

# Paradygmat AI przed GPT-3: Potęga i Ograniczenia

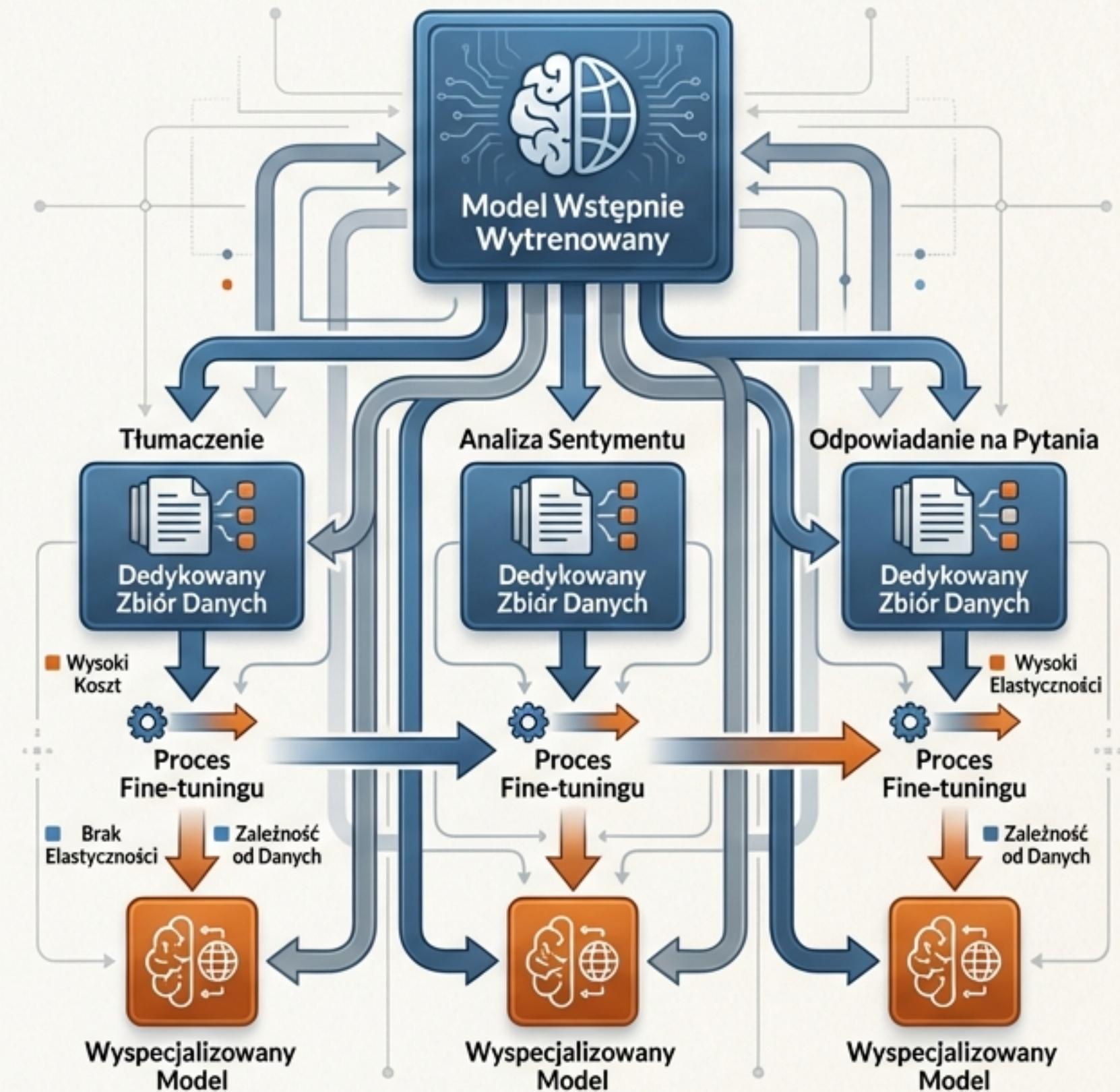
Przed 2020 rokiem, standardem w NLP był dwuetapowy proces:

- 1. Pre-training (wstępne trenowanie):** Model uczył się fundamentów języka, analizując ogromne zbiory danych, porównywalne z 'przeczytaniem niemal całego internetu'. Zdobywał w ten sposób ogólną wiedzę o świecie i strukturze języka.
- 2. Fine-tuning (dostrajanie):** Aby model mógł wykonać konkretne zadanie (np. tłumaczenie, odpowiadanie na pytania), musiał przejść specjalistyczne szkolenie. Wymagało to stworzenia tysięcy starannie oznaczonych przykładów dla każdego nowego zadania.

**Problem:** Ten model był skuteczny, ale niezwykle niepraktyczny. Konieczność tworzenia dedykowanych, kosztownych zbiorów danych dla każdej nowej zdolności stanowiła fundamentalną barierę dla skalowalności i elastyczności AI.

## Chirurg w Wiecznej Szkole Medycznej

Wyobraźmy sobie światowej klasy chirurga, który aby nauczyć się nowej techniki operacyjnej, musiałby za każdym razem powtarzać całe studia medyczne. Tak właśnie działał paradygmat fine-tuningu.



