Imperium skalowania: od metafizyki mocy do polityki danych

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Tekst stanowi pogłębioną analizę współczesnego paradygmatu rozwoju sztucznej inteligencji, znanego jako doktryna skalowania. Autor dekonstruuje techniczne założenia, według których wykładniczy wzrost mocy obliczeniowej, wolumenu danych i liczby parametrów prowadzi do przełomowych momentów emergencji. Analiza wykracza poza ramy inżynieryjne, badając metafizyczne aspekty dążenia do AGI oraz polityczne konsekwencje koncentracji kapitału infrastrukturalnego w rękach technologicznych gigantów. Omówione zostają mechanizmy 'data flywheel', zmiana paradygmatu naukowego w stronę epistemologii korelacji statystycznej oraz etyczne ryzyka związane z iteracyjnym wdrażaniem modeli w ramach globalnego wyścigu zbrojeń AI. To kluczowa lektura dla zrozumienia mechanizmów kształtujących cyfrową przyszłość.

This text provides an in-depth analysis of the contemporary paradigm of AI development, known as the scaling doctrine. The author deconstructs the technical assumptions that exponential growth in computing power, data volume, and parameter counts leads to breakthrough moments of emergence. The analysis goes beyond engineering, examining the metaphysical aspects of the pursuit of AGI and the political consequences of the concentration of infrastructure capital in the hands of tech giants. It discusses the mechanisms of the data flywheel, the shift in the scientific paradigm toward statistical correlation epistemology, and the ethical risks associated with iterative model deployment within the global AI arms race. This is crucial reading for understanding the mechanisms shaping the digital future.

Artykuł analizuje doktrynę skalowania w kontekście rozwoju sztucznej inteligencji (AI), demaskując ją jako więcej niż tylko techniczną heurystykę. Ukazuje, jak obiecując przewidywalny wzrost jakości modeli AI poprzez zwiększanie zasobów obliczeniowych, danych i parametrów, doktryna ta przekształca się w normatywny program cywilizacyjny. Ten program, zdaniem autora, prowadzi do koncentracji władzy, kolonizacji zasobów i marginalizacji

alternatywnych ścieżek badawczych. Krytyczna analiza, oparta na pracy Karen Hao, ujawnia, że wyścig w skalowaniu generuje znaczące koszty społeczne, środowiskowe i ekonomiczne, jednocześnie narzucając epistemologię korelacji, która spłaszcza złożoność ludzkiego doświadczenia. Artykuł stawia tezę, że bez świadomej kontroli i regulacji, doktryna skalowania może przekształcić się w imperialny projekt, który zagraża pluralizmowi poznawczemu i demokracji.

SŁOWA KLUCZOWE

doktryna skalowania | moc obliczeniowa | AGI | próg emergencji | data flywheel | modele fundamentalne | epistemologia korelacji | iteracyjne wdrażanie | prawo OpenAI | architektura transformerów | parametry modeli | kanibalizacja projektów | reżimy poufności | przestrzeń wektorowa | kapitał infrastrukturalny

Doktryna skalowania: inżynieryjna metafizyka mocy

W swoim najbardziej technicznym, inżynierskim wydaniu doktryna skalowania jest prostym, empirycznym prawem, zgodnie z którym błąd modeli głębokiego uczenia maleje w przewidywalny sposób wraz ze wzrostem trzech kluczowych zasobów: wolumenu danych, liczby parametrów i mocy obliczeniowej. W interpretacji kulturowej mamy jednak do czynienia z czymś znacznie więcej niż tylko z wykresem funkcji błędu. Do tej technicznej relacji wprowadza się bowiem dodatkowe założenie o istnieniu progu emergencji, a więc momentu, w którym czysto ilościowe powiększanie sieci neuronowej generuje nie tylko lepsze wyniki, lecz jakościowo nowe, niezaprogramowane wcześniej zdolności, takie jak wieloetapowe rozumowanie, pisanie kodu czy quasi-refleksyjna analiza tekstu.

Ilya Sutskever, przedstawiany w relacji Hao jako profetyczny głos tego ruchu, uosabia wiarę w to, że „compute is king”. Oznacza to, że spośród wszystkich zmiennych to właśnie nieograniczony dostęp do mocy obliczeniowej jest najpewniejszą drogą do przełamywania kolejnych barier w rozwoju sztucznej inteligencji. Doktryna ta, choć wyrosła z konkretnych eksperymentów z architekturą transformerów, przybiera w ten sposób postać szerzej pojętej metafizyki ilości. W jej ramach inteligencja staje się funkcją gęstości połączeń i energii przepływającej przez system, co spycha na margines pytania o architekturę, strukturę czy interpretowalność modeli jako kwestie drugorzędne.

Z perspektywy historyka idei nietrudno dostrzec tu powtórzenie znanego schematu nowoczesności. Najpierw pojawia się skuteczna technologia, która w wąskim obszarze demonstruje imponujące rezultaty. Następnie wokół tej technologii rodzi się teoria uzasadnienia, nadająca jej charakter uniwersalny i niemal historycznie

konieczny. Wreszcie, wewnątrz tej teorii horyzont poznawczy kurczy się tak bardzo, że alternatywy przestają być widoczne, a to, co było jednym z możliwych kierunków badań, zostaje przekształcone w dziejową teleologię. Doktryna skalowania w ujęciu Hao ma dokładnie taki status: zgrabna heurystyka dla inżynierów po kilku udanych demonstracjach staje się całościową metaforą świata, w której wszystko, co istotne, da się sprowadzić do formuły: więcej danych, więcej parametrów, więcej energii.

W tym miejscu wyłania się pierwsza fundamentalna aporia współczesnego paradygmatu AI. Z jednej strony literatura techniczna i ekonomiczna potwierdza, że skalowanie rzeczywiście zapewnia przewidywalny wzrost jakości na wielu benchmarkach, co czyni je niezwykle atrakcyjną strategią dla firm dysponujących odpowiednim kapitałem i infrastrukturą. Z drugiej strony socjologia nauki i krytyczne studia nad technologią alarmują, że tak radykalna koncentracja zasobów wokół jednej metody prowadzi do zdławienia konkurencyjnych programów badawczych, które mogłyby rozwijać inne typy AI: mniej zasobochłonne, bardziej symboliczne czy lepiej osadzone w lokalnych kontekstach wiedzy. Hao podkreśla, że wyścig w skalowaniu doprowadził do kanibalizacji innych projektów w korporacjach, jak choćby wstrzymania prac nad AI do odkrywania leków, by uwolnić czipy na potrzeby trenowania kolejnych modeli konwersacyjnych.

