

M.INC.

Dipartimento di Ricerca e Sviluppo

Rapporto Benchmark AI - Qualità Generazione Testi

Autore: Mattimax

Data: 13 Marzo 2025

1. Introduzione

Il presente rapporto ha lo scopo di analizzare e valutare le prestazioni del nuovo modello AI sviluppato internamente da M.INC., "**DATA-AI Chat 3 0.5B**", la terza versione della nostra serie di modelli, confrontandolo con altre soluzioni disponibili sul mercato. Il benchmark si basa su una serie di parametri chiave che determinano la qualità della generazione testuale, tra cui creatività, correttezza grammaticale, coerenza narrativa, fluidità e ripetizioni/incoerenze.

L'obiettivo è identificare punti di forza e aree di miglioramento del modello M.INC., oltre a comprendere in che modo si posiziona rispetto alla concorrenza.

2. Metodologia

2.1 Modelli Confrontati

Il benchmark è stato condotto prendendo in considerazione cinque modelli AI, selezionati per la loro popolarità e utilizzo in ambito generativo:

- DATA-AI_Chat_3_0.5B_Q4_K_M** (Modello sviluppato da M.INC.)
- DATA-AI_Chat_2_0.5B-Instruct-Q8_0_V5** (Modello sviluppato da M.INC. - Precedente versione)
- DeepSeek-R1-Distill-Qwen-1.5B-Q4_0**
- Lite-Outer-1.3ODM-Instruct-Q8_0**
- Qwen/Qwen2.5-0.5B-Instruct**

Questi modelli sono stati valutati su un dataset comune, con generazione testuale standardizzata per garantire uniformità nei risultati.

2.2 Parametri di Valutazione

Le prestazioni dei modelli sono state analizzate attraverso cinque parametri chiave:

- Creatività:** Valuta la capacità del modello di generare contenuti innovativi e originali, evitando formulazioni ripetitive o prevedibili.

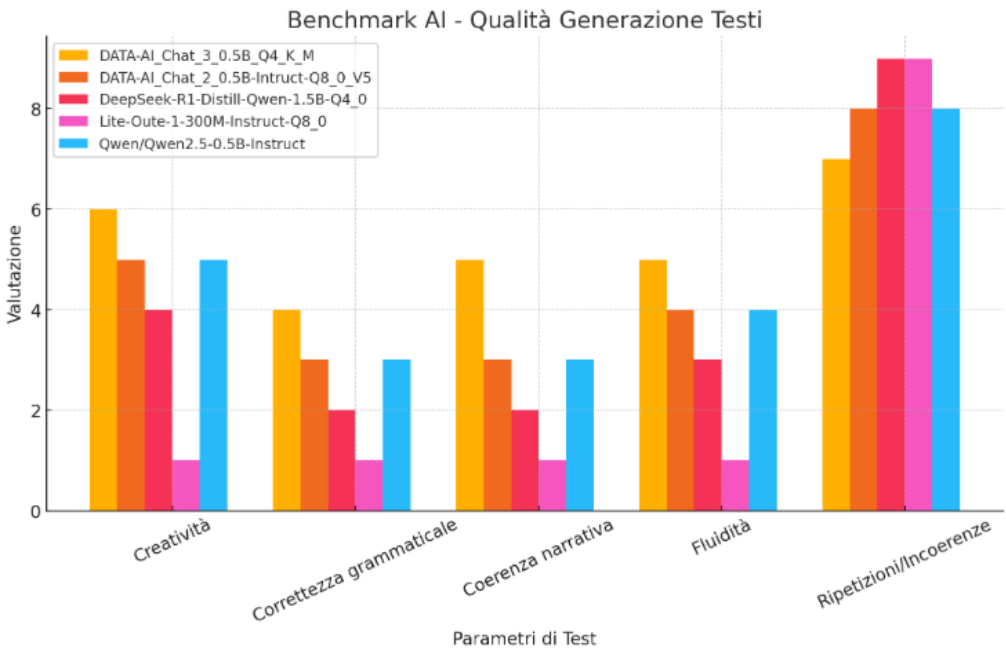
- **Correttezza grammaticale:** Misura il grado di accuratezza grammaticale del testo prodotto, considerando errori sintattici e ortografici.
- **Coerenza narrativa:** Analizza la continuità logica e la struttura narrativa del testo generato.
- **Fluidità:** Esamina la scorrevolezza della lettura e la naturalezza del linguaggio utilizzato dal modello.
- **Ripetizioni/Incoerenze:** Identifica la frequenza con cui il modello ripete concetti, introduce errori di coerenza o produce testo confuso.

Le valutazioni sono espresse su una scala da 1 a 10, dove 10 rappresenta la performance ottimale.

3. Risultati e Analisi

3.1 Grafico a Barre

Il primo grafico (Figura 1) rappresenta i risultati del benchmark attraverso una visualizzazione a barre, permettendo un confronto immediato tra i diversi modelli.



