

Chatbot (GPT) na vyhľadávanie v štruktúrovaných dátach - agregovanie viacerých chatbotov

Zápisnica stretnutia č. 19

Dátum : 9.04.2025

Prítomní :

Ing. Marián Lekavý, PhD.

Marko Matúška, Patrik Kaprinay, Ľuboš Sekerák, Klára Murínová, Michal Čopjan

Program:

1. Prezentovanie úloh zadaných na minulom stretnutí :
 - a. Dáta:
 - i. Skontrolovať dáta (priradiť ich k správnej kategórii) (úloha splnená)
 - ii. Zosúladiť testovaciu a trénovaciu množinu (úloha splnená) dáta rozdelené **80-20**
 - b. Zjednotiť kategórie (úloha splnená)
 - i. **Search for a part by parameters** : 209 vstupov
 - ii. **Suggest complementary parts**: 42 vstupov
 - iii. **Part number**: 398 vstupov
 - iv. **Find similar or alternative parts**: 129 vstupov
 - v. **Invalid request**: 99 vstupov
 - c. Testovanie modelov (úloha splnená)
 - i. Embeddings:
 1. staré dáta, veľa vektorov (nie priemerne) - **79% úspešnosť**
 2. nové dáta, priemerne vektory - **87% úspešnosť**
 3. nové dáta, veľa vektorov (nie priemerne) - **86% úspešnosť**
 - ii. GPT – **86 % úspešnosť**

Zápisnicu vypracoval/a : Klára Murínová

- iii. Luis – **86,23 % úspešnosť**
 - d. Otestovať regex pre extrahovanie OPN/WPN na nových dátach (úloha splnená)
 - e. Otestovať regex pre extrahovanie filtrov na nových dátach (úloha splnená)
 - f. Napísať dokumentáciu (úloha splnená)
2. Úlohy do budúceho stretnutia:
- a. Analýza testovacích dát 3 modelov:
 - i. Excel – prompt , správna kategória, GPT , Luis , Embeddings
 - b. Úspešnosť regex pre extrakciu OPN/WPN
 - i. 2 kategórie
 - 1. Patri do kategórie 1 (neobsahujú žiadne OPN/WPN)
 - 2. Patrí do 2,3,4 (musia obsahovať OPN/WPN)
 - c. Doplniť dokumentáciu
 - i. konfúznú maticu pre všetky modely na základe testovacích dát
 - ii. Zdokumentovať rozhodovacie modely