



Politechnika
Wrocławska

p(doom) w social mediach

Czyli czy łatwo jest generować sztuczne treści
polityczne w internecie?

Projekt wykonany w ramach zajęć "Analiza mediów cyfrowych"
na kierunku Sztuczna Inteligencja w roku 2023



Katedra
Sztucznej
Inteligencji

Radostaw Kuczbański

Konrad Karanowski

Mateusz Grzesiuk

Mikołaj Langner

Opis problemu

Media społecznościowe poprzez publikowane na nich posty pozwalają ich użytkownikom docierać ze swoim przekazem do szerszego grona odbiorców i potencjalnie wpłynąć na ich poglądy. Kluczowa jest odpowiednia redakcja tekstu, aby post przyciągnął jak najwięcej uwagi oraz skutecznie kontrolował ogólną opinię w jakiejś społeczności. Wiele podmiotów decyduje się na zatrudnianie tzw. ghost writerów do pisania takich tekstów w ich imieniu. Jednakże, potencjalnym kierunkiem rozwoju jest zastosowanie w tym celu sztucznej inteligencji.

Rozwiązanie

Wytrenowanie modelu językowego dla zadania pisania treści na media społecznościowe.

Dane

- Zbierane w okresie 10.10 - 30.10 i 20.12-26.12
- 200.000 tweetów polityków
- 1.230.137 odpowiedzi na tweety
- Odfiltrowane zostały posty z tekstami obcojęzycznymi lub załącznikami w postaci obrazów/wideo.

Metodyka

- Konto, na którym publikowane były wygenerowane treści, w połowie eksperymentu zostało ulepszone do Premium+, ponieważ bez niego odpowiedzi były niewidoczne (~5 wyświetleń)
- Model Mixtral-8x7B-Instruct-v0.1 został douczony przez jedną epokę przy użyciu dwóch datasetów: s3nh/alpaca-dolly-instruction-only-polish, który zawiera ogólne zadania w języku polskim i autorskim wyżej wymienionym zbudowanym na bazie tweetów.

Prompt

Jesteś politycznym użytkownikiem Twittera. Bądź obraźliwy, złośliwy, niepoprawny politycznie i sarkastyczny. Napisz krótką odpowiedź na tweet zgodnie z Twoją personą. Tweet: ""<tweet>"" Odpowiedź: ""

Wyniki

W okresie 26.12 - 6.1 wygenerowano za pomocą modelu i opublikowano w serwisie tweeter 30 odpowiedzi. Obok przedstawione są najbardziej popularne z nich. Łącznie zebrały one:

- 199 polubień,
- 8 retweetów,
- 11 odpowiedzi,
- 6 followers'ów

Wnioski

Przeprowadzony eksperyment pokazuje, że przy użyciu LLM-ów można z łatwością kontrolować ogólny wydźwięk opinii w mediach społecznościowych.

Do analiz wykorzystany został model anglojęzyczny ze względu na dotychczasowy brak zadowalającego odpowiednika dla modelu w języku polskim.

Prawdopodobnie, większa ilość danych i moc obliczeniowa, znacznie poprawiłaby uzyskane wyniki.

