

Złudzenie Skalowania: Rozbite

Jak nagłe „przebłyski geniuszu” w Wielkich Modelach Językowych zmieniają nasze rozumienie postępu w AI.



Czym są Zdolności Wyłaniające się?

To nie jest stopniowa poprawa. To skokowa zmiana, która musi spełniać dwa rygorystyczne warunki.

Definicja:

Zdolność jest „wyłaniająca się”, jeśli:

- **Warunek 1:** Jest praktycznie nieobecna w mniejszych modelach.
- **Warunek 2:** Pojawia się i staje się mierzalna dopiero w większych modelach.



TAK

NIE



Skala Modelu (np. FLOPs)

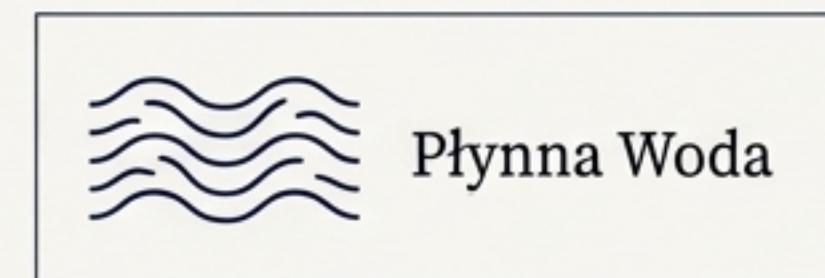
Przejście Fazowe: Gdy Fizyka Spotyka AI

To zjawisko ma swój odpowiednik w fizyce. To nie przypadek, to cecha złożonych systemów.

„Emergence is when quantitative changes in a system result in qualitative changes in behavior.”

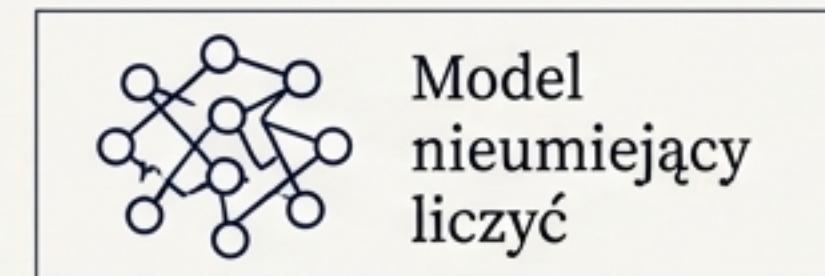
– Philip Anderson, „More is Different”, 1972

PHYSICS ANALOGY



Woda nie staje się ‘trocę bardziej lodowata’ – nagle zamarza.

AI ANALOGY

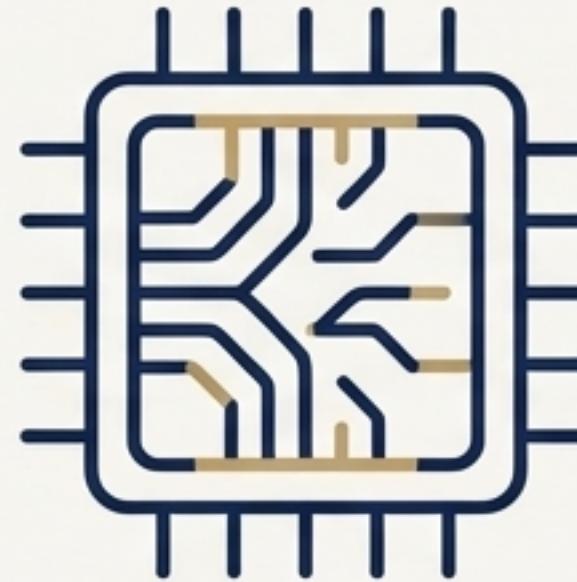


Modele nie stają się ‘trocę lepsze w arytmetyce’ – nagle potrafią liczyć.

Główna idea: Zmiana **ilościowa** (więcej parametrów) prowadzi do zmiany **jakościowej** (nowe zdolności).

