

# Generati e rigenerati. Uno studio preliminare sul trattamento di testi scientifici fra ChatGPT e abilità di studio

Alessandro Puglisi (Università per Stranieri di Siena)

puglisi(at)unistrasi.it

## Abstract

L'interazione tra uomo e macchina è entrata in una nuova fase dallo sviluppo dei Large Language Models, dal 2020 con GPT-3 e poi, soprattutto, dalla fine del 2022 con il rilascio dell'interfaccia conversazionale ChatGPT. Questi modelli hanno progressivamente acquisito un ruolo sempre più pervasivo in numerosi campi. Nel campo educativo le possibilità offerte dall'utilizzo di questi strumenti sono numerose, sebbene non manchino potenziali criticità. Il contributo si propone di esaminare il comportamento di ChatGPT 4o mini nell'esecuzione di tre tipi di task, legati alle abilità di studio, su testi scritti: riassunto, individuazione di punti chiave e semplificazione. I testi generati vengono successivamente analizzati dal punto di vista lessicale e sintattico, anche attraverso gli indici di leggibilità READ-IT. Tre tipologie di attività didattiche vengono dunque suggerite allo scopo di integrare l'intelligenza artificiale generativa nello sviluppo e mantenimento delle abilità di studio in ambito universitario.

The interaction between humans and machines has entered a new phase since the emergence of Large Language Models, starting with GPT-3 in 2020 and, more notably, with the release of the conversational interface ChatGPT at the end of 2022. These models have progressively taken on an increasingly pervasive role across numerous fields. In the educational domain, the possibilities offered using such tools are substantial, although potential challenges are not absent. This study aims to examine the behavior of ChatGPT 4o mini in performing three types of tasks related to study skills on written texts: summarization, key point identification, and simplification. The texts generated are subsequently analyzed from lexical and syntactic perspectives, including through the application of the READ-IT readability indices. Finally, three types of educational activities are proposed to integrate generative artificial intelligence into the development and maintenance of study skills in the university context.

## Parole chiave

intelligenza artificiale generativa, ChatGPT, LLM, abilità di studio

## Keywords

generative artificial intelligence, ChatGPT, LLM, study skills

## 1 Introduzione

La storia più recente dell'interazione tra essere umano e macchina è fortemente caratterizzata dalla comprensione e gestione del linguaggio naturale. I Large Language Models (da ora in avanti LLM) rappresentano strumenti che, tramite approcci di natura prevalentemente, ma non esclusivamente, probabilistica, riescono a *rappresentare* con efficacia la lingua utilizzata in grandi moli di dati

**AILing**

*AI-Linguistica.*

*Linguistic Studies on AI-Generated Texts and Discourses*

CC-BY-NC-SA 4.0

Puglisi, Alessandro. 2025.

Special Issue: *AI-Driven Linguistic Studies*, Vol. 2. No. 2.

DOI: 10.62408/ai-ling.v2i2.17

ISSN: 2943-0070

testuali, al fine di generare, rispondendo a precise richieste dell'utente, poste attraverso un *chatbot*, nuovi testi. Parallelamente, l'università come agenzia educativa affronta una quantità significativa di sfide formative, una delle quali è relativa alle cosiddette abilità di studio, contraddistinte da un forte carattere *operativo*, trasversale e transdisciplinare (Mariani 1988). È noto come la didattica universitaria tenda a subire un sistematico riverbero dello scarso possesso di tali abilità da parte degli studenti, abilità che affondano con ogni probabilità le proprie radici nel percorso scolastico. Leggere e comprendere un testo, guardare a esso criticamente, riuscire a riassumerlo o a individuare le informazioni più rilevanti, fino alla sua semplificazione, sono solo alcune delle abilità necessarie per uno studio efficace ed efficiente. All'incrocio fra LLM e abilità di studio si pone il presente contributo, il quale si propone di ragionare sull'opportunità di far utilizzare a chi apprende strumenti di intelligenza artificiale generativa basati su LLM. In questo senso, vari pregevoli contributi scientifici, anche se di impronta valutativa, sono già apparsi, tra cui Tavosanis (2024) e, prima ancora, Cicero (2023) e De Cesare (2023). La lente attraverso la quale guardiamo qui ai modelli di linguaggio è, invece, non valutativa, non comparativa, ma descrittiva e orientata all'utilizzo didattico. In questo senso, non mancano gli studi recenti dedicati alle possibilità offerte dai grandi modelli di linguaggio in ambito educativo, dei quali una nutrita e aggiornata rassegna è offerta in Mai et al. (2024). In merito all'adozione di questi strumenti e alle percezioni di studenti e docenti a essi legate, alcune indagini significative sono quelle di Uğraş et al. (2024) sui docenti delle *primary schools* e Zhu, Zhang e Wang (2024) su oltre 300 studenti dal Grado 7 al Grado 12 (equivalenti alle secondarie di primo e secondo grado) negli Stati Uniti, dalla quale è emersa la tendenza a utilizzare i LLM in maniera trasversale alle discipline, seppur con considerevoli oscillazioni percettive causate per lo più dalle cosiddette allucinazioni (o, come forse sarebbe più opportuno chiamarle, in accordo con Cristianini (2024), confabulazioni). L'ambito universitario, del resto, non sembra fare eccezione, come dimostra lo studio di Cavazos et al. (2024), che evidenzia tanto la percezione di utilità da parte degli studenti per raccogliere informazioni, fare *brainstorming*, acquisire spunti, quanto il loro complesso rapporto con usi che vanno verso il *cheating*. Per quanto concerne applicazioni più specifiche dei LLM nella didattica, due ricerche rilevanti, in ambiti diversi, sono Yilmaz e Karaoğlan Yilmaz (2023) e Guo e Wang (2023), rispettivamente per la didattica della programmazione e per il feedback dell'insegnante nell'inglese come lingua straniera. Non sono mancate le indagini sul ruolo dell'intelligenza artificiale generativa nel fare emergere, potenzialmente, prospettive argomentative e culturali diverse, oltre a promuovere il pensiero critico (Chan e Hu 2023; Bitzenbauer 2023), sebbene alcune voci critiche, tra cui Selwyn (2023), abbiano evidenziato possibili criticità nell'utilizzo dell'intelligenza artificiale in campo educativo, derivanti dal ruolo che i dati ricoprono nello sviluppo di queste tecnologie, e legate in particolare a possibili *bias*, sottorappresentazioni delle minoranze, visioni riduzionistiche e dispendio energetico.

