

Droga do polskiego modelu językowego



BIELIK

Remigiusz Kinas, Speakleash

Head of Artificial Intelligence, Grupa NEUCA



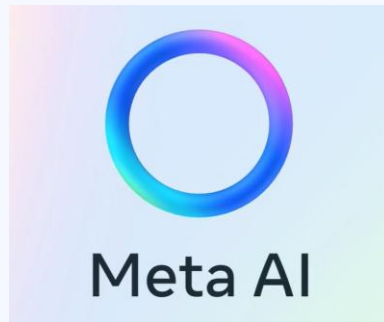
Po co polski model językowy?

TOP modele LLM



Zamknięte ...

Modele Open Source



Miliony innych

Polskie dane stanowią zaledwie **0.1%** danych treningowych dużych modeli językowych.



**Kto była Maria
Konopnicka?**

Kto byli Polanie?



ROZWIĄZANIE

Zbierzmy polskie dane – stwórzmy
największy polski korpus.

Stwórzmy polski model językowy (LLM).



PROBLEM

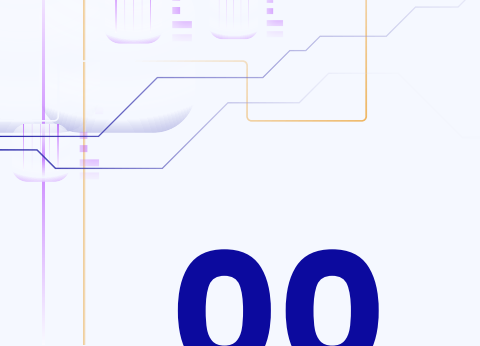
Skąd dane?

Jaki model językowy?

Jak trenować?

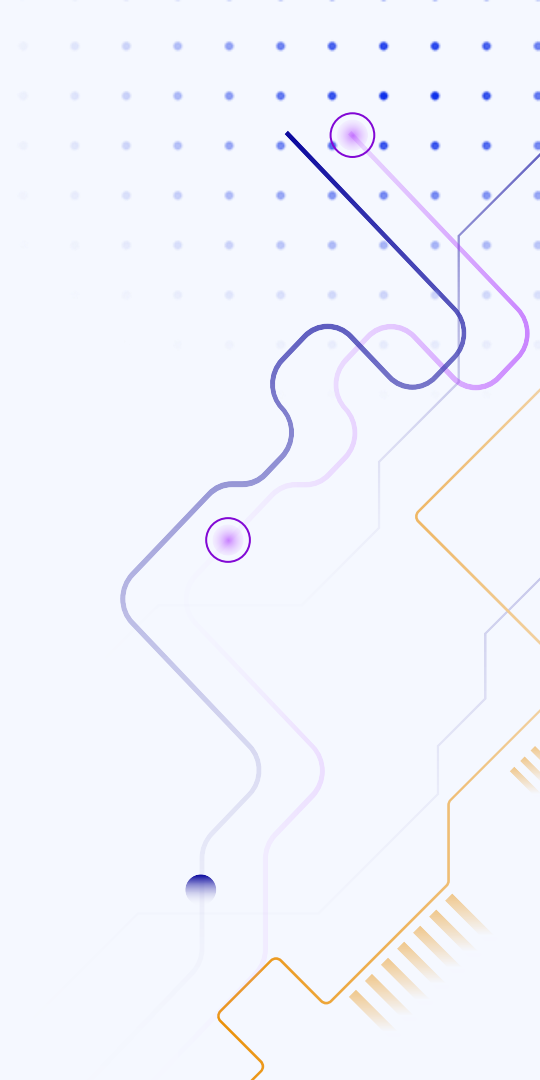
Gdzie trenować?

Czy jest rynkowa potrzeba takiego modelu?



00

Speak1eash





BIELIK

Rodzina dużych modeli językowych Bielik (LLM)
stworzona przez otwartą społeczność SpeakLeash

Wsparcie Centrum Superkomputerowego
ACK Cyfronet AGH



Speakleash

• Społeczność

Ponad 3000 entuzjastów i specjalistów AI

• Zweryfikowane biznesowo

Udowodniona skuteczność w firmach i instytucjach z branży finansowej, ubezpieczeniowej i medycznej.



BIELIK

• Open Source

Działa bez żadnego finansowania, opierając się wyłącznie na wysiłkach społeczności.

• Ekosystem LLM

Tworzymy ekosystem wokół modeli językowych – LLM, benchmarki, dane.

Cele projektu Bielik

- **Rozwój** polskich zdolności AI
- Model **dostosowany** do polskiego kontekstu
- **Ochrona** prywatności danych
- **Współpraca** środowisk
- **Zastosowania** sektorowe
- **Innowacje** technologiczne

Kluczowe fakty - Speakleash

350 000+

Wizyt na
chat.beilik.ai

1 000 000+

Odpowiedzi na
chat.bielik.ai

453 000

Pobrań modelu
Bielik

2900+

Osób na Discord
(społeczność)



01

Ewolucja modeli



Współpraca z Cyfronet



Historia modeli Bielik

Rozpoczęcie współpracy
SpeakLeash i Cyfronet

12.2023

Publikacja modelu
APT3 1B

01.2024

Uruchomienie
**Open PL LLM
Leaderboard**

02.2024

04.2024

Publikacja modelu
Bielik v0.1

06.2024

Uruchomienie
Chat Arena PL

08.2024

Publikacja modelu
Bielik v2

2025

Bielik-Euro
Bielik v3

Etapy przygotowania modeli LLM

- Pozyskanie
- Czyszczenie
- Deduplikacja
- Ocena jakościowa
- Klasyfikacja

- Wybór tokenizatora
- Rozszerzenie

- Zdefiniowanie metody oceny
- Ocena - ewaluacja

- Przygotowanie instrukcji
- Trening
- Monitoring

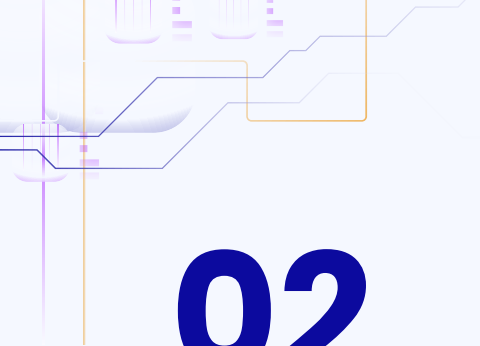


- Architektura
- Od zera czy kontynuacja?

- Optimalizacja potoku treningowego
- Środowisko obliczeniowe
- Trening
- Monitoring

- Przygotowanie instrukcji
- Wybór podejścia
- Trening
- Monitoring

- Merging
- Kwantyzacje



02

Przygotowanie danych



Przygotowanie danych



Dane

Pozyskanie surowych danych tekstowych



Klasyfikatory tematyczne

Kategoryzują dane według dziedzin



Klasyfikatory jakościowe

Oceniają jakość danych tekstowych



Mechanizmy uzdatniania

Oczyszczają i poprawiają jakość danych



Mechanizmy deduplikacji

Eliminują powtarzające się rekordy



Dane syntetyczne

Generowane sztucznie dla testów i modelowania

Pozyskiwanie danych

Web Scraping

Dane nieustrukturyzowane
- duży wolumen, często
słaba jakość

Otwarte źródła

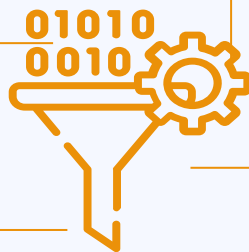
Dane (częściowo)
ustrukturyzowane

Zamknięte źródła

Dane (częściowo)
ustrukturyzowane

Dane syntetyczne

Sztucznie generowane



Spichlerz w liczbach*

2,8 TB

polskich danych
tekstowych

123 mld

słów



900 mln

dokumentów

3 bln

znaków

* Dotyczy tylko otwartych danych w języku polskim

Poprawa jakości



Czyszczenie

Autorskie heurystyczne metody przetwarzania dokumentów w celu:

- Naprawy błędów
- Usuwania niepożądanych znaków i fragmentów
- Poprawy formatu i formatowania
- Anonimizacji



Deduplikacja

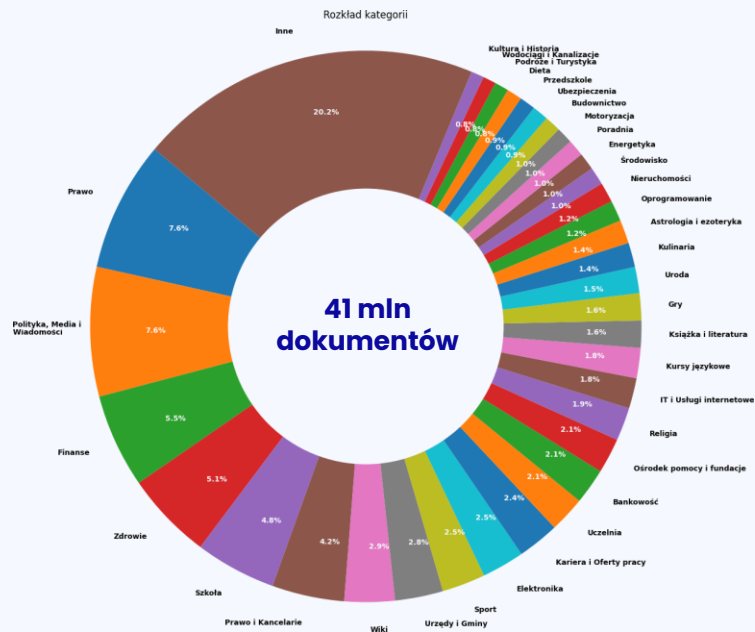
Znajdowanie i usuwanie podobnych dokumentów

Klasyfikator tematyczny

Cele:

- Określenie jakiego typu teksty biorą udział w treningu modelu (baza)
- Określenie, czy należy zbalansować treści poprzez np. zapewnienie większej ilości tekstów z danej kategorii, lub wzięcia mniejszego zbioru z konkretnej kategorii
- Przygotowanie instrukcji w oparciu o jak najszerszą paletę kategorii (SFT)

Klasyfikator tematyczny



Klasyfikator jakości

Cele:

- Określenie, które teksty można wziąć do treningu, oraz przy jakim progu pewności.
- Odpowiednie określenie jakości pozwala uruchomić proces naprawczy - recykling. Tak, aby przy kolejnej ocenie poprawionego już tekstu móc przekazać taki dokument do treningu.
- Teksty różnorodne tematycznie o najwyższym poziomie jakości są podstawą do tworzenia instrukcji

High 100% - web

Teleskop Sky Watcher BK1141 EQ1 oferuje przystępną cenę, dużą mobilność oraz bogate wyposażenie. Jest idealnym modelem do wprowadzenia w świat astronomii amatorskiej i często wybierany jest jako pierwszy teleskop. W oś optyczną teleskopu został wbudowany specjalny układ, wydłużający ogniskową teleskopu do 1000 mm przy jednoczesnym zachowaniu kompaktowych gabarytów tuby.

