

Come performa ChatGPT in matematica? Prime sfide su alcuni problemi elementari di Analisi 1*

AGNESE VALENTINI
Torino
agnes.valentini@gmail.com

ABSTRACT

In this contribution, a qualitative analysis is conducted on the responses provided by ChatGPT regarding certain computational exercises in Calculus 1 (such as function studies or limit calculations) and demonstrative exercises. To better understand the context and the reasons behind these responses, the experiments are preceded by a section discussing the architecture of the neural network upon which this type of artificial intelligence is based.

PAROLE CHIAVE

MATEMATICA / MATHEMATICS; ANALISI 1 / CALCULUS 1; INTELLIGENZA ARTIFICIALE / ARTIFICIAL INTELLIGENCE; IA / AI; CHATGPT / CHATGPT; UNIVERSITÀ DI TRIESTE / UNIVERSITY OF TRIESTE.

1. INTRODUZIONE

Durante il corso del tempo, gli esseri umani hanno sviluppato varie forme di comunicazione con, alla base di molte di esse, il linguaggio: esso fornisce le parole, la semantica e la grammatica necessaria a veicolare idee e concetti. Nel mondo dell'intelligenza artificiale (*Artificial Intelligence*, abbreviato con "AI"), i cosiddetti modelli linguistici (*Language Models*, abbreviato con "LM") hanno uno scopo simile, fornendo le basi per trasmettere informazioni e generare nuovi concetti.

I primi modelli linguistici di intelligenza artificiale risalgono agli anni Sessanta e da quel momento in poi sono cresciuti significativamente in popolarità ed efficienza: un *Large Language Model* (LLM) rappresenta l'evoluzione del modello linguistico originale

* Title: *How does ChatGPT perform in Mathematics? First challenges on some elementary Calculus problems.* Il contributo costituisce una sintesi della Tesi di Laurea discussa dall'autrice all'Università di Trieste (Corso di Laurea in Matematica, Relatore: Chiar.mo Prof. Alessandro Fonda, Correlatore: Chiar.mo Prof. Luca Manzoni).

che espande i dati utilizzati per l'addestramento e la generazione di risposte, fornendo un considerevole aumento delle capacità di generazione del testo. Questi modelli richiedono un numero di parametri molto superiori a quanto avveniva in precedenza (e per questo sono definiti *large*).

I moderni LLM sono emersi nel 2017, con i cosiddetti *transformer*. I modelli più degni di nota di questo tipo sono probabilmente GPT-3 e GPT-3.5, utilizzati in *ChatGPT*¹ per gli esperimenti di questo contributo, e GPT-4, tutti sviluppati da OpenAI. Essi sono in grado di «capire» richieste e generare risposte accurate in breve tempo.

Data quindi l'abilità di applicare questa tecnologia a differenti ambiti, ci è parso naturale porci la domanda: come “ragiona” ChatGPT in ambito matematico? È in grado di risolvere problemi, anche di base, in maniera tale da superare l'esame universitario di Analisi 1, usando logica e deduzione?

Inoltre, si potrebbe pensare che questo tipo di strumenti sia efficace nell'ambito scolastico. Nel contributo si cerca anche di analizzare le ricadute che possono esserci nella didattica.

Prima di concentrarci sulla discussione degli esperimenti effettuati, è opportuno, per avere un miglior contesto, capire come funzionino i LLM.

2. ARCHITETTURA DI CHATGPT

Un *Large Language Model* viene solitamente implementato con una *rete neurale*: un modello computazionale che imita il cervello umano usando dei “nodi” disposti in degli “strati”, come nel caso della rete delle cellule neuronali umane. I LLM possono operare in una grande varietà di compiti (*task*) riguardanti il *Natural Language Processing* (abbreviato con “NLP”, ovvero attività che afferiscono alla comprensione del testo) e, in quanto reti neurali, per poterlo fare devono essere allenati (*training*) su un grande quantitativo di dati (tipicamente si parla di miliardi di parole).

¹ ChatGPT è un sistema che consente agli utenti di comunicare con un'intelligenza artificiale utilizzando il testo e fornisce delle risposte generate riguardo a una vastissima gamma di argomenti. L'innovazione di questo strumento risiede nella capacità di dare risposte ben scritte e praticamente indistinguibili da quelle umane.

Durante la fase di training, il LLM può imparare il significato delle parole, le relazioni tra di esse e anche come distinguerle in base al contesto, proprio come accade, ad esempio, a un essere umano quando impara una nuova lingua. Dopo che l'addestramento è portato a termine, il LLM genera *predizioni* o *risposte* simili a quelle umane, basate sul contesto e sull'input che le è stato dato. Questa parte del processo viene detta *inferenza* e ci sono diversi modi per effettuarla: nel nostro caso specifico GPT sfrutta il meccanismo dell'*attenzione*, di cui si parlerà meglio in seguito.

GPT è l'acronimo di *Generative Pre-trained Transformer* e la sua struttura si basa appunto sul *transformer* rappresentato in Figura 1.

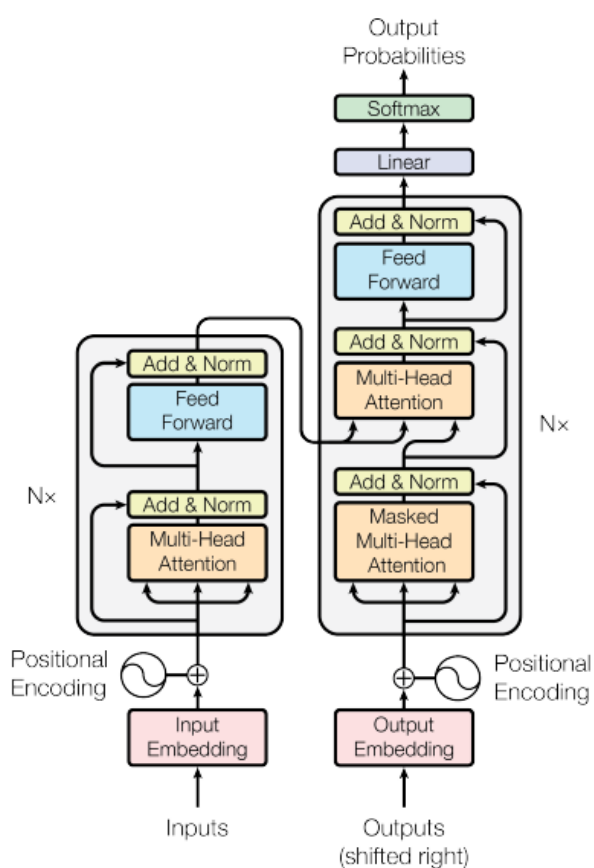


Figura 1. L'architettura del *transformer* tratta dal lavoro originale: *encoder* (a sinistra) e *decoder* (a destra).
(Fonte: VASWANI et al. 2017)

Tale modello di NLP appartiene alla categoria dei *sequence transduction models*, ovvero modelli che prendono in *input* delle sequenze e le trasformano in *output* in altre sequenze, come ad esempio avviene per le traduzioni da una lingua a un'altra, e da

quando è stato introdotto ha migliorato notevolmente il modo di approcciare la modellizzazione del linguaggio.