Questo interesse scientifico, articolato e trasversale, per l'utilizzo didattico dei LLM, del resto, ricalca la tumultuosa diffusione di queste tecnologie, che sembrano percorrere la cosiddetta *curva di adozione* (Rogers 1962) assai rapidamente; ciò consiglia, inevitabilmente, la massima cautela nel trarre conclusioni, seppur provvisorie e parziali.

Il secondo paragrafo del contributo è dedicato a qualche necessaria riflessione sull'intelligenza artificiale generativa; nel terzo si evidenzia l'obiettivo generale dello studio e le domande di ricerca. Il quarto paragrafo specifica meglio la prospettiva di ricerca alla quale si è accennato e precisa il disegno dello studio. Successivamente, vengono descritti e discussi i risultati della ricerca, i quali preludono ad alcune proposte operative. Le brevi conclusioni evidenziano il carattere di inevitabile apertura di ogni ricerca che si muova nell'ambito dell'intelligenza artificiale generativa.

## **2 Cosa è (e soprattutto cosa non è) l'intelligenza artificiale generativa**

L'etichetta *intelligenza artificiale generativa* si applica a una varietà di sistemi di intelligenza artificiale in grado di generare, per l'appunto, contenuti di diversa natura (testuale, visiva, sonora e audiovisiva) tramite richieste formulate dall'utente in linguaggio naturale. Senza volere né, in verità, potere, tracciare una storia dello sviluppo di tali sistemi, converrà limitarsi a tracciare qualche confine. Per farlo, proviamo dunque a dire cosa queste tecnologie *non* sono. È facilmente comprensibile come non si tratti di depositari di ogni sapienza in qualsivoglia dominio della conoscenza, oppure oracoli (per uno studio sulle trasformazioni delle forme digitali del sapere in direzione dell'oracolarità, Roncaglia 2023). Altrettanto inadeguata, quantunque suggestiva a tutta prima, appare la definizione *pappagalli stocastici* (Bender et al. 2021), poiché la componente probabilistica, sì preponderante, è a ogni modo associata a (e temperata da) elementi deterministici (basti pensare ai cosiddetti *guardrail*, meccanismi interni di controllo atti a mitigare il comportamento del sistema). D'altro canto, non si tratta nemmeno di enciclopedie, poiché a mancare è l'organizzazione strutturata di conoscenze prodotte da esperti e verificate, e pure perché nelle opere enciclopediche non si dà generazione di contenuti nuovi né alcun possibile adattamento alle richieste di chi ne fruisce. L'intelligenza artificiale generativa, inoltre, non ha caratteristiche da motore di ricerca (sebbene siano presenti alcune funzionalità simili, per esempio all'interno di ChatGPT), anche in questo caso in ragione di meccanismi di funzionamento assai diversi. Ciò premesso, si dovrà dire inoltre che confronti stringenti tra *performance* umane e artificiali nei più disparati campi, alla ricerca di presunte superiorità di una delle due parti sull'altra, dovrebbero puntare a rilevare le capacità di questi sistemi di IA di svolgere dei compiti, senza con ciò inferire la possibilità, nonché la convenienza, di un confronto diretto tra "intelligenze" che non sono confrontabili, come ha rilevato limpidamente Cristianini (2024). Per converso, le caratteristiche precipue dell'intelligenza artificiale generativa, che stanno per lo più nella capacità di individuare regolarità nei dati e generarne di

nuovi, rendono a nostro avviso sistemi come ChatGPT particolarmente adatti per l'integrazione nella didattica. Proveremo dunque, nel prosieguo, a compiere un percorso che dall'osservazione del *comportamento* di un modello giunga a proporre specifiche attività didattiche nelle quali lo sviluppo e il mantenimento di abilità di studio è informato dall'interazione con strumenti di intelligenza artificiale generativa.

