

[Akceptuję](#)

W ramach naszej witryny stosujemy pliki cookies w celu świadczenia państwu usług na najwyższym poziomie, w tym w sposób dostosowany do indywidualnych potrzeb. Korzystanie z witryny bez zmiany ustawień dotyczących cookies oznacza, że będą one zamieszczone w Państwa urządzeniu końcowym. Możecie Państwo dokonać w każdym czasie zmiany ustawień dotyczących cookies. Więcej szczegółów w naszej [Polityce Prywatności](#)

[Portal](#) [Informacje](#) [Katalog firm](#) [Praca](#) [Szkozenia](#) [Wydarzenia](#) [Porównania międzylaboratoryjne](#)
[Kontakt](#)



[Laboratoria](#)
[.net](#)
[Innowacje](#)
[Nauka](#)
[Technologie](#)



[Logowanie](#) [Rejestracja](#) [pl](#)

Newsletter

zapisz się

Naukowy styl życia

Nauka i biznes

- [Nowe technologie](#)
- [Felieton](#)
- [Tygodnik "Nature"](#)
- [Edukacja](#)
- [Artykuły](#)
- [Przemysł](#)

[Strona główna](#) > [Informacje](#)

Współautorka benchmarku OneRuler

Media obiegra niedawno wiadomość, że „język polski jest najlepszy do promptowania”. To nie jest wniosek z naszych badań - dementuje Marzena Karpińska z Microsoft, współautorka pracy, z której polskie media zbyt pochopnie wyciągnęły taki wniosek.

OneRuler to wielojęzyczny benchmark (zestaw testów) służący do oceny, jak dobrze modele językowe radzą sobie z przetwarzaniem bardzo długich tekstów. Porównano w nim, jak różne modele AI radzą sobie z wyszukiwaniem informacji w 26 różnych językach.

Wprawdzie, średnio rzecz biorąc, w benchmarku modele AI najlepiej radziły sobie z językiem polskim, różnice te jednak nie były znaczne pomiędzy językiem polskim a angielskim, a wynik nie został przez autorów pracy wyjaśniony. W polskojęzycznych mediach podsumowujących te analizy pochopnie wyciągnięto z tych badań wnioski, że „język polski jest najlepszy do promptowania”.

Marzena Karpińska z Microsoft, współautorka badania, zapytana przez PAP, czy to prawda, że według tej pracy polski język jest najlepszy do promptowania, odpowiedziała: - Nie. Wcale tego nie badaliśmy. Stworzyliśmy narzędzie do diagnozowania modeli językowych, sprawdzające, jak dobrze są one w stanie wydobyć informacje z bardzo długich tekstów.

Wyjaśniła, że zadanie, jakie otrzymywały różne modele w ramach benchmarku, polegało na czymś w rodzaju odszukania igły w stogu siana: w książkę w danym języku włożono jakieś zdanie z informacją. A model - poinstruowany w tym samym języku - musiał właśnie to zdanie wskazać jako prawidłową odpowiedź. W edytorach tekstowych czy przeglądarkach z tego typu zadaniem radzi sobie zwykła funkcja CTRL+F (znajdź w dokumencie). Modele AI jednak nie mają jej wbudowanej.

Inne zadanie w ramach benchmarku polegało na opracowaniu listy najpopularniejszych słów z książki.

- Spodziewaliśmy się, że modele w wielu językach będą miały 100 proc. skuteczności. Nie miały. Zauważyliśmy, że modele zaczynają błędzić zwłaszcza wtedy, kiedy w instrukcji przypominaliśmy, że odpowiedź może nie być zawarta w wyszukiwanym tekście. A wtedy model powinien napisać, że odpowiedzi brak - powiedziała współautorka badania. Na tym zadaniu skuteczność znacznie spadała zapewne dlatego, że wymagało ono użycia całego kontekstu, a nie tylko przeszukania książki.