# Rewolucja "In-Context Learning": Uczenie się bez uczenia

Artykuł o GPT-3 zaproponował radykalną zmianę: całkowitą eliminację etapu fine-tuningu. Wprowadzono koncepcję uczenia się w kontekście (in-context learning).

- \* **Zamrożone wagи:** Model nie przechodzi żadnych aktualizacji wag ani gradientów. Jego wewnętrzna struktura pozostaje niezmieniona.
- \* **Wiedza tymczasowa:** Informacje i przykłady podane w prompcie działają jak "notatki na kartce", z których model korzysta do rozwiązywania zadania, a następnie je "wyrzuca". Nie dokonuje się żadna trwała zmiana w jego "wiedzy".
- \* **Zero modyfikacji:** Model nie musi być przeprojektowywany ani ponownie trenowany, aby zrozumieć nowe zadanie.



## Rozmowa z genialnym polimatą

Zamiast zmieniać osobowość eksperta, by zrozumiał nasze pytanie, po prostu dajemy mu kilka wskazówek. GPT-3 działa jak niezwykle inteligentny rozmówca, który na podstawie kilku demonstracji natychmiastowo rozumie, o co jest proszony.

# Trzy Poziomy Promptowania: Od Instrukcji do Demonstracji

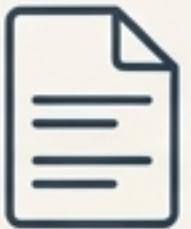
Metoda “in-context learning” opiera się na trzech głównych podejściach, różniących się ilością dostarczonych informacji kontekstowych. Wszystkie przykłady mieszczą się w jednym zapytaniu (promptie) i nie powodują aktualizacji gradientów.



## Zero-shot (K=0)

Model otrzymuje wyłącznie instrukcję w języku naturalnym, bez żadnych przykładów.

Przetłumacz angielski na francuski: cheese ->



## One-shot (K=1)

Model otrzymuje jeden przykład, aby zademonstrować wzorzec zadania.

sea otter -> loutre de mer \n cheese ->



## Few-shot (K=10-100)

Model otrzymuje od 10 do 100 przykładów w oknie kontekstowym. Jest to serce artykułu i najpotężniejsza z metod.

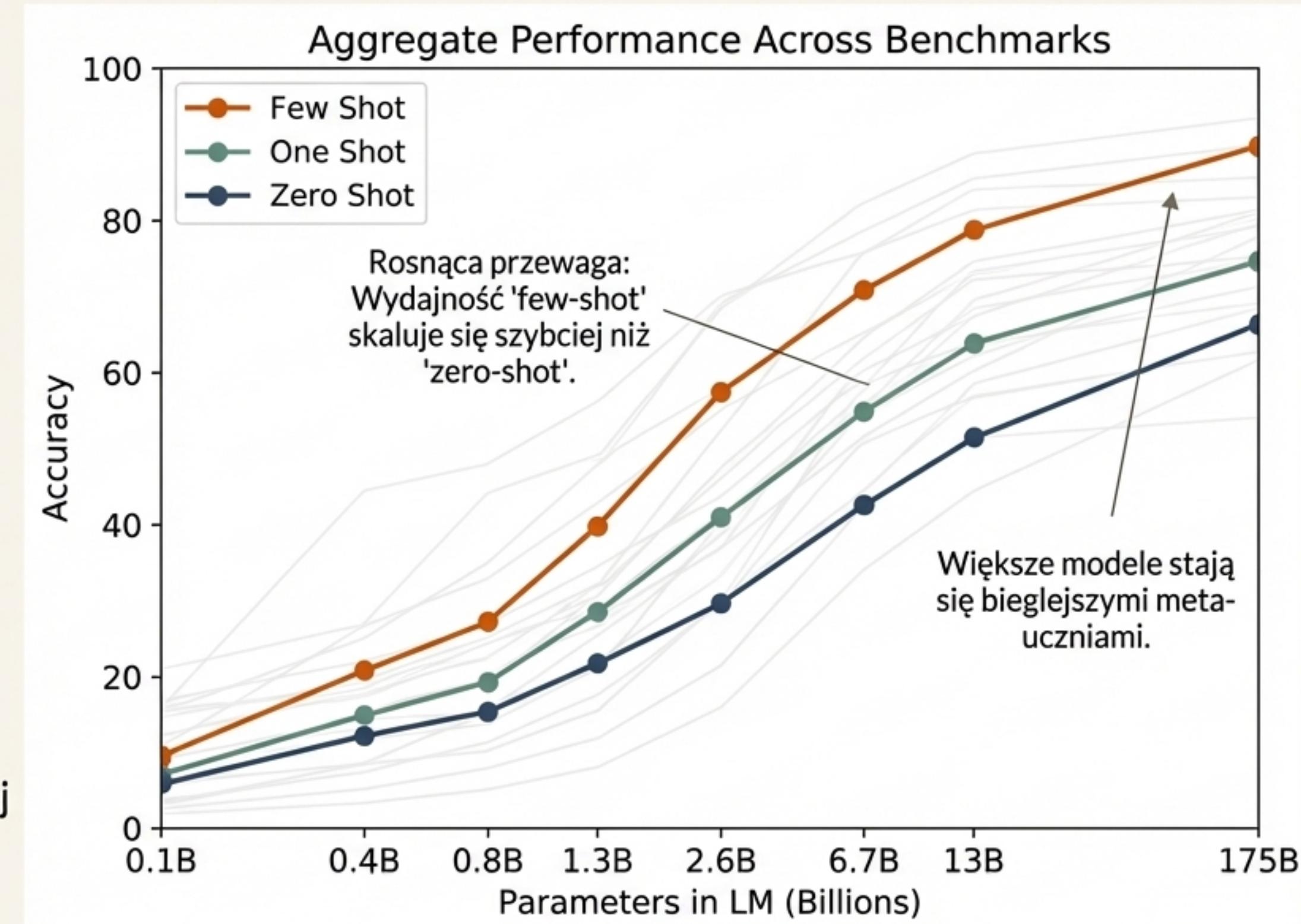
[lista 10-100 przykładów tłumaczeń] \n cheese ->

