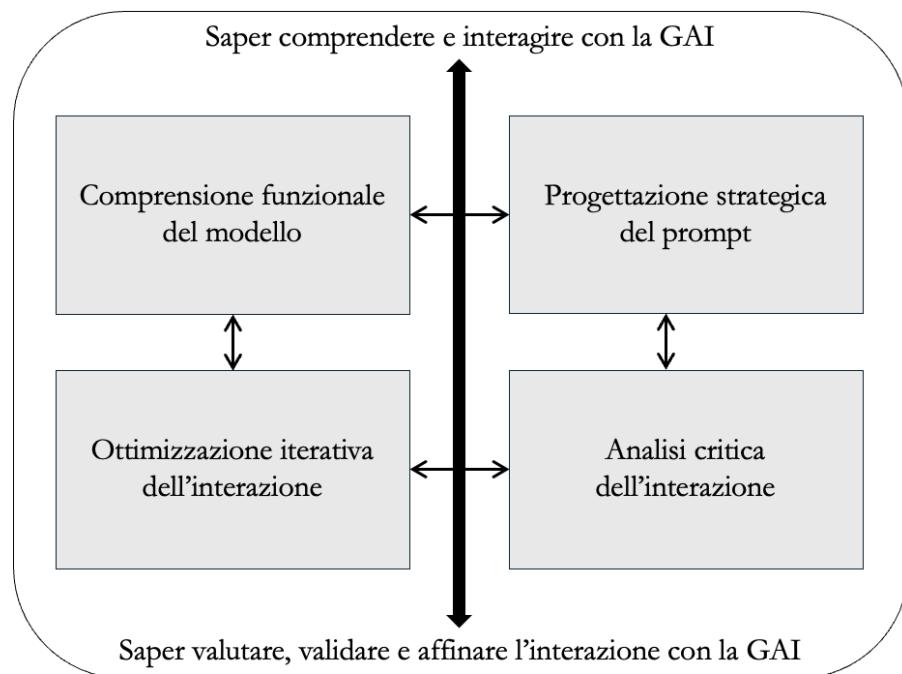
Nell'esempio (12) il *prompt* iniziale presenta variazioni strutturali che ostacolano ChatGPT nella comprensione dell'attività richiesta. Il risultato è una riscrittura del testo argomentativo non in linea con gli obiettivi prefissati dallo studente, che riformula il *prompt* con l'istruzione di valutare il testo argomentativo. Nonostante la risposta di ChatGPT includa un commento sulla coesione testuale e sugli esempi utilizzati, notiamo una tendenza del modello a fornire una versione rivisitata del testo. Non riuscendo a ottimizzare il *prompt* per avere un *feedback* esaustivo sulla struttura argomentativa e gli elementi linguistici, ChatGPT, pertanto, continua a generare la stessa tipologia di *output*.

Un caso simile è osservabile nell'esempio (13), in cui il partecipante formula il *prompt* originale senza istruzioni sul *task* e, ricevendo un riassunto dei contenuti, riformula i due *prompt* successivi, utilizzando tuttavia un linguaggio vago in relazione alla valutazione e alla tipologia testuale. Così, ChatGPT si concentra sul riassumere il testo e soltanto alla fine dell'interazione il *feedback* contiene un riferimento generico alla coesione testuale, alle argomentazioni e agli esempi forniti. Le difficoltà nell'ottimizzare la qualità delle risposte rispetto alla valutazione del testo argomentativo trovano riscontro in quanto emerso dalla letteratura sul *prompt engineering* rispetto alla necessità di promuovere competenze e strategie di *prompting* da parte di apprendenti e insegnanti al fine di integrare in modo critico ed efficace i LLM nel contesto educativo (cfr. tra gli altri: Ou *et al.*, 2024; M. Liu *et al.*, 2025).

6. VERSO UN MODELLO OPERATIVO DELLE COMPETENZE DI PROMPT ENGINEERING

Alla luce dei risultati emersi dall'analisi delle interazioni con ChatGPT, il *prompt engineering* si configura come un'abilità trasversale necessaria per promuovere una collaborazione efficace tra l'umano e la IAG nell'Educazione Linguistica. L'approccio "implicativo" adottato nella presente ricerca ha permesso di individuare quattro competenze chiave che apprendenti ed educatori sono chiamati a sviluppare per integrare in modo consapevole i modelli linguistici generativi nei processi di apprendimento e insegnamento delle lingue. Di seguito si propone una rappresentazione illustrativa delle competenze trasversali associate al *prompt engineering* (Schema 1).