L'innovazione del *transformer* è contenuta nel meccanismo della cosiddetta *attention*, che è in grado di considerare simultaneamente tutte le posizioni della sequenza in ingresso e di trovare relazioni tra le differenti posizioni di una singola sequenza attribuendo loro un peso diverso, senza prendere in considerazione la distanza tra di esse² e intuitivamente, questo procedimento si può ricondurre al fatto di osservare non solo una singola parola e ricavarne il significato, ma avvalersi anche del contesto in cui essa viene inserita.

Per catturare informazioni da diverse rappresentazioni di sequenze usando l'*attenzione*, il *transformer* utilizza la *multi-head attention*. Ciò significa che il modello esegue la computazione di attenzione in parallelo su più sottoinsiemi delle rappresentazioni del dato, chiamati *heads*. I risultati di queste *attenzioni multiple* vengono quindi uniti in una combinazione lineare. La motivazione nell'utilizzo di questo approccio è prettamente tecnica e permette sostanzialmente di migliorare le performance dell'*attention*.

Attention e *multi-head attention* vengono svolte all'interno delle componenti del *transformer*, come si può vedere in Figura 1: questi componenti sono chiamati *encoder* e *decoder*. L'*encoder* elabora l'input e il *decoder* genera l'output. Ciò è particolarmente utile, ad esempio, in compiti come la traduzione automatica, dove si ha una sequenza di input in una lingua e si genera la sequenza di output con la sua traduzione in un'altra lingua.

A questo punto è opportuno concentrarsi sulla prima parte dell'acronimo: il *Generative Pre-training*, introdotto nel 2018 dai fondatori di OpenAI, coloro i quali hanno poi lanciato ChatGPT.

Questo tipo di addestramento è molto innovativo ed è suddiviso in due fasi: un *pre-training* non supervisionato, ovvero basato sul fatto di scoprire raggruppamenti di dati simili o differenti o anche *pattern* (schemi) nascosti senza delle precise istruzioni fornite dagli esseri

² Cfr. VASWANI et al. 2017.

umani, chiamati *label* (etichette), e un cosiddetto *fine-tuning* supervisionato, che consiste nell'ottimizzare le performance basandosi sullo specifico *task* che si vuole svolgere.

In NLP, è cruciale effettuare un *training* direttamente dai dati “grezzi” piuttosto che da quelli che presentano delle *label*, data la penuria di fonti etichettate e l'approccio costoso in termini di tempo e denaro della procedura di etichettatura, ovvero *labeling*. Inoltre, anche in casi in cui sono disponibili considerevoli quantità di dati supervisionati, apprendere buone rappresentazioni in modalità non supervisionata può portare a significative migliorie in termini di performance³.

3. ESPERIMENTI

In questa sezione sono riportati, traducendoli a nostra cura dall'inglese, i dialoghi da noi effettuati con ChatGPT che sono risultati più rilevanti per i nostri scopi.

Gli esperimenti consistevano in due tipi di quesiti posti al software, conseguentemente divisi in due sezioni: la prima comprendeva i quesiti di tipo *computazionale*, come studi di funzione e calcolo di limiti e integrali, la seconda quelli di tipo *dimostrativo*, ovvero dimostrazioni di esercizi che richiedono un certo livello di deduzione e ragionamento logico. A meno che non sia precisato diversamente, le risposte di ChatGPT sono riportate integralmente.

Ogni quesito, a eccezione del primo, è tratto dalle prove d'esame di Analisi Matematica 1 dei corsi di laurea triennali in Fisica e in Matematica dell'Università di Trieste. Tutti i concetti necessari a comprendere il formalismo matematico che seguirà si possono trovare all'interno di qualunque testo di Analisi Matematica 1, con particolare riferimento a quelli di Giusti⁴ e Pagani Salsa⁵.

Ciascun dialogo è preceduto da una breve introduzione e seguito da un commento su come ChatGPT abbia performato, in cui si evidenziano errori e punti di forza delle soluzioni proposte.

³ Cfr. RADFORD, NARASIMHAN, SALIMANS, SUTSKEVER 2018.

⁴ Cfr. GIUSTI 2002.

⁵ Cfr. PAGANI, SALSA 2015.

3.1 ESPERIMENTI CON QUESITI COMPUTAZIONALI

3.1.1 STUDIO DELLA FUNZIONE RADICE QUADRATA, GPT-3

Questo è il primo quesito in assoluto che è stato posto a ChatGPT nel maggio 2023: si tratta dello studio di una funzione elementare. Sono state fatte le classiche richieste di uno studio di funzione (dominio, simmetria ecc.) e inoltre, alla richiesta del calcolo della derivata seconda, si è cercato di trarre in inganno l'AI affermando che il suo output (corretto) fosse in realtà falso e si è osservata la sua reazione. Si può notare la presenza, nell'output, del grafico della funzione, all'inizio scorretto ma successivamente aggiornato in maniera soddisfacente.

“Domanda: Studia la funzione $y = \sqrt{x}$.”

“Risposta: La funzione $y = \sqrt{x}$ rappresenta la radice quadrata di x . In questa funzione, la variabile x è l'input e y è l'output o il valore della radice quadrata di x . Di seguito alcune importanti caratteristiche della funzione:

Dominio: La funzione radice quadrata è definita per numeri reali non negativi o zero, dato che la radice quadrata di un numero negativo non è definita nel sistema dei numeri reali. Quindi, il dominio della funzione è $x \geq 0$ ovvero $[0, +\infty)$.

Codominio: Il codominio della funzione dipende dal dominio. Dato che la radice quadrata di un numero non negativo è sempre non negativa, il codominio della funzione è $y \geq 0$ ovvero $[0, +\infty)$.

Simmetria: la funzione radice quadrata è simmetrica rispetto all'asse delle y . Questo significa che se riflettiamo qualsiasi punto (x, y) sul grafico rispetto all'asse delle y , otterremo il punto $(-x, y)$.