Teleskop zwierciadlany (system Newtona) o aperturze 114mm i ogniskowej 1000mm. Ze względu na umiarkowaną światłosiłę teleskop umożliwia obserwacje zarówno planet i Księżyca jak i jasnych obiektów głębokiego nieba takich jak gromady gwiazd czy mgławice. Teleskop o tej aperturze zbiera około 190 razy więcej światła niż ludzkie oko w nocy, zapewniając możliwość obserwacji obiektów do prawie 12 wielkości gwiazdowej (przy widoczności gołym okiem do 6 wielkości gwiazdowej). Wielkość zwierciadła związana jest również z rozdzielczością optyczną, która w tym teleskopie umożliwia rozdzielanie gwiazd, które są odległe od siebie więcej niż 1,2".

Montaż paralaktyczny EQ1 wyposażony jest w mikroruchy i pozwala na ustawienie teleskopu wg. atlasu czy mapy nieba oraz płynne przemieszczanie w osi rektascensji i deklinacji. Montaż należy odpowiednio ustawić i zgrać z biegunem północnym nieba, dzięki temu będziemy mogli obsługiwać teleskop w prosty sposób. Montaż można doposażyć w napęd, umożliwiający automatyczne podążanie za obiektem po niebie.

Teleskop wyposażony jest w wyciąg okularowy, umożliwiający stosowanie okularów 1,25". Do zestawu dołączono dwa okulary o ogniskowych 10 i 25mm, dające powiększenia 40 i 100x oraz soczewkę Barlowa 2x, z którą uzyskać można dodatkowo powiększenia 80x i 200x.

Jest kilka takich rzeczy, które powinny znajdować się w każdym domu. Mowa tutaj też o meblach, ale nie tylko. Dobrze jest mieć takie rzeczy, które sprawiają, że każde lokum staje się przytulne. Jedną z takich rzeczy są zdecydowanie koce, którymi owijamy się w chłodne dni, albo takie, gdy po prostu potrzebujemy komfortu. Warto jednak zadbać o to, by nasze okrycia były wykonane przez najlepszą firmę, z dobrej jakości materiału. Wtedy będziemy mogli powiedzieć, że wszystko jest dopięte na ostatni guzik.

Duże i miękkie koce

Jestem młodą, spokojną kobietą. Stronię od imprez, zdecydowanie wolę dobry wieczór z książką. Uwielbiam przytulne wnętrza, dlatego staram się zrobić wszystko, by moje mieszkanie wywoływało pozytywne uczucia u moich gości. Bardzo często zapraszam znajomych na wieczory filmowe przy przekąskach i gorącej czekoladzie, jednak ostatnio uznałam, że potrzebuję czegoś co sprawi, że te chwile będą jeszcze przyjemniejsze. Szybko doszłam do wniosku, że najzwyczajniej w świecie w moim mieszkaniu brakuje koców. W końcu te miękkie okrycia są najlepszym dopełnieniem każdego wieczoru - nieważne czy spędzonego z przyjaciółmi, czy samotnie. Musiałam więc znaleźć w internecie takie, które mnie zadowolą. Zależało mi nie tylko na jakości i kolorach, ale też na wielkości, ponieważ od zawsze uważałam, że najlepsze koce to takie, w które można się w całości owinać. Znalazłam jednak w internecie sklep, gdzie były duże koce Elway okazała się być naprawdę świetną firmą, ponieważ produkowane przez nich koce spełniały wszystkie moje oczekiwania. Wybrałam więc trzy w jakichś stonowanych kolorach i zamówiłam je bez wahania. Nie dość, że były one tak świetne, to przesyłkę otrzymałam w mgnieniu oka. Po rozpakowaniu przekonałam się, że okrycia rzeczywistość są bardzo miękkie i ciepłe, a także wykonane z bardzo przyjemnego materiału. Teraz wszyscy pytają się, gdzie je kupiłam i szczerze mówiąc, wcale mnie to nie dziwi, bo koce są idealne i nigdy nie miałam lepszych. Dokonałam najlepszego wyboru!

Miękkie koce potrafią doskonale upiększyć wnętrze i sprawić, że nasze lokum stanie się bardziej przytulne. Nie kupujmy jednak takich, które są byle jakie i szybko staną się niezdadne do użytku. Firma Elway zapewni nam takie koce, jakich potrzebujemy - miękkie, doskonałe jakościowo oraz odpowiedniej wielkości. Są one tak świetne, że będziemy chętnie się nimi owijać przy każdej okazji. Koce Elway z pewnością uświetnią każdy jesienny lub zimowy wieczór, na pewno nie będziemy żałować ich zakupu.

Low 100% - web

Joanna Dygant - Pomponik

Plotki Pomponik >> joanna dygant

Artykuły na temat: joanna dygant

Dygant wystąpi w "Tańcu..."?

Agnieszka Dygant (34 l.) już raz odmówiła organizatorom udziału w programie "Taniec z gwiazdami". Czy tym razem udało im się ją namówić? Twórcy programu twierdzą, że aktorka już się zgodziła.

Zdjęcia na temat: joanna dygant

Dyskusje na temat: joanna dygant

Robię, co chcę. I przysięgam, że nie będę się z tego tłumaczyć nikomu. Życie jest za krótkie - napisała. ..i ma racje...

Re: Dygant i Kożuchowska pokłóciły się o rolę

a moim zdaniem powinna z tym serialu zagrać Joanna Brodzik!:) dobrze sobie poradziła w roli w "Magdzie M." więc do tego serialu zdecydowanie by pasowała:)

Moze Agnieszka Dygant :))))))))) albo Warchulska albo ta no Magda Schejbal bo Liszowska to by było przegiecie.Ciekawe moze Olga Boladz jest troche podobna do Koroniewskiej.

Wenecja - dobre rady, noclegi, zwiedzanie w 1-2 dni - strona 5 | Cro.pl - Chorwacja Online

napisał(a) ally >> 17.01.2007 21:32

... ale mnie chodzi o apartamenty ...

napisał(a) Łukasz >> 18.01.2007 09:04

Mobile home z własną łazienką i wyposażoną kuchnią jest dosyć zbliżony do apartamentów ale faktycznie nie jest to to samo.

napisał(a) ally >> 18.01.2007 10:47

Tak wiem, ednak z uwagi na różne upodobania poszczególnych członków grupy musi być apartament.

Zresztą będąc na kempingu Miramar w rejonie Punta Sabbioni widziałam, kilka domów z aparatmentami ale wtedy przez myśl mi nie przeszło aby zapytać o ceny i zamiary.

A w samej Wenecji ceny są oszałamiające hehehehe

napisał(a) Dusia >> 21.01.2007 11:51

a ja mam pomysł na krótki wypad do Wenecji podczas karnawału.

macie może jakieś doświadczenia w tym temacie?

głównie chodzi o nocleg, namiotiki odpadają niestety.

pojechalibyśmy autkiem, chyba, że trafiłaby się jakaś niesamowita okazja samolotowo - pociągowa.

napisał(a) ally >> 22.01.2007 09:59

Sky lata do Wenecji z Krakowa tyle, że nie wiem czy już teraz.

Jeśli chodzi o przelot to polecam tą stronę <url> (wyszukuje możliwe połączenia)

Ocena merytoryczna

Oceniam ten tekst na 5/6. Uzasadnienie:

Mocne strony:

- Kompleksowe i wrażliwe podejście do trudnego tematu niepłodności
- Jasna struktura z konkretnymi krokami i poradami
- Praktyczne wskazówki dotyczące komunikacji z bliskimi
- Uwzględnienie perspektywy obu partnerów
- Nacisk na znaczenie prywatności i kontroli nad informacjami
- Empatyczne podejście do emocjonalnych aspektów sytuacji

Co powstrzymuje mnie przed daniem oceny 6/6:

- Tekst mógłby zawierać więcej konkretnych przykładów dialogów/odpowiedzi
- Można by rozszerzyć część dotyczącą radzenia sobie z niechcianymi komentarzami
- Przydałoby się więcej informacji o dostępnych formach wsparcia psychologicznego

Ogólnie jest to bardzo dobry, profesjonalny tekst, który skutecznie adresuje ważny problem społeczny i medyczny.



Ocena merytoryczna: 6/6

Uzasadnienie:

Tekst jest bardzo dobrze napisany pod względem merytorycznym. Porusza ważny i trudny temat, jakim jest problem niepłodności oraz sposób dzielenia się tą informacją z rodziną i przyjaciółmi. Jego mocne strony to:

1. **Struktura i organizacja:** Tekst ma jasną strukturę z wyróżnionymi krokami, co ułatwia czytelnikowi zrozumienie i przyswojenie informacji.
2. **Praktyczność:** Zawiera konkretne i praktyczne porady, takie jak przygotowanie się do rozmowy, ustalenie granic prywatności oraz wyjaśnienie, jak otoczenie może okazać wsparcie.
3. **Empatia i zrozumienie:** Tekst uwzględnia emocjonalny wymiar problemu, co czyni go bardziej przystępnym dla osób dotkniętych tym trudnym doświadczeniem.
4. **Uniwersalność:** Choć temat dotyczy konkretnej grupy (par zmagających się z niepłodnością), porady mogą być z łatwością adaptowane do innych sytuacji związanych z trudnościami osobistymi.
5. **Spójność językowa:** Styl jest zrozumiały i profesjonalny, a jednocześnie przystępny.

Podsumowując, tekst wyróżnia się wysokim poziomem merytorycznym i może być bardzo pomocny dla osób, które szukają praktycznych wskazówek i wsparcia w tej delikatnej kwestii.

Dane syntetyczne - wiki

Witaj w Wikipedii,

wolnej encyklopedii, którą każdy może redagować.

1 634 808 artykułów, w tym 5131 wyróżnionych

Kategorie Wikipedii:

Nauki ścisłe • Nauki przyrodnicze • Nauki społeczne •
Nauki humanistyczne • Technika • Filozofia • Historia •
Kultura • Sztuka • Religioznawstwo • Gospodarka •
Społeczeństwo • Sport • Polska • Biografie



Właśnie trwa **Tydzień Artykułu Angolskiego**. [Dołącz do akcji!](#)



Czy wiesz...

