

Modele językowe

Ćwiczenia 1

Zajęcia 3

Każde zadanie warte jest 1 punkt.

Zadanie 1. Wymyśl rodzaj stosunkowo łatwego dla ludzi zadania (nieodwołującego się do wiedzy faktograficznej), z jasno sprecyzowaną odpowiedzią,¹ z którym ChatGPT sobie nie radzi (alternatywnie: pokaż trzy nietrywialne zadania, które Chat umie robić, a Cię to trochę zdziwiło)

Zadanie 2. Wybierz 10, raczej trudniejszych, pytań z zadania 4 z Pracowni 1, a następnie sprawdź, jak wiele z nich jest poprawnie rozwiązanych przez ChataGPT.

Zadanie 3. Znak wodny w przypadku modeli językowych jest nadaniem generowanym tekstom pewnej własności, która jest niewidoczna bez użycia specjalistycznych metod wykrywania znaku wodnego. Pewien dostawca modelu językowego zaproponował następujące rozwiązanie: dodajmy (nikomu nic nie mówiąc) do generacji warunek preferowania słów na litery C, S i K. Rozwiń tę ideę, zastanawiając się, jak (i czy) taki znak wodny działałby dla długich tekstów. A dla krótkich?

Zadanie 4. Na Olimpiadzie Sztucznej Inteligencji było zadanie Riddles². W skrócie, chodziło o przypisanie słów tekstom opisującym te słowa, przy czym odpowiedzi należały do znanego zbioru (zawierającego około 10K słów). Przykładowym tekstem było:

kobieta podróżująca środkiem transportu, np. samolotem, pociągiem, statkiem

dla którego odpowiedzią jest *pasażerka*³.

Uczestnicy Olimpiady rozwiązujący mieli do dyspozycji definicje słów z Wiktionary, ale my założymy, że mamy dostęp jedynie do modelu językowego. Odpowiedz na pytania:

- a) Czy da się zmusić któryś z naszych modeli do tego, żeby rozwiązywał to zadanie w trybie generacji. Jaka strategia wydaje Ci się obiecująca?
- b) Czy funkcja oceniania prawdopodobieństwa zdania może być z sukcesem użyta w tym zadaniu? (jaki jest główny problem z jej użyciem)

Zadanie 5. Opisz procedurę, która, wykorzystując podstawowy interfejs do generacji tekstu modelu językowego, możliwie efektywnie rozwiązuje problem generacji **dokładnie** jednego wyrazu (dla zadanego prefiksu, składającego się z pełnych wyrazów).

Zadanie 6. Przeczytaj i opowiedz o uprzedzeniach (bias) modelu papuGaPT. Jak były one badane? Jakie z tych badań wynikają konkluzje? (sekcja Bias Analysis na stronie <https://huggingface.co/flax-community/papuGaPT2>)

Zadanie 7. Okazuje się, że model polka umie tłumaczyć proste angielskie zdania na polski. Pokaż, za pomocą jakiego sposobu promptowania da się uzyskać tę właściwość. Pokaż dwa przykładowe zdania, jedno przetłumaczone poprawnie, a drugie błędnie. Załóżmy, że mamy dostęp do słownika angielsko-polskiego. Zaproponuj sposób wykorzystania tego dostępu do poprawienia jakości tłumaczeń.

Zadanie 8. Załóżmy, że mamy dwa modele językowe i chcemy generować tekst korzystając z wiedzy obu modeli. Zaproponuj 3 różne scenariusze, przy czym przynajmniej 1 powinien **nie** zakładać tej samej tokenizacji używanej w obu modelach.

Zadanie 9. Zaproponuj inne niż sugerowane na liście pracowniowej rozwiązanie zadania z wyborem permutacji wyrazów, które częściowo usuwa problemy związane z zachłannym łączeniem słów. Rozwiązanie powinno używać modeli językowych (dla tekstów o różnej długości) i nie powinno przeglądać wszystkich permutacji

¹Odpadają zawiłe filozoficzne kwestie, musi istnieć prosta i niekwestionowalna odpowiedź

²https://github.com/OlimpiadaAI/I-OlimpiadaAI/blob/main/first_stage/riddles/zagadki.ipynb

³Tak, zagadki czasem były niejednoznaczne. Wymyślał je ChatGPT