

Titolo del Progetto: Implementazione della Metrica CHATEVAL per la Valutazione Comparativa di Dialoghi Open-Domain

Obiettivi:

- Comprendere la metrica CHATEVAL proposta nell'articolo "CHATEVAL: Towards Better LLM-Based Evaluators Through Multi-Agent Debate".
- Implementare il framework CHATEVAL per valutare risposte generate attraverso un processo collaborativo multi-agente.
- Esplorare l'efficacia di CHATEVAL nella valutazione di risposte in task di generazione di dialoghi e domande aperte, analizzando metriche di correlazione con giudizi umani.

Metodologia di Implementazione:

1. Analisi della Metrica CHATEVAL:

- Studiare l'articolo per comprendere i componenti chiave di CHATEVAL: agenti debater, ruoli diversificati e strategie di comunicazione.

2. Implementazione del Framework:

- **Agenti Debater:** Configurare un insieme di agenti basati su modelli linguistici per generare risposte e discutere su quale sia la migliore.
- **Ruoli Diversificati:** Definire prompt personalizzati per ogni agente, assegnando ruoli distinti (ad esempio, Critico, Autore, Psicologo) per favorire una valutazione multi-prospettiva.
- **Strategie di Comunicazione:** Implementare le strategie di comunicazione "One-by-One", "Simultaneous-Talk" e "Simultaneous-Talk-with-Summarizer" per simulare discussioni realistiche.

3. Valutazione:

- Eseguire esperimenti su dataset di benchmark che vi verranno forniti per misurare la correlazione tra i giudizi di CHATEVAL e quelli umani utilizzando metriche statistiche come Kappa di Cohen, Spearman, Pearson e Kendall-Tau.
- Analizzare le prestazioni delle diverse strategie di comunicazione.

Risultati Attesi:

- Implementazione funzionante di CHATEVAL per la valutazione collaborativa.
- Analisi delle correlazioni tra i punteggi di CHATEVAL e i giudizi umani.
- Confronto dettagliato con metriche di riferimento e discussione delle sfide affrontate.

Risorse:

- **Articolo Base:** "CHATEVAL: Towards Better LLM-Based Evaluators Through Multi-Agent Debate".
- **Codice Open Source:** Framework e prompt forniti nell'articolo (se disponibili).
- **API Modelli:** <https://github.com/selfsff/GPT4ALL-Free-GPT-API>