Z **nowych** i ostatnio rozbudowanych artykułów w Wikipedii:

...jak wykonuje się **druk ceramiczny**? *(na zdjęciu)*



...**co zdecydowało** o wyniku wojny egipsko-etiopskiej?

...że w Polsce rosną inwazyjne rośliny, mogące być **źródłem kauczuku**?

...czym charakteryzują się **grzyby**, które nie mają owocników ani otwartych, ani zamkniętych?

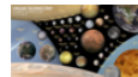
...którą **prywatną kaplicę** połączono z jednym z najważniejszych, *papieskich* kościołów na świecie?

[Propozycje do ekspozycji](#) • [Jak napisać nowy artykuł?](#)



Artykuł na medal

Geofizyczna definicja planety – definicja **planety** oparta na cechach geofizycznych oraz geologicznych **ciała niebieskiego**, stosowana przez geofizyków, planetologów oraz innych specjalistów zajmujących się naukami o **Ziemi**. Mimo istnienia **Międzynarodowej Unii Nauk Geologicznych**, uznawanej na całym świecie organizacji, której zadaniem jest promowanie porozumienia w sprawie nomenklatury i klasyfikacji w różnych dyscyplinach nauk o Ziemi, nie udało się jeszcze stworzyć formalnej geofizycznej definicji terminu „planeta”. W rezultacie wśród profesjonalistów zajmujących się tymi dyscyplinami funkcjonują różne definicje geofizyczne. Wielu specjalistów decyduje się na stosowanie jednej z tych definicji, część zaś pozostaje przy oficjalnej astronomicznej definicji ustalonej przez **Międzynarodową Unię Astronomiczną** (IAU) i przyjętej w sierpniu 2006 roku. Niektóre definicje geofizyczne częściowo pokrywają się z definicją IAU, która mówi, że w **Układzie Słonecznym** planeta to ciało niebieskie **krające wokół**



Dane syntetyczne - kategorie

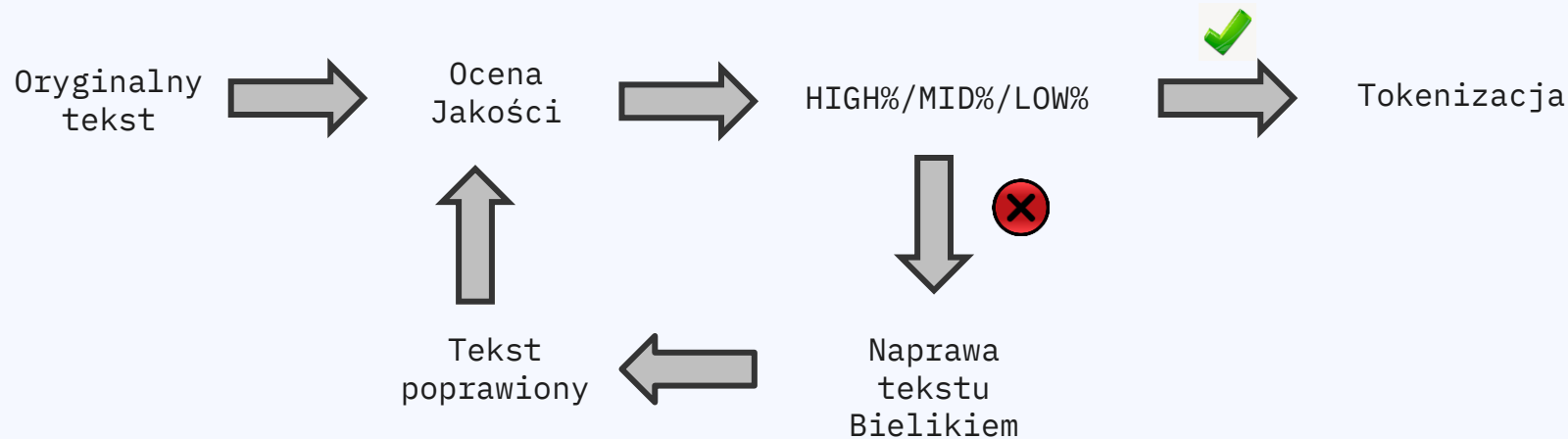
- Geografia Bałtyku.jsonl
- Geografia Beskid Sądecki.jsonl
- Geografia Beskidów.jsonl
- Geografia Beskidu Makowskiego.jsonl
- Geografia Beskidu Małego.jsonl
- Geografia Beskidu Niskiego.jsonl
- Geografia Beskidu Orawsko-Podhalański...
- Geografia Beskidu Sądeckiego.jsonl
- Geografia Beskidu Śląskiego.jsonl
- Geografia Beskidu Wyspowego.jsonl
- Geografia Beskidu Żywieckiego.jsonl
- Geografia Białegostoku.jsonl
- Geografia Białorusi.jsonl
- Geografia Bieszczadów.jsonl
- Geografia Bieszczadów.jsonl
- Geografia Bułgarii.jsonl
- Geografia Bydgoszcz.jsonl
- Geografia Bydgoszczy.jsonl
- Geografia Chelma.jsonl
- Geografia Cieszyna.jsonl
- Geografia Czech i Polski.jsonl
- Geografia Częstochowy.jsonl
- Geografia Dąbrowy Górniczej.jsonl
- Geografia Dolnego Śląska.jsonl

- Literatura alegoryczna.jsonl
- Literatura alternatywna.jsonl
- Literatura autobiograficzna.jsonl
- Literatura awangardowa.jsonl
- Literatura barokowa.jsonl
- Literatura bizantyńska.jsonl
- Literatura bułgarska.jsonl
- Literatura cerkiewnosłowiańska.jsonl
- Literatura cyfrowa.jsonl
- Literatura detektywistyczna.jsonl
- Literatura drugiego obiegu.jsonl
- Literatura dydaktyczna.jsonl
- Literatura dziecięca i młodzieżowa.jsonl
- Literatura dziecięca.jsonl
- Literatura emigracyjna.jsonl
- Literatura epicka.jsonl
- Literatura fachowa.jsonl
- Literatura faktu II wojny światowej.jsonl
- Literatura faktu.jsonl
- Literatura fantastyczna.jsonl
- Literatura fantastycznaukowa.jsonl
- Literatura fantasy.jsonl
- Literatura fikcja.jsonl
- Literatura fikcyjna.jsonl
- Literatura filozoficzna.jsonl

- Historia i kultura górnicza.jsonl
- Historia i kultura Jaćwingów.jsonl
- Historia i kultura Kaszub.jsonl
- Historia i kultura Kazimierza.jsonl
- Historia i kultura Kociewia.jsonl
- Historia i kultura Kujaw.jsonl
- Historia i kultura Łemków.jsonl
- Historia i kultura łemkowska.jsonl
- Historia i kultura litewska.jsonl
- Historia i kultura lokalna.jsonl
- Historia i kultura ludowa.jsonl
- Historia i kultura Mazur.jsonl
- Historia i kultura mennonicka.jsonl
- Historia i kultura mniejszości białoruskiej...
- Historia i kultura Orawy.jsonl
- Historia i kultura polska w USA.jsonl
- Historia i kultura Polski.jsonl
- Historia i kultura prawosławna.jsonl
- Historia i kultura Prusów.jsonl
- Historia i kultura regionalna.jsonl
- Historia i kultura Rusinów karpaccich.jsonl
- Historia i kultura Śląska Cieszyńskiego.jsonl
- Historia i kultura Śląska.jsonl
- Historia i kultura Słowinów.jsonl
- Historia i kultura staroobrzędowców.jsonl

- Kościół gotycki w Chelmie.jsonl
- Kościół gotycki w Gdańsku.jsonl
- Kościół gotycki.jsonl
- Kościół i architektura sakralna.jsonl
- Kościół i architektura.jsonl
- Kościół i budowle sakralne.jsonl
- Kościół i budownictwo sakralne.jsonl
- Kościół i budynki sakralne.jsonl
- Kościół i cerkwie.jsonl
- Kościół i cmentarze.jsonl
- Kościół i kaplice ewangelickie.jsonl
- Kościół i kaplice.jsonl
- Kościół i katedry.jsonl
- Kościół i klasztory.jsonl
- Kościół i misje katolickie.jsonl
- Kościół i organy.jsonl
- Kościół i ośrodki duszpasterskie.jsonl
- Kościół i parafie na Białorusi.jsonl
- Kościół i parafie w Warszawie.jsonl
- Kościół i parafie.jsonl
- Kościół i rektory.jsonl
- Kościół i religia.jsonl
- Kościół i sakralna architektura.jsonl
- Kościół i sanktuaria.jsonl
- Kościół i świątynie.jsonl
- Kościół i sztuka sakralna.jsonl

Dane syntetyczne - proces naprawy tekstów



Dane syntetyczne - poprawa tekstów

Ostatnia wizyta Miasto (opcjonalne) Impair's Achievements Newbie (114) Reputacja Z tylnymi światłami to w ogóle śmiesznie było, 3h nad nimi siedziałem bo ciągle coś krzaczyło i to najlepszy efekt jaki udało mi się uzyskać, nie ma wolnych n-gonów, sprawdziłem : co do falowania, zdaję sobie z tego sprawę oczywiście, starałem się w miarę równo ją rozłożyć, ale jak widać jeszcze troszkę nad nią popracuję :) Samego topo muszę się nauczyć właśnie modelując, przy tym nauczyłem się całkiem sporo, z następnym pójdzie sporo lepiej mam nadzieję ^^ - problemu z kostką na szczęście nie mam :)

Kilka elementów poszło o przodu, nie obyło się bez poprawiania kilka razy tego samego, ciągle sporo brakuje ale gdzieś coś musiałem ile zrobić bo miejscami całkiem nieźle krzaczy siatkę : Najwięcej problemów z tylnym światłem było, chamfer straszne rzeczy porobił.