Powstaje zatem głębokie napięcie między empiryczną skutecznością jednego programu a pluralizmem poznawczym całej dziedziny nauki. Doktryna skalowania rozwiązuje tę sprzeczność w sposób, który można określić jako imperialny. Skoro ten jeden, jedyny program obiecuje dostęp do AGI – technologii mającej rzekomo rozwiązać problemy klimatyczne, zdrowotne i ekonomiczne – to poświęcenie alternatywnych ścieżek badawczych staje się kosztem dopuszczalnym. Nawet degradacja środowiska i prekaryzacja pracy zostają wpisane w mit o przyszłej, wielkiej rekompensacie. Jak zauważa Hao

OpenAI: cyfrowe imperium nowej formacji

Gdyby spróbować uchwycić istotę doktryny skalowania i iteracyjnego wdrażania w jednym zdaniu, należałoby stwierdzić, że współczesne imperium sztucznej inteligencji nie jest już zbiorem algorytmów. Jest raczej zinstytucjonalizowaną władzą nad kierunkiem, tempem i kosztem cywilizacyjnego uczenia się, gdzie postęp definiują ci, którzy kontrolują przepływy mocy obliczeniowej, danych i globalnej wyobraźni. Karen Hao, opisując OpenAI jako nowy typ mocarstwa, zauważa, że nie działa ono jak zwykła spółka, lecz jak imperialna formacja. Przekracza ona granice gospodarki, polityki i kultury, kolonizując pracę, zasoby i społeczną przyszłość pod sztandarem nieuchronnej ogólnej sztucznej inteligencji, czyli AGI.

W tym ujęciu doktryna skalowania przestaje być wyłącznie techniczną hipotezą, wedle której wydajność modeli rośnie przewidywalnie wraz z ilością danych, liczbą parametrów i zużytą mocą obliczeniową. Staje się normatywnym programem cywilizacyjnym, który zakłada, że rozsądne, a nawet moralnie konieczne, jest przepalenie przez ludzkość niewyobrażalnych zasobów materialnych, energetycznych i ludzkich. Celem jest

dotarcie do progu, za którym inteligencja maszynowa zacznie rozwiązywać problemy szybciej, niż jesteśmy w stanie je tworzyć. W logice tej narracji wzrost mocy obliczeniowej jest nie tyle środkiem, ile samą miarą sensu: Sam Altman i Ilya Sutskever uczynili z niego quasi-teologiczną zasadę, którą Greg Brockman ochrzcił mianem Prawa OpenAI, wskazując, że ilość mocy obliczeniowej w przełomowych eksperymentach podwajała się co kilka miesięcy, znacznie szybciej niż w klasycznym prawie Moore'a.

Aby jednak zrozumieć polityczny ciężar tej strategii, trzeba porzucić perspektywę laboratoryjną i spojrzeć na doktrynę skalowania jako na specyficzny wariant późnokapitalistycznej racjonalności. Wzrost staje się w niej samoistnym imperatywem, zaś iteracyjne wdrażanie jako eksperyment na społeczeństwie służy naturalizacji kosztów ubocznych jako nieuniknionych ofiar w marszu ku technologicznemu zbawieniu. Hao wprost sugeruje, że OpenAI, reorganizując się z idealistycznej organizacji non-profit w hybrydową strukturę z ograniczonym zyskiem, ale praktycznie nieograniczonym apetytem na kapitał, wytworzyło nową formę budowania imperium. W tej nowej formule religia AGI zastępuje dawne misje cywilizacyjne.

Pytanie o doktrynę skalowania nie dotyczy już tego, czy modele językowe są imponujące. Dotyczy raczej tego, jakie warunki brzegowe społecznego uznania i politycznej legitymizacji muszą zostać spełnione, aby tak radykalna koncentracja władzy – poznawczej, infrastrukturalnej i finansowej – mogła uchodzić za racjonalny projekt, a nie za monstrualny eksperyment na gatunku ludzkim.

Prawo OpenAI: monopolizacja rozwoju technologii

Jeżeli w pierwszej części można było jeszcze postrzegać doktrynę skalowania i iteracyjnego wdrażania jako radykalną, lecz wewnętrznie spójną strategię inżynieryjną, to perspektywa nauk społecznych i krytycznej teorii technologii zmusza nas do rewizji tego poglądu. Staje się bowiem jasne, że mamy do czynienia z czymś znacznie bardziej fundamentalnym: z projektem przebudowy samej infrastruktury wiedzy w skali planetarnej. Karen Hao w swojej rekonstrukcji historii OpenAI proponuje intuicję na pozór prostą, lecz w istocie druzgocącą. Obecny boom na generatywną AI to nie tylko opowieść o naukowym przełomie, ale także o nowym „pograniczu imperialnym", gdzie centra danych, łańcuchy dostaw chipów i korporacyjne reżimy poufności zastępują dawne formy ekspansji terytorialnej.

Paradygmat naukowy, który wyrasta z tej doktryny, można opisać jako radykalne przesunięcie od epistemologii interpretacji ku epistemologii korelacji. Tradycyjny ideał naukowy, dziedzictwo Oświecenia, zakładał, że wiedza polega na odkrywaniu struktur przyczynowych, formułowaniu weryfikowalnych teorii i poddawaniu ich zinstytucjonalizowanym procedurom falsyfikacji. Logika modeli fundamentalnych opiera się natomiast na innym założeniu: jeśli wystarczająco dużą sieć neuronową nakarmimy wystarczająco obszernym korpusem danych, to wytworzone przez nią relacje korelacyjne okażą się w praktyce równie użyteczne jak jawne modele przyczynowe, a w wielu zastosowaniach wręcz je przewyższą.

Nauki społeczne reagują na tę zmianę dwojako. Z jednej strony część badaczy dostrzega w generatywnej AI potężną maszynę do symulacji, pozwalającą modelować scenariusze polityczne, ekonomiczne czy kulturowe z niespotykaną dotąd szczegółowością. Z drugiej strony narasta nurt krytyczny, który w modelach fundamentalnych diagnozuje powtórzenie klasycznego błędu urzeczowienia, czyli traktowania abstrakcji jak realnych bytów, tyle że w wersji turbodoładowanej. W tym ujęciu złożone relacje społeczne, normatywne konflikty i historyczne traumy zostają spłaszczone i zredukowane do punktów w przestrzeni wektorowej. To, co Habermas nazwałby proceduralną jednością uzasadniania – wspólnym dochodzeniem do prawdy przez dialog – zostaje zastąpione przez statystyczną jedność aproksymacji, a roszczenia do prawdziwości, słuszności i autentyczności przepakowuje się w formę generatywnych prawdopodobieństw.

Hao brutalnie demaskuje tę transformację, pokazując, jak w praktyce wygląda owa „epistemologia bagna danych". Z jej relacji wynika, że nawet wewnątrz OpenAI wiedza o tym, co dokładnie znajduje się w zbiorach treningowych, jest fragmentaryczna i często po prostu niedostępna. Dane są zbyt obszerne, by można je było ręcznie skatalogować czy zweryfikować. Oznacza to, że fundamentem współczesnych modeli nie jest przejrzysta biblioteka ludzkości, lecz mroczne archiwum, w którym obok traktatów naukowych leżą nawoływania do ludobójstwa, a obok podręczników medycyny – pseudonaukowe teorie spiskowe. Modele te ucieleśniają zatem radykalne roszczenie: że można budować krytyczne kompetencje poznawcze na bazie nieprzejrzystej, niepoddającej się audytowi masy danych, pod warunkiem że będziemy wystarczająco agresywnie skalować.