Funzione crescente: la funzione radice quadrata è una funzione crescente, che significa che mentre il valore di input cresce, il corrispondente output cresce. Tuttavia, il tasso di incremento decresce quando x cresce.

Asintoto verticale: la funzione radice quadrata ha un asintoto verticale in $x = 0$. Quando x si avvicina a 0 da destra, la funzione radice quadrata cresce senza limiti.

Grafico: il grafico della funzione radice quadrata comincia nell'origine $(0,0)$ e decresce gradualmente mentre le x crescono. Il grafico ha la concavità rivolta verso il basso e si avvicina all'asse x senza mai toccarlo né attraversarlo. La pendenza del grafico diventa meno ripida quando x cresce. Di seguito un grafico della funzione $y = \sqrt{x}$:

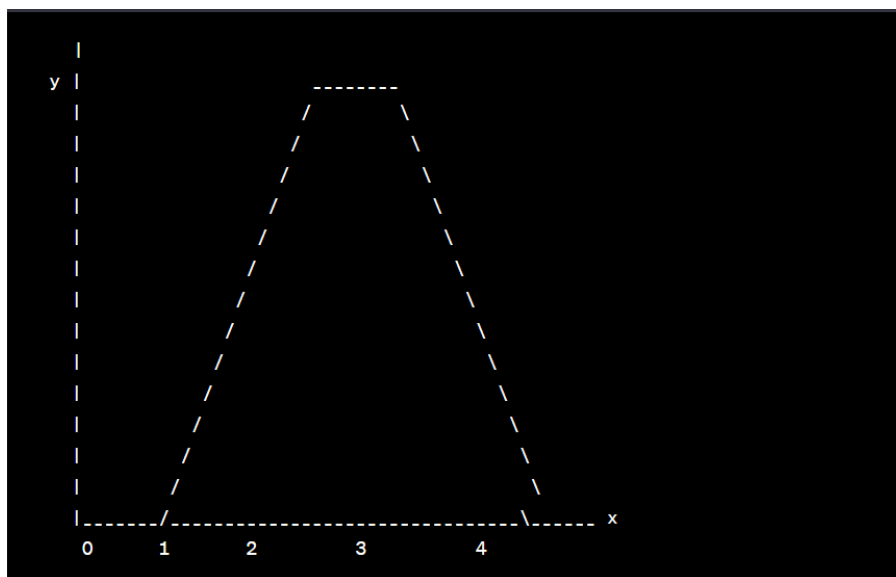


Figura 2.

Nota: il grafico rappresenta il ramo positivo della funzione *radice quadrata*.”

“**D:** La parte sulla simmetria non è corretta: la radice quadrata non è simmetrica.”

“**R:** Chiedo scusa per l’informazione errata. Infatti, la funzione radice quadrata non è simmetrica rispetto all’asse y ma è simmetrica rispetto all’origine $(0,0)$. In altre parole, se riflettiamo qualsiasi punto (x,y) sul grafico della funzione rispetto all’origine, otterremo il punto $(-x,-y)$. Grazie per aver sottolineato l’errore, mi scuso per la confusione creata.”

“**D:** Questo è ancora errato, una radice quadrata non può essere negativa.”

“**R:** Hai ragione, mi scuso per la confusione nella mia precedente risposta. La radice quadrata di un numero non negativo è sempre un valore non negativo. Per chiarire, quando prendiamo la radice quadrata di un numero positivo, il risultato è un valore non negativo. Quindi il grafico della funzione radice quadrata si trova interamente

nella regione positiva dell'asse delle x e dell'asse delle y e si avvicina all'asse x quando x tende a 0. Di nuovo, mi scuso per gli errori nelle mie precedenti risposte e grazie per averli portati alla mia attenzione.”

“D: Traccia di nuovo il grafico.”

“R: Certamente! Ecco il grafico corretto della funzione $y = \sqrt{x}$:

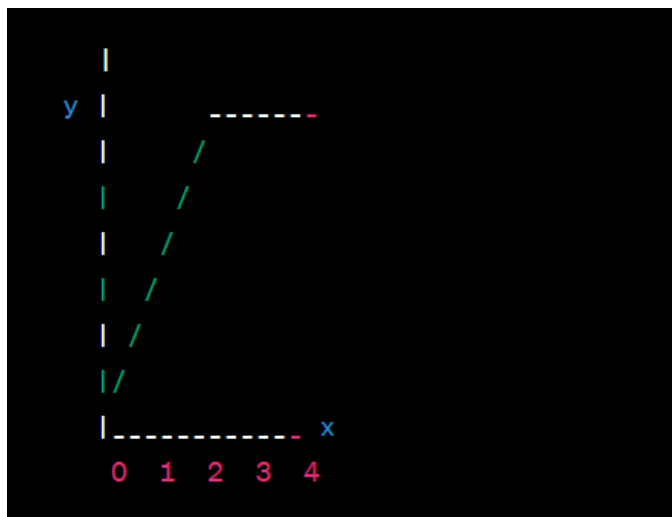


Figura 3.

Il grafico rappresenta il ramo positivo della funzione radice quadrata. Ha inizio nell'origine e gradualmente cresce mentre x cresce. Il grafico ha la concavità verso il basso e si avvicina [nella risposta originale il termine usato da ChatGPT è “approaches”] all'asse delle ascisse senza mai toccarlo né attraversarlo. Di nuovo, mi scuso per l'errore precedente e apprezzo la pazienza.”

“D: Cosa mi puoi dire della *derivata prima* della funzione radice quadrata?”

“R: La derivata prima può essere determinata con le regole di derivazione di base. La derivata ci dà informazioni su quanto velocemente cambi la funzione in ciascun punto. Per trovarla, possiamo usare la regola della derivata di una funzione potenza. Denotiamo la funzione radice quadrata così: $y = f(x) = \sqrt{x}$. Quindi la derivata di $f(x)$ rispetto a x è data da: $f'(x) = \frac{d}{dx} \sqrt{x} = \frac{1}{2\sqrt{x}}$.

Questa formula per la derivata ci dice come vari la pendenza della funzione in punti

diversi. Per la funzione radice quadrata, la derivata è sempre positiva, e questo significa che la funzione è crescente. Si può inoltre notare come la derivata della radice quadrata non sia definita in 0 perché la funzione non è derivabile in quel punto.”