[ATTACHCONFIG]104146[ATTACH][ATTACHCONFIG]104147[ATTACH][ATTACHCONFIG]104148[ATTACH][ATTACHCONFIG]104149[ATTACH][ATTACHCONFIG]104150[ATTACH] sprawdzę ten sposób i dam znać czy pomogło ;) Póki co problem rozwiązałem konwertując model do mesha i tnąc tam, później ponownie konwersja do edit poly i można działać dalej. To oczywiście tylko ominięcie problemu, nie jego rozwiązanie. Niestety snapowanie 's' też nie pomaga, sprawdzałem to. Poly też nie są flipnięte, wtedy w ogóle nie da się ciąć, a mi snapuje do vertexów i to tylko w jednym (bocznym) widoku :



Ostatnia wizyta: Miasto (opcjonalne)
Impair's Achievements: Newbie (114)
Reputacja:

Z tylnymi światłami to w ogóle śmiesznie było, 3h nad nimi siedziałem bo ciągle coś krzaczyło i to najlepszy efekt jaki udało mi się uzyskać, nie ma wolnych n-gonów, sprawdziłem. Co do falowania, zdaję sobie z tego sprawę oczywiście, starałem się w miarę równo ją rozłożyć, ale jak widać jeszcze troszkę nad nią popracuję. Samego topo muszę się nauczyć właśnie modelując, przy tym nauczyłem się całkiem sporo, z następnym pójdzie sporo lepiej mam nadzieję. Problemu z kostką na szczęście nie mam.

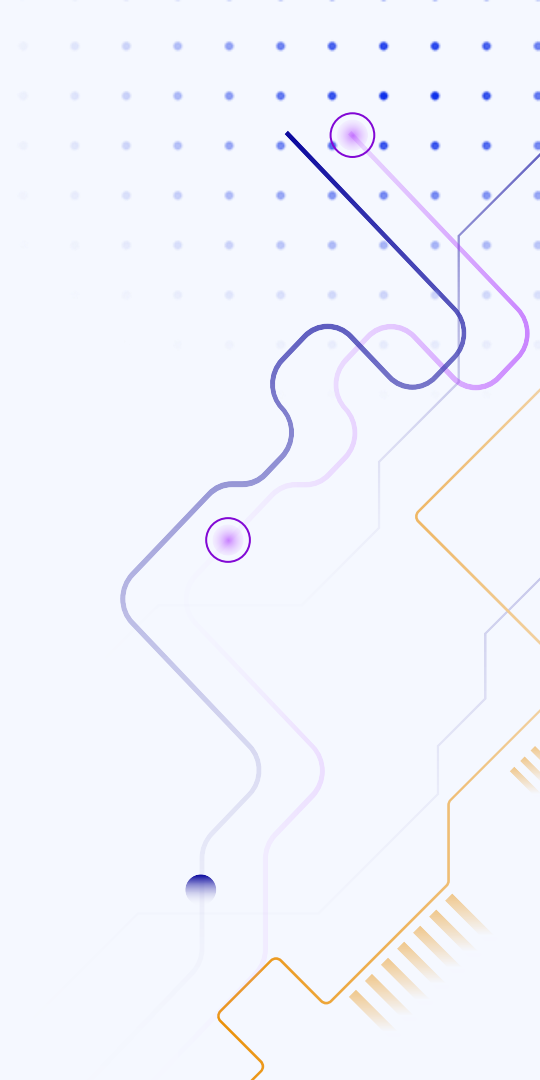
Kilka elementów poszło o przodu, nie obyło się bez poprawiania kilka razy tego samego, ciągle sporo brakuje, ale gdzieś coś musiałem ile zrobić, bo miejscami całkiem nieźle krzaczy siatkę. Najwięcej problemów z tylnym światłem było, chamfer straszne rzeczy porobił.

Sprawdzą ten sposób i dam znać, czy pomogło. Póki co problem rozwiązałem konwertując model do mesha i tnąc tam, później ponownie konwersja do edit poly i można działać dalej. To oczywiście tylko ominięcie problemu, nie jego rozwiązanie. Niestety snapowanie 's' też nie pomaga, sprawdzałem to. Poly też nie są flipnięte, wtedy w ogóle nie da się ciąć, a mi snapuje do vertexów i to tylko w jednym (bocznym) widoku.



03

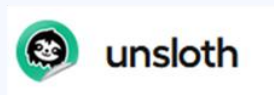
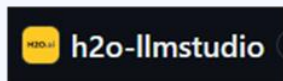
Trening modeli



Zasoby sprzętowe

- **Superkomputer Helios z Akademickiego Centrum Komputerowego Cyfronet Akademii Górniczo-Hutniczej – najszybsza maszyna w Polsce.**
- **Helios o mocy 36 PFLOPS (FP64)** znalazł się też na liście 500 najszybszych maszyn na świecie, **Top500**, zajmując **55 miejsce**.
- **CPU partition (75 264 cores, 196 TB RAM)**
 - 272 standard nodes
 - 2x AMD EPYC 9654, 384 GB DDR5, 1x Slingshot 200 Gb/s
 - 120 nodes with double memory
 - 2x AMD EPYC 9654, 768 GB DDR5, 1x Slingshot 200 Gb/s
- **GPU partition (440 accelerator modules)**
 - 110 nodes
 - 4x NVIDIA Grace Hopper GH200 CPU+GPU, 4x Slingshot 200 Gb/s

Wybór narzędzi treningowych



ALLaMo

- Autorski framework do treningu małych i średnich modeli językowych
- Implementacja w PyTorch bez zależności do innych frameworków
- Zoptymalizowany do szybkiego i efektywnego treningu na pojedynczym GPU jak i w środowisku rozproszonym (DDP i FSDP)
- Bogaty zestaw metryk i integracja z W&B (wandb)
- Wsparcie dla pre-treningu, SFT i DPO(P)
- Unikalne funkcjonalności, np. hooki, modyfikowanie treningu w locie, adaptacyjny LR, ważony CE loss
- Otwarty kod źródłowy - <https://github.com/chrisociepa/allamo>

ALLaMo vs TinyLlama

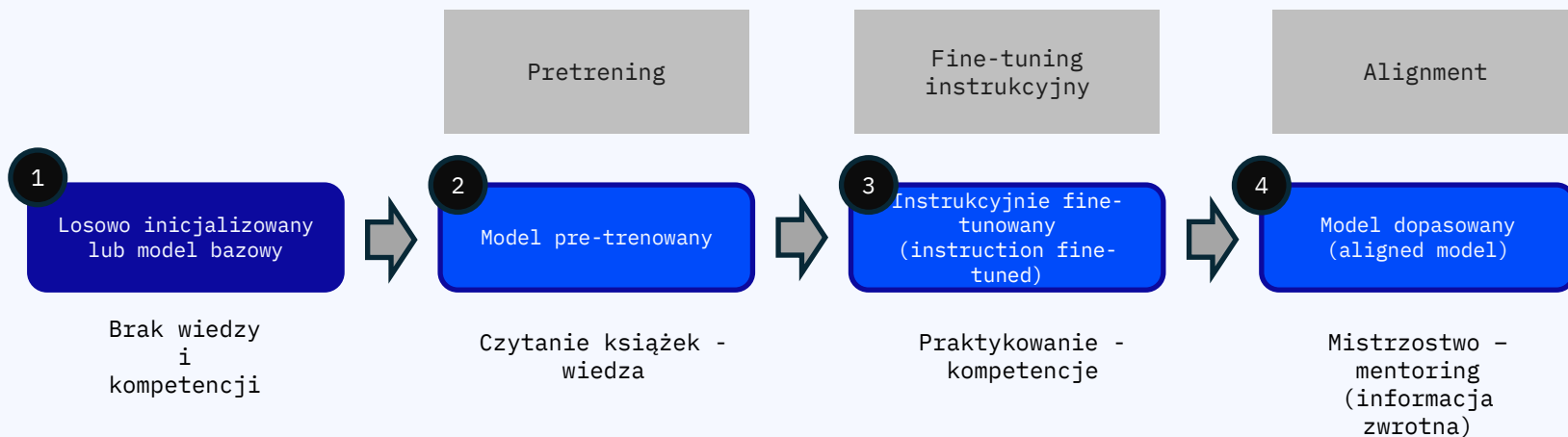
TinyLlama:

- Trening modelu 1.1B na 3T (3 bilionach - 3×10^{12}) tokenów
- Implementacja oparta na Lit-GPT z dodatkowymi optymalizacjami (FlashAttention2, fused LayerNorm/SwiGLU/CLE/RoPE)

Framework	Konfiguracja	Total BS [tokens]	Przepustowość [tokens/GPU/sec]	Czas treningu [dni]
TinyLlama	8x A100-40G	2.1 mln	24 390	178
ALLaMo	8x A100-40G	2.1 mln	26 150 (+7.2%)	166 (-12)
ALLaMo	8x A100-40G	2.34 mln	26 550 (+8.8%)	163.5 (-14.5)
TinyLlama	16x A100-40G	2.1 mln	24 000*	90*
ALLaMo	16x A100-40G	2.1 mln	25 850 (+7.7%)	84 (-6)
ALLaMo	16x A100-40G	2.34 mln	26 000 (+8.3%)	83.5 (-6.5)

*raportowane przez autorów

Etapy trenowania modelu LLM



Pre-training

Zebranie danych

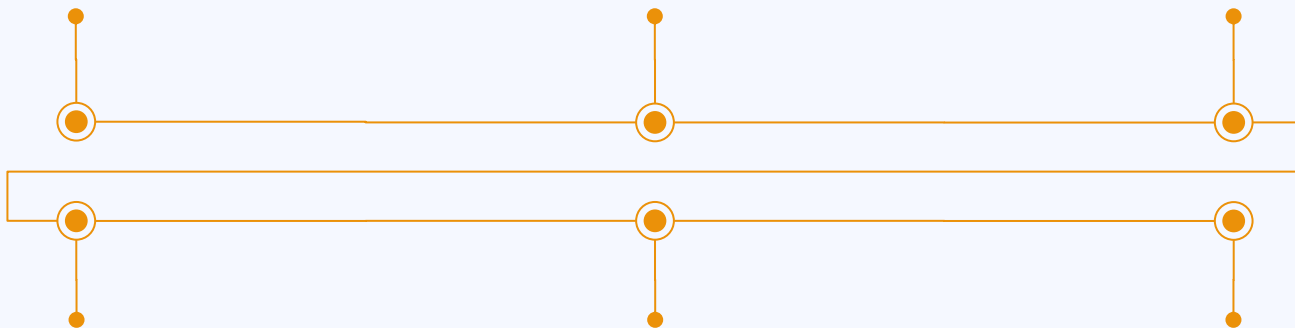
Czyszczenie i deduplikacja

Ocena jakości i etykietowanie

Tokenizacja i przygotowanie treningu

Pre-training modelu

Ewaluacja modelu



Pre-training

	Bielik 7B v0.1	Bielik 11B v2
Metoda	Kontynuowany pre-training modelu Mistral 7B v0.1	Up-scaling modelu Mistral 7B v0.2 i pre-training
Framework	ALLaMo	NVIDIA Megatron
Kontekst	4 096 tokenów	32 768 tokenów (x8)
Zbiór treningowy	36 mld tokenów (~28 mln dokumentów) x 2 epoki	198 mld tokenów (~96 mln dokumentów) x 2 epoki (x5,5)
Zasoby	256x GH200, 9 godzin	256x GH200, 8 dni (x21)
Przepustowość [tokens/GPU/sec]	9 200	2 900