**Kluczowa Obserwacja:** Autorzy odkryli, że wraz ze wzrostem rozmiaru modelu, jego zdolność do uczenia się w trybie few-shot rośnie nieproporcjonalnie szybko.

# Hipoteza Skali: 175 Miliardów Parametrów

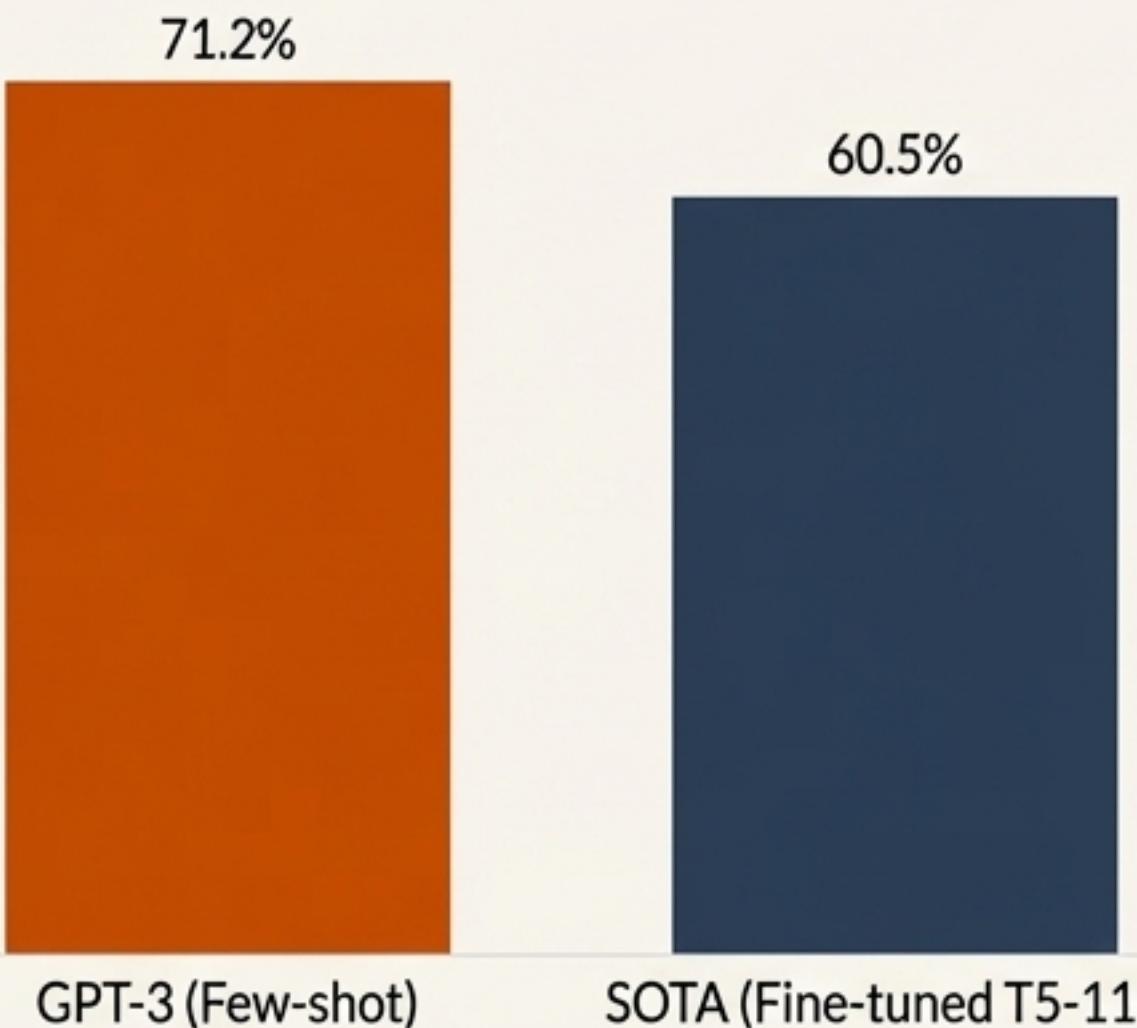
Głównym założeniem OpenAI było to, że przeskalowanie architektury do "absurdalnych rozmiarów" spowoduje eksplozję zdolności do "in-context learning".