### **3 Obiettivo dello studio**

Lo studio è parte di una ricerca di più ampio respiro sul trattamento dei testi mediante modelli di linguaggio. In esso ci si propone di osservare il *comportamento* di ChatGPT 4o mini nella generazione di riassunti, nell'individuazione di punti chiave di un testo e nella semplificazione del testo stesso, allo scopo di verificare l'utilizzabilità di tali strumenti, sotto la guida dell'insegnante e/o in autonomia da parte degli apprendenti, nello studio universitario. Esaminare i risultati forniti dal modello di linguaggio consente di evidenziare tanto le potenzialità quanto i limiti, come vedremo. Da ciò potranno dunque scaturire risposte, seppur provvisorie, alle seguenti domande di ricerca:

DR 1: è possibile utilizzare strumenti come ChatGPT per supportare lo studio universitario in ambito linguistico?

DR 2: quali tipologie di attività, svolte tramite strumenti come ChatGPT, potrebbero efficacemente sostenere lo studio universitario in ambito linguistico?

Per poter rispondere, riteniamo sia necessario porsi in una prospettiva di ricerca, descritta in dettaglio nel prossimo paragrafo, che eviti un confronto stretto, peraltro scarsamente utile, fra umano e macchina in quanto tali.

### **4 Prospettiva di ricerca e disegno dello studio**

La prospettiva di ricerca che adottiamo è quella dell'osservazione, qui intesa nel senso assunto da questo termine nell'epistemologia costruttivista, come atto situato, interno, non neutrale e, nei fatti, costitutivo del mondo (fra gli altri, von Glasersfeld 1995; Maturana e Varela 1980; von Foerster 2003). Il fenomeno che osserviamo non è dato di per sé, non esiste al di fuori del processo di osservazione, ma l'osservatore partecipa attivamente, osservando non solo il fenomeno ma anche la propria stessa esperienza dell'osservare e finanche generando il fenomeno, e quindi co-costruendolo. Tale processo consente qui di *rilevare* alcuni aspetti del comportamento del modello relativamente a un caso di studio di dimensioni ridotte, verificando se e quanto si possano compiere generalizzazioni o, meglio, percepire regolarità. È necessario ribadire come l'osservatore sia, in questo contesto, direttamente parte in causa, con riferimento alla metodologia che adotta (inclusa la scrittura di appositi *prompt*) e agli strumenti di rilevazione che sceglie e con i quali

costruisce, appunto, categorie di analisi. Inoltre, un'operazione di questo genere non può non considerare l'opacità, se non oscurità, dei meccanismi interni di funzionamento dei modelli di linguaggio, in particolare di quelli *closed-source* che, in quanto tali, non possono essere ispezionati, come invece accade, almeno parzialmente, nei modelli *open weights*, vale a dire quei modelli dei quali vengono resi pubblici i "pesi" del modello, cioè i parametri appresi dal modello durante l'addestramento. Va altresì tenuto presente il fatto, non secondario, di non conoscere l'entità precisa, né tantomeno la composizione esatta del dataset testuale di addestramento di GPT-4 (anche se alcune informazioni, relative a GPT-3, vengono riportate in Cristianini 2024 sulla base del contributo di Brown et al. 2020). In questo scenario, a nostro avviso, un'analisi valutativa fondata non può darsi, a maggior ragione per la lingua italiana, plausibilmente poco rappresentata nel *dataset* di GPT-4. Si ha ragione di credere che neanche un imponente numero di interazioni, con numerosissimi *prompt* differenti, possa consentire di compiere generalizzazioni sulla lingua italiana prodotta da questi strumenti; del resto, anche concedendo che un procedimento di questa natura possa condurre eventualmente a qualche generalizzazione, ragionando per assurdo vi sarebbe il rischio, a nostro avviso, di giungere a considerare che le *caratteristiche* della lingua generata risiedano, più semplicemente, nella lingua alla quale il modello è stato esposto durante l'addestramento.

Il disegno dello studio ha previsto la selezione di brevi estratti da quattro testi scientifici di ambito linguistico: *Storia linguistica dell'Italia unita* (De Mauro 2008 (1963): 12-14), *Storia linguistica dell'Italia repubblicana* (De Mauro 2014: 19), *Scripta manent* (Serianni 1987: 182-184) e *Italiani scritti* (Serianni 2003: 24-25). Si è scelto di analizzare quattro testi, più e meno recenti, di due autori che nella loro scrittura hanno unito la profondità del pensiero alla chiarezza del dettato, senza mai rinunciare ad affrontare la complessità. I primi due estratti fanno parte di trattazioni ben più estese e articolate, il terzo è un articolo preso nella sua interezza, mentre il quarto estratto deriva da un volume di impronta manualistica. Il modello utilizzato è ChatGPT 4o mini, versione più leggera, veloce ed efficiente rispetto a 4o, nonché attuale versione di default per gli account non a pagamento, con cui si è interagito tramite *web client*. I *task* sono tre, con i relativi *prompt* di tipologia *Act As* riportati di seguito: riassunto (*prompt* 1); individuazione di punti chiave (*prompt* 2); semplificazione (*prompt* 3).