Dla dyskursu publicznego jest to moment przełomowy. Dopóki dane, modele i decyzje były rozproszone między uniwersytety, think tanki i inne instytucje eksperckie, istniała szansa na pluralizm paradygmatyczny – na wielość współzawodniczących programów badawczych, które nawzajem się k

Skalowanie kanibalizuje alternatywne programy badawcze

W tym miejscu odsłania się pierwsza, fundamentalna sprzeczność współczesnego paradygmatu naukowego. Z jednej strony literatura techniczna z obszaru modeli fundamentowych i ekonomii modeli bazowych potwierdza, iż skalowanie faktycznie przynosi przewidywalny wzrost jakości na wielu benchmarkach, co czyni je narzędziem niezwykle pociągającym dla korporacji dysponujących odpowiednim kapitałem i infrastrukturą. Z drugiej jednak strony socjologia nauki oraz krytyczne studia nad technologią alarmują, że tak skrajna koncentracja zasobów wokół jednej metody badawczej prowadzi do zdławienia konkurencyjnych programów, które mogłyby rozwijać inne, mniej zasobochłonne, bardziej symboliczne czy lepiej osadzone w lokalnych formach wiedzy typy sztucznej inteligencji. Jak dobitnie pokazuje Hao, wyścig w skalowaniu doprowadził do kanibalizacji innych projektów badawczych, czego przykładem jest wstrzymanie prac nad AI

służącą do odkrywania leków, aby uwolnić układy scalone niezbędne do trenowania kolejnych modeli konwersacyjnych.

Rodzi się zatem głębokie napięcie między brutalną skutecznością jednego programu a pluralizmem poznawczym, który stanowi o zdrowiu całego pola nauki. Doktryna skalowania rozwiązuje tę sprzeczność w sposób właściwy dla imperium. Skoro ten jeden, dominujący program obiecuje dostęp do AGI – technologii mającej rzekomo rozwiązać problemy klimatyczne, zdrowotne i ekonomiczne – to kosztem w pełni dopuszczalnym staje się poświęcenie alternatywnych ścieżek badawczych. Co więcej, nawet degradacja środowiska i struktur pracy zostaje wpisana w mit o przyszłej, wspaniałej rekompensacie. Jest to, jak zauważa Hao, retoryka niebezpiecznie bliska argumentacji dawnych imperiów kolonialnych, które bezwzględną eksploatację zasobów i ludzkiej pracy uzasadniały obietnicą przyszłego dobra cywilizacyjnego.

Iteracyjne wdrażanie AI: globalny eksperyment społeczny

Jeśli doktryna skalowania odpowiada na pytanie, jak technicznie powiększać moc modeli, to iteracyjne wdrażanie jest odpowiedzią na pytanie, jak wprowadzić je do tkanki społecznej, by zminimalizować ryzyko katastrofy, nie tracąc przy tym rynkowej przewagi. OpenAI konsekwentnie przedstawia swoją strategię jako proces stopniowego, kontrolowanego uwalniania coraz potężniejszych systemów, udostępnianych coraz szerszej publiczności. Kluczem do zrozumienia tej logiki jest koncepcja data flywheel, czyli koła zamachowego danych, mechanizmu, w którym każda interakcja użytkownika staje się paliwem dla kolejnej, doskonalszej wersji modelu.

W tej konstrukcji produkt i eksperyment stają się nierozróżnialne. ChatGPT, anonsowany jako skromny „podgląd badawczy”, był w istocie wehikułem do masowego zbierania danych oraz poligonem do testowania społecznych reakcji na zupełnie nowy typ interfejsu. Jak pokazuje Hao, prawdziwym celem nie było stworzenie gotowego narzędzia, lecz uruchomienie pętli zwrotnej, w której miliony rozmów z ludźmi stają się bezcennym materiałem treningowym dla następnych iteracji, przede wszystkim dla GPT-4. W ten sposób społeczeństwo, nie do końca świadomie, wzięło udział w jego tworzeniu.

Z perspektywy prawa i teorii demokracji jest to sytuacja radykalnie nowa. Tradycyjna regulacja technologii wysokiego ryzyka opierała się na logice ex ante, czyli na wymogu przeprowadzenia testów, analiz i certyfikacji, zanim produkt trafi do masowego użytku. Iteracyjne wdrażanie przesuwa środek ciężkości na logikę ex post i ongoing, gdzie to rynek staje się główną przestrzenią wykrywania błędów, a użytkownicy pełnią funkcję nieświadomych uczestników globalnego eksperymentu. Dopiero gdy pojawiają się konkretne

nadużycia – jak generowanie materiałów pedofilskich w grach opartych na GPT-3 – następuje reaktywna korekta parametrów i wzmocnienie filtrów.

Co znamienne, strategia ta jest przedstawiana jako etycznie odpowiedzialna. Sam Altman argumentuje, że społeczeństwo musi stopniowo oswajać się z coraz potężniejszymi systemami, a inżynierowie muszą uczyć się zarządzać ich ryzykiem, ponieważ nadejście AGI jest nieuniknione. W tej optyce każda bieżąca szkoda – dezinformacja, utrwalone uprzedzenia czy automatyzacja pracy – zostaje wchłonięta przez metanarrację o przygotowaniach na jeszcze większe zagrożenie, jakim jest potencjalnie niekontrolowana superinteligencja. Iteracyjne wdrażanie jest więc zarazem techniką uczenia maszynowego, strategią rynkową i praktyką pedagogiczną wobec ludzkości, która ma zostać oswojona z myślą, że ten cywilizacyjny eksperyment nie ma przycisku pauzy.

Z perspektywy krytycznej teorii społecznej taka konstrukcja rodzi fundamentalny problem legitymizacji. Racjonalność nowoczesnych demokracji wymaga, aby ryzyka systemowe podlegały publicznej debacie, transparentnym procedurom i zewnętrznemu nadzorowi. Tymczasem w imperium AI kluczowe decyzje o tempie wdrażania i zakresie testów zapadają wewnątrz korporacji, a społeczeństwo zapraszane jest do roli beta-testerów, nie współdecydentów. Rodzi to silne napięcie między iteracyjnym wdrażaniem jako rzekomo najbezpieczniejszą metodą oswajania technologii a iteracyjnym wdrażaniem jako narzędziem narzucania społeczeństwu faktów dokonanych.

W tym właśnie miejscu pojawia się reakcja europejskiego paradygmatu regulacyjnego. Unia Europejska, wprowadzając Akt o sztucznej inteligencji, próbuje ograniczyć tę spontaniczność, ustanawiając ramy dla modeli ogólnego przeznaczenia. Wymogi dotyczące dokumentacji technicznej, zgodności z prawem autorskim czy obowiązku streszczania zbiorów danych treningowych to próba przywrócenia elementarnych warunków dla publicznego dyskursu w sferze, którą amerykański paradygmat korporacyjny oddał logice „poruszaj się szybciej niż inni”.

Problem polega jednak na tym, że nawet najbardziej ambitne regulacje przychodzą za późno, w obliczu dokonanej już konsolidacji rynkowej i infrastrukturalnej. Iteracyjne wdrażanie zadziałało tu jak strategia blitzkriegu. Zanim zdążył powstać dojrzały dyskurs normatywny na temat konsekwencji modeli fundamentalnych, społeczeństwa przyzwyczaiły się do ich codziennego użycia, a państwa i firmy zaczęły uzależniać od nich swoje procesy. Z tej perspektywy krytyka Hao pełni funkcję spóźnionego, lecz koniecznego demaskowania, które przypomina, że coś, co wprowadzono jako neutralne narzędzie, jest w istocie rezultatem serii wysoce kontrowersyjnych decyzji o kształcie globalnego ładu.