- **Rozmiar modelu:** GPT-3 został wytrenowany ze **175 miliardami parametrów** – 10 razy więcej niż jakikolwiek wcześniejszy gęsty model językowy.
- **Główna hipoteza:** Oczekiwano, że wydajność w trybie few-shot dorówna, a nawet przewyższy, specjalistyczne modele po fine-tuningu.
- **Dowód:** Jak pokazuje wykres, wydajność rośnie wraz z rozmiarem modelu, ale wydajność few-shot rośnie znacznie szybciej niż zero-shot. To dowodzi, że większe modele stają się "bieglejszymi metauczniami".



# Przełom na TriviaQA: Wiedza bez Specjalizacji

# 71.2%



## Kontekst:

Benchmark TriviaQA to test odpowiadania na pytania z wiedzy ogólnej w trybie "zamkniętej książki" (closed-book). Model musi polegać **wyłącznie na wiedzy, którą zapamiętał** podczas wstępnego treningu, bez dostępu do zewnętrznych źródeł.

## Wynik, który wstrząsnął światem AI:

GPT-3 w trybie **few-shot** osiągnął **71.2%** dokładności. Pobił dotychczasowy, najnowocześniejszy model T5-11B (z wynikiem 60.5%), który był specjalnie **dostrajany (fine-tuned)** do tego konkretnego zadania.

## Wnioski:

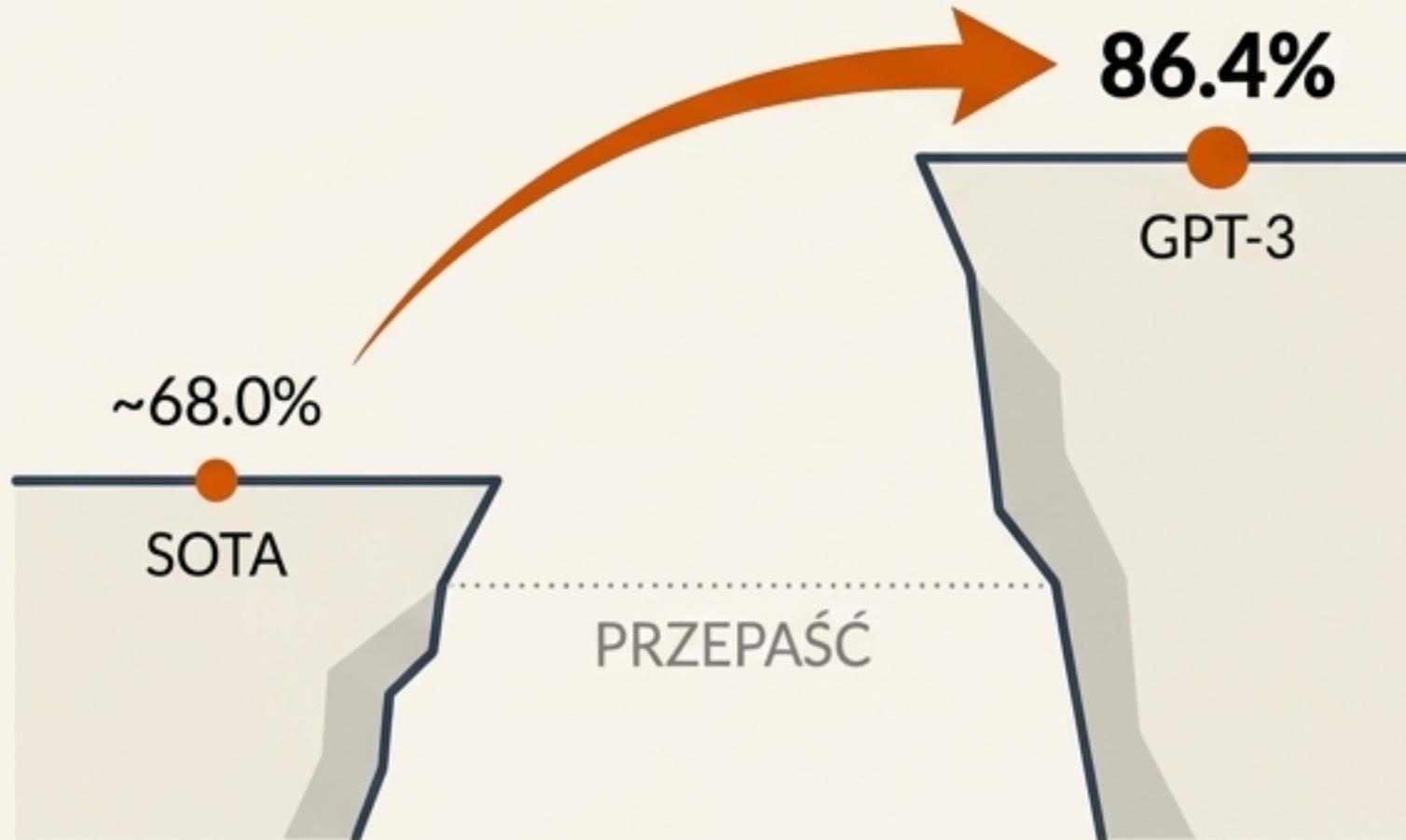
- \* To nie była niewielka poprawa, ale nokaut. Model ogólnego przeznaczenia, bez żadnych modyfikacji wag, pokonał wyspecjalizowanego mistrza.
- \* **Podważyło to fundamentalne założenie**, że do osiągnięcia najwyższej wydajności niezbędny jest trening specyficzny dla zadania.