Tym, co mogło zaważyć na tym, jaki język wypadł w benchmarku najlepiej, mogła być metodologia, a konkretnie - dobór lektur wykorzystywanych w próbach. Dla każdego języka wykorzystywano bowiem inną książkę. I tak np. dla języka polskiego - informacje były wyszukiwane w 3. tomie książki „Noce i dnie”, dla języka hiszpańskiego - w „Don Kichocie”, dla angielskiego - w „Małych kobietkach”, a dla niemieckiego w „Czarodziejskiej Górze”. Nie były to więc te same teksty przetłumaczone na różne języki, ale zupełnie inne książki z różnych okresów historycznych. Możliwe więc, że to wcale nie język polski okazał się w czymś lepszy niż angielski, ale „Noce i dnie” okazały dla modeli z jakiegoś powodu łatwiejsze do przeczesań niż „Małe kobietki”.

Badaczka wyjaśniła, że taki dobór książek wynikał z konieczności wyboru tekstów w każdym języku, do których prawa autorskie już wygasły.

- W badaniu tym jest tak dużo różnych czynników, że na pewno nie możemy na jego podstawie stwierdzić, że polski jest najlepszym językiem do „promptowania” - zapewniła Karpińska.

Zdaniem rozmówczyni PAP to, że żaden z modeli nie osiągnął 100 proc. skuteczności w tak prostym zadaniu, powinno być przestrogą dla wszystkich użytkowników modeli językowych. - Ludzie ładują do chata GPT całe góry dokumentów i zadają pytania o te treści. A trzeba pamiętać, że modele językowe ciągle jeszcze mają bardzo ograniczone zdolności przetwarzania tekstu. Czasami są niesamowicie dobre, a chwilę później - popełniają ogromne błędy. Trzeba pytać drugi raz, upewniać się w innym modelu. No i przede wszystkim trzeba uważać, jakie dokumenty się do modeli ładuje, szczególnie jeśli chodzi o treści wrażliwe i prywatność - zakończyła badaczka.

Praca OneRuler została zrecenzowana i zaprezentowana w październiku br. na konferencji naukowej CoLM (Conference on Language Modeling).

Źródło: pap.pl

<https://laboratoria.net/aktualnosci/32650.html>



23-06-2026

Flexicon FPC50 w dydaktyce pracy laboratoryjnej

Dostawca szkoleń aptaskil przygotowuje wykwalifikowanych specjalistów.



22-06-2026

Blisko 2,8 mln zł na badania nad terapią

Opracowanie strategii leczenia nowotworów odpornych na terapię.



22-06-2026

Studenci AGH zaprezentowali swój najnowszy bolid elektryczny

Pojazd powstał z myślą o udziale w zawodach inżyniersko-wyścigowych.



22-06-2026

[Naukowcy sprawdzili, czy protony są wieczne](#)

W badaniach uczestniczyły polskie ośrodki.



22-06-2026

[Polska wśród krajów z najniższym poziomem stresu psychicznego](#)

Wśród ukraińskich uchodźców.



22-06-2026

[Życie seksualne coraz częściej przenosi się do świata technologii](#)

Sfera ta rośnie szybciej niż wiedza o jej wpływie na ludzką seksualność.



22-06-2026

Przyjemnych snów życzy anestezyjolog

Wystarczy przestrzegać protokołu znieczulenia.



22-06-2026

Za mało siedzenia także może szkodzić

Od lat lekarze i naukowcy powtarzają, że należy mniej siedzieć i więcej się ruszać.

Informacje dnia: [Flexicon FPC50 w dydaktyce pracy laboratoryjnej](#) [Blisko 2,8 mln zł na badania nad terapią](#) [Studenci AGH zaprezentowali swój najnowszy bolid elektryczny](#) [Naukowcy sprawdzili, czy protony są wieczne](#) [Polska wśród krajów z najniższym poziomem stresu psychicznego](#) [Życie seksualne coraz częściej przenosi się do świata technologii](#) [Flexicon FPC50 w dydaktyce pracy laboratoryjnej](#) [Blisko 2,8 mln zł na badania nad terapią](#) [Studenci AGH zaprezentowali swój najnowszy bolid elektryczny](#) [Naukowcy sprawdzili, czy protony są wieczne](#) [Polska wśród krajów z najniższym poziomem stresu psychicznego](#) [Życie seksualne coraz częściej przenosi się do świata technologii](#)

Partnerzy