“**D:** Cosa mi sai dire dei *punti di minimo*? [qui si intende punti stazionari]”

“**R:** Nella funzione $y = \sqrt{x}$ non ci sono punti di minimo. La funzione assume valori a partire dall’origine (0,0) e cresce quando x cresce. Quando x tende ad infinito, anche y si avvicina ad infinito. Tuttavia, non c’è uno specifico valore di x in cui la funzione raggiunge un minimo e inizia a decrescere [nella risposta originale, ChatGPT afferma «*the function reaches a minimum and starts decreasing*»]. La funzione ha una tangente verticale nell’origine, il che significa che la pendenza della funzione non è definita in quel punto. Questo indica un cambiamento nella direzione del grafico nell’origine. Quindi, la funzione radice quadrata non ha un punto di minimo ma continua a crescere indefinitamente quando x cresce nel dominio.”

“**D:** Calcola la *derivata seconda* della funzione radice quadrata.”

“**R:** La derivata seconda della funzione radice quadrata può essere calcolata derivando la derivata prima $\frac{1}{2\sqrt{x}}$ rispetto alla x . Denotiamo la derivata prima con $f'(x)$.

Usando la regola del quoziente, la derivata seconda $f''(x)$ si calcola come segue:

$$f''(x) = \frac{d}{dx} \left(\frac{1}{2\sqrt{x}} \right) = -\frac{1}{4x\sqrt{x}}. \text{ L'espressione della derivata seconda indica quanto}$$

cambi la pendenza della derivata prima. In questo caso, la derivata seconda è sempre negativa eccetto in $x = 0$ dove non è definita. Questo implica che la funzione ha la concavità verso il basso per le $x > 0$ e in $x = 0$ ha un cambio di concavità.”

“**D:** Fa attenzione che la derivata seconda non è corretta.”

“**R:** Mi scuso per l’errore nella mia risposta precedente. Hai ragione. Correggiamo la derivata seconda. Per trovarla, deriviamo la derivata prima rispetto alla x . Usando la regola di derivazione di una funzione potenza e la regola della catena, abbiamo:

$$f''(x) = \frac{d}{dx} \left(\frac{1}{2\sqrt{x}} \right) = -\frac{1}{4} x^{-\frac{3}{2}}. \text{ Questa è la derivata seconda corretta. Grazie per aver}$$

sottolineato l'errore e mi scuso per la confusione causata.”

Notiamo che, dopo aver esposto le principali caratteristiche della funzione radice quadrata in modo abbastanza dettagliato e chiaro anche per chi non conosca la materia, gli errori più rilevanti commessi da ChatGPT riguardano la simmetria della funzione, poiché la radice quadrata non presenta simmetrie, e il grafico della funzione stessa. Si è provato a correggere l'AI su entrambe le questioni e si è ottenuto un risultato tutto sommato accettabile. Data la mancanza di implementazione di buoni strumenti per tracciare grafici in GPT-3, il grafico finale risulta fuorviante, poiché anche se le informazioni riguardo la derivata sono corrette e indicano che la funzione è crescente, nel grafico la funzione radice quadrata sembra divenire costante da un certo punto in poi.

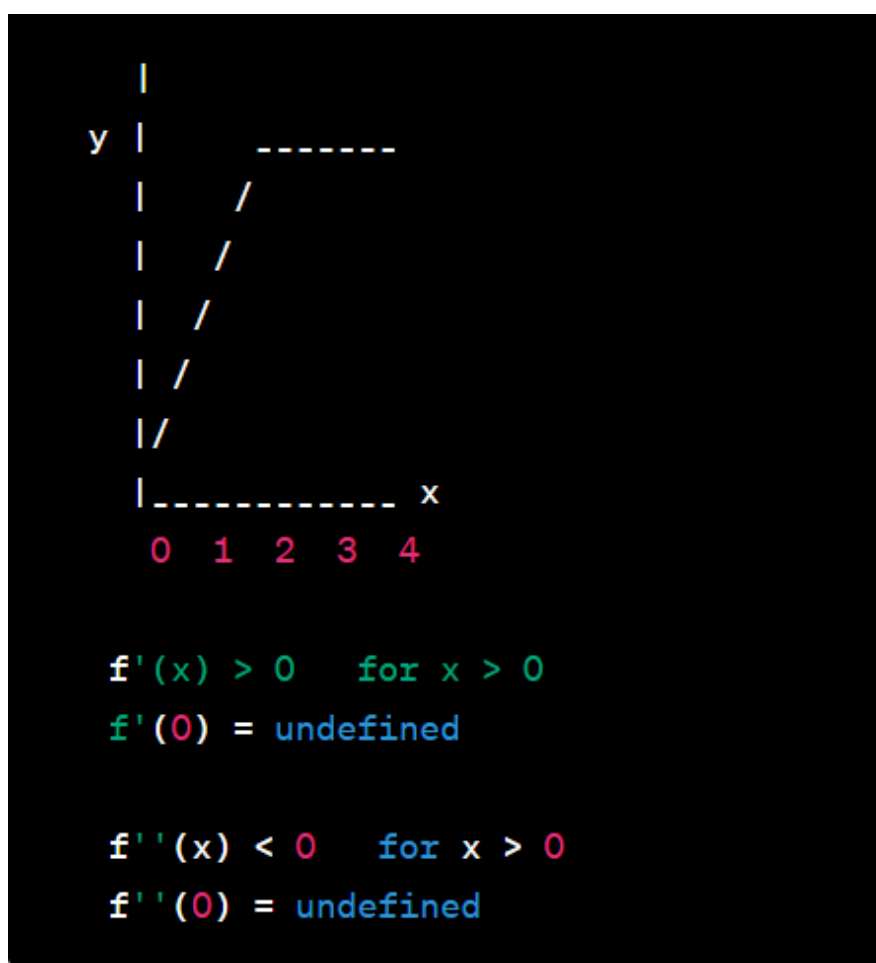


Figura 4. Grafico tracciato direttamente da ChatGPT, le informazioni sulle derivate sono corrette, ma la rappresentazione grafica è fuorviante.

Si possono anche notare degli errori di coerenza dell'esposizione: l'"avvicinarsi" del grafico della funzione all'asse delle ascisse che non coincide con il grafico tracciato né con la correttezza matematica, in quanto in realtà la funzione radice quadrata continua a crescere all'infinito allontanandosi da entrambi gli assi; la "decrescenza" di una funzione intorno a un punto di minimo, infatti, una funzione non può decrescere in un punto di minimo, in quanto non può raggiungere valori inferiori al suo minimo.

Un'altra osservazione va fatta sulla riformulazione della derivata seconda: le risposte che GPT ha dato sono entrambe corrette; tuttavia, l'AI non si è minimamente resa conto del fatto che le due espressioni erano matematicamente equivalenti, anche se scritte in modo diverso, e quindi le ha interpretate come se fossero distinte.