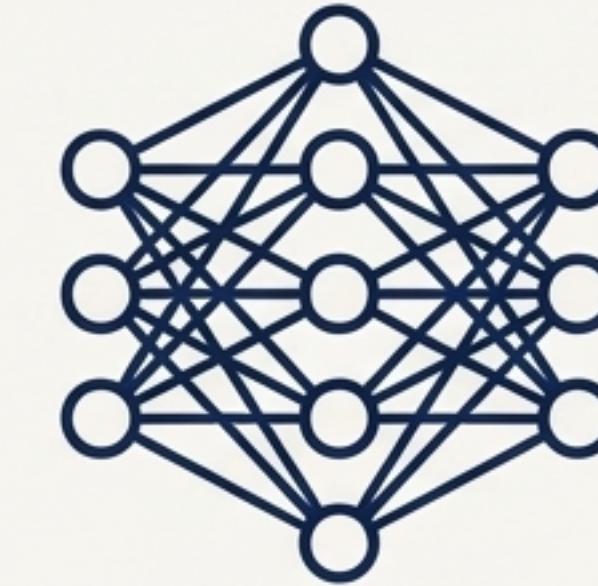
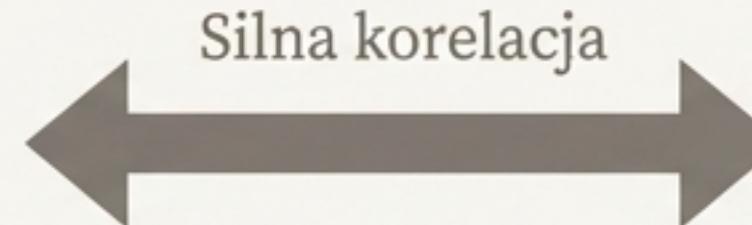
Mierzenie Skali: Oś X Wyłaniania się

„Większy” to konkretne, mierzalne wartości, które napędzają tę rewolucję.



Training FLOPs

Całkowita moc obliczeniowa (Floating Point Operations per Second) zużyta podczas treningu.



Liczba Parametrów

Miliardy (a nawet biliony) wag w sieci neuronowej, które są dostosowywane w procesie uczenia.

Ważne uwagi

- Modele z większą liczbą parametrów zazwyczaj wymagają proporcjonalnie więcej mocy obliczeniowej do treningu.
- Krytyczny próg, przy którym pojawia się dana zdolność, jest różny dla różnych zadań i architektur modeli.

