

Strona główna - Premiera Bielika v2. Co nowego w największym polskim modelu językowym?

Premiera Bielika v2. Co nowego w największym polskim modelu językowym?

Dodane: 28.08.2024



Mam Startup



Bielik v2, owoc współpracy Fundacji SpeakLeash i Academic Computer Centre CYFRONET AGH, ma dziś swoją premierę. Ten mówiący po polsku model języka, w oczach twórców stać się ma „kolejnym kamieniem milowym w rozwoju sztucznej inteligencji” w naszym kraju.

gigantów technologicznych.”

Co nowego w Bieliku v2?

1. **Zwiększony rozmiar:** 11 miliardów parametrów zapewnia jeszcze lepsze zrozumienie i generowanie języka polskiego.
2. **Szerokie okno kontekstowe:** 32 768 tokenów pozwala na przetwarzanie dłuższych tekstów i bardziej złożonych zadań.
3. **Ulepszona baza treningowa:** Większa ilość wysokiej jakości danych przekłada się na poprawę wyników w różnorodnych zadaniach.
4. **Poprawione NLP:** Lepsze wyniki w zadaniach takich jak podsumowywanie tekstów i rozpoznawanie jednostek nazewniczych (NER).
5. **Elastyczność wdrożenia:** Dostępne są wersje skwantyzowane, umożliwiające efektywne wykorzystanie modelu na różnych platformach sprzętowych.
6. **Wysoka jakość:** Pomimo niewielkich rozmiarów, Bielik v2 dorównuje jakością w języku polskim dużo większym modelom, co potwierdzają benchmarki (m.in. Open LLM PL Leaderboard) oraz ranking ELO Chat Areny.

Więcej o Bieliku dowiedzie się z naszej rozmowy [z Sebastianem Kondrackim, jednym z twórców Bielika](#). W innym naszym tekście przeczytacie o [superkomputerze, który posłużył do stworzenia tego języka](#).

Autorzy Bielika zachęcają do testów swojego rozwiązania – [można to zrobić tutaj](#).

BIELIK-11B-v2

Large Language Model

Bezpieczne przetwarzanie

Pełna kontrola

Kompaktowa moc

Porozmawiaj z Bielikiem



Helios i Athena – moce obliczeniowa dla nauki

Współpraca kadry z Akademii Górniczo-Hutniczej w Krakowie z fundacją Speakleash umożliwiła wykorzystanie odpowiednich mocy obliczeniowych niezbędnych do stworzenia modelu i wsparcie zespołu SpeakLeash niezbędną wiedzą ekspercką oraz naukową gwarantując sukces wspólnego projektu. Wsparcie zespołu ACK Cyfronet dotyczyło optymalizacji i skalowania procesów treningowych, prac nad potokami przetwarzania danych oraz rozwoju i działania metod generowania danych syntetycznych, a także prac w zakresie metod testowania modeli. Wynikiem tego jest Polski ranking modeli (Polish OpenLLM Leaderboard). Cenne doświadczenia i wiedza zebrane w wyniku tej współpracy umożliwiły zespołowi ekspertów PLGrid przygotowanie wytycznych oraz zoptymalizowanych rozwiązań w tym środowisk obliczeniowych do prac z modelami językowymi na bazie klastrów Athena i Helios dla potrzeb użytkowników naukowych.

– Zasoby Heliosa, najszybszej aktualnie maszyny w Polsce, wykorzystaliśmy do uczenia modeli językowych – precyzuje **Marek Magryś, zastępca dyrektora ACK Cyfronet AGH ds. Komputerów Dużej Mocy**. – Nasza rola polega na wsparciu wiedzą ekspercką, doświadczeniem i przede wszystkim mocą obliczeniową procesu katalogowania, zbierania, przetwarzania danych oraz na wspólnym przeprowadzeniu procesu uczenia modeli językowych. Dzięki pracy zespołu SpeakLeash i AGH udało nam się stworzyć Bielika, model LLM, który doskonale radzi sobie z naszym językiem oraz kontekstem kulturowym i który może być kluczowym elementem

rankingowych dla języka polskiego – dodaje Magryś.

Moc obliczeniowa Heliosa i Atheny w tradycyjnych symulacjach komputerowych to łącznie ponad 44 PFLOPS, a dla obliczeń z zakresu sztucznej inteligencji w niższej precyzji to aż 2 EFLOPS.

– Jeśli operujemy tak dużymi danymi jak w przypadku projektu Bielik to oczywiście infrastruktura potrzebna do pracy przekracza zdolności zwykłego komputera. Musimy dysponować mocą obliczeniową potrzebną tylko do tego żeby przygotowywać dane, porównywać je ze sobą, trenować modele. Bariera dostępności tego typu superkomputerów powoduje, że mało która firma jest w stanie takie prace prowadzić samodzielnie. Szczęśliwie AGH dysponuje takim zapleczem – wyjaśnia **prof. Kazimierz Wiatr, dyrektor ACK Cyfronet AGH.**

Równolegle z zasobów superkomputerów z ACK Cyfronet AGH korzysta kilka tysięcy naukowców reprezentujących wiele dziedzin. Zaawansowane modelowanie i obliczenia numeryczne są wykorzystywane głównie w zakresie: chemii, biologii, fizyki, medycyny i technologii materiałowej, a także astronomii, geologii i ochrony środowiska. Superkomputery w Cyfronecie dostępne w ramach infrastruktury PLGrid są również wykorzystywane na potrzeby:

- fizyki wysokich energii (projekty ATLAS, LHCb, ALICE i CMS),
- astrofizyki (CTA, LOFAR),
- nauk o Ziemi (EPOS),
- europejskiego źródła spalacyjnego (ESS),
- badań fal grawitacyjnych (LIGO/Virgo),
- biologii (WeNMR).