- (1) Agisci come se fossi l'insegnante di italiano. Realizza un riassunto del seguente testo riducendo la sua estensione del 50% e mantenendo tutte le informazioni fondamentali, senza ricorrere a elenchi puntati o elenchi numerati.
- (2) Agisci come se fossi l'insegnante di italiano. Individua i tre punti chiave all'interno del seguente testo e realizza un elenco puntato. Ciascun punto dell'elenco deve essere al massimo di venticinque parole.

- (3) Agisci come se fossi l'insegnante di italiano. Semplifica il seguente testo usando un lessico più comune e una sintassi più semplice, mantenendo un'estensione simile a quella del testo originale.

Le analisi effettuate tanto sui testi originali quanto su quelli generati di volta in volta riguardano l'indice di leggibilità READ-IT (Dell'Orletta, Montemagni e Venturi 2011), con riferimento particolare ai sottoindici lessicale e sintattico, il numero di periodi, parole e la loro lunghezza media, oltre alla percentuale di termini afferenti al Vocabolario di Base (VdB)<sup>1</sup> e alla densità lessicale. Sono state inoltre effettuate rilevazioni di natura sintattica; oltre al relativo indice READ-IT già citato, la percentuale di sostantivi, aggettivi, verbi e congiunzioni, il numero medio di proposizioni per periodo e il numero medio di parole per proposizione. Le analisi sono state svolte grazie all'ausilio dello strumento DyLan TextTools (v2.1.9) dell'ILC-CNR, liberamente utilizzabile *online*.<sup>2</sup>

## 5 Risultati

Presentiamo in questa sezione i risultati delle analisi condotte sui testi originali e sui conseguenti testi generati da ChatGPT, mettendo in evidenza gli aspetti più significativi.<sup>3</sup> In Tabella 1 e Tabella 2 troviamo i risultati relativi al primo dei quattro testi, l'estratto dalla *Storia linguistica dell'Italia unita*. Si osserva che dal punto di vista del READ-IT lessicale (Tabella 1), l'unico cambiamento, di per sé molto significativo è rilevabile nel caso del *task 3* (da 99% a 3,5%). Per il resto, i periodi sono 3, come richiesto, nel *task 2*, mentre aumentano nel *task 3* a 10, pur con un minor numero di parole complessivo. La percentuale di termini dal VdB aumenta lievemente nei *task 1* e *3*, mentre diminuisce nel *2*, tenendo sempre presente la brevità dei testi in esame. Dal punto di vista sintattico (Tabella 2) notiamo, per questo testo, una netta diminuzione dell'indice READ-IT sintattico nei *task 2* e *3* (da 99% a 3% e a 24,3%, rispettivamente) con un aumento percentuale, non inatteso, dei sostantivi e degli aggettivi nel *task 2* (di cui discuteremo nel paragrafo successivo).

---

<sup>1</sup> Si noti che sarebbe ancora più utile sfruttare il Nuovo Vocabolario di Base, più aggiornato; tuttavia, per ragioni di opportunità in fase di analisi, utilizzando lo strumento DyLan TextTools, non è stato possibile procedere a un confronto automatizzato puntuale col Nuovo VdB, confronto che, comunque, sarebbe auspicabile.

<sup>2</sup> <https://www.ilc.cnr.it/dylanlab/apps/texttools/>

<sup>3</sup> Le percentuali indicate in corrispondenza dell'indice READ-IT lessicale e READ-IT sintattico indicano, nella logica adottata dallo strumento di analisi, la "probabilità" che il testo risulti di difficile comprensione.

Tabella 1: *Storia linguistica dell'Italia unita*: profilo di base e lessicale.

Testo	READ-IT lessicale (%)	Periodi	Parole	Lunghezza media periodi	Lunghezza media parole	VdB (%)	Dens. less.
originale	99,0	6	369	61,5	5,47	70,6	0,567
1	92,0	6	197	32,8	5,5	72,6	0,547
2	100,0	3	62	20,7	5,9	64,4	0,552
3	3,5	10	254	25,4	5,1	74,8	0,550

Tabella 2: *Storia linguistica dell'Italia unita*: profilo sintattico.