Schema 1. *Modello operativo delle competenze di prompt engineering*



Riteniamo opportuno introdurre una riflessione preliminare che consenta di contestualizzare il modello operativo proposto, anche in riferimento ai principali contributi della letteratura scientifica sul tema. In primo luogo, le quattro competenze riflettono i fondamenti teorici dell'alfabetizzazione critica all'IA (cfr. Ranieri *et al.*, 2023) e si articolano in due dimensioni operative fondamentali. La prima riguarda il saper comprendere e interagire con i LLM (e più in generale con la IAG), pianificando e formulando il *prompt* in relazione alle caratteristiche dei modelli linguistici e agli obiettivi prefissati dall'utente. La seconda dimensione, invece, consiste nel saper valutare, validare o eventualmente affinare l'interazione con i LLM, analizzando criticamente le risposte generate dal modello, verificandone la pertinenza rispetto al compito richiesto e, se necessario, riformulando il *prompt* per migliorare la qualità dell'interazione. In secondo luogo, come riscontrato anche da Wang e Wang (2025), lo sviluppo di tali competenze non segue un processo sequenziale, ma si realizza attraverso un processo dinamico e iterativo, in cui le diverse abilità e conoscenze interagiscono tra loro in maniera complementare e interconnessa. Infine, il modello operativo intende essere un punto di riferimento flessibile e trasferibile (cfr. Eager, Brunton, 2023), fondato su competenze che non si legano a un contesto d'uso predefinito, ma che possono essere acquisite, messe in pratica e adattate a seconda dei contesti e degli obiettivi educativi. Il modello, pertanto, offre un quadro di orientamento teorico-pratico per supportare apprendenti e educatori nello sviluppo di un'interazione mirata e critica con i LLM.

6.1. *Saper comprendere e interagire con la LAG*

La prima competenza chiave per interagire efficacemente con i LLM riguarda lo sviluppo di una comprensione dei meccanismi interni che ne regolano il funzionamento (Gao *et al.*, 2021). Ciò implica avere familiarità con alcuni concetti fondamentali che caratterizzano il comportamento dei modelli, tra cui:

- La natura opaca del loro funzionamento, che rende difficile tracciare in modo chiaro e trasparente i processi che conducono alla generazione di una determinata tipologia di *output* (Ribeiro *et al.*, 2016);
- La sensibilità dei modelli alle variazioni minime nella formulazione dei *prompt*, che può determinare la generazione di risposte diverse anche in presenza di modifiche strutturali e/o linguistiche lievi (Zhao *et al.*, 2021; Lu *et al.*, 2022);
- La presenza e le cause delle allucinazioni, ovvero la generazione di informazioni errate o fittizie, nelle risposte dei modelli, che, seppur plausibili, non risultano coerenti con le richieste dell'utente (Ji *et al.*, 2023).

Comprendere tali fenomeni permette ad apprendenti ed educatori non solo di acquisire una maggiore consapevolezza delle limitazioni connesse a un utilizzo acritico dei LLM (Cain, 2024), ma anche di adottare strategie di *prompting* efficaci al fine di allineare la qualità delle risposte dei modelli alle loro esigenze e agli obiettivi richiesti.

Inoltre, una conoscenza funzionale dei LLM riveste un ruolo fondamentale per la progettazione strategica dei *prompt*, permettendo di sfruttare in maniera consapevole ed efficace il potenziale generativo dei LLM (Cain, 2024). Ciò comporta l'importanza di esplicitare, durante il processo di costruzione del *prompt*, tutte le informazioni rilevanti che possano orientare l'interazione del modello rispetto agli obiettivi e alle aspettative dell'utente (Eager, Brunton, 2023; Mondal *et al.*, 2024). Facendo riferimento ai risultati emersi dalla ricerca, questo processo si traduce operativamente nel fornire al modello tutte le componenti essenziali, adeguatamente strutturate per svolgere il *task*. Per esempio, è necessario fornire istruzioni (e dati di *input*) dettagliate ed esaustive sull'attività didattica (Korzynski *et al.*, 2023), privilegiando parole come “valuta” e “testo argomentativo” rispetto a formulazioni ambigue o generiche (“cosa ne pensi di questo testo?” oppure “è scritto bene?”) che da una lato, possono compromettere la qualità del *feedback* e dall'altro, possono indurre il LLM fornire risposte non pertinenti rispetto all'obiettivo, come commenti soggettivi, riassunti dei contenuti o una misinterpretazione della tipologia testuale.