RLHF: eksploatacja pracowników globalnego Południa

W ten sposób, w cieniu Doliny Krzemowej, rodzi się globalny proletariat afektywny. Jego niewidzialni członkowie – anotatorzy z Kenii, Wenezueli czy Filipin, zatrudniani przez platformy takie jak Scale AI – otrzymują wynagrodzenie często niższe niż dwa dolary za godzinę. Ich zadaniem jest przyjmowanie na siebie psychicznego ciężaru oczyszczania modeli, co w praktyce oznacza wielogodzinny kontakt z najbardziej traumatycznymi treściami, jakie internet jest w stanie wygenerować: opisami gwałtów, kazirodztwa, tortur czy samobójstw. Wszystko po to, aby użytkownik z klasy średniej w Ameryce czy Europie mógł doświadczyć gładkiej, uprzejmej i rzekomo bezpiecznej interakcji z maszyną.

W sensie antropologicznym mamy tu do czynienia z radykalnie nierównym podziałem pracy nad utrzymaniem porządku symbolicznego. Uprzywilejowani użytkownicy delegują na niewidzialnych pracowników z globalnego Południa funkcję strażników bram, którzy stają między nieprzetworzoną przemocą danych a oswojoną, sterylną powierzchnią produktu. Karen Hao trafnie zauważa, że to właśnie ten ludzki filtr, a nie abstrakcyjna troska korporacji, czyni współczesne modele pozornie bezpiecznymi. Z tej perspektywy iteracyjne wdrażanie i doktryna skalowania wymagają nie tylko miliardów dolarów zainwestowanych w karty graficzne, ale także systematycznego zużywania zasobów emocjonalnych i zdrowia psychicznego tych, którzy nigdy nie pojawią się na konferencjach o przyszłości AGI.

Trening wielkich modeli: ukryty koszt środowiskowy

Doktryna skalowania, choć przyobleczona w szaty naukowej empirii, generuje koszty, których skala wymyka się standardowym kategoriom rachunku ekonomicznego. Z jednej strony mamy bowiem bezpośrednie nakłady finansowe na trening kolejnych generacji modeli, liczone w dziesiątkach milionów dolarów za pojedynczy eksperyment i w miliardach w skali rocznej całego przedsiębiorstwa. Z drugiej zaś piętrzą się koszty infrastrukturalne, obejmujące budowę centrów danych, zakup zaawansowanych układów scalonych i utrzymanie sieci energetycznej zdolnej obsłużyć nienasycony apetyt na moc obliczeniową. Sojusz OpenAI z Microsoftem oraz analogiczne alianse w branży dowodzą, że bez wsparcia globalnych gigantów chmurowych całe to przedsięwzięcie nie byłoby w ogóle wykonalne.

Jeszcze bardziej problematyczne okazują się koszty środowiskowe, które Karen Hao opisuje z bezlitosną precyzją. Trening modelu GPT-4 w stanie Iowa doprowadził w ciągu jednego miesiąca do zużycia ponad jedenastu milionów galonów wody w regionie dotkniętym suszą, co stanowiło znaczący odsetek lokalnego zużycia. Centra danych wznoszone w Chile, na Bliskim Wschodzie czy w innych obszarach o ograniczonych zasobach wodnych powielają ten destrukcyjny wzorzec, pochłaniając wodę niezbędną lokalnym społecznościom do życia i uprawy roli. Jednocześnie cała infrastruktura AI opiera się na wydobyciu

surowców takich jak miedź i lit, co drenuje ekosystemy i nierzadko wiąże się z naruszeniami praw ludności autochtonicznej.

Hao, sięgając do języka historii kolonialnej, argumentuje, że jesteśmy świadkami nowej formy imperializmu infrastrukturalnego. Zamiast flot handlowych i kompanii kolonialnych pojawiają się sieci centrów danych, sieci energetyczne i kontrakty na dostawy układów scalonych, jednak fundamentalna logika pozostaje niezmienna. Imperia AI akumulują zasoby materialne i polityczne na globalnej Północy, jednocześnie eksternalizując, czyli przerzucając, koszty środowiskowe i społeczne na globalne Południe.

Z perspektywy ekonomii środowiska można by oczywiście twierdzić, że każdy przełom technologiczny pociąga za sobą okres wzmożonej konsumpcji zasobów, po którym następuje faza optymalizacji i większej efektywności. Problem z doktryną skalowania polega jednak na tym, że sam mechanizm przewagi konkurencyjnej jest nierozerwalnie sprzężony z nieustannym zwiększaniem skali, co wyklucza istnienie naturalnego punktu nasycenia. Skoro przewaga jednej firmy nad drugą wyraża się w liczbie parametrów i ilości mocy obliczeniowej zużytej do treningu, to każda próba ograniczenia tego wzrostu jest w tym paradygmacie równoznaczna z przyznaniem się do porażki w wyścigu do AGI.

Dla porządku należy dodać, że podobny sposób myślenia przenika dziś również strategie państwowe, co doskonale ilustrują ambicje krajów Zatoki Perskiej. Arabia Saudyjska i Zjednoczone Emiraty Arabskie inwestują w gigantyczne klastry centrów danych, zamawiają dziesiątki tysięcy najbardziej zaawansowanych układów scalonych i budują narodowe strategie AI, które mają przekształcić ich gospodarki w obliczu przewidywanego szczytu popytu na ropę. W ten sposób doktryna skalowania zostaje przeszczepiona na grunt geopolityczny, a parametry infrastruktury sztucznej inteligencji stają się nową miarą potęgi państw, obok produktu krajowego brutto, wydatków militarnych i rezerw walutowych.

USA, Europa i kraje arabskie: geopolityczne wizje AI

Różnice w podejściu do doktryny skalowania i iteracyjnego wdrażania w Ameryce, Europie i świecie arabskim to coś więcej niż spór o technologię. To zderzenie trzech współczesnych odmian tego, co kiedyś nazywano świecką religią postępu, z których każda ma własnych kapłanów, własne dogmaty i własne wyobrażenie zbawienia.

W świecie amerykańskim dominującą gramatyką jest splot libertariańskiej wiary w rynek jako ostateczny mechanizm selekcji z mesjanistyczną narracją o technologicznym odkupieniu. Gdy Sam Altman przedstawia AGI jako siłę zdolną „położyć kres ubóstwu”, wpisuje się w długą tradycję myślenia, w której innowacja zastępuje dawne figury Opatrzności czy Dziejów jako gwarant ostatecznego porządku. Nieprzypadkowo sekretarz generalny ONZ, krytykując w Davos „nieodpowiedzialne ściganie zysku” przez Big Tech,

porównuje ryzyka AI do kryzysu klimatycznego. W obu przypadkach prywatne podmioty wykorzystują strukturalną bezradność wspólnoty politycznej, by narzucić wszystkim swój horyzont czasowy i własną definicję katastrofy.

Amerykański paradygmat, choć coraz bardziej świadomy potrzeby regulacji, wciąż traktuje iteracyjne wdrażanie jako podstawowy mechanizm społecznego uczenia się w warunkach niepewności. To właśnie logika „move fast and fix things later", w której błędy są akceptowalnym kosztem postępu, o ile zapewniają przewagę w globalnym wyścigu. Krytycy, tacy jak Karen Hao, próbują w tym reżimie przywrócić widoczność ukrytych kosztów, jednak ich głos z trudem przebija się przez instytucjonalną grawitację Doliny Krzemowej, gdzie zarówno kapitał, jak i wyobraźnia polityczna są głęboko zakorzenione w micie o konieczności „odważnych eksperymentów".

