

Brief Zwolnienia — sześćostronicowe streszczenie apelu Anthropic

Anthropic opublikował w tym tygodniu coś, co powinno dać do myślenia każdemu liderowi biznesu. To jest wersja skondensowana. Sześć stron. Kluczowe liczby, trzy scenariusze przyszłości, punkt styku z unijnym AI Act i lista działań na dziewięćdziesiąt dni. Podziel się z dyrektorem technicznym, z zarządem, z osobą odpowiedzialną za compliance — albo zatrzymaj dla siebie, kiedy znajdziesz dziesięć cichych minut.

Brief wykonawczy

Ład w AI

~10 min czytania

v1 · 2026-06-06

SEKCJA · 01 · TEZA

01. Co właściwie powiedział Anthropic

Przez większość historii AI to ludzie wykonywali pracę rozwojową — pisali kod, testowali, projektowali eksperymenty, debugowali, budowali infrastrukturę, trenowali modele, interpretowali wyniki. Anthropic publicznie twierdzi, że to się zmienia.

Claude wykonuje dziś coraz większą część tej pracy wewnątrz samego Anthropic. Możliwość, którą wskazują na przyszłość, to system AI zdolny **samodzielnie zaprojektować, zbudować, przetestować, wytrenować i ulepszyć swojego następcę**. To jest rekursywne samodoskonalenie. Są przy tym ostrożni: *twierdzą, że jeszcze do tego nie doszło i że nie jest to nieuniknione, ale może nadejść szybciej, niż instytucje są w stanie się przygotować.*

JEDNO ZDANIE DO ZAPAMIĘTANIA

AI nie jest już budowane wyłącznie przez ludzi. AI w coraz większym stopniu pomaga budować kolejne AI. Pętla zaczyna się zamykać. Instytucje nie są gotowe na to, co stanie się wtedy, kiedy zamknie się w pełni.

02. Co rzeczywiście zmierzyl

Trzy liczby niosą większość ciężaru argumentacji w eseju Anthropic. Każda z nich ma zastrzeżenie, do którego autorzy uczciwie się przyznają; razem opisują krzywą, której trudno nie zauważyć.

80%+

SCALONEGO KODU ANTHROPIC

Tyle kodu wprowadzonego do bazy Anthropic w maju 2026 zostało napisane przez Claude'a.

8x

WYDAJNOŚĆ INŻYNIERSKA

Inżynierowie Anthropic dostarczają w jednym kwartale około 8x więcej kodu niż w latach 2021–2025 (liczba linii kodu zawiąza realną produktywność, ale sam trend jest faktyczny).

4x

WZROST ZGŁASZANY PRZEZ BADACZY

Mediana samooceny w wewnętrznej ankiecie z marca 2026 wśród 130 badaczy korzystających z Mythos Preview.

03. Podwajanie długości zadań — wizualnie

Najbardziej uderzający trend, który opisuje Anthropic, to długość zadań, jakie modele AI są w stanie wykonać niezawodnie — **podwaja się ona mniej więcej co cztery miesiące**. Poniższy wykres jest najprostszym sposobem, żeby to zobaczyć.

DŁUGOŚĆ ZADAŃ, KTÓRE AI POTRAFI WYKONAĆ NIEZAWODNIE



Istotny jest sam kształt tej krzywej. Nawet jeśli skrócimy dwa ostatnie słupki o połowę, kierunek pozostaje ten sam: *jednostki pracy, jakie AI potrafi wykonać bez nadzoru, robią się coraz większe — i to szybko.*

04. Od asystenta do pracownika — i przesuujące się wąskie gardło

Zmiana operacyjna, którą nazywa Anthropic, to przejście od **asystenta** do **pracownika**. Claude jest dziś w stanie wziąć niedoprecyzowany problem inżynierski — cel bez metody — i wypracować rozwiązanie. Człowiek wyznacza cel; AI dobiera sposób.