Anche le informazioni contestuali rivestono un ruolo fondamentale, non solo nel migliorare l'esaustività e l'accuratezza delle risposte dei LLM (O'Connor, Andreas, 2021), ma anche nel ridurre il rischio di *output* errati, incompleti o fittizi (Anderson *et al.*, 2024). Nel caso specifico della valutazione di un testo argomentativo, ciò può significare fornire nel *prompt* indicazioni mirate sul processo valutativo, come l'inclusione di una griglia di criteri per la valutazione e descrizioni dettagliate degli aspetti formali e di contenuto da analizzare (cfr. Su *et al.*, 2023). Un ulteriore elemento che può contribuire a migliorare la qualità del *feedback* consiste nell'esplicitare le istruzioni operative che gli studenti hanno ricevuto per svolgere l'attività di scrittura. Questo tipo di informazioni aiutano, infatti, il LLM verso una valutazione più esaustiva e in linea con gli obiettivi di apprendimento.

Infine, l'organizzazione delle informazioni nel *prompt* rappresenta un fattore essenziale per ottimizzare l'interazione. Nello specifico, variazioni strutturali non adeguate possono influenzare negativamente il modello, il quale, a causa delle limitazioni intrinseche nel suo funzionamento, può generare errori, risposte parziali o incorrere in fenomeni di allucinazione. Per esempio, i *prompt multi-task*, in cui vengono esplicitate più istruzioni contemporaneamente (come valutare, riscrivere e spiegare) tendono a generare una ristruttura del testo, anziché alla generazione di un *feedback* articolato e pertinente sulla struttura argomentativa. Una strategia utile per migliorare la qualità delle risposte nei *task* che richiedono più fasi intermedie, come in questo caso, consiste nel formulare *prompt* come richieste distinte ma logicamente interconnesse (Lo, 2023), in modo che ciascun passo si costruisca sul precedente e guidi progressivamente il modello verso il completamento del *task* (Wu *et al.*, 2022).

6.2. Saper valutare, validare e affinare l'interazione con la LAG

La terza competenza chiave riguarda il saper valutare criticamente l'interazione con i LLM (cfr., tra gli altri: M. Liu *et al.*, 2025; Wang, Wang, 2025), analizzando in modo integrato sia la formulazione del *prompt* sia la qualità delle risposte ottenute, in quanto elementi strettamente interdipendenti. Questa competenza si fonda sullo sviluppo del pensiero critico (Cain, 2024), inteso come la capacità di riflettere consapevolmente sull'interazione tra l'utente e il LLM attraverso la valutazione, la verifica e la validazione dell'affidabilità e della pertinenza dell'*output* rispetto all'*input*. Come sottolinea anche Lo (2023), saper valutare criticamente l'interazione richiede anche l'abilità di identificare e apportare eventuali modifiche nella costruzione del *prompt*, che consentano di migliorare le risposte del modello rispetto agli obiettivi del *task*. Le interazioni analizzate nello studio suggeriscono l'importanza di questa competenza: la mancanza di componenti fondamentali, una struttura inadeguata o l'uso di un linguaggio generico o impreciso all'interno del *prompt* tendono a generare *feedback* incompleti o non pertinenti alla valutazione del testo argomentativo. Analizzare criticamente l'interazione è pertanto un processo esteso, che attraversa tutte le fasi di pianificazione, progettazione e validazione del *prompt*.