Testo	READ-IT sintattico (%)	Sost. (%)	Agg. (%)	Verbi (%)	Cong. (%)	Num. medio proposizioni	Num. medio parole
originale	99	19,0	9,5	11,7	3,5	5,667	10,853
1	97,6	21,3	7,1	12,7	4,1	3,167	10,368
2	3,0	24,2	14,5	11,3	3,2	1,667	12,400
3	24,3	18,5	6,3	14,2	5,5	3,100	8,194

Per quel che riguarda il secondo testo, estratto dalla *Storia linguistica dell'Italia repubblicana*, il profilo di base e lessicale (Tabella 3) mostra scostamenti significativi tra l'originale e i testi generati solo nel caso dei periodi e della loro lunghezza media, mentre la percentuale di termini dal VdB è significativamente inferiore nel *task 2*. Questo fenomeno è spiegabile a partire dalla necessità di rispettare un vincolo molto stretto di lunghezza (in parole) di ciascun punto e, quindi, dall'impossibilità di rendere i medesimi concetti con termini più comuni o perifrasi. La situazione è invece più variegata se si guarda al profilo sintattico (Tabella 4). In questo caso, infatti, l'indice READ-IT sintattico crolla dal 100% allo 0,2% nel *task 1*, e al 29,6% nel *task 2*. Stando a considerazioni di superficie, nel *task 1* (e, in misura leggermente minore, nel *task 2*) si è avuta, dunque, una significativa semplificazione sintattica, unita a un deciso processo di nominalizzazione.

Tabella 3: *Storia linguistica dell'Italia repubblicana*: profilo di base e lessicale.

Testo	READ-IT lessicale (%)	Periodi	Parole	Lunghezza media periodi	Lunghezza media parole	VdB (%)	Dens. less.
originale	98,5	4	260	65	5,6	74,3	0,585
1	100,0	5	135	27	6,2	77,3	0,625
2	100,0	3	75	25	6,8	54,3	0,741
3	99,2	6	199	33,2	5,5	80,7	0,565

Tabella 4: *Storia linguistica dell'Italia repubblicana*: profilo sintattico.

Testo	READ-IT sintattico (%)	Sost. (%)	Agg. (%)	Verbi (%)	Cong. (%)	Num. medio proposizioni	Num. medio parole
originale	100,0	20,4	13,8	10,4	5,4	6	10,833
1	0,2	23,7	17,0	7,4	8,1	2	13,500
2	29,6	36,0	16,0	1,3	5,3	0,333	75,000
3	72,5	21,6	11,1	9,5	6,0	2,667	12,438

Il terzo testo in esame, tratto dall'articolo di Luca Serianni intitolato *Scripta manent*, mostra un aumento dell'indice READ-IT lessicale (Tabella 5), che parte dal 73,5%, quando si genera un riassunto (*task 1*, 99,5%) e quando si individuano i punti chiave (*task 2*, 96,0%), mentre la tendenza è opposta nel caso del *task 3* (9,1%). Per tutti e tre i compiti si osserva, inoltre, un aumento della percentuale di parole appartenenti al VdB. In quanto al profilo sintattico (Tabella 6), la tendenza sembra opposta a quella appena descritta per gli aspetti lessicali con, in particolare il passaggio del READ-IT sintattico dal 99,3% al 51,2% ma, in maniera controintuitiva, al 100% nel *task 3* di semplificazione.

Tabella 5: *Scripta manent*: profilo di base e lessicale.

Testo	READ-IT lessicale (%)	Periodi	Parole	Lunghezza media periodi	Lunghezza media parole	VdB (%)	Dens. less.
originale	73,5	7	314	44,9	5,5	67,6	0,594
1	99,5	8	192	24	5,8	76,7	0,596
2	96,0	3	82	27,3	5,3	82,0	0,600
3	9,1	8	272	34	5,1	77,7	0,579

Tabella 6: *Scripta manent*: profilo sintattico.

Testo	READ-IT sintattico (%)	Sost. (%)	Agg. (%)	Verbi (%)	Cong. (%)	Num. medio proposizioni	Num. medio parole
originale	99,3	18,5	8,9	11,5	5,1	4	11,214
1	51,2	20,3	12,5	15,6	6,2	3	8
2	91,4	23,2	9,8	12,2	7,3	2,667	10,250
3	100,0	18,4	9,6	11,0	7,4	3,125	10,880

L'ultimo estratto analizzato, dal volume *Italiani scritti*, ancora di Luca Serianni, mostra un miglioramento dell'indice READ-IT lessicale (Tabella 7) tanto nel *task 1* (dal 93,7% al 74,3%) quanto nel 3, giungendo al 56,6%, unitamente a un aumento più significativo delle parole dal VdB nei *task 2* e 3 mentre, ad esempio, nel primo testo si era rilevata una tendenza opposta per il *task 2*. Infine, dal punto di vista della sintassi (Tabella 8), il primo testo generato (riassunto) mostra un deciso aumento dell'indice READ-IT sintattico (da 15,5% a 74,4%) mentre nel *task 2* si osserva un aumento minimo (16,1%) e una certa diminuzione nel 3 (9,8%), in linea con la richiesta espressa nel *prompt*.

Tabella 7: *Italiani scritti*: profilo di base e lessicale.