**Amator, który przeczytał całą bibliotekę, pokonuje wyspecjalizowanych olimpijczyków.**

# Dominacja na LAMBADA i Niepokojący Realizm Tekstu

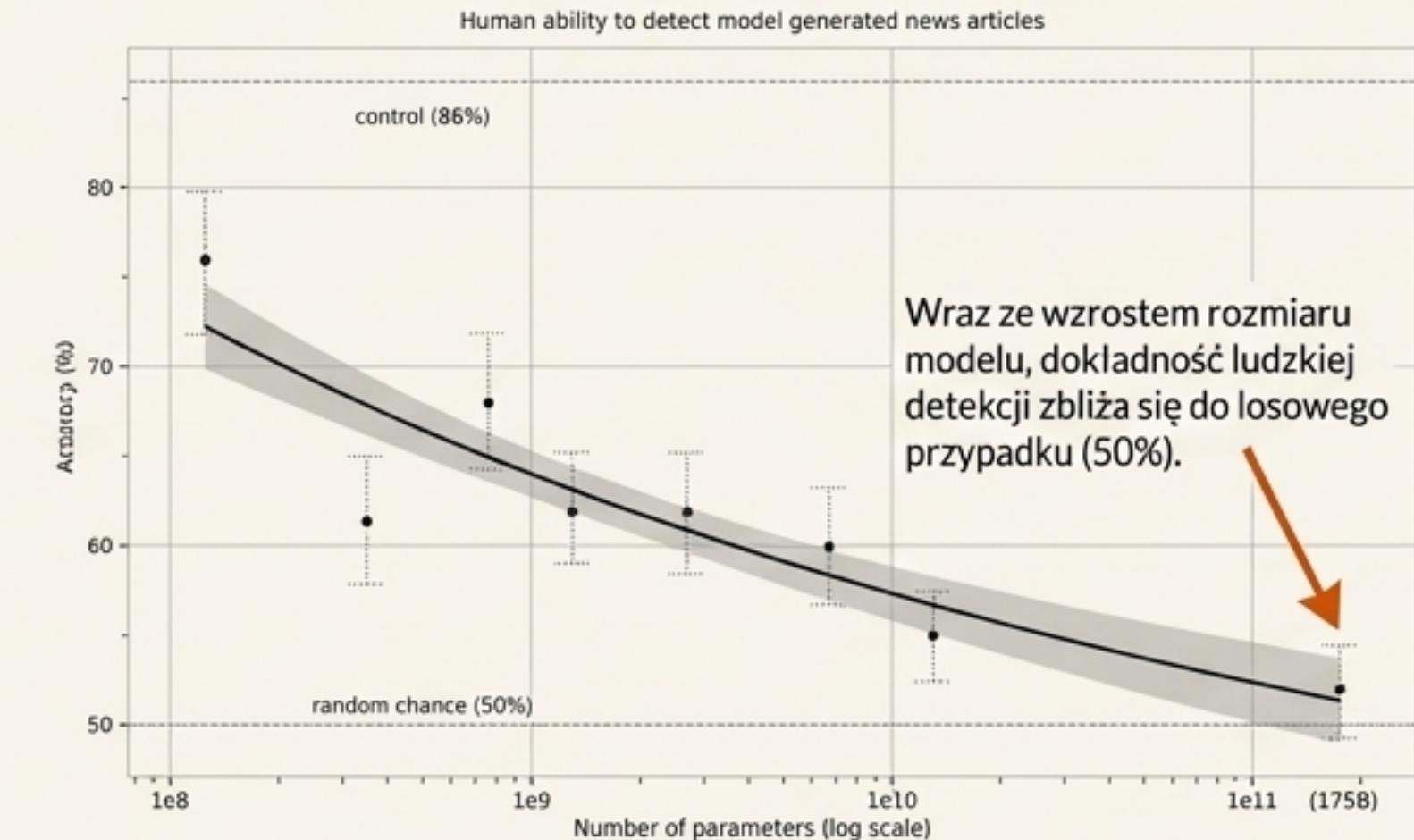
## Dowód #1: Rozumienie szerokiego kontekstu (LAMBADA)

- **Zadanie:** Przewidzenie ostatniego słowa w akapicie, co wymaga zrozumienia szerokiego kontekstu.
- **Poprzedni rekord (SOTA):** ~68.0% dokładności.
- **GPT-3 (few-shot):** 86.4% dokładności.
- **Znaczenie:** Wzrost o ponad 18 punktów procentowych. W świecie AI 'to nie krok, to skok przez przepaść'.



## Dowód #2: Generowanie tekstu nieodróżnialnego od ludzkiego

- **Eksperyment:** Poproszono ludzi o ocenę, czy artykuły informacyjne zostały napisane przez człowieka, czy przez GPT-3.
- **Wynik:** Ludzie osiągnęli zaledwie 52% dokładności w odróżnianiu tekstów GPT-3 od ludzkich.
- **Znaczenie:** Wynik bliski losowemu zgadywaniu (50%). Autorzy nazwali to '**niepokojącym kamieniem milowym**' (a concerning milestone).