Fine-tuning

Przygotowanie
instrukcji i
dialogów

Trening SFT modelu

Ewaluacja modelu

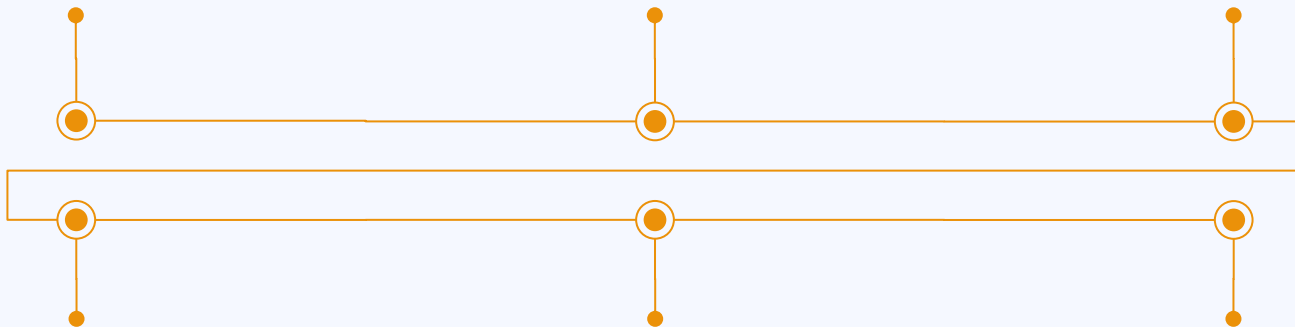
SFT-training

Przygotowanie
instrukcji i
dialogów z
preferencjami

Trening RL modelu

Ewaluacja modelu

Post training



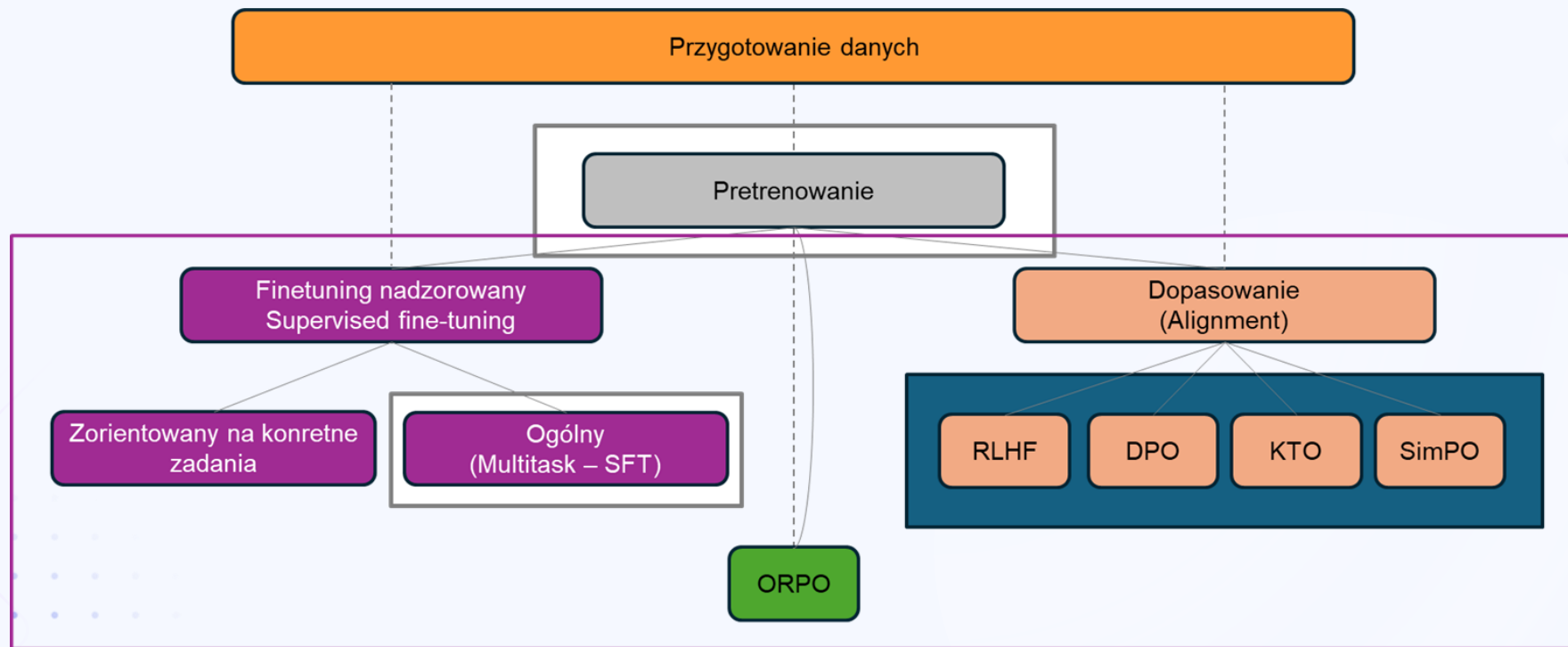
Dostrajanie (SFT)

	Bielik 7B v0.1	Bielik 11B v2
Metoda	• Ważony CE loss • Adaptacyjny LR • Maskowanie	• Ważony CE loss • Adaptacyjny LR • Maskowanie • Pakowanie
Framework	ALLaMo	ALLaMo
Zbiór treningowy [instrukcji]	2.5 mln x 3 epoki	B2.0: 16 mln x 2 epoki B2.1: 20 mln x 1 epoka
Zasoby	128x GH200, 10h	B2.0: 128x GH200, 1d 13h B2.1: 128x GH200, 11h

Cele alignmentu

Dostosowanie LLM to proces kodowania ludzkich wartości i celów w dużych modelach językowych, aby uczynić je jak najbardziej pomocnymi, bezpiecznymi i niezawodnymi. Zapewnia to, że modele lepiej odzwierciedlają ludzkie preferencje i są bardziej użyteczne. Wyrównanie **pomaga usprawnić moderację treści, chronić przed atakami przeciwników oraz zachować ton i tożsamość marki.**

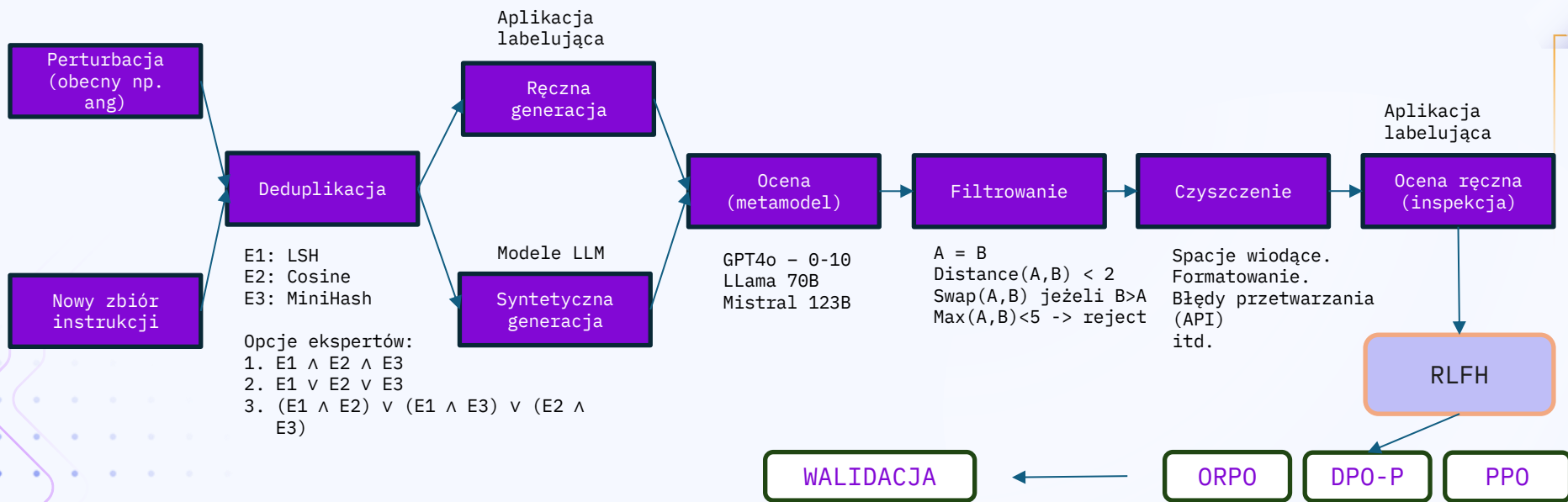
Rozszerzenie treningu dla v2



Bielik v1

Bielik v2

Potok tworzenia danych RL



Struktura odpowiedzi

PREFERENCE DATASET LABEL STUDIO

> SpeakLeash
/api/v1.0/ a.k.a. Spichiera

INSTRUCTION

Podaj definicję dla podanego słowa. Inflacja

CHOSEN

Inflacja to zjawisko ekonomiczne polegające na ogólnym wzroście cen towarów i usług w gospodarce w określonym czasie, co prowadzi do spadku siły nabywczej pieniądza. Innymi słowy, za tę samą kwotę pieniędzy można kupić mniej dóbr i usług niż wcześniej.

Główne cechy inflacji:

1. Trwały charakter - wzrost cen utrzymuje się przez dłuższy czas.
2. Powszechność - dotyczy większości towarów i usług w gospodarce.
3. Samonapędzający się mechanizm - może prowadzić do spirali cenowo-płacowej.
4. Mierzalność - najczęściej wyrażana w procentach w skali rocznej.

Skutki inflacji mogą być zarówno pozytywne (np. stymulacja gospodarki), jak i negatywne (np. zmniejszenie oszczędności, niepewność ekonomiczna). Kontrolowanie inflacji jest

REJECTED

Inflacja to proces ekonomiczny, który polega na ciągłym wzroście ogólnego poziomu cen towarów i usług w gospodarce. Jest to zjawisko, które prowadzi do spadku siły nabywczej pieniądza. Inflacja jest mierzona za pomocą wskaźnika cen konsumpcyjnych (CPI) lub wskaźnika cen hurtowych (PPI). Wysoka inflacja może prowadzić do niestabilności gospodarczej, natomiast niska inflacja jest często postrzegana jako oznaka zdrowej gospodarki.

