

Подходи за обработка на естествен език

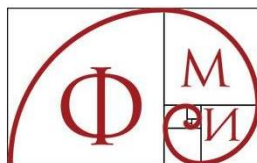
SemEval-2025 Task-3 — Mu-SHROOM

Кристиян Симов, ф.н. 4МІ3400288

Цветан Цветанов, ф.н. 4МІ3400570

Изкуствен интелект

7 февруари 2025 г.



1 Въведение

Езиковите модели са вещи в естествените езици и звучат убедително, но понякога генерират грешни и лъжовни твърдения, които нямат реална подкрепа с факти.

Халюцинирането е един от големите все още неразрешени проблеми на големите езикови модели. Задачата Mu-SHROOM в изданието на SemEval от 2025 е продължение на SHROOM от предишното издание. През 2024 година задачата пред участниците е била да класифицират дали даден текст е халюцинация (да или не). Промяната в сегашното издание е, че се очаква да се предскаже началото и края на халюцинация в изходния текст на конкретен модел.

2 Данни

Използвани са единствено публично достъпни данни от страницата на задачата. Не е правен опит да се генерират синтетични данни. Организаторите са подготвили текстови данни за трениране, тестване и валидация на 14 различни езика в jsonl формат.

3 Метод

3.1 Обработка на данните

3.2 Алгоритми

4 Експерименти

5 Резултати

Публично Github хранилище с кода от проекта
Официално класиране

6 Заключение и бъдеща работа

7 Индивидуален принос