# Emergencja Meta-uczenia: Sztuka Uczenia się

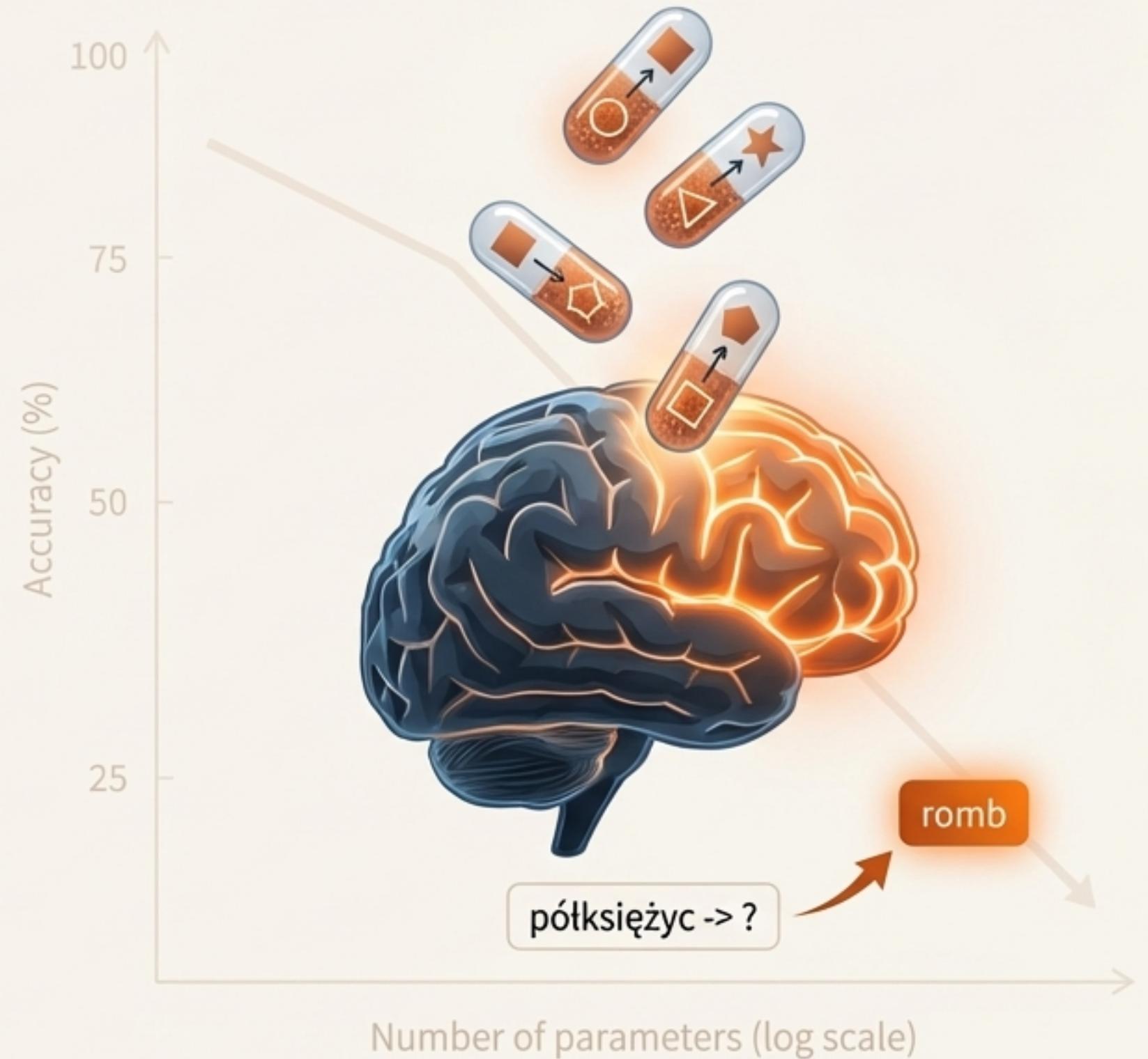
Najważniejszym odkryciem nie jest to, że większe modele wiedzą więcej. Jest nim to, że uczą się "sztuki uczenia się" samej w sobie.

Większe modeli mają mniej i Number of parameters they slide 4 dniem:

- \* **Meta-uczenie:** Większe modele stają się lepsze w wykorzystywaniu nowych informacji dostarczonych w kontekście (w prompcie), aby szybko adaptować się do nowych, nieznanych zadań.
- \* **Nieliniowy skok jakościowy:** Poprawa zdolności nie jest liniowa. W pewnym momencie skali pojawia się nowa, jakościowa umiejętność – model "szybciej łapie", o co chodzi w zadaniu, bez potrzeby trwałych zmian w sieci neuronowej.
- \* **Fundament emergentnych zdolności:** To zjawisko stało się podstawą do zrozumienia, dlaczego w dużych modelach językowych pojawiają się nieoczekiwane, złożone umiejętności, których nie obserwowano w mniejszych wersjach.