W Europie krajobraz jest zupełnie inny. Tutaj pamięć o katastrofach XX wieku oraz specyficzny model społecznej gospodarki rynkowej zrodziły głęboką wrażliwość na to, że techniczna racjonalność musi być osadzona w ramach prawnych, które chronią godność jednostki i integralność procesu demokratycznego. Unijny Akt o sztucznej inteligencji jest próbą uchwycenia modeli fundamentowych jako nowego rodzaju „infrastruktury ryzyka systemowego", porównywalnego nie z pojedynczym produktem, lecz z siecią energetyczną czy systemem bankowym, których awaria pociąga za sobą całe społeczeństwo.

Europejski paradygmat jest więc mniej skłonny do bezkrytycznego przyjmowania doktryny skalowania, a bardziej zainteresowany ograniczaniem jej najbardziej agresywnych przejawów poprzez wymogi transparentności, oceny ryzyka i mechanizmy odpowiedzialności. W tym sensie Unia Europejska próbuje „zracjonalizować" iteracyjne wdrażanie, przenosząc część ciężaru decyzyjnego z korporacyjnych zarządów na sformalizowane procedury audytu i notyfikacji. Różnica jest fundamentalna: tam, gdzie amerykańska logika widzi w skalowaniu obietnicę technologicznej hegemonii, europejska dostrzega przede wszystkim konieczność zapobiegania koncentracji władzy nad światem przeżywanym obywateli.

Najbardziej intrygująca, zwłaszcza w kontekście analiz Hao, jest jednak sytuacja świata arabskiego, a szczególnie państw Zatoki Perskiej. Tutaj doktryna skalowania została przyjęta nie tylko jako instrument przyspieszonej modernizacji, ale też jako narzędzie geopolitycznego pozycjonowania się wobec USA i Chin. Arabia Saudyjska oraz Zjednoczone Emiraty Arabskie zamawiają setki tysięcy najnowocześniejszych procesorów Nvidii, budują gigantyczne centra danych i zawierają sojusze, w których dostęp do infrastruktury AI staje się potężną kartą przetargową.

W tej konfiguracji iteracyjne wdrażanie przestaje być strategią firmy, a staje się elementem

Kompresja danych: reprodukcja uprzedzeń w modelach AI

W optyce doktryny skalowania uczenie maszynowe jawi się jako proces niemal naturalny, uruchamiany automatycznie po dostarczeniu odpowiedniej ilości danych i mocy obliczeniowej. Jednak z perspektywy antropologii wiedzy ta sama operacja okazuje się niezwykle kłopotliwą formą kompresji świata. Sama metoda, polegająca na tym, że modele głębokiego uczenia zyskują swoje zdolności poprzez przewidywanie kolejnych słów w gigantycznych korpusach tekstu, jest w interpretacji Ilyi Sutskevera sposobem na zmuszenie ich do destylacji esencji ludzkiej wiedzy.

Cały problem w tym, że świat poddawany tej destylacji nie jest abstrakcyjną esencją racjonalności, lecz konkretnym, historycznie uwarunkowanym archiwum przemocy symbolicznej, utrwalonych uprzedzeń, asymetrii władzy i niewidzialnej pracy opiekuńczej. Internet, na którym trenują się modele OpenAI i ich konkurenci, to przecież nie tylko Wikipedia i podręczniki akademickie, ale także fora pełne rasistowskich obelg, mizoginicznych fantazji, teorii spiskowych i kampanii propagandowych. W świetle tej, jak określa to Karen Hao, epistemologii bagien danych, modele są strukturalnie predysponowane do reprodukcji dyskryminujących wzorców, nawet gdy ich twórcy usiłują wzbogacić zbiory o filtrujące reguły.

Z perspektywy ekonomii politycznej wiedzy rodzi to fundamentalny paradoks. Z jednej strony duże modele językowe przedstawia się jako narzędzia demokratyzacji dostępu do informacji, zdolne udzielać wyjaśnień w stylu osobistego tutora na niemal każde pytanie. Z drugiej jednak ich wewnętrzne reprezentacje świata nasycone są strukturalnymi uprzedzeniami odziedziczonymi po danych wejściowych. Symptomatyczne przykłady generowania obrazów, na których inżynier jest domyślnie mężczyzną, a piękna osoba zazwyczaj młodą, białą kobietą, ledwie dotykają głębszych konsekwencji tego stanu rzeczy, jakimi są różnicowanie politycznej widzialności grup, amplifikacja stereotypów w debacie publicznej czy zacieranie granicy między rzetelną informacją a fabularyzowaną konfabulacją.

Demontaż metafizyki skali: nowe paradygmaty badawcze

Pierwszy ruch polega na świadomym zdegradowaniu doktryny skalowania z pozycji niemal metafizycznej do roli zwykłego narzędzia. Oznacza to odrzucenie pokusy, by traktować krzywe skalowania jako nieuchronne prawo dziejowe, a zamiast tego przyjęcie, że są one zaledwie jednym z wielu sposobów doskonalenia modeli. Ich zastosowanie musi być zatem podporządkowane normom zewnętrznym, takim jak

sprawiedliwość środowiskowa, pluralizm epistemiczny – czyli poszanowanie dla różnych form wiedzy – oraz ogólne bezpieczeństwo systemowe.

W praktyce pociągnęłoby to za sobą wprowadzenie do debaty regulacyjnej kategorii, które dotąd pozostawały w sferze wewnętrznych planów technologicznych korporacji. Zamiast więc dyskutować wyłącznie o mglistych „ryzykach zastosowań", należałoby wprost postawić kwestię twardych limitów mocy obliczeniowej przeznaczanej na pojedyncze eksperymenty, globalnych budżetów na zasoby obliczeniowe oraz mechanizmów rejestracji i monitorowania wielkich procesów uczenia w czasie rzeczywistym. Jak sygnalizuje Hao, w środowisku badaczy bezpieczeństwa AI już dziś pojawiają się propozycje stworzenia międzynarodowego rejestru eksperymentów, które przekraczają określony próg zasobów, połączonego z obowiązkiem wcześniejszej notyfikacji i poddania się niezależnym audytom.

Tego rodzaju rozwiązanie przeniosłoby część logiki zarządzania skalą z poziomu poszczególnych firm na poziom instytucji przypominającej nieco Międzynarodową Agencję Energii Atomowej. Różnica jest oczywiście ilościowa – zamiast kilkudziesięciu reaktorów jądrowych mielibyśmy do czynienia z setkami potężnych procesów trenowania modeli rocznie – jednak idea pozostaje ta sama: im większe potencjalne oddziaływanie technologii, tym silniejsza potrzeba ponadnarodowego nadzoru i transparentności.

Z perspektywy ekonomisty można spodziewać się natychmiastowego kontrargumentu, że takie ograniczenia spowolnią postęp i osłabią konkurencyjność Zachodu w starciu z autorytarnymi reżimami technologicznymi. Odpowiedź, którą sugeruje Hao, nie polega na naiwnym zaprzeczeniu, lecz na odwróceniu pytania: czy naprawdę chcemy budować porządek globalny, w którym za liderów uważa się te państwa, które najszybciej i najbardziej agresywnie spalą swoje zasoby wodne i energetyczne w imię wyścigu o moc obliczeniową, a te, które odmówią udziału w tym szaleństwie, zostaną uznane za zacofane? Jeżeli odpowiedź brzmi „nie", wówczas hamulce instytucjonalne nie są zdradą racjonalności, lecz jej koniecznym rozszerzeniem – z wąsko pojmowanej racjonalności instrumentalnej na szerszą racjonalność społeczną.