Konsekwencją jest to, że wąskie gardło się przesuwa. Kiedy AI pisze kod szybciej, niż ludzie są w stanie go przejrzeć, wąskim gardłem staje się **przegląd**. Kiedy AI proponuje więcej eksperymentów, niż badacze są w stanie ułożyć w priorytety, wąskim gardłem staje się **wybór tego, który eksperyment naprawdę ma znaczenie**. To prawo Amdahla zastosowane do organizacji.

PRAKTYCZNA KONSEKWENCJA DLA TWOJEJ ORGANIZACJI

Firmy nie potrzebują wyłącznie narzędzi AI. Potrzebują **modeli operacyjnych pracy przyspieszonej przez AI** — świadomych decyzji o tym, gdzie człowiek dokonuje przeglądu, co wybiera i kiedy eskaluje. Jeśli świadomie nie zaprojektujesz nowego wąskiego gardła, ono samo się ułoży — i zwykle źle.

05. Co Anthropic widzi dalej — trzy scenariusze

Najbardziej uczciwa część eseju szkicuje trzy możliwości i nie udaje, że wiadomo, która z nich nadejdzie.

SCENARIUSZ A

Trend wyhamowuje

Możliwości układają się w krzywą S. Moc obliczeniowa, energia albo jakaś nowa bariera daje znać o sobie. Nawet wtedy — samo dzisiejsze AI wystarczy, żeby istotnie zmienić gospodarkę.

SCENARIUSZ B

Wzrost składany, sterowany przez ludzi

Małe zespoły wykonują pracę dużych organizacji. Wspaniale dla produktywności. Niebezpiecznie dla nadzoru, manipulacji, nadużyć w cyber, masowych operacji wpływu. Ten sam silnik — obie strony jednocześnie.

SCENARIUSZ C

Rekursywne samodoskonalenie

Systemy AI budują swoich następców. Rola człowieka kurczy się do nadzoru, walidacji, weryfikacji. Najtrudniejszy do przewidzenia. Największe ryzyko niedopasowania (alignment), jeśli rozjazdy kumulują się przez kolejne pokolenia systemów.

Anthropic nie twierdzi, że wie, który scenariusz się ziszczy. Argumentuje, że czas, w którym trzeba przygotować instytucje na B i C, jest *teraz* — póki A jeszcze pozostaje możliwy.

06. O co właściwie proszą resztę z nas

Anthropic argumentuje, że dla świata byłoby dobrze **mieć opcję** spowolnienia albo czasowego wstrzymania rozwoju AI na granicy możliwości — żeby badania nad bezpieczeństwem i struktury społeczne mogły nadrobić dystans. Przyznają, że jednostronna pauza jednej firmy niewiele zmienia; sensowna pauza wymagałaby *wielu laboratoriów na granicy w wielu krajach, wspólnych warunków oraz wiarygodnej weryfikacji*. Porównują to do kontroli zbrojeń, z wyraźnym zastrzeżeniem, że trening modeli AI jest znacznie trudniejszy do wykrycia niż silosy rakietowe.

Ich końcowe przesłanie warto wzmocnić najmocniej: **w tej rozmowie muszą uczestniczyć ludzie spoza firm AI**. Ustawodawcy, badacze, społeczeństwo obywatelskie, firmy AI niedziałające na granicy, zwykli praktycy budujący realne systemy. Rozmowa nie może zostać zamknięta wewnątrz laboratoriów. Decyzje są zbyt duże.

07. Jeśli prowadzisz działalność w Europie lub na rynek europejski

Jeśli Twoja działalność dotyka użytkowników w Europie, unijny AI Act ma zastosowanie niezależnie od kraju, w którym jesteś zarejestrowany. Większość obowiązków dla dostawców modeli ogólnego przeznaczenia wchodzi w życie w **sierpniu 2026**. Kilka wątków eseju Anthropic bezpośrednio nakłada się na przepisy aktu:

Artykuł 12 (prowadzenie rejestrów) — systemy AI wysokiego ryzyka muszą prowadzić dzienniki zdarzeń. Problem przesuwanego się wąskiego gardła, który nazywa Anthropic, czyni ten obowiązek jednocześnie pilniejszym i bardziej złożonym. **Artykuł 14 (nadzór człowieka)** — wymagany dla systemów wysokiego ryzyka, a zarazem dokładnie ta kwestia, której scenariusz rekursywnego samodoskonalenia najmocniej zagraża. **Artykuł 15 (dokładność, odporność, cyberbezpieczeństwo)** — postawa bezpieczeństwa jest dziś również artefaktem regulacyjnym. **Nakładka RODO** — dane osobowe przechodzące przez AI dalej są danymi osobowymi; AI nikogo nie zwalnia z obowiązków ochrony danych.

UCZCIWE ODCZYTANIE

Unijny AI Act był projektowany pod krzywą możliwości, która rośnie wolniej. Esej Anthropic jest w praktyce publicznym przyznaniem, że krzywa, którą oni sami obserwują, jest na tyle szybka, że napina założenia wpisane w sam akt. Zespoły compliance powinny spodziewać się nowelizacji. Zespoły operacyjne nie powinny na nie czekać.

08. Dziewięćdziesiąt dni praktycznej pracy

Trzy konkretne ruchy, które każdy lider technologii powinien wykonać w najbliższych dziewięćdziesięciu dniach — niezależnie od tego, czy wierzy w najbardziej ambitne projekcje Anthropic.

- 01 **Zmapuj, gdzie dziś w Twoich operacjach siedzi AI.** Gdzie dotyka danych wrażliwych, gdzie podejmuje decyzje w imieniu ludzi, gdzie skok możliwości mógłby przełamać dzisiejsze mechanizmy kontrolne. Nie da się zareagować na zmianę, której się wcześniej nie zmapowało.
- 02 **Zbuduj model operacyjny, nie tylko narzędzia.** Wybierz jeden zespół i jeden proces. Przeprojektuj go przy założeniu, że wykonanie przez AI jest szybkie, a ograniczeniem jest ludzki osąd. Co ten zespół przegląda? Co wybiera? Kiedy eskaluje? Zaprojektuj wąskie gardło świadomie.
- 03 **Upewnij się, że ład w organizacji potrafi nadążyć za tempem AI.** Jeśli komitet ds. zmian zbiera się raz na kwartał, a możliwości AI podwajają się co cztery miesiące, masz strukturalne rozjechanie. Tradycyjna postawa ład nie wytrzyma zderzenia z systemami, które rosną.

09. Jak używam tego eseju we własnej pracy w terenie

W pracy z klientami esej Anthropic krystalizuje rozmowę, która od miesięcy stoi na każdym poważnym stole. Będę go używał jako głównego punktu odniesienia w trzech kontekstach.

Przeglądy gotowości pod unijny AI Act — gdzie luka między „tempem zmiany możliwości” a „tempem zmiany ładu” jest najbardziej niedoszacowanym ryzykiem. **Przeglądy architektoniczne dla zespołów inżynierskich przyspieszonych przez AI** — gdzie problem przesuwającego się wąskiego gardła widać już w czasie rzeczywistym. **Sesje wykonawcze o tym, jak wygląda autentyczny ład w AI** — kiedy system, który podlega ładowi, jest współbudowany przez samo AI.

Nic z tego nie jest katastrofizmem. Nic z tego nie jest też zachwytem. To po prostu praca polegająca na potraktowaniu krzywej poważnie, póki jest jeszcze czas, żeby ją zaprojektować.

Jeśli ten brief okazał się pomocny, przekaz go dalej.

Podziel się z jednym liderem, któremu ufasz. Prześlij dyrektorowi technicznemu, zarządowi, osobie od compliance. Rozmowa zostaje użyteczna tylko wtedy, kiedy wychodzi poza laboratoria.

Jeśli Twój zespół potrzebuje bardziej bezpośredniego wsparcia — gotowość pod unijny AI Act, przegląd architektoniczny, projektowanie ładu w AI, sesje strategiczne dla zarządu — to jest właśnie ta praca w terenie, którą wykonuję. Porozmawiajmy.