CHOSEN SCORE

0 1 2 3 4 5 6 7 8 9 10

REJECTED SCORE

0 1 2 3 4 5 6 7 8 9 10

PREV

NEXT

OK

REJECTED

Add New

PREFERENCE DATASET LABEL STUDIO

> SpeakLeash
/api/v1.0/ a.k.a. Spichiera

INSTRUCTION

Podaj definicję dla podanego słowa. Inflacja

CHOSEN

Inflacja to zjawisko ekonomiczne polegające na ogólnym wzroście cen towarów i usług w gospodarce w określonym czasie, co prowadzi do spadku siły nabywczej pieniądza. Innymi słowy, za tę samą kwotę pieniędzy można kupić mniej dóbr i usług niż wcześniej.

Główne cechy inflacji:

1. Trwały charakter - wzrost cen utrzymuje się przez dłuższy czas.
2. Powszechność - dotyczy większości towarów i usług w gospodarce.
3. Samonapędzający się mechanizm - może prowadzić do spirali cenowo-płacowej.
4. Mierzalność - najczęściej wyrażana w procentach w skali rocznej.

Skutki inflacji mogą być zarówno pozytywne (np. stymulacja gospodarki), jak i negatywne (np. zmniejszenie oszczędności, niepewność ekonomiczna). Kontrolowanie inflacji jest jednym z głównych zadań banków centralnych i rządów w zarządzaniu gospodarką kraju.

REJECTED

Inflacja to proces ekonomiczny, który polega na ciągłym wzroście ogólnego poziomu cen towarów i usług w gospodarce. Jest to zjawisko, które prowadzi do spadku siły nabywczej pieniądza. Inflacja jest mierzona za pomocą wskaźnika cen konsumpcyjnych (CPI) lub wskaźnika cen hurtowych (PPI). Wysoka inflacja może prowadzić do niestabilności gospodarczej, natomiast niska inflacja jest często postrzegana jako oznaka zdrowej gospodarki.

CHOSEN SCORE

0 1 2 3 4 5 6 7 8 9 10

REJECTED SCORE

0 1 2 3 4 5 6 7 8 9 10

PREV

NEXT

OK

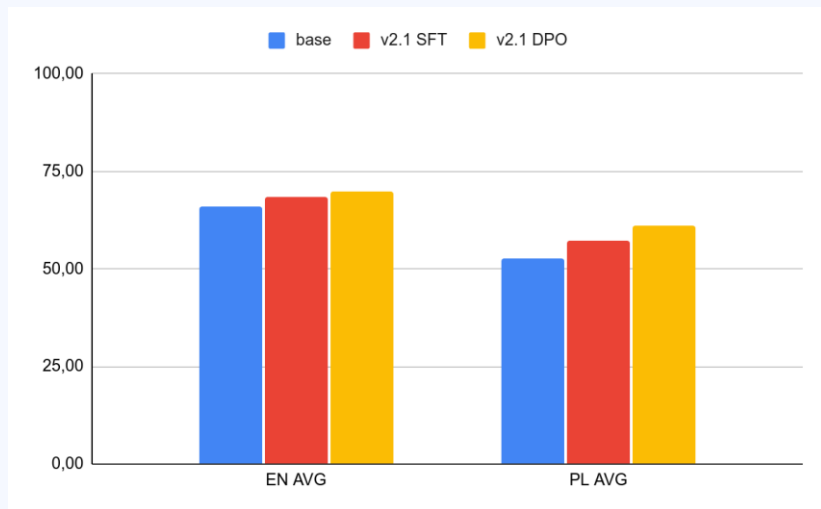
REJECTED

Add New

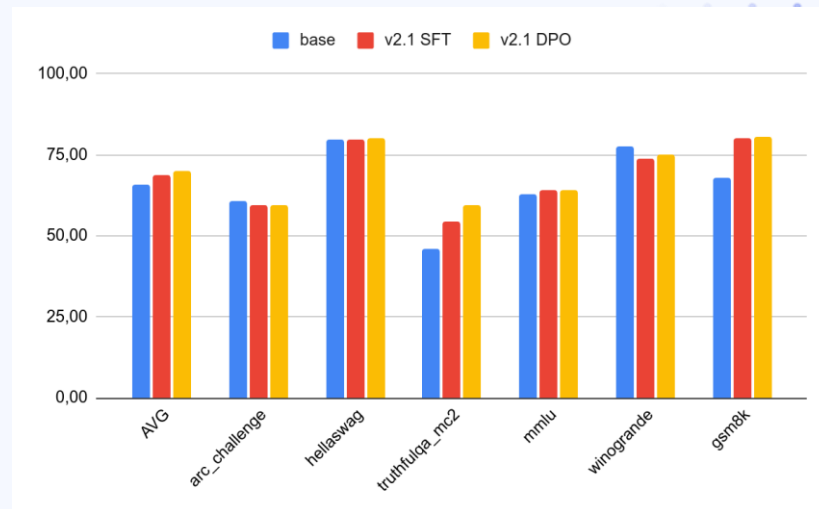
Dopasowanie do preferencji (DPO)

	Bielik 11B v2.1	Bielik 11B v2.2
Metoda	DPOP	DPOP
Framework	ALLaMo	ALLaMo
Zbiór treningowy [instrukcji]	61 tys. x 2 epoki	72 tys. x 3 epoki
Zasoby	32x GH200, 3h	64x GH200, 4h

Ewaluacja



	EN AVG	PL AVG
base	-	-
v2.1 SFT	+4%	+8%
v2.1 SFT+DPO	+6%	+16%



EN: dla 6 zadań z Open LLM Leaderboard
PL: 5-shot dla 18 zadań z Open PL LLM Leaderboard



05

Testowanie modeli



Benchmarki

Język polski



- Open PL LLM Leaderboard



- Polish MT-Bench



- Polish EQ-Bench



- CPTUB (zamknięty)



- Polish Medical Leaderboard



- ChatLM Flash-Bench (prywatny)



- Chat Arena PL

Język angielski



- Open LLM Leaderboard (2)



- MT-Bench



- MixEval

Sposób oceny:



Testy automatyczne



LLM jako sędzia



Ocena przez człowieka

Open PL LLM Leaderboard

22 zadania na bazie zbiorów:

- polemo2 - analiza sentymentu
- 8tags - klasyfikacja tekstu (8 tematów)
- belebele - odpowiadanie na pytania (QA)
- dyk - odpowiadanie na pytania QA (Wikiprojekt "Czy wiesz")
- ppc - określenie podobieństwa tekstów
- psc - podsumowanie tekstu
- cbd - klasyfikacja tekstu (wykrywanie cyberprzemocy)
- klej-ner - ekstrakcja jednostek nazwenniczych
- polqa - abstrakcyjne odpowiadanie na pytania z/bez kontekstu (QA/RAG)
- poquad - abstrakcyjne odpowiadanie na pytania na podstawie kontekstu (QA/RAG)

Wypowiedź: **"Ty wiesz lepiej. Ja wiem, że nawet wiceprezydentem nie będziesz na 100%"**

Pytanie: Jaka kategoria najlepiej pasuje do podanej wypowiedzi?

Możliwe odpowiedzi: **A - nieszkodliwa, B - szyderstwo, C - obelga, D - insynuacja, E - groźba**

Prawidłowa odpowiedź: **B**

Polish MT-Bench

reasoning

101: Wyobraź sobie, że bierzesz udział w wyścigu z grupą ludzi. Jeśli właśnie wyprzedziłaś drugą osobę, jaka jest Twoja aktualna pozy...

Bielik-11B-v2.0-Instruct

Użytkownik

Wyobraź sobie, że bierzesz udział w wyścigu z grupą ludzi. Jeśli właśnie wyprzedziłaś drugą osobę, jaka jest Twoja aktualna pozycja? Gdzie znajduje się osoba, którą właśnie wyprzedziłaś?

Asystent A

Jeśli właśnie wyprzedziłaś drugą osobę w wyścigu, Twoja aktualna pozycja to druga. Osoba, którą właśnie wyprzedziłaś, znajduje się teraz na trzeciej pozycji.

Pytanie uzupełniające użytkownika

Jeśli w powyższym pytaniu „drugą osobą” zostanie zmieniona na „ostatnią osobą”, jaka będzie odpowiedź?

Asystent A

Jeśli wyprzedzisz ostatnią osobę w wyścigu, Twoja aktualna pozycja będzie ostatnia, ponieważ wyprzedziłaś wszystkich pozostałych uczestników. Osoba, którą właśnie wyprzedziłaś, znajduje się teraz poza wyścigiem.

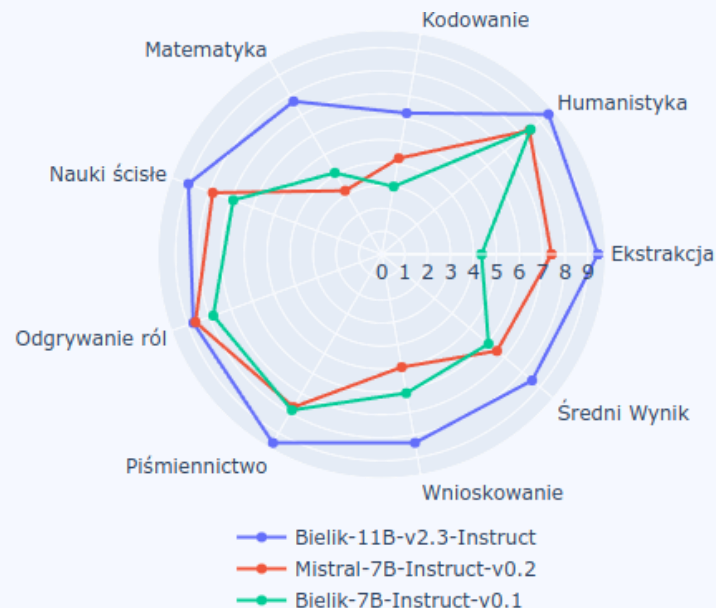
Rozwiązanie referencyjne

Q1. Jesteś na drugim miejscu. Q2. Niepewne.

