

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

Zápisnica stretnutia č. 8

Dátum : 18.11.2024

Prítomní :

Ing. Marián Lekavý, PhD.

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

Program:

1. Predstavenie úloh zadaných z minulého týždňa:
 - a. Vypracovanie testovacích dát (splnené)
 - i. Všetky testovacie dáta by ale mali byť zaradené do 1. kategórie z dôvodu že obsahujú údaje pre filtrovanie.
 - b. Presnosť na testovacích dátach jednotlivých modelov :
 - i. **gpt-4-1106-preview** – 85.14 % (4 krát drahšia ako gpt-4o)
 - ii. **gpt-4o** – 83.78 % (najlacnejšia)
 - iii. **gpt-4** – 56.76% (najdrahšia)
 - iv. **Luis** – 100% (nedostatok dát pre trénovanie a testovanie)
 - v. **Embedding** – 45%
 - c. Integrácia GPT do existujúceho kódu (úloha splnená, každý ma svoju verziu kódu s konkrétnym modelom)
 - d. Upraviť diagram tried (úloha splnená, diagram sa nachádza v dokumentácii)
2. Úlohy do budúceho stretnutia:
 - a. Upraviť prompt pre GPT model, rozšíriť ho (Klára, Ľuboš)

Zápisnicu vypracoval/a : Klára Murínová

- b. Upraviť testovacie dáta pre každú kategóriu tak aby :
 - i. Dáta ktoré budú obsahovať údaje pre vyplňanie filtrov budú zaradené do 1. kategórie .
 - ii. Dáta ktoré budú obsahovať OPN (jedinečné identifikačné číslo súčiastky) budú na základe požiadavky zaradené do 2,3,4 alebo 5 kategórie.
- c. Zjednotiť kód s rozhodovacími modelmi, každý model bude mať vlastnú triedu , následne sa bude len meniť ktorú triedu zavoláme na základe modelu.
- d. Embeding, vypočítať priemer pre všetky testovacie dáta danej kategórie , a následne porovnať ku ktorej kategórii je daný prompt najbližšie. (Patrik)
- e. Vložiť do dokumentácie diagram z minuloročnej dokumentácie, pre porovnanie ktorá časť bola neimplementovaná a ktorú časť sme pridali my. (Michal)