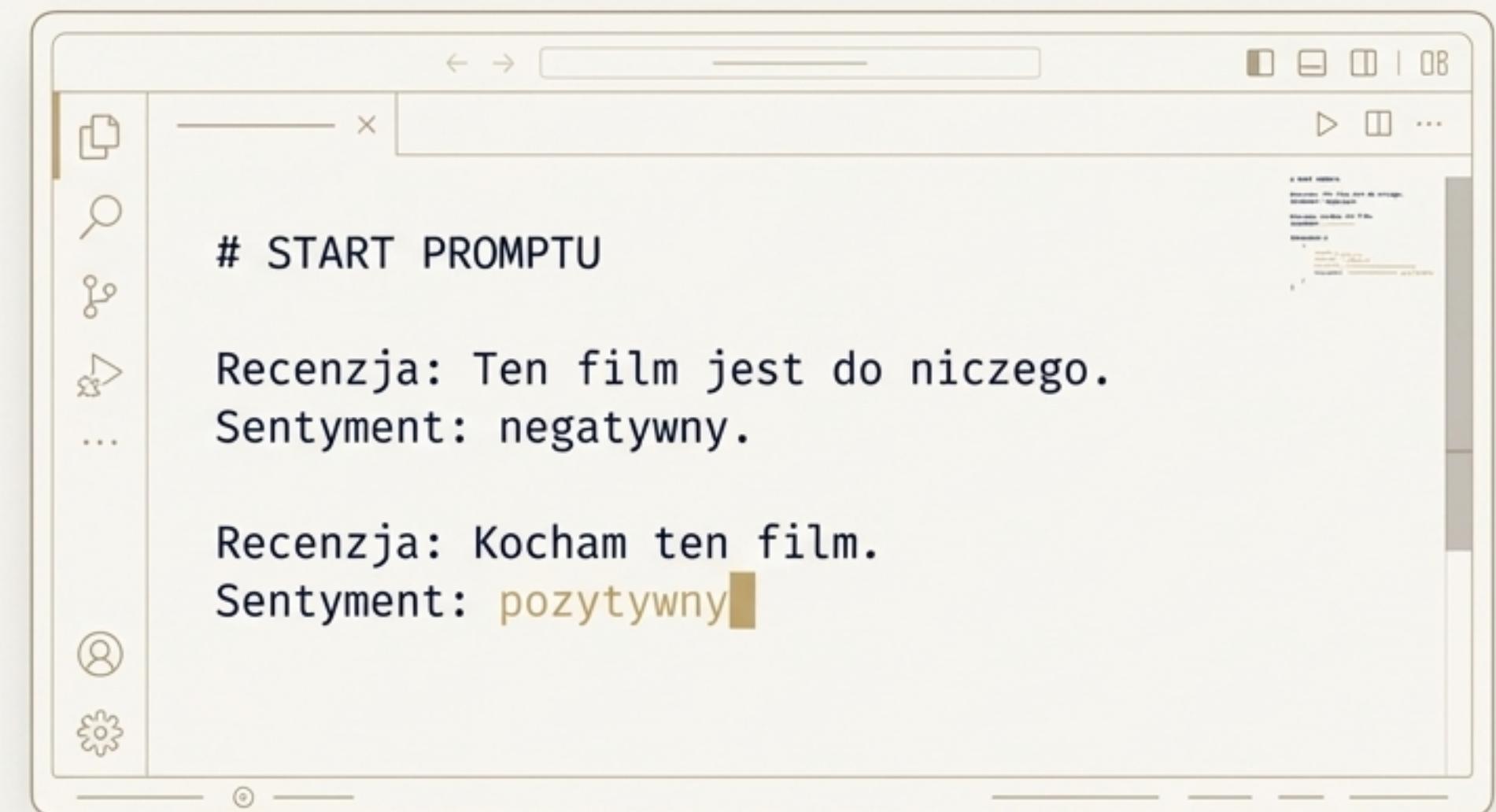
Few-Shot Prompting: Protokół Testowy

Tak sprawdzamy, czy model „nauczył się” czegoś nowego bez dodatkowego treningu.

Definicja

Technika, w której modelowi podaje się w prompcie kilka przykładów zadania (wejście → wyjście), a następnie prosi się go o rozwiązanie nowego, niewidzianego wcześniej przykładu.

Żadnych aktualizacji wag ani fine-tuningu.
Model musi sam wywnioskować wzorzec z podanych przykładów.



Kluczowa uwaga:

Sama zdolność do efektywnego uczenia się w ten sposób ([in-context learning](#)) jest również zdolnością wyłaniającą się – działa znacznie lepiej w modelach o dużej skali.

Wyłanianie się Arytmetyki: Liczby Nie Kłamią

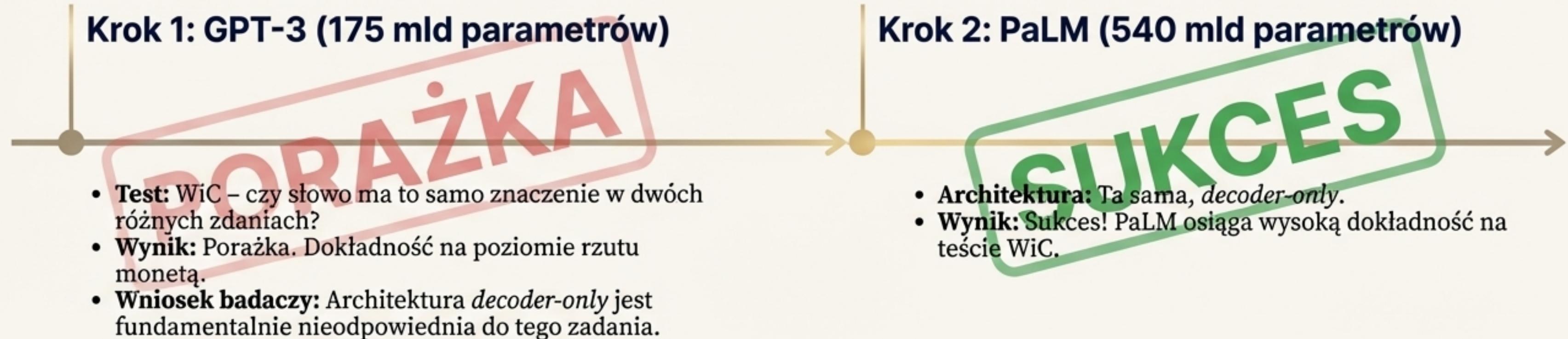
Zdolność do prostego dodawania pojawia się znikąd przy określonym progu skali.



