

Titolo del Progetto: Implementazione della Metrica GPTScore per la Valutazione di Testi Generati

Obiettivi:

- Comprendere la metrica GPTScore proposta nell'articolo "GPTScore: Evaluate as You Desire".
- Implementare il framework GPTScore per la valutazione di output nel task di generazione di dialoghi

Metodologia di Implementazione:

1. Analisi della Metrica GPTScore:

- Studiare l'articolo per comprendere i componenti chiave: costruzione del prompt, definizione degli aspetti di valutazione, utilizzo di modelli pre-addestrati.

2. Implementazione del Framework:

- **Definizione del Protocollo di Valutazione:**
 - Creare prompt per il task specifico e l'aspetto desiderato, utilizzando definizioni fornite nell'articolo.
- **Integrazione Modello:**
 - Utilizzare API di modelli pre-addestrati (es. GPT-3 o Flan-T5) per calcolare le probabilità condizionate associate agli output generati.

3. Valutazione delle Performance:

- Calcolare la correlazione tra i punteggi GPTScore e i giudizi umani utilizzando metriche statistiche come Kappa di Cohen, Spearman, Pearson e Kendall-Tau.

Risultati Attesi:

- Implementazione completa del framework GPTScore per il task di generazione di dialoghi.
- Analisi della correlazione tra GPTScore e giudizi umani, con discussione delle sue capacità di personalizzazione e limiti.

Risorse:

- Articolo di Riferimento: "GPTScore: Evaluate as You Desire".
- Codice Open Source: [GPTScore GitHub](https://github.com/evalplus/gptscore).
- **API Modelli:** <https://github.com/selfsff/GPT4ALL-Free-GPT-API>