Testo	READ-IT lessicale (%)	Periodi	Parole	Lunghezza media periodi	Lunghezza media parole	VdB (%)	Dens. less.
originale	93,7	11	306	27,8	5,4	76,9	0,597
1	74,3	6	171	28,5	5,5	75,8	0,635
2	100,0	3	63	21	5,8	84,8	0,649
3	56,6	11	220	20	5,2	80,6	0,649

Tabella 8: *Italiani scritti*: profilo sintattico.

Testo	READ-IT sintattico (%)	Sost. (%)	Agg. (%)	Verbi (%)	Cong. (%)	Num. medio proposizioni	Num. medio parole
originale	15,5	20,9	11,4	12,1	4,2	2,818	9,871
1	74,4	20,5	12,9	12,9	6,4	3,333	8,550
2	16,1	27,0	12,7	14,3	4,8	2,333	9
3	9,8	17,3	11,4	17,3	7,3	2,727	7,333

## 6 Discussione

I risultati emersi meritano qualche ulteriore considerazione. In linea generale, il modello esegue i compiti richiesti a partire dal prompt utilizzato per ciascun *task*, come abbiamo visto. Dall'analisi degli indici di leggibilità READ-IT (lessicale e sintattico), nonché dall'osservazione dei valori relativi, per esempio, al VdB e alla percentuale di sostantivi, aggettivi, verbi e congiunzioni, le acquisizioni che possiamo trarre non sono numerose, oltre a presentarsi come poco sistematiche. In generale, potremmo dire che tanto il *task* di riassunto quanto quello di semplificazione vengono eseguiti in maniera soddisfacente; inoltre, nel primo sembra esservi talvolta qualcosa della seconda (soprattutto in termini sintattici), anche se non richiesta, e viceversa. Possiamo dedurre che, almeno limitatamente ai testi analizzati e ai *prompt* utilizzati, compiti di riassunto portano con sé una certa semplificazione, e compiti di semplificazione conducono a testi che rappresentano anche un riassunto rispetto all'originale. In quanto al compito di individuazione dei punti chiave nei brevi testi forniti, del resto, il vincolo inserito nel *prompt* viene generalmente rispettato. Dal punto di vista sintattico, vi è una certa emergenza di nominalizzazioni, in particolare in *Storia linguistica dell'Italia repubblicana* e *Scripta manent* in relazione al *task* di riassunto e all'individuazione dei punti chiave. In relazione a questo secondo *task*, però, il fenomeno è spiegato facilmente dalla natura della richiesta espressa nel *prompt*, vincolo incluso. Un'ultima considerazione riguarda la semplificazione lessicale a confronto con quella sintattica: non si rilevano regolarità particolari, giacché in due casi su quattro l'indice READ-IT lessicale subisce un calo più deciso rispetto al READ-IT sintattico, mentre nei restanti due avviene il processo inverso.

Tutto quanto emerso, considerando dati quantitativamente modesti (un modello, quattro brevi testi, tre *prompt*, ventiquattro *task* complessivi eseguiti) non siamo autorizzati, nei fatti, a trarre alcuna conclusione. Ciò nonostante, è proprio questo tipo di risultati che dovrebbe contribuire a scoraggiare analisi episodiche, impressionistiche, indirizzate alla rilevazione e descrizione di una presunta *lingua dell'intelligenza artificiale*. Tuttavia, la complessiva efficacia dello strumento nell'esecuzione dei *task* richiesti, pur nella variabilità dei risultati e nell'assenza di regolarità, consente di avanzare proposte di sfruttamento didattico di questi strumenti per lo sviluppo delle abilità di studio. La loro utilità in chiave di utilizzo didattico sarebbe infatti scarsa, se essi funzionassero deterministicamente, sempre nello stesso modo, a parità di *prompt*, con regolarità osservabili.

Non essendo così, di conseguenza, potremo proporre alcune tipologie di attività che prevedano l'interazione, il dialogo, tra apprendente e modello di

linguaggio. Tre proposte operative saranno perciò descritte brevemente nel prossimo paragrafo.

### **7 Tre proposte di attività didattiche per sviluppare le abilità di studio**

Una considerazione preliminare, tanto ovvia quanto importante, riguarda la già significativa penetrazione di strumenti come ChatGPT nell'uso quotidiano. Una tale pervasività, raggiunta in un lasso di tempo invero assai limitato, può incoraggiare utilizzi poco etici presso in differenti contesti educativi. A tal proposito, il buon senso, prima che le scienze pedagogiche, sembra suggerire che possa essere più vantaggioso abbandonare atteggiamenti di divieto e tentare, per quanto possibile, una proficua integrazione di strumenti di intelligenza artificiale generativa nella pratica didattica.

Le proposte che seguono sono immaginate in un contesto universitario e, particolarmente, nell'ambito umanistico; ciò non toglie che vi siano significative possibilità di adattamento a contesti differenti.