La capacità di valutare criticamente l'interazione non riguarda soltanto la progettazione strategica del *prompt*, tenendo presente i punti di forza e di debolezza dei LLM, ma anche la sua ottimizzazione attraverso un processo iterativo che consenta di migliorarne l'efficacia in funzione degli obiettivi del task. In particolare, i LLM possono manifestare un calo nella qualità dell'interazione durante la generazione di testi estesi. Questo fenomeno, noto come *bias di esposizione* (Tan *et al.*, 2021), può manifestarsi quando imprecisioni o errori nella formulazione del *prompt* si accumulano progressivamente nel corso dell'interazione. Per esempio, nel caso del presente studio, il *bias di esposizione* si osserva quando i partecipanti incontrano difficoltà nel riconoscere e nel modificare gli elementi poco funzionali del *prompt* iniziale (cfr. Anderson *et al.*, 2024), continuando a ricevere *output* simili nonostante i tentativi di riformulazione. Questo fenomeno si collega allo sviluppo di conoscenze sulle caratteristiche intrinseche dei LLM, che integrate all'acquisizione di competenze digitali necessarie alla formulazione e revisione critica del *prompt* risultano essenziali per ottimizzare progressivamente l'interazione con i modelli linguistici. Queste competenze, centrali nel modello operativo proposto, mirano alla promozione di una collaborazione produttiva tra l'umano e la IAG in ambito edulinguistico, orientata alla risposta ai bisogni degli apprendenti e al raggiungimento degli obiettivi formativi prefissati.

7. RIFLESSIONI CONCLUSIVE

In questo contributo sono stati presentati gli esiti di una ricerca condotta su un corpus di 58 interazioni tra studenti e ChatGPT al fine di esplorare come l'interdipendenza tra le variazioni linguistiche e strutturali nella formulazione del *prompt* possano influenzare la qualità e l'esaustività delle risposte generate dal modello. È stato osservato che la mancanza di componenti chiave nel *prompt*, le variazioni strutturali e l'utilizzo di un linguaggio impreciso o ambiguo possono compromettere la capacità dei LLM di fornire un *feedback* mirato degli aspetti formali e di contenuto durante la valutazione di un testo argomentativo. Inoltre, l'analisi delle interazioni ha messo in luce le criticità riscontrate dai partecipanti nel valutare criticamente l'efficacia e la pertinenza delle risposte ricevute, nell'ottimizzare iterativamente i *prompt* per allineare l'*output* agli obiettivi prefissati, orientando con maggiore precisione il comportamento del modello rispetto allo

svolgimento dell'attività richiesta. In particolare, i tentativi di riformulare i *prompt* messi in atto dagli studenti suggeriscono un utilizzo poco strategico dei LLM. Questo dato suggerisce la necessità di promuovere una più solida comprensione delle caratteristiche intrinseche di tali strumenti e favorire l'acquisizione di competenze trasversali per una mediazione efficace tra l'umano e la IAG. Alla luce di questi risultati, sono state individuate le competenze trasversali necessarie allo sviluppo delle abilità di *prompt engineering* all'interno dei processi di apprendimento e insegnamento delle lingue. Queste competenze, rappresentate nel modello operativo proposto, riflettono due dimensioni fondamentali dell'alfabetizzazione critica all'IA e devono essere promosse attraverso un percorso dinamico e iterativo, in cui conoscenze, abilità e pensiero critico convergono nel favorire una mediazione efficace con i modelli linguistici generativi.

In conclusione, riteniamo opportuno riflettere sulle possibili traiettorie di sviluppo future a partire dai risultati di questa ricerca. Emergono due dimensioni, in particolare, che meritano di essere approfondite. Un primo aspetto concerne il campo della *Human-AI Interaction* e riguarda la necessità di ampliare e diversificare l'esplorazione sulle strategie di mediazione tra l'umano e la IAG, adottando un approccio sistematico di sperimentazione al *prompt engineering* che coinvolga tipologie di utenti differenti in contesti educativi eterogenei. Lo studio da noi condotto, analogamente agli studi riportati nella letteratura scientifica di riferimento (cfr. § 3.2), di fatto esamina le interazioni tra LLM e studenti inesperti durante attività finalizzate al potenziamento delle abilità di scrittura. A nostro avviso, questi risultati rappresentano un primo importante contributo verso la comprensione delle potenzialità e dei limiti dell'uso della IAG nei processi di apprendimento linguistico. Allo stesso tempo, progetti di ricerca futuri potrebbero ampliare il campo di indagine su più fronti. Da un lato, si potrebbero esplorare le interazioni tra i LLM e diverse tipologie di apprendenti e educatori, le cui competenze trasversali sul *prompt engineering* risultano meno esplorate. Dall'altro, sarebbe interessante condurre l'analisi delle interazioni su un ventaglio più ampio di attività didattiche, al fine di esplorare come i LLM possano supportare lo sviluppo di diverse competenze linguistiche e strategiche. In quest'ottica, una sperimentazione più articolata ed esaustiva potrebbe contribuire ad ampliare la prospettiva sulle dinamiche di interazione e collaborazione tra l'umano e l'IA, restituendo una visione d'insieme delle potenzialità e limitazioni di questi strumenti nei processi di apprendimento e insegnamento delle lingue.