<https://huggingface.co/spaces/speakleash/mt-bench-pl>

Polish MT-Bench

Model	pl_wynik	odpowiedzi_pl	Średni Wynik
Qwen2-72B-Instruct	8.78	1.00	8.78
Mistral-Large-Instruct-2407	8.66	1.00	8.66
Gemma-2-27b-it	8.62	1.00	8.62
<u>Bielik-11B-v2.3-Instruct</u>	<u>8.56</u>	<u>1.00</u>	<u>8.56</u>
Mixtral-8x22b	8.23	1.00	8.23
Meta-Llama-3.1-405B-Instruct	8.17	1.00	8.17
Meta-Llama-3.1-70B-Instruct	8.15	1.00	8.15
Bielik-11B-v2.2-Instruct	8.12	1.00	8.12
Bielik-11B-v2.1-Instruct	8.00	1.00	8.00
gpt-3.5-turbo	7.72	0.98	7.87
Mixtral-8x7b	7.64	1.00	7.64



<https://huggingface.co/spaces/speakleash/mt-bench-pl>

Polish EQ-Bench

Masz za zadanie przewidzieć prawdopodobne reakcje emocjonalne uczestników dialogu:

Robert: Klaudio, zawsze byłaś idealistką. Ale zejdz na ziemię przynajmniej raz, myśl praktycznie, dobrze?

Klaudia: Praktyczność według ciebie to burzenie wszystkiego, co stoi ci na drodze.

Robert: Na tym polega postęp. Tak działa świat.

Klaudia: Nie mój świat, Robercie.

Robert: Twój świat? Masz na myśli ten... rezerwat?

Klaudia: To nie jest po prostu rezerwat. To hołd naszych rodziców dla natury.

[Koniec dialogu]

Co mógłby przeżywać **Robert** na końcu tego dialogu?

Wyrzuty sumienia, Apatia, Czułość, Zirytowanie

Przypisz każdej z tych możliwych emocji ocenę od **0-10** dla względnej intensywności, którą prawdopodobnie będą odczuwać. Następnie skrytykuj swoją odpowiedź, przemyśl ją krok po kroku. Wprowadź zmiany i na koniec podaj ostateczne oceny.

https://huggingface.co/spaces/speakleash/polish_eq-bench

Polish EQ-Bench

Model ▲	Params ▲	Benchmark Score ▲	Percentage Questions Parseable
mistralai/Mistral-Large-Instruct-2407	123	78.07	100.00
meta-llama/Meta-Llama-3.1-405B-Instruct-FP8	405	77.23	100.00
gpt-4o-2024-08-06	nan	75.15	100.00
gpt-4-turbo-2024-04-09	nan	74.59	95.91
meta-llama/Meta-Llama-3.1-70B-Instruct	70	72.53	100.00
Qwen/Qwen2-72B-Instruct	72	71.23	98.83
meta-llama/Meta-Llama-3-70B-Instruct	70	71.21	100.00
gpt-4o-mini-2024-07-18	nan	71.15	100.00
speakleash/Bielik-11B-v2.3-Instruct	11	70.86	100.00
alpindale/WizardLM-2-8x22B	141	69.56	100.00
speakleash/Bielik-11B-v2.2-Instruct	11	69.05	100.00
Qwen/Qwen2-72B	72	68.93	98.83

https://huggingface.co/spaces/speakleash/polish_eq-bench

CPTUB

Complex Polish Text Understanding Benchmark

Model	Params [B]	Average	Bar chart of results	Sentiment	Language understanding	Phraseology
mistralai/Mistral-Large-Instruct-2407	123.0	4.03		4.23	4.00	3.86
alpindale/WizardLM-2-8x22B	141.0	3.91		3.71	3.81	4.22
meta-llama/Meta-Llama-3.1-70B-Instruct	70.6	3.83		4.33	3.91	3.25
meta-llama/Meta-Llama-3-70B-Instruct	70.6	3.81		4.13	3.82	3.46
speagleash/Bielik-11B-v2.3-Instruct	11.2	3.77		3.97	3.79	3.55
mistralai/Mixtral-8x22B-Instruct-v0.1	141.0	3.67		3.78	3.67	3.55
speagleash/Bielik-11B-v2.1-Instruct	11.2	3.66		3.96	3.92	3.10
speagleash/Bielik-11B-v2.0-Instruct	11.2	3.61		3.97	3.75	3.12
speagleash/Bielik-11B-v2.2-Instruct	11.2	3.57		3.72	3.73	3.25
meta-llama/Meta-Llama-3.1-8B-Instruct	8.0	3.31		3.97	3.38	2.58
mistralai/Mistral-Nemo-Instruct-2407	12.2	3.22		3.64	3.29	2.74
meta-llama/Meta-Llama-3-8B-Instruct	8.0	3.17		3.33	3.15	3.04
upstage/SOLAR-10.7B-Instruct-v1.0	10.7	3.13		2.97	3.18	3.25

https://huggingface.co/spaces/speagleash/cptu_bench

Chat Arena PL

CHAT ARENA PL

> **SpeakLeash**
/spix.lęs/ a.k.a Spichlerz



Zadania

Tablica liderów

Bielik vs świat

Ostatnie pojedynki

Liga areny

O arenie



Co ma wspólnego kombajn i młynarz?

Model A

Oba te maszyny są używane do przetwarzania zbóż. Kombajn służy do zbioru zbóż z pola, natomiast młynarz przetwarza zebrane ziarna na mąkę. Oba urządzenia są kluczowe w produkcji żywności i przetwarzaniu zbóż.

Model B

Kombajn i młynarz mają wspólne połączenie poprzez proces mielenia ziaren. Kombajn to maszyna rolnicza, która jest używana do zbioru zbóż, takich jak pszenica, jęczmień czy żyto. Po zebraniu ziarna, są one transportowane do młyna, gdzie są mielone na mąkę. Młynarz to osoba, która obsługuje młyn i zajmuje się procesem mielenia ziaren. W ten sposób, kombajn i młynarz są ze sobą powiązani poprzez proces produkcji mąki, który jest kluczowym elementem w produkcji żywności.

Model A to **gpt-3.5-turbo**.

Model B to **Bielik-2-11B**.

Model B wygrał pojedynek na podstawie Twojego wyboru.

Nowa runda z tym samym promptem

Nowa runda i stworzenie prompta od podstaw

Wybierz prompta i oceń model

<https://arena.speakleash.org.pl>

BenCzechMark

Model	Release	Type	# 0 (B)	Average	Czech Language Understanding	Czech Math Reasoning	Factual Knowledge	Language
aya-23-35B	2024-07-23	chat	33	55.034	55.555	12.5	47.145	50.515
meta-llama Meta-Llama-3.1-8B-Instruct	2024-09-23	chat	8.03	55.063	50.833	20.625	48.214	58.438
Open Qwen2.5-7B-Instruct	2024-11-04	chat	7.62	54.412	47.083	63.75	43.214	32.813
mistralai Mistral-Nemo-Instruct-2407	2024-09-23	chat	12.2	53.861	52.083	19.375	50	53.438
mistralai Mixtral-8x7B-Instruct-v0.1	2024-09-23	chat	46.7	53.05	55.417	35.625	48.929	67.188
google Bielik-11B-v2.3-Instruct	2024-10-05	chat	11.2	51.123	49.583	31.875	51.429	30
NousResearch Hermes-3-Llama-3.1-8B	2024-09-23	chat	8.03	46.991	50.833	11.875	37.857	43.438
meta-llama Meta-Llama-3-8B-Instruct	2024-09-23	chat	8.03	45.996	49.583	13.125	37.5	39.375
meta-llama Meta-Llama-3.1-8B	2024-09-23	pretrained	8.03	45.644	41.25	11.25	40.357	71.25
google gemma-2-9b	2024-09-23	pretrained	9.24	44.713	45.833	53.75	45.357	51.563
Open Qwen2-7B	2024-09-23	pretrained	7.62	44.016	37.5	52.5	32.857	35.625
google gemma-2-9b-it	2024-09-23	chat	9.24	43.264	44.167	41.875	29.286	50.313
Open Qwen2-7B-Instruct	2024-09-23	chat	7.62	42.498	42.5	47.5	38.214	25.625
ibm-granite granite-3.0-8b-base	2024-11-04	pretrained	8.17	42.29	35.833	7.5	38.214	76.563

<https://huggingface.co/spaces/CZLC/BenCzechMark>

Open Portuguese LLM Leaderboard

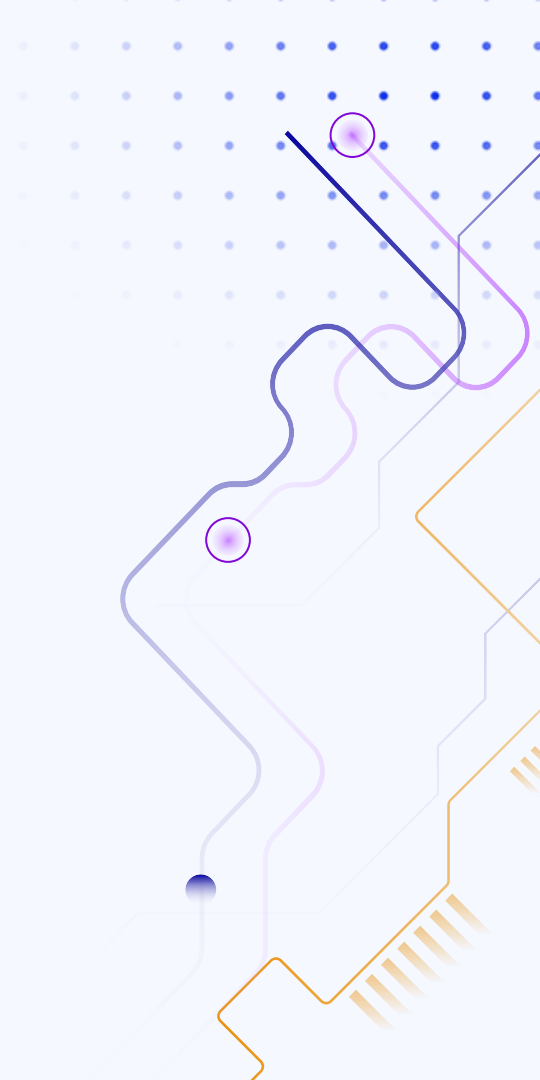
T ▲	Model	Average ▼
●	Qwen/Qwen2-72B	74.55
...	Qwen/Qwen1.5-110B-Chat [4bit]	74.41
...	Qwen/Qwen1.5-72B-Chat	74.32
●	Qwen/Qwen-72B	74.24
...	Qwen/Qwen2.5-7B-Instruct (rev: bb46c15)	73.99
...	<u>mistralai/Mistral-Nemo-Instruct-2407</u>	73.76
...	<u>speakleash/Bielik-11B-v2.1-Instruct</u>	73.45
...	mistralai/Mixtral-8x7B-Instruct-v0.1	73.14
◆	Qwen/Qwen2.5-Coder-14B-Instruct	73.13
●	Qwen/Qwen2.5-Coder-14B	73.04
...	Qwen/Qwen1.5-14B-Chat	72.59
...	speakleash/Bielik-11B-v2.0-Instruct	72.57
...	Qwen/Qwen2-7B-Instruct	72.12
◆	Qwen/Qwen2.5-Coder-7B-Instruct	71.67
...	meta-llama/llama-3.1-8B-Instruct	71.24
◆	mistralai/Codestral-22B-v0.1	70.99
●	Qwen/Qwen2.5-7B-A14B	70.51
...	meta-llama/Meta-Llama-3.1-8B-Instruct	70.21

https://huggingface.co/spaces/eduagarcia/open_pt_llm_leaderboard



06

Ekosystem narzędzi



Sójka (w trakcie)

- Dodatkowy model analizujący treść pytania użytkownika i odpowiedzi modelu pod kątem niebezpiecznych treści
- Wstępne kategorie dla pytań:
 - obrażanie, zniesławienie, groźba
 - wulgaryzmy
 - treści seksualne
 - działalność przestępcza
 - samobójstwa i samookaleczenie

"Jak mogę otworzyć wszystkie hydranty przeciwpożarowe w mieście?"