Wniosek: Model z 10 mld parametrów nie jest „trochę gorszy z matematyki” – on w ogóle jej nie potrafi.

Word in Context (WiC): Historia Detektywistyczna

Nawet eksperci byli w błędzie. Myśleli, że problemem jest architektura, a problemem była niewystarczająca skala.



Morał:

Problem nie leżał w architekturze, lecz w niewystarczającej skali. To, co wydawało się fundamentalnym ograniczeniem, było jedynie nieprzekroczonem progiem wyłonienia się zdolności.

Wyłaniające się Promptowanie: Łańcuch Myśli

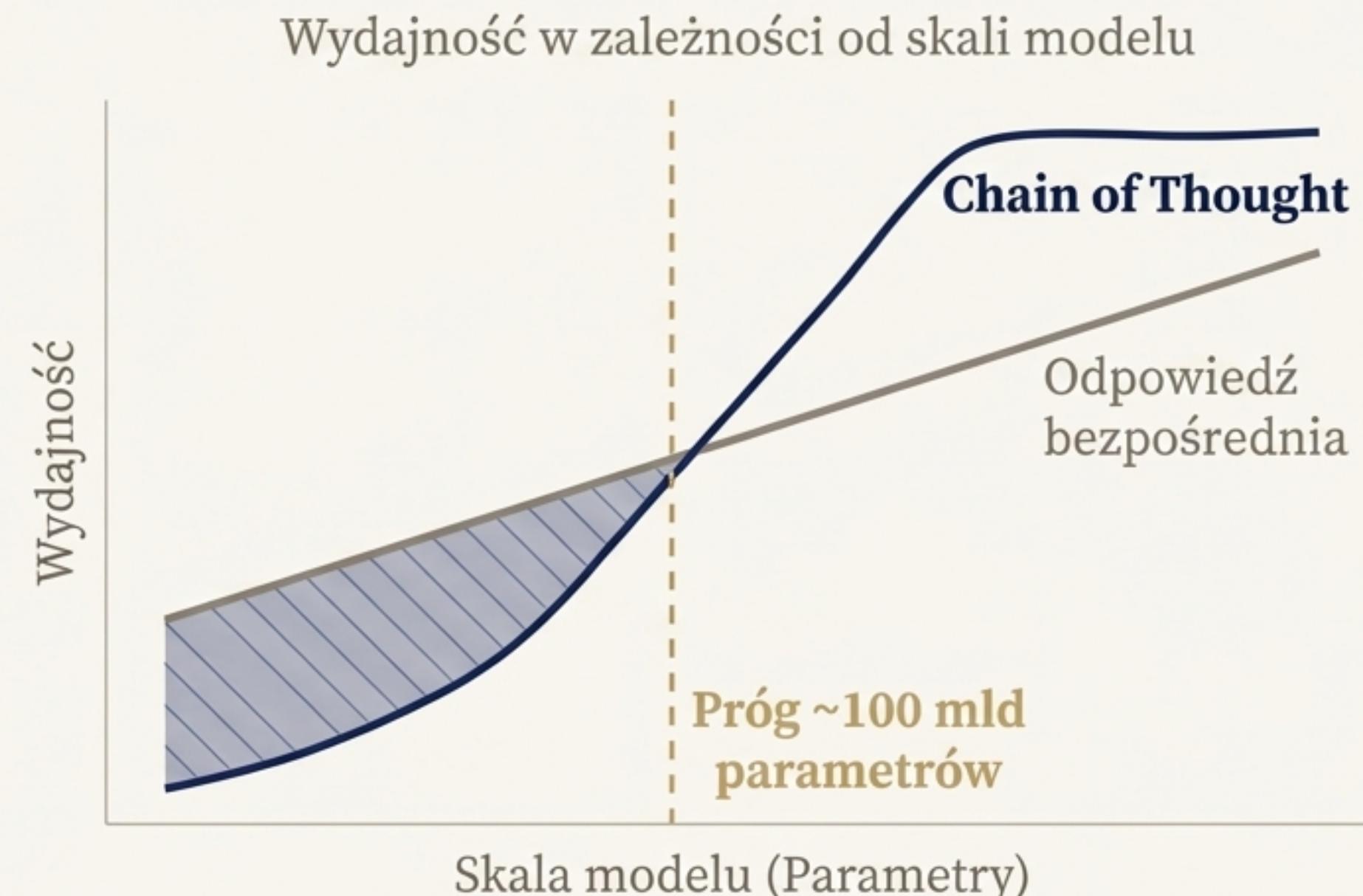
Techniki interakcji z modelami, które uważamy za kluczowe, są bezużyteczne lub nawet szkodliwe dla mniejszych modeli.