La seconda direzione interessa il campo di indagine sull'alfabetizzazione critica all'IA nei contesti educativi. Come riscontrato anche dalla letteratura scientifica di riferimento (cfr. § 2), iniziano ad emergere diversi studi che riconoscono il *prompt engineering* come un'abilità emergente, fondamentale per implementare efficacemente gli strumenti basati sulla IAG nelle pratiche didattiche. In questa prospettiva, il modello operativo proposto in questo contributo intende offrire una cornice teorico-pratica delle competenze trasversali di *prompt engineering* in ambito edulinguistico, adottando un approccio di tipo *evidence-based* che esplora l'impatto delle strategie di *prompting* sui comportamenti dei LLM. Alla luce di queste considerazioni, ad oggi, si aprono interessanti prospettive di ricerca, tra cui la necessità di costruire strumenti per la misurazione delle competenze trasversali di *prompt engineering* (Lee, Palmer, 2025) e promuovere percorsi di formazione mirati, rivolti ad apprendenti ed educatori, per un utilizzo efficace dei LLM nei processi di apprendimento e insegnamento. Per queste ragioni, riteniamo indispensabile continuare ad approfondire le competenze trasversali che sottendono il *prompt engineering*, in quanto prerequisito per lo sviluppo di una *prompt literacy* che consenta una collaborazione dinamica e sinergica tra l'umano e la IAG nell'Educazione Linguistica.

RIFERIMENTI BIBLIOGRAFICI

- Adiguzel T., Kaya M. H., Cansu F. K. (2023), "Revolutionizing education with AI: Exploring the transformative potential of ChatGPT", in *Contemporary Educational Technology*, 15, 3, ep429: <https://doi.org/10.30935/cedtech/13152>.
- Anderson N., McGowan A., Galway L., Hanna P., Collins M., Cutting D. (2024), "Implementing generative AI and Large Language Models in education", in *ISAS 2023: Proceedings of the 7th International Symposium on Innovative Approaches in Smart Technologies*, Institute of Electrical and Electronics Engineers Inc., pp. 1-6: <https://doi.org/10.1109/isas60782.2023.10391517>.
- Barrot J. S. (2024), "ChatGPT as a language learning tool: An emerging technology report", in *Technology, Knowledge and Learning*, 29, pp. 1151-1156: <https://doi.org/10.1007/s10758-023-09711-4>.
- Bender E. M., Koller A. (2020), "Climbing towards NLU: On meaning, form, and understanding in the age of data", in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL, pp. 5185-5198: [10.18653/v1/2020.acl-main.463](https://doi.org/10.18653/v1/2020.acl-main.463).
- Cain W. (2024), "Prompting change: Exploring prompt engineering in Large Language Model AI and its potential to transform education", in *TechTrends*, 68, pp. 47-57: <https://doi.org/10.1007/s11528-023-00896-0>.
- Chen B., Zhang Z., Langrené N., Zhu S. (2025), "Unleashing the potential of prompt engineering for Large Language Models", in *Patterns*, 6, 6, pp. 1-44.
- Chiu T. K. F., Moorhouse B. L., Chai C. S., Ismailov M. (2023), "Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot", in *Interactive Learning Environment*, 32, 7, pp. 3240-3256: <https://doi.org/10.1080/10494820.2023.2172044>.
- Dang H., Mecke L., Lehmann F., Goller S., Busheck D. (2022), "How to prompt? Opportunities and challenges of zero- and few-Shot learning for human-AI interaction in creative applications of generative models", in *ArXiv*, abs/2209.01390.
- Eager B., Brunton R. (2023), "Prompting higher education towards AI-augmented teaching and learning practice", in *Journal of University Teaching and Learning Practice*, 20, 5, pp. 1-19: <https://doi.org/10.53761/1.20.5.02>.
- Farrokhnia M., Banihashem S. K., Noroozi O., Wals A. (2023), "A SWOT analysis of ChatGPT: Implications for educational practice and research", in *Innovations in Education and Teaching International*, 61, 3, pp. 460-474: <https://doi.org/10.1080/14703297.2023.2195846>.
- Gao T., Fisch A., Chen D. (2021), "Making pre-trained language models better few-shot learners", in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, ACL, pp. 3816-3830: [10.18653/v1/2021.acl-long.295](https://doi.org/10.18653/v1/2021.acl-long.295).
- Ge T. (2024), "Assessing the acceptance and utilization of ChatGPT by Chinese university students in English writing education", in *International Journal of Learning and Teaching*, 10, pp. 166-170: [10.18178/ijlt.10.1.166-170](https://doi.org/10.18178/ijlt.10.1.166-170).
- Ji Z., Lee N., Frieske R., Yu T., Su D., Xu Y., Ishii E., Bang Y., Chen D., Dai W., Chan H. S., Madotto A., Fung P. (2023), "Survey of hallucination in Natural Language Generation", in *ACM Computer Survey*, 55, 12, pp. 1-38: <https://doi.org/10.1145/3571730>.
- Jiang T., Sun Z., Fu S., Lv Y. (2024), "Human-AI interaction research agenda: A user-centered perspective", in *Data and Information Management*, 8, 4, pp. 1-10: <https://doi.org/10.1016/j.dim.2024.100078>.

- Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F., Gasser U., Groh G., Günemann S., Hüllermeier E., Krusche S., Kutyniok G., Michaeli T., Nerdel C., Pfeffer J., Poquet O., Sailer M., Schmidt A., Seidel T., Stadler M., Weller J., Kuhn J., Kasneci G. (2023), "ChatGPT for good? On opportunities and challenges of Large Language Models for Education", in *Learning and Individual Differences*, 103, pp. 1-12: <https://doi.org/10.1016/j.lindif.2023.102274>.
- Knott N., Tolzin A., Janson A., Leimeister J. M. (2024), "AI literacy and its implications for prompt engineering strategies", in *Computers and Education: Artificial Intelligence*, 6, pp. 1-14: <https://doi.org/10.1016/j.caai.2024.100225>.
- Korzynski P., Mazurek G., Krzypkowska P., Kurasinski A. (2023), "Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT", in *Entrepreneurial Business and Economics Review*, 11, 3, pp. 25-37: <https://doi.org/10.15678/EBER.2023.110302>.
- Lai C. Y., Cheung K. Y., Chan C. S. (2023), "Exploring the role of intrinsic motivation in ChatGPT adoption to support active learning: An extension of the technology acceptance model", in *Computers and Education: Artificial Intelligence*, 5, pp. 1-13: <https://doi.org/10.1016/j.caai.2023.100178>.
- Lee D., Palmer E. (2025), "Prompt engineering in higher education: A systematic review to help inform curricula", in *International Journal of Education Technology in Higher Education*, 22, 7, pp. 1-22: <https://doi.org/10.1186/s41239-025-00503-7>.
- Lee U., Jung H., Jeon Y., Sohn Y., Hwang W., Moon J., Kim H. (2023), "Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education", in *Education and Information Technologies*, 29, 9, pp. 11483-11515: <https://doi.org/10.1007/s10639-023-12249-8>.
- Linardatos P., Papastefanopoulos V., Kotsiantis S. (2021), "Explainable AI: A review of machine learning interpretability methods", in *Entropy*, 23, 1, pp. 1-18: <https://doi.org/10.3390/e23010018>.
- Lim W. M., Gunasekara A., Pallant J. L., Pallant J. I., Pechenkina E. (2023), "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators", in *The International Journal of Management Education*, 21, 2, pp. 1-13: <https://doi.org/10.1016/j.ijme.2023.100790>.
- Liu L. (2023), "Analyzing the text contents produced by ChatGPT: Prompts, feature-components in responses, and a predictive model", in *Journal of Educational Technology Development and Exchange*, 16, 1, pp. 49-70: <https://doi.org/10.18785/jetde.1601.03>.
- Liu M., Zhang L. J., Zhang D. (2025), "Enhancing student GAI literacy in digital multimodal composing through development and validation of a scale", in *Computers in Human Behavior*, 166, pp. 1-16: <https://doi.org/10.1016/j.chb.2025.108569>.
- Liu P., Yuan W., Fu J., Jiang Z., Hayashi H., Neubig G. (2021), "Pre-train, prompt, and predict: A systematic survey of prompting methods in Natural Language Processing", in *ACM Computer Surveys*, 55, 9., pp. 1-35: <https://doi.org/10.1145/3560815>.
- Liu Y., Deng G., Xu Z., Li Z., Zheng Y., Zhang Y., Zhao L., Zhang T., Wang K. (2024), "A hitchhiker's guide to jailbreaking ChatGPT via prompt engineering", in *SEA4DQ 2024: Proceedings of the 4th International Workshop on Software Engineering and AI for Data Quality in Cyber-Physical Systems/Internet of Things*, ACM, pp. 12-21: <https://doi.org/10.1145/3663530.3665021>.
- Lo L. S. (2023), "The CLEAR path: A framework for enhancing information literacy through prompt engineering", in *The Journal of Academic Librarianship*, 49, 4, pp. 1-3: <https://doi.org/10.1016/j.acalib.2023.102720>.