Warto w tym miejscu przywołać myśl Immanuela Wallersteina, który pisał, że każdy system światowy wynajduje ideologię racjonalizującą jego dominujące formy akumulacji. Korporacyjna doktryna skalowania jest właśnie taką ideologią, usprawiedliwiającą bezprecedensową koncentrację mocy obliczeniowej i kapitału pod hasłem nieuchronnego nadejścia AGI. Zadaniem krytyki nie jest jednak wyśmianie tej ideologii jako naiwnej, lecz wykazanie, że jest ona w istocie samowywrotna, ponieważ prowadzi do systemu, który zużywa warunki własnej możliwości: stabilne środowisko, zaufanie społeczne i pluralizm wiedzy.

Demokratyzacja wdrożeń: oddolna kontrola technologii

Drugi ruch polega na odzyskaniu kontroli nad iteracyjnym wdrażaniem. Skoro w warunkach radykalnej niepewności pewien rodzaj publicznych eksperymentów jest nieunikniony, bo żaden model nie przewidzi wszystkich skutków ubocznych nowej technologii, to rozwiązaniem nie jest powrót do iluzji pełnej kontroli. Jest nim natomiast przesunięcie ciężaru iteracji z prywatnej decyzji korporacji na publicznie nadzorowaną procedurę, która traktuje iteracyjne wdrażanie jako eksperyment na społeczeństwie.

Oznaczałoby to, po pierwsze, włączenie użytkowników i społeczeństwa obywatelskiego w projektowanie kryteriów sukcesu takich wdrożeń, a po drugie, stworzenie instytucjonalnych kanałów, którymi doświadczenia i skargi obywateli wpływałyby na proces korekty modeli. Wreszcie, po trzecie, wymagałoby to ustanowienia realnych, a nie tylko deklaratywnych, progów zatrzymania eksperymentu w razie pojawienia się określonych szkód. Przypomina to poniekąd dawne spory o procedury dopuszczania leków na rynek, z tą fundamentalną różnicą, że w przypadku AI toksyczność nie dotyczy biochemii, lecz sfery symbolicznej, politycznej i psychicznej.

W praktyce mogłoby to przybrać formę zobowiązania firm do prowadzenia zewnętrznie nadzorowanych faz pilotażowych. Odbywałyby się one z ograniczoną liczbą użytkowników, poddanych ścisłemu monitoringowi i wymogowi raportowania skutków ubocznych. Różnica względem obecnego modelu „research preview” byłaby zasadnicza: nie byłaby to jednostronna deklaracja firmy, lecz element formalnego systemu, w którym zewnętrzne ciało, umocowane być może przy ONZ lub Radzie Europy, posiadałoby uprawnienia do wstrzymania eksperymentu, gdy określone progi szkód zostaną przekroczone.

Karen Hao dostarcza tu ważnych przykładów, które pokazują, że brak takich procedur prowadzi do patologii. Historia wykorzystania GPT-3 w grze AI Dungeon, gdzie model generował treści o charakterze pedofilskim, została ujawniona dopiero po nagłośnieniu sprawy przez media. Reakcja firmy sprowadziła się do wprowadzenia doraźnych filtrów i przerzucenia odpowiedzialności na zewnętrznych dostawców. Gdyby istniała formalna procedura z obowiązkiem uprzedniego zgłoszenia, tego rodzaju użycie modelu fundamentalnego mogłoby zostać zakwestionowane znacznie wcześniej, zanim gra zyskała masową popularność.

Z perspektywy teorii prawa taka instytucjonalizacja iteracji przypominałaby wprowadzenie zasady ostrożności do sfery cyfrowej. Przy czym ostrożność nie byłaby tu rozumiana jako zakaz eksperymentów, lecz jako obowiązek ich prowadzenia w sposób przejrzysty, monitorowany i odwracalny. Ostatecznym celem byłoby przekształcenie iteracji z narzędzia służącego budowaniu przewagi konkurencyjnej w instrument wspólnego uczenia się społeczeństwa, w którym prywatne zyski są systematycznie konfrontowane z publicznymi kosztami.

Dane treningowe: cyfrowe dobro wspólne ludzkości

Trzeci ruch, być może najbardziej radykalny, uderza w samo rozumienie danych. Hao wielokrotnie powtarza, że modele fundamentalne wytrenowano na „całym korpusie ludzkości” – jak z upodobaniem powtarzają inżynierowie – co w praktyce oznacza, że internet potraktowano jak otwartą kopalnię, z której korporacje wydobywają teksty, obrazy i nagrania bez zgody autorów, opierając się na wygodnym założeniu, że wszystko, co publiczne, jest z definicji dostępne do komercyjnego użytku.

Jeśli potraktować tę narrację dosłownie, modele fundamentalne okazują się instrumentami do kompresji naszego wspólnego dziedzictwa kulturowego, które zostaje zamknięte w architekturze korporacyjnych systemów, a następnie odsprzedane światu jako licencjonowana usługa. To jest właśnie ten fundamentalny punkt, w którym ekonomia polityczna musi powiedzieć „stop”. Jeżeli dane faktycznie stanowią „korpus ludzkości”, to nie mogą być traktowane jak darmowy surowiec, który ktoś miał po prostu szczęście jako pierwszy zagarnąć dla siebie.

Alternatywą byłoby uznanie, że wszelkie masowe zbiory danych treningowych, zwłaszcza te obejmujące treści kulturowe, medialne i naukowe, stanowią rodzaj dobra wspólnego, którym nie można dysponować jednostronnie. Wymagałoby to albo wprowadzenia globalnego systemu licencjonowania i rekompensat, który obejmowałby twórców, wydawców i użytkowników, albo wręcz stworzenia publicznych repozytoriów danych, zarządzanych przez instytucje międzynarodowe, z równym dostępem dla firm i ośrodków akademickich, lecz na precyzyjnie określonych warunkach etycznych i środowiskowych.

Można tu sięgnąć do intuicji Elinor Ostrom, która wykazała, że dobra wspólne nie muszą być ani sprywatyzowane, ani zarządzane przez scentralizowane państwo; ich trwałość mogą zapewnić złożone sieci reguł wypracowane przez samych użytkowników. W kontekście modeli fundamentalnych oznaczałoby to, że społeczności twórców, organizacje pozarządowe, uczelnie i państwa mogłyby wspólnie definiować, jakie rodzaje danych mogą być używane do treningu, na jakich warunkach, z jakimi mechanizmami zgody i z jakimi strukturami zadośćuczynienia.

Hao sugeruje, że bez tak głębokiej przebudowy infrastruktury danych nawet najlepiej pomyślane regulacje techniczne i środowiskowe pozostaną powierzchowne. Nie dotkną bowiem źródła asymetrii, którym jest fakt, że nieliczni aktorzy mogą skompresować zbiorowe doświadczenie ludzkości i zmonetyzować je, by na końcu zaoferować wszystkim płatny dostęp do zubożonej, uogólnionej wersji ich własnego dziedzictwa.