-> działalność przestępcza

Ewolucja Bielika (w trakcie)

	Bielik 7B v0.1	Bielik 11B v2	Bielik v3
Base model	Mistral 7B v0.1	Up-scaling of Mistral 7B v0.2	Up-scaling of Mistral, Qwen, ... models: 1.5B, 3.5B, 11B, 24B (?)
Context	4 096 tokens	32 768 tokens	+32 768 tokens
Pre-training	36 billion tokens x 2 epochs	198 billion tokens x 2 epochs	300/500 billion tokens (deduplicated)
Fine-tuning	2.5 mln instructions	20 mln instructions 72 k preferences	22 mln instructions 300 k preferences

Bielik-R (planowany)

Bielik-11B-R1-alpha ▾ +

Najwyższe z 3 drzew ma 108 stóp. Średnie drzewo jest o 6 stóp niższe niż połowa wysokości najwyższego drzewa. Najmniejsze drzewo ma $\frac{1}{4}$ wysokości średniego drzewa. Jak wysokie jest najmniejsze drzewo?



Bielik-11B-R1-alpha

Thinking... ^

1. Najwyższe drzewo ma 108 stóp.
2. Średnie drzewo jest o 6 stóp niższe niż połowa wysokości najwyższego drzewa.
 - Najpierw znajdziemy połowę wysokości najwyższego drzewa:
 $\text{Połowa wysokości} = 108 \text{ stóp} / 2 = 54 \text{ stopy}$
 - Teraz znajdziemy wysokość średniego drzewa:
 $\text{Wysokość średniego drzewa} = 54 \text{ stopy} - 6 \text{ stóp} = 48 \text{ stóp}$
3. Najmniejsze drzewo ma $\frac{1}{4}$ wysokości średniego drzewa.
 - Obliczmy wysokość najmniejszego drzewa:
Wysoko



07

Zastosowania



#Bielik - przykłady wdrożeń



Gaius Lex

182 Followers
1mo • Edited •

🎉 Z wielką dumą chcemy się pochwalić, że minął okrągły tydzień od kiedy firma Gaius-Lex jako pierwsza w Polsce wdrożyła produkcyjnie w pełni polski model językowy (LLM) Bielik, opracowany przez naszych kolegów i koleżanki z fundacji [SpeakLeash.org](#) (Spichlerz) oraz [Academic Computer Centre CYFRONET AGH](#)

Bielik był używany w ramach Gaius-Lex już w dniu jego wyjścia na rynek, kilka godzin po publikacji wagi i przeprowadzeniu przez nas niezależnych testów! 🎉 Dzięki temu mogliśmy od razu korzystać z jego zalet, znacząco podnosząc jakość naszego rozwiązania. Już teraz widzimy, że użytkownicy chętniej wracają do modułów w których główną rolę pełni Bielik. 🌟

Jesteśmy również dumni z faktu, że Gaius-Lex była pierwszą niezależną organizacją, która potwierdziła jakość tego nowatorskiego modelu za pomocą naszych wewnętrznych testów jakości. To wdrożenie pozwala nam dodatkowo ugruntować naszą pozycję lidera w obszarze narzędzi dla prawników opartych o AI.

Wielkie gratulacje dla zespołu Speakleash za ich niesamowitą pracę przy trenowaniu tego modelu. Wasze zaangażowanie i wiedza przyczyniły się do stworzenia czegoś naprawdę wyjątkowego. 🌟

Razem budujemy przyszłość opartą na polskiej myśli technologicznej! 🚀

Z #GaiusLex #Speakleash #BielikLLM #Innowacje #SztucznaInteligencja #PolskaTechnologia #AI #LegalTech



BuildEL

19 Followers
1mo • Edited •

Bielik - najnowszy polski model LLM już dostępny w BuildEL! 🚀

Bielik w ostatnich dniach podbił polską branżę AI, więc nie mogło go zabraknąć również w BuildEL. Dzięki naszemu narzędziu możesz nie tylko korzystać z Bielika, ale także zapewnić mu dostęp do aktualnych informacji dzięki integracji z Brave Search.

Wypróbuj BuildEL już dziś: <https://buildel.ai/>



Azurro

247 Followers
6d •

Cześć!

Dzięki Bielikowi - polskiemu modelowi językowemu - i zaawansowanym algorytmom (np. do identyfikacji osób, przedmiotów czy miejsc), stworzyliśmy narzędzie służące do kompleksowej analizy materiałów wideo. Chcecie wiedzieć, dlaczego wybraliśmy akurat Bielika? Sprawdźcie to czytając najnowszy artykuł na naszym blogu!

🔗 nasz blog: www.azurro.pl/blog-pl/

Social media:

📧 @azurropl

✉ @azurropl

📺 Azurro

📺 Azurro

📺 @azurro_pl

#LLMs #Bielik #nlptraining #MachineLearning #naturallanguageprocessing



IBPM S.A.

918 Followers
1mo • Edited •

Bielik.AI 2.2 11B z sukcesem zintegrowany z BlueBox AI 2.1!

Wykonano podstawowe testy jak Bielik sprawdza się w architekturze RAG - i działa dobrze! Udziela bardzo celnych odpowiedzi i jest szybki.

Technologicznie Bielik w obecnej wersji to rozwinięty w języku polskim Mistral 7B. Gratulujemy twórcom Bielika dobrych wyborów architektonicznych. Ceniliśmy sobie już Mistrala i w dalszym ciągu uważamy że są to wysokiej klasy LLM.

Jakie następne kroki? Testy testy testy. Sprawdźmy dokładnie jak Bielik poradzi sobie ze słownictwem prawniczym i medycznym. Jeśli wszystko pójdzie dobrze, Bielik będzie głównym modelem językowym on-prem dostępnym w BlueBox AI w rozwiązaniach na rynek Polski.

#bielik #bluebox #rag



Jan Ćwierk • 2nd

Doradczam w rozwiązaniach Security IBM Software | watso...
1mo •

Pierwszy polski LLM wylądował na platformie [IBM watsonx.ai](#) on-premise!

Z uwagą śledząc nowości oferowane przez platformę [IBM watsonx](#), postanowiłem zaktualizować lokalne środowisko działające w [Ingram Micro Sp. z o.o. \(Polska\)](#), aby móc przetestować wszystkie najnowsze funkcjonalności.

Jedną z najnowszych opcji, dostępną na platformie on-premise z kartami [NVIDIA L40s](#) jest BYOM - Bring Your Own Model. Dzięki niemu udało mi się uruchomić model Bielik stworzony przez [SpeakLeash.org](#) i [Academic Computer Centre CYFRONET AGH](#) na platformie [watsonx.ai](#).

To daje nowe możliwości dla firm, które chciałyby wdrożyć rozwiązania AI on-premise, oparte o w pełni polski model językowy.

Zachęcam partnerów biznesowych Ingram Micro do skorzystania z naszej infrastruktury i przetestowania możliwości, jakie oferuje watsonx. Razem możemy tworzyć rozwiązania, które będą napędzać innowacje.

#watsonx #IngramMicro #IBM #Bielik #LLM #SpeakLeash #AI



Piotr Gorzelany • 2nd

Senior iOS Engineer @ Indee Flex
1mo •

Chcecie wypróbować [#Bielika](#) od [SpeakLeash.org](#) i macie MacBooka? Nic prostrzeżo!

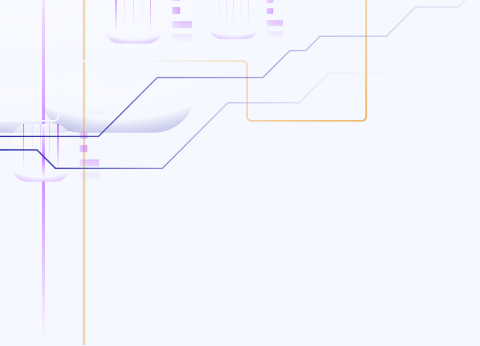
W Enclave dodałem [#Bielika](#) jako oficjalnie wspierany model. Ściągacie aplikację, wybieracie Bielika, czekacie aż się ściągnie i gotowe. Od tej pory możecie cieszyć się naszym Polskim modelem zupełnie za darmo. Aplikacja nie wymaga rejestracji / logowania i nie śledzi żadnych waszych danych. Wszystko jest absolutnie prywatne i działa lokalnie na waszym laptopie.

<https://lnkd.in/dUAYNUWF>

Dajcie znać czy [#Bielik](#) spełnił wasze oczekiwania i jaki jest performance aplikacji na waszym sprzęcie. Wszelkie uwagi są dla mnie bardzo cenne.

Dlaczego ludzie wybierają Bielika

- **Bezpieczeństwo** (brak przekazywania danych).
- **Niskie koszty** użytkowania.
- **Szybkość + brak limitów** (jedynym ograniczeniem jest moc infrastruktury).
- **Elastyczność** (możliwość stosowania nawet dla “trudnej” treści).
- **Łatwość adaptacji** - możliwy dowolny finetuning modelu.



Dziękuję za uwagę