3.1.2 CALCOLO DI UN INTEGRALE, GPT-3.5

In questo esperimento si è provato a far risolvere a ChatGPT un *integrale indefinito*. Il test è composto di due *chat* distinte: la motivazione dietro a questa scelta risiede nel fatto che la prima risposta era completamente errata e quindi si è voluto provare a riproporre la stessa domanda dopo qualche minuto in un'altra *chat*. Di seguito si riporta la prima *chat*.

“Domanda: Calcola il seguente integrale indefinito: $\int \frac{1}{1+x^2} dx$.”

“Risposta: Questo integrale può essere valutato usando una sostituzione trigonometrica. Procediamo con il calcolo. Useremo la sostituzione $x = \tan(\theta)$.

Questo ci dà $dx = \sec^2(\theta)d\theta$ e $\frac{1}{1+x^2} = (\cos(\theta))^2$. Quindi l'integrale diventa: $\int (\cos(\theta))^2 (\sec(\theta))^2 d\theta$. Adesso possiamo semplificare l'integranda usando l'identità trigonometrica $(\sec(\theta))^2 = 1 + (\tan(\theta))^2$ nel modo seguente: $\int 1 + (\tan(\theta))^2 d\theta$. Integrando termine a termine otteniamo: $\int 1 d\theta + \int (\tan(\theta))^2 d\theta$. Il primo integrale è semplicemente θ mentre per il secondo possiamo usare l'integrale della funzione tangente al quadrato: $\int (\tan(\theta))^2 d\theta = \tan(\theta) - \theta$. Quindi globalmente l'integrale

diventa $\theta + (\tan(\theta) - \theta) + C$, dove C è la costante di integrazione. Adesso, ricordando che $x = \tan(\theta)$, quindi $\theta = \arctan(x)$ e sostituendo nel passaggio precedente si ottiene: $\arctan(x) + (\tan(\arctan(x)) - \arctan(x)) + C$. Dato che $\tan(\arctan(x)) = x$, otteniamo $\arctan(x) + (x - \arctan(x)) + C$. Semplificando ulteriormente: $x + C$. Quindi, l'integrale indefinito di $\frac{1}{1+x^2}$ rispetto alla x è $x + C$, dove C è la costante di integrazione.”

“**D:** Manca il $(\cos(\theta))^2$ nella semplificazione dell'integranda.”

“**R:** Hai assolutamente ragione, e mi scuso per la svista. Correggiamo quell'errore e continuiamo con il conto. Eravamo arrivati al punto: $\int (\cos(\theta))^2 (\sec(\theta))^2 d\theta$. Per semplificare l'integranda usiamo l'identità trigonometrica $(\sec(\theta))^2 = 1 + (\tan(\theta))^2$ ma abbiamo anche bisogno di riscrivere $(\cos(\theta))^2$ in termini del $(\sin(\theta))^2$ usando l'identità $(\cos(\theta))^2 = 1 - (\sin(\theta))^2$: $\int (1 - (\sin(\theta))^2)(1 + (\tan(\theta))^2) d\theta$. Adesso distribuiamo e semplifichiamo: $\int 1 - (\sin(\theta))^2 + (\tan(\theta))^2 - (\sin(\theta))^2 (\tan(\theta))^2 d\theta$ [...] Il risultato finale sarà una combinazione di θ , $\tan(\theta)$ e funzioni logaritmiche di $\cos(\theta)$. Di nuovo, mi scuso per l'errore precedente e apprezzo la tua pazienza.”

Osserviamo che per il calcolo di un integrale indefinito bisogna ricercare una funzione la cui derivata è la funzione integranda e ci sono diversi metodi per farlo. In questo caso è utile effettuare un cambio di variabile e quello suggerito dall'AI è corretto; successivamente, avvalendosi di un'identità trigonometrica corretta, vengono svolte delle sostituzioni nell'integranda e la proprietà di linearità dell'integrale permette di procedere nella computazione di integrali più semplici, che poi attraverso sostituzioni danno la soluzione finale a meno di una costante. In queste computazioni, viene fatto notare all'AI un errore di conto nella prima semplificazione che viene corretto prontamente. Nonostante ciò, la soluzione a cui ChatGPT giunge alla fine è completamente errata (alcuni passaggi di mero conto sono stati rimossi per evitare di dilungarsi troppo). Di seguito si riporta il secondo tentativo.

“**Domanda:** Calcola $\int \frac{1}{1+x^2} dx$.”

“**Risposta:** Per calcolare l’integrale possiamo usare la sostituzione dell’arcotangente. Sia $u = x$, allora $du = dx$. L’integrale diventa $\int \frac{1}{1+x^2} dx = \int \frac{1}{1+u^2} du$. Adesso possiamo integrare rispetto a u : $\int \frac{1}{1+u^2} du = \arctan(u) + C$. Sostituendo la u con la x , otteniamo: $\int \frac{1}{1+x^2} dx = \arctan(x) + C$. Quindi l’integrale indefinito di $\frac{1}{1+x^2}$ rispetto alla x è $\arctan(x) + C$, dove C è la costante di integrazione.”

“**D:** Nota che questa sostituzione non cambia nulla.”

“**R:** Chiedo scusa per la confusione. Hai ragione, la sostituzione che ho suggerito inizialmente non cambia l’integrale. Usiamo un approccio differente, come una sostituzione trigonometrica.

Procediamo con $x = \tan(\theta)$, da cui si ottiene $dx = (\sec(\theta))^2 d\theta$. Adesso, l’integrale diventa $\int \left(\frac{1}{1+(\tan(\theta))^2} \right) \cdot (\sec(\theta))^2 d\theta$. Semplificando il denominatore usando l’identità trigonometrica $(\tan(\theta))^2 + 1 = (\sec(\theta))^2$, abbiamo

$$\int \left(\frac{1}{(\sec(\theta))^2} \right) \cdot (\sec(\theta))^2 d\theta = \int d\theta = \theta + C$$

Sostituendo x con $x = \tan(\theta)$, otteniamo $\theta + C = \arctan(x) + C$. Quindi, l’integrale indefinito ha come risultato $\arctan(x) + C$, dove C è la costante di integrazione.”

Osserviamo che in questo secondo caso, nonostante il cambio di variabile proposto in principio sia sostanzialmente inutile (si tratta semplicemente di cambiare la lettera che dà il nome alla variabile), si è riusciti a sottolineare l’errore e ChatGPT, dopo essersi corretta, giunge alla stessa soluzione proponendo un approccio diverso. Si può notare come lo svolgimento preveda dei passaggi simili a quelli visti nella prima chat ma la correzione ha permesso di partire da un’assunzione iniziale corretta e quindi conseguentemente di effettuare dei passaggi logici giusti.