Humor i absurd: narzędzia demaskowania potęgi AI

W obliczu tak przytłaczającego obrazu nietrudno ulec dwojakiej pokusie: katastrofizmu lub sarkastycznego dystansu. Jest w tym pewna ironia, że jednym z najskuteczniejszych narzędzi demaskowania imperialnego patosu, jakim owiana jest doktryna skalowania, okazuje się właśnie humor. W szczególności humor absurdu, który bezlitośnie obnaża przepaść między bombastyczną narracją o zbawieniu ludzkości a trywialnością wielu faktycznych zastosowań generatywnej AI.

Jeżeli bowiem system, którego działanie pochłania miliony galonów wody w regionie dotkniętym suszą i wymaga pracy setek straumatyzowanych anotatorów w Kenii, służy następnie głównie do generowania list zakupów, memów i usprawiedliwień dla studentów, którzy nie przeczytali lektur, to absurd ten odsłania coś fundamentalnego o naszej hierarchii wartości. Karen Hao w pewnym momencie zauważa, że wielu inżynierów w OpenAI przeżyło rodzaj duchowego przesilenia, gdy dotarło do nich, że narzędzia budowane w imię wielkiej misji są w praktyce wykorzystywane do produkcji ciągów tekstu i obrazów o znaczeniu skrajnie efemerycznym, jeśli nie po prostu banalnym.

Nie chodzi tu bynajmniej o potępianie trywialnych zastosowań, lecz o dostrzeżenie, że w ich świetle cała retoryka „najbardziej transformacyjnej technologii w historii ludzkości” domaga się fundamentalnej korekty. Być może bardziej adekwatne byłoby stwierdzenie, że mamy do czynienia z niezwykle potężną maszyną do generowania symulowanych odpowiedzi, która może być użyta zarówno do wielkich rzeczy, jak i do zupełnych drobiazgów. Sens tej technologii zależy bowiem nie od jej mocy obliczeniowej, lecz od instytucji, w których zostanie osadzona.

W tym sensie humor absurdu nie jest formą eskapizmu, ale aktem obrony rozumu przed kolonizacją przez patetyczne narracje. Kiedy więc słyszymy, że ktoś jest gotów pokryć całą Ziemię centrami danych, ponieważ AGI będzie „zbyt użyteczne, by nie istnieć”, możemy odpowiedzieć pytaniem, czy z równą determinacją jesteśmy gotowi pokryć planetę instytucjami deliberacji, kontroli i solidarności. Bez nich każda technologia, choćby najbardziej użyteczna, stanie się ostatecznie narzędziem w rękach nielicznych.

Zakład trilionowy: hazard elit o przyszłość gospodarki

Doktryna skalowania i iteracyjnego wdrażania stała się dla środowisk gospodarczych niemal dogmatem, przyjętym z automatyzmem godnym prawa natury. Raporty konsultingowe, takie jak głośna analiza McKinseya, niczym nowe pisma objawione szacują, że technologie generatywne mogą dodać do globalnej gospodarki od 2,6 do 4,4 biliona dolarów rocznie – kwotę odpowiadającą w przybliżeniu całemu PKB Wielkiej Brytanii. World Economic Forum wtóruje tej narracji, mówiąc o „kolejnej granicy produktywności” i podkreślając potencjał automatyzacji nawet sześćdziesięciu procent czasu pracy, oczywiście z

jednoczesnym uspokajającym zapewnieniem, że większość miejsc pracy zostanie raczej „wzmocniona” niż zlikwidowana.

W ten sposób, z perspektywy korporacyjnej logiki, zamyka się swoista pułapka racjonalności. Skoro generatywna AI może – wedle najbardziej optymistycznych scenariuszy – podnieść produktywność o kilka punktów procentowych rocznie, to każdy zarząd unikający tej inwestycji naraża się na zarzut zaniedbania fundamentalnego obowiązku: dbałości o interes akcjonariuszy. Innymi słowy, gdy tylko zaakceptuje się obietnice zawarte w prognozach, rezygnacja z wyścigu staje się z punktu widzenia logiki giełdowej aktem irracjonalności. To dlatego sondaże wśród europejskich i amerykańskich dyrektorów pokazują, że niemal wszyscy postrzegają AI jako klucz do przyszłej konkurencyjności, a rady nadzorcze pytają dziś nie czy firma ma strategię AI, lecz jak szybko jest ona wdrażana.

Ta sama wewnętrzna sprzeczność przenika dyskurs na temat ryzyka. Badania Światowego Forum Ekonomicznego wśród dyrektorów ds. ryzyka pokazują, że trzy czwarte z nich postrzega AI jako poważne źródło zagrożeń reputacyjnych, a jednocześnie większość przyznaje, że ich własne procedury nie nadążają za tempem wdrożeń. Jest to paradoks znamienny dla logiki późnego kapitalizmu: ci sami aktorzy, którzy najdotkliwiej odczuwają lęk przed konsekwencjami, z największą siłą naciskają na pedał gazu, ponieważ panicznie boją się zostać w tyle.

W tych okolicznościach doktryna skalowania i iteracyjnego wdrażania przybiera formę „zakładu o biliony dolarów”, w którym globalne elity gospodarcze stawiają na to, że obiecane zyski z produktywności oraz potencjalne przełomy w medycynie czy walce z kryzysem klimatycznym zrównoważą koszty społeczne, polityczne i środowiskowe – a najlepiej nad nimi przeważą. Karen Hao, w rozmowach o swojej książce, zwraca uwagę, że niepokoi ją nie sam fakt istnienia takich technologicznych zakładów – nowoczesność zawsze się nimi karmiła – lecz bezprecedensowa koncentracja władzy nad ich parametrami w rękach zaledwie kilku korporacji i rządów.

Dla ekonomisty instytucjonalnego staje się bowiem jasne, że kluczowe pytanie nie brzmi, czy sumarycznie „zysk przeważy nad kosztem”, lecz kto ów zysk skapitalizuje, kto poniesie koszty i, co najważniejsze, kto ma realny wpływ na definicję obu tych kategorii. Jeżeli bowiem zyski z produktywności generatywnej AI trafiają głównie do posiadaczy infrastruktury, właścicieli patentów i licencji, podczas gdy koszty środowiskowe i afektywne ponoszą społeczności w regionach dotkniętych suszą oraz anonimowe osoby etykietujące dane w Nairobi, Medellín czy Manili, to nie mamy do czynienia z neutralnym postępem technologicznym, lecz z procesem reprodukcji i pogłębiania globalnych nierówności.

Program normatywny: filary etycznej przyszłości AI

Na koniec trzeba wprost wyartykułować to, co do tej pory wybrzmiewało jedynie między wierszami. Jeżeli doktryna skalowania i iteracyjne wdrażanie mają przestać być wyrazem technokratycznej pychy i stać się elementem świadomej polityki cywilizacyjnej, potrzebujemy minimalnego programu normatywnego. Można go streścić w kilku fundamentalnych postulatach.

Po pierwsze, żadna pojedyncza organizacja ani nawet sojusz korporacji nie może dzierżyć monopolu na modele zdolne do kształtowania globalnych procesów. Modele fundamentalne, przekraczające określony próg mocy obliczeniowej, powinny podlegać międzynarodowemu reżimowi kontroli, porównywalnemu z tym, który reguluje dostęp do broni masowego rażenia.