La prima proposta di lavoro riguarda attività in cui gli studenti producono, anche in coppie o piccoli gruppi, dei testi, per esempio una sintesi, in maniera guidata senza l'ausilio di alcuno strumento e, successivamente, interagiscono con un'intelligenza artificiale generativa per far *giudicare* il testo. In tal senso, la prospettiva valutativa, che abbiamo rifiutato nella nostra analisi, potrebbe essere chiamata in causa con efficacia. Questa proposta può essere gestita in maniera più o meno strutturata da parte dell'insegnante, incoraggiando (meta)riflessioni su quali aspetti far giudicare e attraverso quali *prompt*. Ciò consentirebbe di svolgere attività significative, oltre che utili per lo sviluppo della competenza linguistica, metalinguistica e per l'abilità di riassumere testi. Tale proposta potrebbe essere meglio esemplificata attraverso i seguenti passi:

1. Insegnante e studenti riflettono insieme, discutendo, su quali siano gli aspetti che determinano, in misura maggiore o minore, l'efficacia di una sintesi;
2. L'insegnante propone agli studenti di realizzare, senza utilizzare alcuno strumento informatico, la sintesi di un testo argomentativo, rispettando precisi limiti di parole;
3. Gli studenti vengono indirizzati a sottoporre, anche iterativamente, la sintesi prodotta allo strumento di intelligenza artificiale generativa *text-to-text*, attraverso *prompt* che specifichino gli aspetti rilevanti individuati in precedenza e "valutino" il testo;
4. Si conduce in plenaria una riflessione complessiva sulle valutazioni ricevute da parte del LLM e sugli esiti dell'attività.

Una seconda proposta presenta un processo in un certo senso inverso rispetto alla prima, vale a dire far giudicare a chi apprende, anche sulla base di griglie di valutazione costruite collaborativamente in precedenza, testi prodotti da un sistema

come ChatGPT, tra cui sintesi o elenchi di punti chiave di uno scritto. Queste prime due proposte sono naturalmente integrabili, allo scopo di usare l'IA generativa come strumento di *assistenza* o, in altre parole, *companion*.

Una terza tipologia di attività guarda in maniera specifica a un aspetto della competenza linguistica, vale a dire la capacità di gestire più registri linguistici differenti. Interazioni finalizzate a cambi di registro e/o semplificazioni di testi, preesistenti o prodotti da chi apprende o dall'IA generativa, possono incoraggiare un approccio critico ai contenuti disciplinari e alla lingua con cui sono espressi.

Questa tipologia di attività potrebbe prevedere, per esempio, una lista di registri differenti ed esempi di *prompt* che gli apprendenti potrebbero inizialmente limitarsi ad applicare e, successivamente, modificare sperimentando modi diversi di interagire con il LLM.

Ciascun elemento al centro della nostra analisi (riassunto, individuazione di punti chiave, semplificazione) può costituire, dunque, il fulcro di una o più attività interattive, nelle quali l'IA generativa viene utilizzata in maniera critica e costruttiva.

## 8 Conclusioni

Il presente contributo si pone come un tentativo di verifica, invero preliminare, della possibilità di sfruttare didatticamente strumenti di intelligenza artificiale generativa per supportare lo sviluppo e il mantenimento delle abilità di studio in contesto universitario. Attraverso l'analisi quantitativa di testi generati da ChatGPT in tre differenti *task*, a partire da quattro diversi estratti di ambito linguistico, possiamo sostanzialmente concludere che non vi sono sufficienti elementi per stabilire la presenza di regolarità nei testi generati. Ciò fornisce, dunque, ampi spazi entro cui lavorare da un punto di vista didattico e ci consente di suggerire alcune proposte di attività, che sono state descritte nel precedente paragrafo.

Alla prima domanda di ricerca si può perciò rispondere affermativamente, mentre la seconda domanda trova riscontro nelle proposte di attività di cui sopra, le quali attenderebbero il conforto della sperimentazione. Per questa via, auspichiamo di aver contribuito a indicare una direttrice di ricerca che, confrontandosi con l'IA generativa, non potrà non integrare teoria scientifica e pratica didattica.

## Riconoscimenti

Si segnala l'utilizzo di ChatGPT-4o per l'assistenza nell'analisi dei dati raccolti. Si assicura comunque che il contributo aderisce in maniera rigorosa agli standard generali e alle linee guida etiche della scrittura accademica.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this contribution.