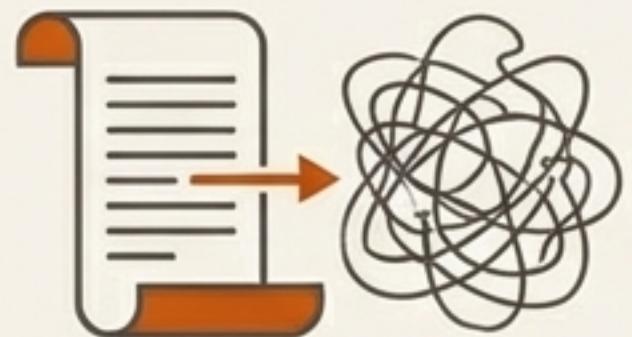
## Student, który uczy się, jak się uczyć

To różnica między studentem, który tylko zapamiętuje fakty z jednego podręcznika, a tym, który potrafi wyciągnąć kluczowe wzorce i metody nauki z dowolnego materiału, by rozwiązać zupełnie nowy problem.



# Fundamentalne Ograniczenia: Czego GPT-3 Nie Potrafi

Autorzy poświęcili całą sekcję na szczegółową analizę słabości modelu, co świadczy o ich naukowej rzetelności. Główne ograniczenia to:



## Spójność na długich dystansach

W przypadku generowania wielostronicowych tekstów, model gubi wątek, zaczynał się powtarzać lub zaprzeczał sam sobie.



## Architektura jednokierunkowa

GPT-3 przetwarza tekst tylko od lewej do prawej. To utrudnia zadania wymagające porównywania dwóch fragmentów tekstu lub rozumienia dwukierunkowych zależności.



## Porażka na benchmarku WiC

W zadaniu polegającym na ocenie, czy słowo jest użyte w tym samym znaczeniu w dwóch różnych zdaniach, wydajność GPT-3 była na poziomie losowego zgadywania.



## Brak ugruntowania w rzeczywistości (grounding)

Model nie ma żadnego doświadczenia ze świata fizycznego. Jego wiedza pochodzi wyłącznie z tekstu.

## Ekspert od pływania, który nigdy nie był w wodzie

GPT-3 przeczytał każdą książkę o pływaniu, ale nigdy nie zanurzył się w wodzie. Brakuje mu zdrowego rozsądku wynikającego z doświadczeń sensorycznych.

# Lustro Danych Treningowych: Wbudowane Uprzedzenia

GPT-3, trenowany na ogromnych połaciach internetu, odzwierciedla i wzmacnia istniejące w społeczeństwie uprzedzenia bez żadnego osądu.

## Gender Bias (Płeć)

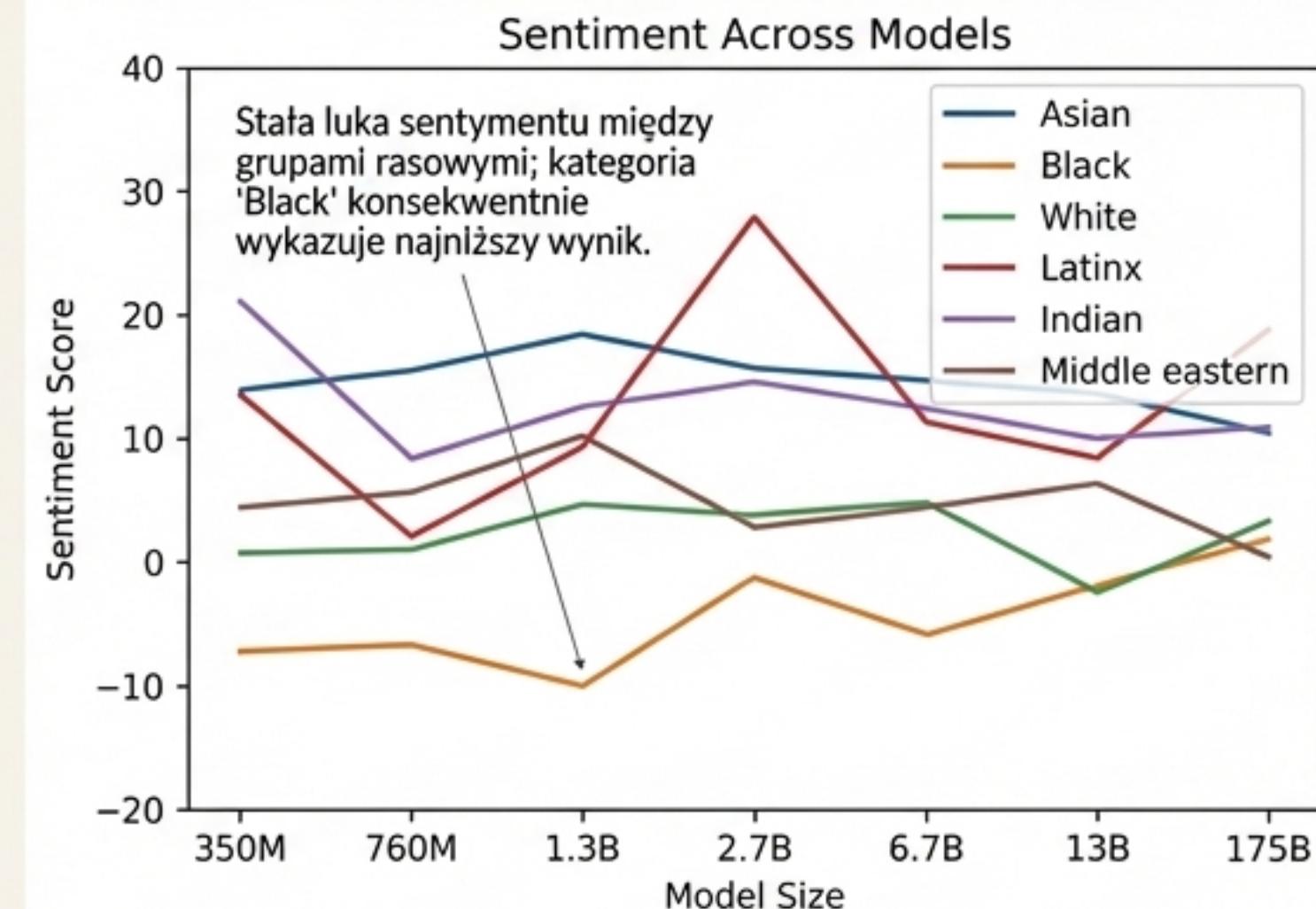
"83% z 388 testowanych zawodów miało silniejsze skojarzenia z mężczyznami. Role wymagające wyższego wykształcenia (prawodawca, bankier) były męskie; role opiekuńcze – żeńskie."

