

Negli ultimi mesi sono stati rilasciati in rapida successione una serie di modelli LLM come ChatGPT o1 - preview, ChatGPT o1 - mini, Claude 3.5 Sonnet, Mistral e Gemini che presentano una serie di caratteristiche che li distinguono dai modelli precedenti. In particolare si è rilevato:

- una migliore comprensione del contesto con la gestione di conversazioni più lunghe e complesse;
- una maggiore precisione nel risolvere problemi matematici, logici e tecnici;
- la capacità di elaborare non solo testo, ma anche immagini, audio e altre forme di dati;
- una sensibile riduzione dei tempi di risposta;
- la capacità di produrre testi creativi di tipologie diversificate conservando un linguaggio naturale e fluente;
- possibilità per alcune versioni avanzate di GPT - 4 (ChatGPT 4o) di accedere a dati in tempo reale.

L'ultima novità di rilievo si è registrata il 3 ottobre con il lancio, da parte di OpenAI, di [Canvas](#), una nuova interfaccia per ChatGPT, attualmente in versione beta e disponibile per gli utenti Plus e Teams.

La nuova funzionalità rappresenta il primo grande cambiamento di design di ChatGPT dal suo lancio nel 2022. L'utente si trova a disposizione uno spazio di lavoro separato per scrittura e codifica, dove sono permesse modifiche dirette senza rigenerare l'intero output. Canvas svolge funzionalità simili a quelle di Artifacts in Claude 3.5 Sonnet. OpenAI prevede, con il rilascio della versione definitiva, di rendere Canvas disponibile anche agli utenti gratuiti.

Nell'[allegato](#) viene proposta la risoluzione degli otto quesiti di matematica assegnati nella [seconda prova dell'Esame di Stato della sessione ordinaria 2024 per il Liceo Scientifico](#).

Gli "esperimenti artificiali" realizzati con cinque chatbot: [ChatGPT con Canvas](#), [ChatGPT o1 - preview](#), [ChatGPT o1 - mini](#); [Mistral Large 2](#), [Gemini](#) hanno permesso di testare i chatbot su un'ampia gamma di argomenti: geometria euclidea, probabilità, geometria analitica nello spazio, algebra, analisi matematica, calcolo integrale, geometria analitica, geometria piana e tassellazioni.

Come in precedenti esperimenti si sono sottoposti i singoli quesiti all'analisi di [Claude 3.5 Sonnet](#) con la richiesta di un'eventuale riformulazione per ottenere dei prompt ottimizzati. La riscrittura dei quesiti operata dall'AI risulta interessante, mira ad una comunicazione diretta ed efficace. I nuovi prompt sono stati sottoposti all'analisi dei chatbot.

Le soluzioni ai quesiti fornite dall'IA sono state messe a confronto con le risposte proposte dal prof. Luigi Tomasi e presenti sul sito matematica.it/tomasi.

Gli output ottenuti, rispetto quelli di versioni precedenti dei chatbot, evidenziano una migliore precisione nei calcoli e una potenziata capacità argomentativa nel descrivere i procedimenti risolutivi. Le allucinazioni e gli errori sono sensibilmente diminuiti. Una lettura comparata delle soluzioni ottenute permette di evidenziare la migliore efficienza dei modelli GPT di OpenAI in particolare di ChatGPT o1 - preview, emergono alcune incertezze da parte di ChatGPT o1 - mini. E' altresì da rilevare un significativo miglioramento nelle prestazioni matematiche da parte dell'ultima versione di Mistral Large 2, alcune difficoltà emergono nelle risposte di Gemini. E' opportuno segnalare che di questi due ultimi modelli si sono utilizzate le versioni gratuite.