- Lu Y., Bartolo M., Moore A., Riedel S., Stenetorp P. (2022), “Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity”, in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, ACL, pp. 8086-8098: [10.18653/v1/2022.acl-long.556](https://doi.org/10.18653/v1/2022.acl-long.556).
- Mezzadri M. (2024), *Apprendere, insegnare e valutare le lingue*, Franco Cesati Editore, Firenze.
- Mezzadri M., Paita M. (2025), “Artificial intelligence in language education: Exploring prompting strategies to foster argumentative writing skills”, in *Journal of Studies in Language, Culture, and Society*, 8 (1), pp. 188-205.
- Mondal S., Bappon S. D., Roy C. K. (2024), “Enhancing user interaction in ChatGPT: Characterizing and consolidating multiple prompts for issue resolution”, in *MSR '24: Proceedings of the 21st International Conference on Mining Software Repositories*, ACM, pp. 222-226: <https://doi.org/10.1145/3643991.3645085>.
- O'Connor J., Andreas J. (2021), “What context features can transform language models use?”, in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, ACL, pp. 851-864: [10.18653/v1/2021.acl-long.70](https://doi.org/10.18653/v1/2021.acl-long.70).
- Oppenlaender J., Linder R., Silvennoinen J. (2024), “Prompting AI art: An investigation into the creative skill of *prompt engineering*”, in *International Journal of Human-Computer Interaction*, 1, pp. 1-23: <https://doi.org/10.1080/10447318.2024.2431761>.
- Ou A. W., Khuder B., Franzetti S., Negretti R. (2024), “Conceptualising and cultivating Critical GAI Literacy in doctoral academic writing”, in *Journal of Second Language Writing*, 66, pp. 1-16: <https://doi.org/10.1016/j.jslw.2024.101156>.
- Ranieri M., Cuomo S., Biagini G. (2023), *Scuola e intelligenza artificiale: percorsi di alfabetizzazione critica*, Carrocci, Roma.
- Rawas S. (2024), “ChatGPT: Empowering lifelong learning in the digital age of higher education”, in *Education and Information Technologies*, 29, pp. 6895-6908: <https://doi.org/10.1007/s10639-023-12114-8>.
- Ribeiro M. T., Singh S., Guestrin C. (2016), “Why should I trust you? Explaining the predictions of any classifier”, in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp. 1135-1144: <https://doi.org/10.1145/2939672.2939778>.
- Sawalha G., Taj I., Shoufan A. (2024), “Analyzing student prompt and their effect on ChatGPT’s performance”, in *Cogent Education*, 11, 1, pp. 1-20: <https://doi.org/10.1080/2331186X.2024.2397200>.
- Shaikh S., Yayilgan S. Y., Klimova B., Pikhart M. (2023), “Assessing the usability of ChatGPT for formal English language learning”, in *Investigation in Health, Psychology and Education*, 13, 9, pp. 1937-1960: <https://doi.org/10.3390/ejihpe13090140>.
- Steele J. L., (2023), “To GPT or not GPT? Empowering our students to learn with AI”, in *Computers and Education: Artificial Intelligence*, 5, pp. 1-5: <https://doi.org/10.1016/j.caai.2023.100160>.
- Su Y., Lin Y., Lai C. (2023), “Collaborating with ChatGPT in argumentative writing classrooms”, in *Assessing Writing*, 57, pp. 1-13: <https://doi.org/10.1016/j.asw.2023.100752>.
- Tan B., Yang Z., Al-Shedivat M., Xing E. P., Hu Z. (2021), “Progressive generation of long text with pretrained language models”, in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, ACL, pp. 4313-4324: [10.18653/v1/2021.nacl-main.341](https://doi.org/10.18653/v1/2021.nacl-main.341).
- 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”, in *Smart Learning Environments*, 10, 15, pp. 1-24: <https://doi.org/10.1186/s40561-023-00237-x>.