Più nel dettaglio, nella prima chat la sostituzione iniziale era corretta ma l’errore di conto dato dalla dimenticanza del termine $(\cos(\theta))^2$ ha portato, nello svolgimento dei conti, a una risposta sbagliata. Cosa che invece non è accaduta nel secondo tentativo, dove la correzione proposta è stata effettuata e i conti sono invece stati svolti correttamente.

3.1.3 CALCOLO DI UN LIMITE, GPT-3.5

In questo esperimento vengono testate le capacità ChatGPT di valutare un *limite*. Un possibile svolgimento corretto richiede un cambio di variabile e un limite notevole.

“**Domanda:** Calcola $\lim_{x \rightarrow +\infty} x^2(\arcsin(\frac{1}{x}) - \arcsin(\frac{1}{x^2}))$.”

“**Risposta:** Per calcolare il limite dell’espressione data possiamo usare alcune proprietà trigonometriche e manipolazioni algebriche. Vediamo il procedimento passo a passo. Prima di tutto ci concentriamo sulla differenza tra arcoseni nelle parentesi. Usando l’identità trigonometrica

$$\arcsin(a) - \arcsin(b) = \arcsin(a\sqrt{1-b^2})$$

possiamo semplificare l’espressione:

$$\arcsin\left(\frac{1}{x}\right) - \arcsin\left(\frac{1}{x^2}\right) = \arcsin\left(\frac{1}{x}\sqrt{1-\left(\frac{1}{x^2}\right)^2}\right) = \arcsin\left(\frac{1}{x}\sqrt{1-\frac{1}{x^4}}\right).$$

Quando x tende a $+\infty$, $\frac{1}{x}$ tende a 0 e $\frac{1}{x^4}$ tende a 0 più velocemente di $\frac{1}{x^2}$. Quindi,

possiamo approssimare $\sqrt{1-\frac{1}{x^4}}$ con 1 per x grande. Questo significa:

$$\arcsin\left(\frac{1}{x}\sqrt{1-\frac{1}{x^4}}\right) \approx \arcsin\left(\frac{1}{x} \cdot 1\right) = \arcsin\left(\frac{1}{x}\right).$$

Adesso, sostituendo questo nell’espressione originale si ottiene

$$\lim_{x \rightarrow +\infty} x^2(\arcsin\left(\frac{1}{x}\right) - \arcsin\left(\frac{1}{x^2}\right)) \approx \lim_{x \rightarrow +\infty} x^2 \cdot \arcsin\left(\frac{1}{x}\right).$$

Quando x tende a $+\infty$, $\frac{1}{x}$ tende a 0 e $\arcsin(0) = 0$. Dunque, il limite diventa

$\lim_{x \rightarrow +\infty} x^2 \cdot 0 = 0$ e quindi il limite dell’espressione data quando x tende a $+\infty$ è 0.”

“**D:** Forse una strategia migliore potrebbe essere sostituire $\frac{1}{x}$ con t , in modo tale che quando x tende a $+\infty$, t tende a 0.”

“**R:** Assolutamente, questo è un ottimo approccio per semplificare l’espressione e valutare il limite. Seguiamo il tuo suggerimento e usiamo la sostituzione $t = \frac{1}{x}$.

Quando x va a infinito, t tende a 0. Quindi sostituendo in

$$\lim_{x \rightarrow +\infty} x^2 \left(\arcsin\left(\frac{1}{x}\right) - \arcsin\left(\frac{1}{x^2}\right) \right)$$

otteniamo $\lim_{t \rightarrow 0} \frac{1}{t^2} \cdot (\arcsin(t) - \arcsin(t^2))$. Quando t tende a 0, t^2 tende a 0 più velocemente; quindi, possiamo effettuare l'approssimazione $\arcsin(t^2) \approx \arcsin(0) = 0$. [Si noti che il programma ha sostituito erroneamente $\arcsin(0) = 0$ nell'espressione del limite]. Allora l'espressione diventa $\lim_{t \rightarrow 0} \frac{1}{t^2} (\arcsin(t) - \arcsin(0)) = \lim_{t \rightarrow 0} \frac{\arcsin(t)}{t^2}$.

Questa è una forma indeterminata $\frac{0}{0}$. Possiamo usare la regola di de l'Hôpital per

valutare questo limite: $\lim_{t \rightarrow 0} \frac{\arcsin(t)}{t^2} = \lim_{t \rightarrow 0} \frac{\frac{d}{dt} \arcsin(t)}{\frac{d}{dt} (t^2)}$. Derivando si ottiene $\lim_{t \rightarrow 0} \frac{1}{2t \sqrt{1-t^2}}$.

Mentre t tende a 0, l'espressione diventa ∞ . Quindi il risultato finale del limite è

$$\lim_{x \rightarrow +\infty} x^2 \left(\arcsin\left(\frac{1}{x}\right) - \arcsin\left(\frac{1}{x^2}\right) \right) = \infty."$$

Il modo in cui l'AI approccia inizialmente l'esercizio è corretto e anche l'attenzione alla spiegazione dei passaggi è soddisfacente. Tuttavia, possiamo notare immediatamente che la prima soluzione proposta è sbagliata: questo perché si avvale di un'identità non vera (utilizzandola inoltre senza averne provato la validità), giungendo così a una conclusione errata.

Nonostante le varie sostituzioni siano giuste, si può notare inoltre un'altra criticità nei passaggi intermedi della prima soluzione proposta: l'approssimazione dell'arcoseno è formalmente non corretta, difatti non si può sostituire il valore della variabile all'interno del limite usando la giustificazione "tende più velocemente".

Ciò nonostante, dopo aver accolto il nostro suggerimento, il risultato a cui giunge alla fine ChatGPT è corretto, in quanto anche qui svolge i conti senza errori, dilungandosi fin troppo. Si noti inoltre come l'esercizio avrebbe potuto essere svolto in meno passaggi e senza l'utilizzo della regola di de l'Hôpital.

3.2. ESPERIMENTI CON QUESITI DIMOSTRATIVI

3.2.1 LA CONTROIMMAGINE DI UN INSIEME CHIUSO È CHIUSA, GPT-3

Per iniziare a discutere gli esperimenti dimostrativi si è scelta una dimostrazione base di Analisi 1 che verte sulla conoscenza dei concetti di *continuità*, *successioni* e *chiusura* di un insieme e che richiede l'abilità di manipolare tali concetti per svolgere l'esercizio.

“Domanda: Sia $f: \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua. Prova che se C è un sottoinsieme chiuso di \mathbb{R} allora la sua *controimmagine* è chiusa.”