– **Wykorzystujemy do trenowania Bielika dwa najszybsze superkomputery w Polsce, Athenę i Heliosa, ale i tak w porównaniu z infrastrukturą światowych liderów mamy dużo mniejsze zaplecze.** Do tego, w tym samym czasie z zasobów superkomputerów korzysta kilkuset innych użytkowników – wyjaśnia Marek Magryś. – Nasze systemy umożliwiają jednak przeprowadzenie w kilka godzin lub dni obliczeń, które na zwykłych komputerach mogłyby trwać lata lub, w

Bielik a chat GPT – podstawowe różnice

– Zbiór danych zasilających Bielika cały czas rośnie, jednak trudno będzie nam się ścigać z zasobami wykorzystywanymi przez inne modele, które funkcjonują w języku angielskim. Poza tym liczba treści w internecie, która funkcjonuje w języku polskim jest znacznie mniejsza niż w angielskim – wyjaśniają twórcy.

Najbardziej popularnym produktem wykorzystującym duży model językowy jest ChatGPT, który powstał w oparciu o zasoby firmy OpenAI. Konieczność opracowywania modeli językowych w różnych innych językach znajduje jednak swoje uzasadnienie.

Marek Magryś zwraca uwagę na istotne zagadnienie: – O ile ChatGPT potrafi mówić w języku polskim, to nasycony jest treściami w języku angielskim. W związku z tym ma nikłe pojęcie na temat np. polskiej kultury czy niuansów polskiej literatury. **Nie do końca też sobie radzi ze zrozumieniem logiki bardziej skomplikowanych tekstów np. prawnych czy medycznych.** Jeśli chcielibyśmy zastosować go w tych właśnie specjalistycznych obszarach i mieć model językowy, który dobrze rozumie w języku polski i odpowiada poprawną polszczyzną, to nie możemy opierać się wyłącznie na zagranicznych modelach językowych.

Wersja, którą mogą testować użytkownicy jest utrzymywana nieodpłatnie w domenie publicznej i jest wciąż udoskonalana. **Autorzy udostępnili, oprócz pełnych wersji opracowanych modeli, także całą gamę wersji skwantyzowanych w najpopularniejszych dostępnych formatach, które umożliwiają uruchomienie modelu na własnym komputerze.**

– Warto wiedzieć, że Bielik będzie bardzo dobrze sprawdzał się w zakresie np. streszczania treści. Już w tym momencie nasz model ma swoją użyteczność w obszarze naukowym oraz biznesowym, może służyć na przykład do usprawnienia komunikacji z użytkownikami podczas obsługi zgłoszeń w helpdesku – wyjaśnia **Szymon Mazurek z ACK Cyfronet AGH.**

Czy warto budować polskie modele językowe?

Twórcy Bielika wyjaśniają, że usługi sztucznej inteligencji funkcjonujące w internecie, w tym te

tekstach, które z różnych powodów nie mogą opuścić firmy, np. są poufne, to jedyną możliwością jest uruchomienie takiego modelu u siebie. Ten model nie będzie tak doskonały jak ChatGPT, ale nie musi też być tak bardzo ogólny.

Dodatkową korzyścią płynącą z uruchomienia modeli językowych typu Bielik jest wzmocnienie pozycji Polski w obszarze innowacji w sektorze AI. Ponadto, co podkreślają twórcy, warto zabiegać o to, aby budować własne narzędzia i tym samym uniezależnić się od zewnętrznych firm, które w przypadku zawirowań na rynkach, regulacji czy ograniczeń prawnych mogą np. uniemożliwić dostęp do swoich zasobów. Tym samym rozwijając i udoskonalając narzędzia w Polsce, budujemy stabilne zaplecze i jesteśmy w stanie zabezpieczać wiele naszych sektorów – bankowy, administracyjny medyczny lub prawniczy.

– Intensywne działania nad rozwojem AI, modelami językowymi typu Bielik czy innymi narzędziami opartymi o sztuczną inteligencję są w interesie wszystkich dobrze działających gospodarek. Obserwujemy wzmożone prace nad tego typu rozwiązaniami w wielu krajach – wyjaśnia **Jan Maria Kowalski z Fundacji Speakleash.**

Czytaj także:

- [Jak dobry jest ChatGPT-4o? Przyglądamy się Omni, najnowszej sztucznej inteligencji od OpenAI](#)
- [Umarł król, niech żyje król? Mistral prezentuje nowy model LLM, a OpenAI na drodze do bankructwa](#)



A tak w ogóle, o startupach piszemy na Google News. [Obserwuj nas](#)



Mam Startup

[Czytaj więcej o autorze](#)