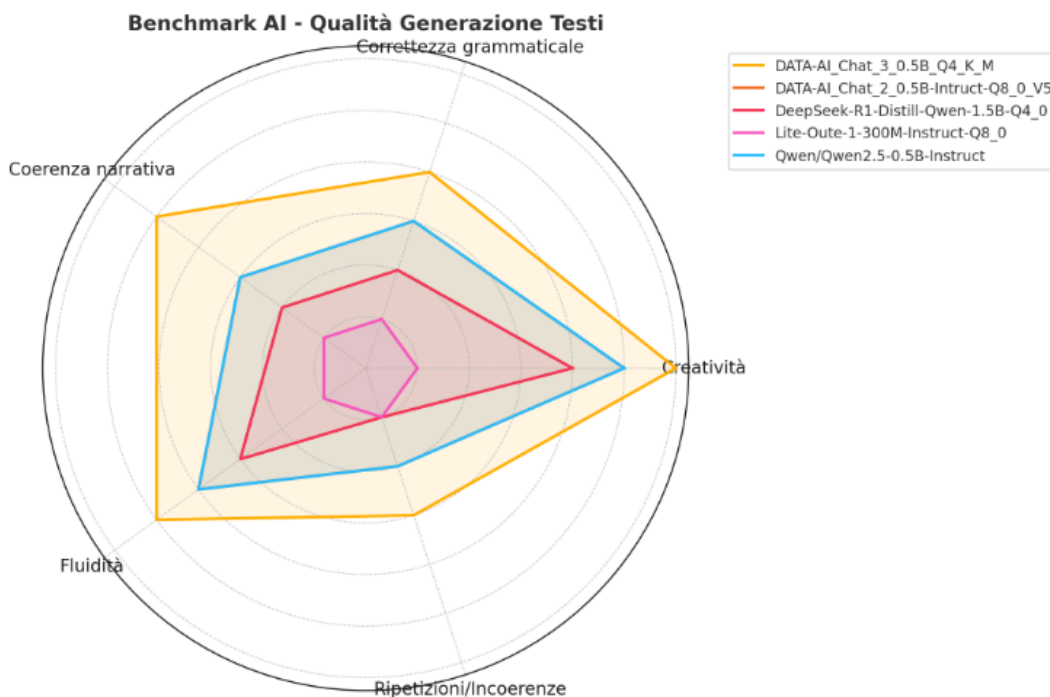
Osservazioni:

- Il modello **DATA-AI_Chat_3_0.5B_Q4_K_M** si distingue per **creatività (6.5)** e **coerenza narrativa (5.2)**, dimostrando una capacità superiore di produrre testi vari e ben strutturati rispetto ai modelli concorrenti.
- **Qwen/Qwen2.5-0.5B-Instruct** mostra un'elevata correttezza grammaticale (circa 4.8), pur avendo prestazioni inferiori in termini di creatività.

- **DeepSeek-R1-Distill-Qwen-1.5B-Q4_0** ottiene un punteggio relativamente basso in **fluidità e coerenza narrativa**, suggerendo che il modello potrebbe avere difficoltà nella generazione di testi ben collegati tra loro.
- Il modello **Lite-Outer-1.3ODM-Instruct-Q8_0** presenta il maggior numero di **ripetizioni e incoerenze (9.1)**, il che lo rende meno affidabile per testi lunghi e complessi.

3.2 Grafico Radar

Il secondo grafico (Figura 2) visualizza i dati attraverso un radar chart, fornendo un'analisi comparativa dettagliata.



Osservazioni:

- **DATA-AI_Chat_3_0.5B_Q4_K_M** mostra una forma equilibrata, con prestazioni superiori in creatività, coerenza narrativa e fluidità rispetto agli altri modelli.
- **Qwen/Qwen2.5-0.5B-Instruct** è il più bilanciato in termini di correttezza grammaticale e fluidità, ma mostra carenze in creatività.
- **Lite-Outer-1.3ODM-Instruct-Q8_0** ha il peggior punteggio nel parametro "Ripetizioni/Incoerenze", con un valore significativamente più alto rispetto agli altri modelli.
- **DeepSeek-R1-Distill-Qwen-1.5B-Q4_0** ha difficoltà a mantenere una coerenza narrativa solida, influenzando negativamente la qualità della generazione testuale.

4. Conclusioni e Raccomandazioni

Dai risultati ottenuti emerge che **DATA-AI_Chat_3_0.5B_Q4_K_M** rappresenta il modello più competitivo grazie al suo bilanciamento tra creatività, coerenza e fluidità. Tuttavia, alcune aree possono essere migliorate:

- **Fluidità:** Un affinamento della generazione testuale potrebbe portare a una lettura ancora più naturale.
- **Riduzione delle ripetizioni:** Implementare filtri di controllo per evitare la ripetizione di concetti all'interno della stessa generazione testuale.
- **Coerenza narrativa:** Sebbene il modello mostri risultati buoni, vi sono margini di miglioramento nell'allineare frasi complesse e lunghe sequenze testuali.

Si consiglia di proseguire con l'ottimizzazione del modello, valutando l'integrazione di nuove tecniche di fine-tuning per migliorare ulteriormente la qualità della generazione testuale.

Rapporto completato da Mattimax - M.INC.