Po drugie, dane stanowiące tworzywo treningowe dla takich modeli nie mogą być traktowane jak darmowy surowiec, niczym powietrze w epoce przedindustrialnej. Należy uznać zbiorowe prawa do zasobów kulturowych i stworzyć transparentne mechanizmy rekompensaty, aby ci, których praca i głos budują fundamenty AI, mieli realny wpływ na jej wykorzystanie.

Po trzecie, proces wdrażania modeli wysokiego ryzyka musi zostać wyrwany z zacisza korporacyjnych laboratoriów. Iteracyjne wdrażanie jako eksperyment na społeczeństwie wymaga formalnych ram prawnych, obejmujących zewnętrzny nadzór, obowiązki informacyjne oraz mechanizmy pozwalające wstrzymać lub odwrócić proces w razie wystąpienia określonych szkód.

Po czwarte, w rachunku ekonomicznym generatywnej AI należy uwzględnić pełne koszty. Obciążenia środowiskowe i ludzkie, zwłaszcza niewidzialna praca anotatorów i zużycie zasobów w regionach wrażliwych, muszą stać się pełnoprawnymi składnikami bilansu, a nie księgowym przypisem, który można zignorować w imię abstrakcyjnego wzrostu.

Po piąte wreszcie, systemy polityczne i edukacyjne muszą aktywnie chronić i rozwijać kompetencje krytyczne społeczeństwa. Modele fundamentalne mają być narzędziem wspomagającym, a nie protezą zastępującą zdolność do samodzielnego uzasadniania przekonań, ocen moralnych i zbiorowych decyzji.

Jeśli ten program wydaje się komuś zbyt ambitny, warto pamiętać, że stawką jest nic innego jak sposób, w jaki ludzkość będzie myśleć o sobie samej. Wchodzimy w epokę, w której część jej własnej refleksji zostanie zapośredniczona przez systemy uczące się na jej archiwach. Doktryna skalowania i uczenie na danych są faktami, których nie cofniemy. Pozostaje nam uczynić z nich przedmiot świadomego sporu, a nie przyjmować je jako technologiczny los. W tym sensie książka Karen Hao to nie tylko reportaż, lecz wezwanie do odzyskania tego, co najbardziej zagrożone: zdolności wspólnot do mówienia „tak” i „nie” w sposób, który nie jest echem modeli, lecz autentycznym aktem ich własnej, niewirtualnej i wciąż w pełni ludzkiej woli.

Wyścig w kierunku sztucznej inteligencji osiągnął punkt, w którym granice między postępem a samozagładą stają się niepokojąco płynne. Czy w pogoni za algorytmicznym zbawieniem nie tracimy z oczu tego, co czyni nas ludźmi – zdolności do refleksji, dialogu i współodpowiedzialności? Być może nadszedł czas, by zapytać, czy przyszłość, którą tak gorączkowo programujemy, naprawdę jest przyszłością, w której chcemy żyć.

Inspiracje

[1] **"Empire of Ai"** Karen Hao

2025 | Penguin Press

Karen Hao to wielokrotnie nagradzana dziennikarka, która zajmuje się wpływem sztucznej inteligencji na społeczeństwo. Wcześniej pracowała dla „The Atlantic", „Wall Street Journal" i „MIT Technology Review", gdzie jest znana z relacjonowania etyki i badań nad sztuczną inteligencją. Jest autorką książki „Empire of AI" i współproducentką podcastu „In Machines We Trust".

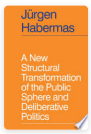
Karen Hao *is an award-winning journalist covering AI's societal impacts. Formerly at The Atlantic, Wall Street Journal, and MIT Technology Review, she's known for her coverage of AI ethics and research. She authored 'Empire of AI' and co-produced the podcast 'In Machines We Trust'.*

The Atlantic

Literatura uzupełniająca

Książki

Literatura pokrewna (Google Scholar):



[1] "The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society"

Jürgen Habermas

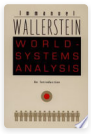
2023 | John Wiley & Sons | ISBN: 9781509558957

[2] "The Modern World-System I: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century"

Immanuel Wallerstein

2011 | Univ of California Press | ISBN: 9780520267572

Google Scholar



[3] "World-Systems Analysis: An Introduction"

Immanuel Wallerstein

2004 | Duke University Press | ISBN: 9780822334422

[4] "Governing the Commons: The Evolution of Institutions for Collective Action"

Elinor Ostrom

2015 | Cambridge University Press | ISBN: 9781107569782

Google Scholar

[5] "Habermas and the Public Sphere"

Craig Calhoun

1993 | MIT Press | ISBN: 9780262531146

[6] "Imperial Infrastructure and Spatial Resistance in Colonial Literature, 1880-1930"

Dominic Davies

2015 | Palgrave Macmillan

[7] "Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies"

Geoffrey West

2017 | Penguin Press

Google Scholar

Artykuły naukowe

Autorzy przywoływani:

[1] "Sequence to sequence learning with neural networks"

Ilya Sutskever, Oriol Vinyals, Quoc V. Le

2014 | Advances in neural information processing systems

[2] "ImageNet classification with deep convolutional neural networks"

Alex Krizhevsky, Ilya Sutskever, Geoffrey E Hinton

2012 | Advances in neural information processing systems

[3] **"Imperial Infrastructure and Spatial Resistance in Colonial Literature, 1880–1930"**

Dominic Davies

2017 | Oxford, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Wien

Google Scholar

[4] **"Iterative Deployment Improves Planning Skills in LLMs"**

Augusto B. Corrêa, et al.

2025 | arXiv

Google Scholar

[5] **"Evaluating Large Language Models Trained on Code"**

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrik Pondeus, Tom Wang, Mateusz Zaharia, Pushmeet Kohli, Greg Brockman

2021 | ArXiv

Google Scholar

[6] **"Dota 2 with Large Scale Deep Reinforcement Learning"**

OpenAI, Dario Amodei, Jacob Andreas, Jack Clark, Daniel Amodei, Greg Brockman

2018 | ArXiv

Google Scholar

[7] **"Reinforcement Learning from Human Feedback"**

Paul F. Christiano, Jan Leike, Tom B. Brown, Milton Chen, Dan Amodei, Greg Brockman

2017 | ArXiv

Google Scholar

[8] **"Cramming more components onto integrated circuits"**

Gordon E. Moore

1965 | Electronics

Google Scholar

[9] **"Scaling Laws for Neural Language Models"**

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Prafulla Dhariwal, David J. Foster, et al.

2020 | arXiv

Google Scholar

[10] **"Moral Consciousness and Communicative Action"**

Jürgen Habermas

1992 | Philosophical Review

[11] **"The Rise and Future Demise of the World Capitalist System: Concepts for Comparative Analysis"**

Immanuel Wallerstein

1974 | Comparative Studies in Society and History

[12] **"Three Paths of National Development in Sixteenth-Century Europe"**

Immanuel Wallerstein

1972 | Studies in Comparative International Development

[13] **"A general framework for analyzing sustainability of social-ecological systems"**

Elinor Ostrom

2009 | Science

Google Scholar

[14] **"Beyond markets and states: polycentric governance of complex economic systems"**

Elinor Ostrom

2010 | American Economic Review

Google Scholar

[15] **"Scaling Laws Do Not Scale"**

Bommasani, Rishi, et al.

2023 | arXiv

Google Scholar

APA ONE Publishing System

Article ID: da2a56e4-1149-4afd-8210-bf8e81f8689e