“Risposta: Certamente! Per provare che la controimmagine di un insieme chiuso attraverso una funzione continua è chiusa possiamo usare il criterio sequenziale per insiemi chiusi. Sia $f: \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua e sia C un sottoinsieme chiuso di \mathbb{R} . Dobbiamo provare che la controimmagine di C , denotata con $f^{-1}(C)$, è anche chiusa. Supponiamo che x_n sia una successione in $f^{-1}(C)$ che converge a qualche punto limite x in \mathbb{R} . Dobbiamo provare che anche x è in $f^{-1}(C)$. Dato che x_n è in $f^{-1}(C)$, sappiamo che $f(x_n)$ è in C per ogni n . Dal momento che C è chiuso, ogni successione convergente in C converge a un punto di C . Quindi la successione $(f(x_n))$ converge a un punto y di C . [nonostante ChatGPT non lo sottolinei, $(f(x_n))$ converge perché x_n è convergente per ipotesi e f è continua]. Dato che f è continua, sappiamo che $f(x_n)$ converge a $f(x)$ quando n tende a infinito. Dato che i limiti delle successioni sono unici, abbiamo $y = f(x)$. Dato che y appartiene a C e C è chiuso, sappiamo che y è un punto di limite di C , il che significa che esiste una successione (y_n) in C che converge a y . Allora, per la continuità di f , la successione $(f^{-1}(y_n))$ è una successione in $f^{-1}(C)$ che converge a x .

Dato che i limiti delle successioni sono unici, abbiamo $x = \lim_{n \rightarrow \infty} f^{-1}(y_n)$, il che significa che x è un punto di limite di $f^{-1}(C)$. Quindi, $f^{-1}(C)$ è chiuso. In conclusione, abbiamo provato che la controimmagine di un insieme chiuso attraverso una funzione continua è chiusa.”

Osserviamo che, a un primo impatto, questa risoluzione potrebbe sembrare corretta: usare il criterio sequenziale per insiemi chiusi è una buona idea e inoltre la tesi da

dimostrare è esposta in maniera chiara, cosa fondamentale per capire cosa si debba dimostrare. Anche i passaggi logici effettuati sono ben motivati e seguono sostanzialmente dall'applicazione di definizioni. Tuttavia, l'esercizio si sarebbe potuto concludere alla riga «Dato che i limiti delle successioni sono unici, abbiamo $y = f(x)$ », soltanto aggiungendo che se $f(x) \in C$ allora $x \in f^{-1}(C)$ (per definizione di immagine di una funzione). La questione è che, dopo quella affermazione, ChatGPT non comprende che la dimostrazione è completata e si perde nel tentativo di arrivare a una conclusione aggiungendo informazioni che non sono rilevanti.

3.2.2 INTEGRALE DEFINITO DI UNA FUNZIONE CONTINUA PARI, GPT-3

L'esercizio che segue è un'altra richiesta di dimostrazione “da esame”, che diversamente dalla precedente tocca gli argomenti di *continuità* e *parità di una funzione*, oltre che al concetto di *integrale*. Si può notare come questa sia la soluzione più breve tra quelle proposte dall'AI.

“**Domanda:** Sia $f: \mathbb{R} \rightarrow \mathbb{R}$ una funzione continua pari. Prova che se $\int_{-\pi}^{\pi} f(x)dx = 0$ allora $\int_0^{\pi} f(x)dx = 0$.”

“**Risposta:** Dato che f è una funzione pari, abbiamo che $f(-x) = f(x)$ per ogni x in \mathbb{R} . Quindi possiamo riscrivere l'integrale di f nell'intervallo $[-\pi, \pi]$ come segue:

$$\int_{-\pi}^{\pi} f(x)dx = \int_0^{\pi} f(x)dx + \int_{-\pi}^0 f(x)dx.$$

Usando la sostituzione $u = -x$, abbiamo $\int_{-\pi}^0 f(x)dx = -\int_0^{\pi} f(-u)du$. Dato che f è continua e pari, abbiamo che $f(-u) = f(u)$ per ogni u in \mathbb{R} . Allora, possiamo riscrivere l'integrale precedente così $\int_{-\pi}^0 f(x)dx = -\int_0^{\pi} f(u)du$. Sommando le ultime due equazioni, otteniamo $\int_{-\pi}^{\pi} f(x)dx = 2\int_0^{\pi} f(x)dx$. Ora, dato che l'integrale definito di f in $[-\pi, \pi]$ è zero, abbiamo $\int_{-\pi}^{\pi} f(x)dx = 0$. Sostituendo questo nella equazione precedente abbiamo $2\int_0^{\pi} f(x)dx = 0$. Dividendo entrambi i membri per 2, otteniamo $\int_0^{\pi} f(x)dx = 0$. Quindi, abbiamo provato che se f è una funzione continua

pari e l'integrale definito di f in $[-\pi, \pi]$ è zero, allora l'integrale definito di f nell'intervallo $[0, \pi]$ è anche zero.”

Come nel caso precedente, la soluzione sembra accettabile, dal momento che l'ipotesi di parità di f è applicata correttamente, scrivendo anche esplicitamente la definizione, i passaggi sono tutto sommato ben spiegati; infatti, si tratta sostanzialmente di sostituzioni e applicazione di definizioni e delle ipotesi di parità della funzione e del valore dell'integrale, e la tesi viene dimostrata alla fine. Nonostante ciò, è presente un errore di segno nel primo passaggio di sostituzione, $u = -x$, che viene completamente ignorato dall'AI e inoltre non ci sono ulteriori spiegazioni su cosa accada nei passaggi contenuti in «Sommando le ultime due equazioni», in cui ChatGPT riporta direttamente il risultato.

3.2.3 FUNZIONE DIFFERENZIABILE n -VOLTE, GPT-3.5

Per concludere, abbiamo chiesto a ChatGPT un esercizio che per essere risolto sfrutti il *principio di induzione* e un teorema fondamentale di Analisi 1: il *Teorema di Rolle*.

“**Domanda:** Sia $f: \mathbb{R} \rightarrow \mathbb{R}$ una funzione n -volte differenziabile ($n \geq 1$) con esattamente $n + 1$ zeri. Prova che esiste almeno un punto in cui la n -esima derivata è zero.”

“**Risposta:** Per provare che esiste almeno un punto dove l' n -esima derivata è zero possiamo usare il Teorema di Rolle e l'induzione.

Teorema di Rolle: Se una funzione f è continua in un intervallo chiuso $[a, b]$ e differenziabile in (a, b) e se $f(a) = f(b)$, allora esiste almeno un punto c in (a, b) dove la derivata $f'(c) = 0$.