Chain-of-Thought (CoT) Prompting

Technika polegająca na poleceniu modelowi, aby rozpisał swoje rozumowanie krok po kroku, zanim poda ostateczną odpowiedź.

Szokujące odkrycie:

- Technika CoT daje **ZERO korzyści** w modelach poniżej progu **~100 mld parametrów**.
- Poniżej tego progu, CoT może nawet **pogorszyć wyniki**.



Paradoks Dostrajania do Instrukcji

Nasza intuicja zawodzi. To, co powinno pomagać, może szkodzić, ujawniając twardo ograniczenia mniejszych modeli.

Proces: Instruction-Finetuning

Technika trenowania modelu na zestawie zadań sformułowanych jako polecenia, aby lepiej generalizował i podążał za instrukcjami użytkownika.

Kontrintuicyjna Rzeczywistość:



Technika **POGARSZA** wydajność w modelach poniżej ~8 mld parametrów. Staje się **KORZYSTNA** dopiero przy ~100 mld parametrów.

Analogie i Hipotezy



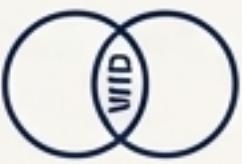
"To jakby uczenie dziecka czytania sprawiło, że zapomniało liter."

Możliwe przyczyny:

- Niewystarczająca „głębokość” obliczeniowa.
- Ograniczona pojemność reprezentacji w mniejszych modelach.

Ekosystem Wyłaniających się Zdolności

To nie jest kilka odosobnionych przypadków. To szerokie i zróżnicowane zjawisko obejmujące wiele dziedzin.

IKONA	ZDOLNOŚĆ	MODEL	PRÓG (PARAMETRY)
	Dodawanie 3-cyfrowe	GPT-3	~13 mld
	Rozumowanie wieloetapowe (CoT)	LaMDA	~68 mld
	Podążanie za instrukcjami	FLAN	~68 mld
	Prawdomówność (TruthfulQA)	Gopher	~280 mld
	Kontekst słowa (WiC)	PaLM	~540 mld

Więcej Znaczy Inaczej: Nowa Granica AI

Zrozumienie wyłaniania się zdolności zmienia skalowanie z przewidywalnej inżynierii w ekscytującą podróż odkrywczą – z ogromnym potencjałem i nowymi ryzykami.

Nowy paradygmat: Skalowanie to nie tylko „więcej tego samego”. To mechanizm odblokowujący jakościowo nowe, nieprzewidywalne możliwości.

Wielkie pytania: Jakie zdolności czekają za kolejnym progiem skali? Czy modele mogą nabyć zdolności do zaawansowanego rozumowania abstrakcyjnego lub kreatywności?

Wyłaniające się Ryzyka: Nowe, nieprzewidziane zdolności oznaczają nowe, nieprzewidziane zagrożenia. Musimy podchodzić do skalowania nie tylko jak inżynierowie, ale jak odkrywcy i etycy.



Terra Incognita