- Tseng W., Warschauer M. (2023), “AI-writing tools in education: if you can’t beat them, join them”, in *Journal of China Computer-Assisted Language Learning*, 3, 2, pp. 258-262: <https://doi.org/10.1515/jccall-2023-0008>.
- Vaccino-Salvadore S. (2023), “Exploring the ethical dimensions of using ChatGPT in language learning and beyond”, in *Languages*, 8, 3, p. 191: <https://doi.org/10.3390/languages8030191>.
- Walter Y. (2024), “Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education”, in *International Journal of Educational Technology in Higher Education*, 21, 15, pp. 1-29: <https://doi.org/10.1186/s41239-024-00448-3>.
- Wang C., Wang Z. (2025), “Investigating L2 writers’ critical AI literacy in AI-assisted writing: An ASPE model”, in *Journal of Second Language Writing*, 67, pp. 1-17: <https://doi.org/10.1016/j.jslw.2025.101187>.
- Wang F., Cheung A. C. K., Chai C. S. (2024), “Language learning development in human-AI interaction: A thematic review of the research landscape”, in *System*, 125, pp. 1-24: <https://doi.org/10.1016/j.system.2024.103424>.
- Wang L., Chen X., Wang C., Xu L., Shadiev R., Li Y. (2024), “ChatGPT’s capabilities in providing feedback on undergraduate students’ argumentation: A case study”, in *Thinking Skills and Creativity*, 51, pp. 1-14: <https://doi.org/10.1016/j.tsc.2023.101440>.
- Wang M., Wang M., Xu X., Yang L., Cai D., Yin M. (2024), “Unleashing ChatGPT’s power: A case study on optimizing information retrieval in flipped classrooms via prompt engineering”, in *IEEE Transactions on Learning Technologies*, 17, pp. 629-641: <https://doi.org/10.1109/TLT.2023.3324714>.
- Warschauer M., Tseng W., Yim S., Webster T., Jacob S., Du Q., Tate T. (2023), “The affordances and contradictions of AI-generated text for writers of English as a second or foreign language”, in *Journal of Second Language Writing*, 63, pp. 1-7: <https://doi.org/10.1016/j.jslw.2023.101071>.
- White J., Fu Q., Hays S., Sandborn M., Olea C., Gilbert H., Elnashar A., Spencer-Smith J., Schmidt D. C. (2023), “A prompt pattern catalogue to enhance prompt engineering with ChatGPT”, in *PLoP ‘23: Proceedings of the 30th Conference on Pattern Languages of Programs*, The Hillside Group, USA, pp. 1-31.
- Wu T., Terry M., Cai C. J. (2022), “AI chains: Transparent and controllable human-AI interaction by chaining Large Language Model prompts”, in *CHI ‘22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ACM, pp. 1-22: <https://doi.org/10.1145/3491102.3517582>.
- Young J. C., Shishido M. (2023), “Investigating OpenAI’s ChatGPT potentials in generating chatbot’s dialogue for English as a Foreign Language Learning”, in *International Journal of Advanced Computer Science and Applications*, 14, 6, pp. 65-72: <http://dx.doi.org/10.14569/IJACSA.2023.0140607>.
- Zamfirescu-Pereira J. D., Wong R. Y., Hartmann B., Yang Q. (2023), “Why Johnny can’t prompt: How non-AI experts try (and fail) to design LLM prompts”, in *CHI ‘23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, ACM, pp. 1-21: <https://doi.org/10.1145/3544548.3581388>.
- Zhao T. Z., Wallace E., Feng S., Klein D., Singh, S. (2021), “Calibrate before use: Improving few-shot performance of language models”, in *PMLR: Proceedings of the 38th International Conference on Machine Learning*, pp. 12697-12706.