## Riferimenti bibliografici

- Bender, Emily M. & Gebru, Timnit & McMillan-Major, Angelina & Shmitchell, Shmargaret. 2021. On the dangers of stochastic parrots: can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency FAccT '21* (Online, 3-10 marzo 2021), 610–623. New York: Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922> (ultimo accesso 27.01.2025)
- Bitzenbauer, Philipp. 2023. ChatGPT in physics education: a pilot study on easy-to-implement activities. *Contemporary Educational Technology* 15(3). <https://doi.org/10.30935/cedtech/13176>
- Brown, Tom B. et al. 2020. Language Models are Few-Shot learners. <https://doi.org/10.48550/arXiv.2005.14165> (ultimo accesso 31.05.2025)
- Cavazos, Jenel T. & Hauck, Keane A. & Baskin, Hannah M. & Bain, Catherine M. 2024. ChatGPT goes to college: exploring student perspectives on artificial intelligence in the classroom. *Teaching of Psychology*. <https://doi.org/10.1177/00986283241268829> (ultimo accesso 15.04.2025)
- Chan, Cecilia Ka Yuk & Hu, Wenjie. 2023. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education* 20. <https://educationaltechnologyjournal.springeropen.com/counter/pdf/10.1186/s41239-023-00411-8.pdf> (ultimo accesso 15.04.2025)
- Cicero, Francesco. 2023. L'italiano delle intelligenze artificiali generative. *Italiano LinguaDue* 15(2). 733–761. <https://doi.org/10.54103/2037-3597/21990>
- Cristianini, Nello. 2024. *Machina sapiens. L'algoritmo che ci ha rubato il segreto della conoscenza*. Bologna: il Mulino.
- Dell'Orletta, Felice & Montemagni, Simonetta & Venturi, Giulia. 2011. READ-IT: Assessing readability of Italian texts with a view to text simplification. In *Proceedings of the Workshop on Speech and Language Processing for Assistive Technologies SLPAT 2011* (Edimburgo, 30 luglio 2011), 73–83. <http://aclweb.org/anthology/W/W11/W11-2308.pdf> (ultimo accesso 27.01.2025)
- De Cesare, Anna-Maria. 2023. Assessing the quality of ChatGPT's generated output in light of human-written texts. A corpus study based on textual parameters. *CHIMERA: Romance Corpora and Linguistic Studies* 10. 179–210. <https://revistas.uam.es/chimera/article/view/17979> (ultimo accesso 27.01.2025)
- De Mauro, Tullio. 2008. *Storia linguistica dell'Italia unita*. Roma-Bari: Laterza [ed. or. 1963].
- De Mauro, Tullio. 2014. *Storia linguistica dell'Italia repubblicana: dal 1946 ai giorni nostri*. Roma-Bari: Laterza.
- Guo, Kai & Wang, Deliang. 2023. To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education*

- and Information Technologies* 29. 8435–8463.  
<https://doi.org/10.1007/s10639-023-12146-0>
- Mai, Duong Thy Thuy & Da, Can Van & Hanh, Nguyen Van. 2024. The use of ChatGPT in teaching and learning: a systematic review through SWOT analysis approach. *Frontiers in Education* 9.  
<https://doi.org/10.3389/feduc.2024.1328769>
- Mariani, Luciano. 1988. Le abilità di studio: una prospettiva transdisciplinare. *Lingua e Nuova Didattica* 3. 56–67.
- Maturana, Humberto R. & Varela, Francisco J. 1980. *Autopoiesis and Cognition. The Realization of the Living*. Dordrecht: Springer.
- Rogers, Everett M. 1962. *Diffusion of Innovations*. New York: Free Press of Glencoe.
- Roncaglia, Gino. 2023. *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*. Roma-Bari: Laterza.
- Selwyn, Neil. 2023. On the limits of AI in education. *Nordisk tidsskrift for pedagogikk og kritikk* 10(3). 3–14. <https://doi.org/10.23865/ntpk.v10.6062> (ultimo accesso 31.05.2025)
- Serianni, Luca. 1987. Scripta manent. *Italiano & Oltre* 4. 182–184.
- Serianni, Luca. 2003. *Italiani scritti*. Bologna: il Mulino.
- Tavosanis, Mirko. 2024. Valutare la qualità dei testi generati in lingua italiana. *AI-Linguistica. Linguistic Studies on AI-Generated Texts and Discourses* 1(1).  
<https://doi.org/10.62408/ai-ling.v1i1.14> (ultimo accesso 27.01.2025)
- Uğraş, Hilal & Uğraş, Mustafa & Papadakis, Stamatios & Kalogiannakis, Michail. 2024. ChatGPT-supported education in primary schools: the potential of ChatGPT for sustainable practices. *Sustainability* 16(22). 9855.  
<https://doi.org/10.3390/su16229855>
- Yilmaz, Ramazan & Karaoğlan Yilmaz, Fatma Gizem. 2023. The effect of generative AI-based tool use on students' computational thinking skills and motivation in programming education. *Computers and Education: Artificial Intelligence* 4. <https://doi.org/10.1016/j.caeai.2023.100147> (ultimo accesso 15.04.2025)
- von Foerster, Heinz. 2003. *Understanding Understanding. Essays on Cybernetics and Cognition*. New York: Springer. <https://doi.org/10.1007/b97451>
- von Glasersfeld, Ernst. 1995. *Radical Constructivism: A Way of Knowing and Learning*. London & Washington: The Falmer Press.
- Zhu, Tiffany & Zhang, Kexun & Wang, William Yang. 2024. Embracing AI in education: understanding the surge in Large Language Model use by secondary students. *arXiv preprint*  
<https://doi.org/10.48550/arXiv.2411.18708> (ultimo accesso 31.05.2025)