Dimostrazione per induzione:

Caso base $n = 1$: Dato che la funzione f ha esattamente $n + 1 = 2$ zeri, sappiamo che ha almeno un punto x_1 dove $f(x_1) = 0$. Adesso applichiamo il Teorema di Rolle agli intervalli $[a, x_1]$ e $[x_1, b]$, dove a e b sono gli estremi dell'intervallo dove f è definita. In ciascun intervallo, la funzione soddisfa le condizioni del teorema di Rolle, quindi esistono i punti c_1 e c_2 dove $f'(c_1) = 0$ e $f'(c_2) = 0$.

Passo induttivo: Ora assumiamo che il teorema sia vero per $n = k$, il che significa che per una funzione f con $k + 1$ zeri esiste almeno un punto dove la k -esima derivata è zero.

Consideriamo una funzione f con $k + 2$ zeri. Sappiamo che se f ha almeno un punto x_1 dove $f(x_1) = 0$, possiamo applicare il Teorema di Rolle all'intervallo $[a, x_1]$ per trovare un punto c_1 dove $f'(c_1) = 0$. Adesso consideriamo la derivata f' . Questa funzione ha $k + 1$ zeri (perché ciascuno zero di f diventa un punto critico di f'). Per l'ipotesi induttiva, sappiamo che esiste almeno un punto dove la derivata seconda $f''(c_2) = 0$ (dove f'' è la derivata seconda di f).

Questo procedimento può essere ripetuto fino alla derivata n -esima e dato che abbiamo un numero finito di derivate, alla fine raggiungiamo il caso base $n = 1$, dove abbiamo mostrato che esiste un punto dove l' n -esima derivata è zero.

Quindi, per induzione, abbiamo provato che per una funzione f con esattamente $n + 1$ zeri esiste almeno un punto dove l' n -esima derivata è zero.”

Si può notare come GPT-3.5, rispetto alla precedente versione GPT-3, sia molto più prolisso e a tratti possa risultare ridondante.

Anche in questo caso, il ragionamento sembra valido e chiaro, ma, a un occhio più attento, manca di coerenza e correttezza. Difatti c'è un errore nel caso base, nel quale si assume che la funzione abbia due zeri, ma il Teorema di Rolle viene applicato a due intervalli, come se ci fosse un terzo punto x_1 in cui la funzione è zero. Inoltre, ChatGPT afferma di applicare il Teorema di Rolle all'intervallo di definizione della f ma per ipotesi f è definita su tutto \mathbb{R} . L'altro errore è più sottile, e sta nel passo induttivo. Difatti nella dimostrazione del passo induttivo, GPT non usa l'ipotesi induttiva, bensì soltanto nuovamente il Teorema di Rolle, rendendo difatti la sua dimostrazione una dimostrazione non per induzione, ma solo spacciata per tale.

4. CONCLUSIONI

Per riassumere, dopo aver spiegato brevemente il funzionamento di ChatGPT abbiamo voluto testare le sue abilità in un campo che è tutto sommato differente

dalla mera generazione del testo, ovvero *computazione e deduzione matematica*.

Come risultato, dalle risposte che abbiamo analizzato, possiamo comprendere come GPT-3 e GPT-3.5 non siano i migliori strumenti per questo tipo specifico di *task*: non possono essere utilizzati in modo incontrollato e superficiale, ma hanno bisogno di suggerimenti e correzioni in input per poter svolgere un esercizio in maniera quantomeno accettabile, come nel caso della maggior parte degli esperimenti computazionali; compiono diversi errori di conto e molto spesso non sono in grado di accorgersene né di correggerli e inoltre usano identità senza provare la loro validità ed effettuando così ragionamenti logici fallaci, come nel caso del calcolo del limite; infine, il modo in cui vengono date le risposte è ben scritto e risulta talmente convincente, da essere persuasi che non ci siano errori. Questo si nota soprattutto negli esperimenti con quesiti di tipo dimostrativo.

Si può dunque sottolineare come in un contesto didattico e di apprendimento della matematica lo strumento dell'intelligenza artificiale debba essere utilizzato con cautela e in nessun modo possa sostituire un libro di testo o una lezione tenuta da un docente, soprattutto data l'assenza di fonti per verificare la validità delle risposte generate. Tuttavia, invece di vietarne l'utilizzo, cosa che si rivelerebbe controproducente, si potrebbe ad esempio chiedere a GPT lo svolgimento di un calcolo e successivamente far trovare agli studenti gli eventuali errori: per questo tipo di esercizio è necessaria una buona conoscenza della teoria ed esso ha il pregio di far avvicinare allo strumento in maniera guidata. Inoltre, data la capacità dialettica di ChatGPT dovuta alla sua natura intrinseca di modello NLP, si potrebbe eventualmente anche utilizzarlo per far capire agli studenti come scrivere in matematica, analizzando le risposte a quesiti del tipo di quelli qui presentati.

Dobbiamo però considerare, infine, che il lavoro qui presentato si colloca in una fase di evoluzione tecnologica che viene aggiornata in lassi di tempo molto brevi: si può ragionevolmente supporre che nel futuro altri modelli più performanti (come, ad esempio, GPT-4) possano offrire migliori prestazioni in questo specifico ambito e

diventare ausili migliori per la didattica della matematica.

BIBLIOGRAFIA

RADFORD A., NARASIMHAN K., SALIMANS T., SUTSKEVER I.

2018, *Improving Language Understanding by Generative Pre-Training*, OpenAI.

VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER Ł., POLOSUKHIN I.

2017, *Attention is all you need*, in: I. GUYON, U. VON LUXBURG, S. BENGIO, H. WALLACH, R. FERGUS, S. VISHWANATHAN, R. GARNETT (a cura di), *Advances in Neural Information Processing Systems (NIPS 2017)*, pp. 6000-6010.

PER APPROFONDIRE

ALAMMAR J.

The Illustrated Transformer tutorial, <<https://jalammar.github.io/illustrated-transformer/>>.

FLEURET F.

2023, *The Little Book of Deep Learning*, Genève, Université de Genève, Version beta-2023.05.03.

GIUSTI E.

2002, *Analisi 1*, Torino, Bollati Boringhieri

PAGANI C. D., SALSA S.

2015, *Analisi Matematica 1*, Seconda Edizione, Bologna, Zanichelli.

RUSSELL S. J., NORWIG P.

2010, *Artificial intelligence. A modern approach*. Third edition, Upper Saddle River (New Jersey) (USA), Pearson Education, inc.