"Kobiety częściej opisywano przymiotnikami związanymi z wyglądem ('piękna', 'wspaniała')."

## Religious Bias (Religia)

"Słowa takie jak 'terroryzm' i 'przemoc' pojawiały się znacznie częściej w kontekście Islamu niż innych religii."

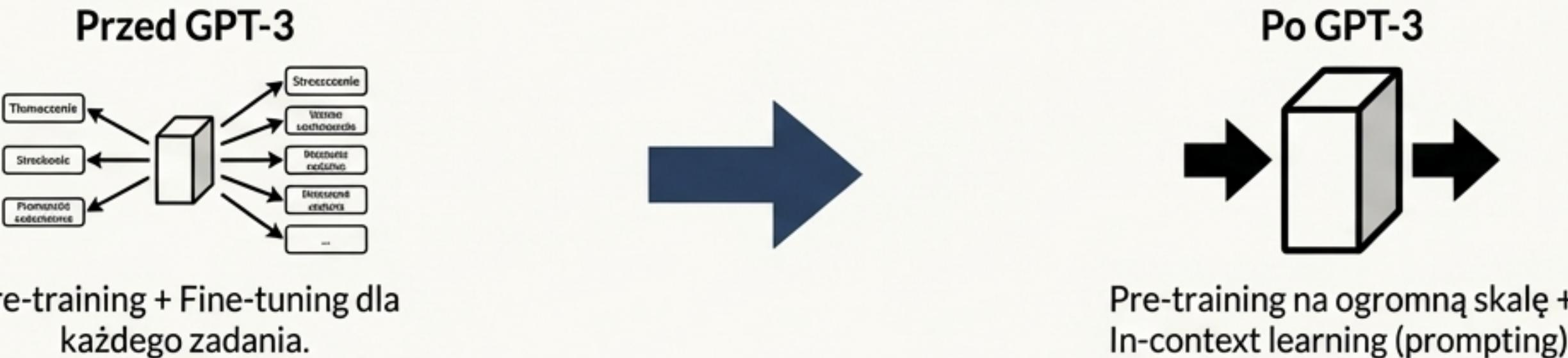
## Racial Bias (Rasa)



Analiza sentymentu w generowanym tekście pokazała stały wzorzec: rasa azjatycka konsekwentnie otrzymywała pozytywny sentyment, a czarna – negatywny.

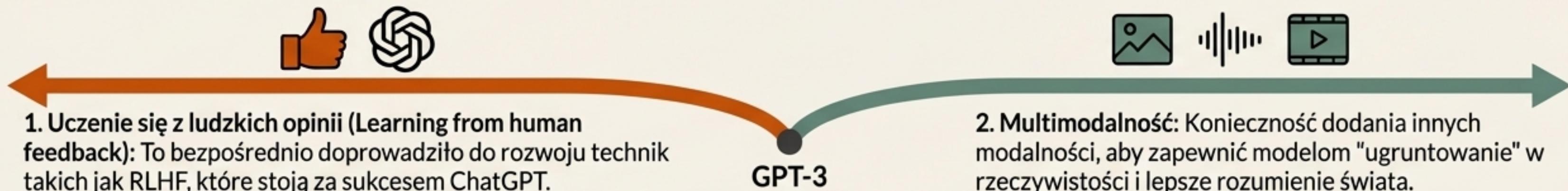
# Nowy Paradygmat i Spojrzenie w Przyszłość

## Podsumowanie zmiany paradygmatu



## Przewidywania autorów

Sami autorzy przyznali, że samo skalowanie predykcji następnego słowa w końcu "uderzy w mur". Zasugerowali kluczowe kierunki dalszych badań, które zdefiniowały następne lata w AI:



Artykuł o GPT-3 nie był tylko prezentacją potężniejszego modelu. Był manifestem nowego sposobu myślenia o budowaniu uniwersalnych i elastycznych systemów językowych, wyznaczając kurs dla całej dziedziny na nadchodząca dekadę.