Nvidia i rewolucja równoległości: technika, władza i los

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Esej rzuca światło na fenomen Nvidii, która z niszowego producenta układów graficznych stała się architektem współczesnej rzeczywistości cyfrowej. Kluczem do tego sukcesu okazało się przejście od przetwarzania sekwencyjnego do równoległego, ucieleśnione w architekturze GPU i platformie CUDA. Tekst analizuje, jak ta techniczna zmiana umożliwiła rozkwit głębokich sieci neuronowych, redefiniując pojęcie mocy obliczeniowej. Autor wykracza poza aspekty techniczne, badając geopolityczne konsekwencje dominacji Nvidii oraz etyczne wyzwania związane z ogromnym zapotrzebowaniem energetycznym centrów danych. To kompleksowe spojrzenie na to, jak krzemowa technologia kształtuje współczesną władzę, gospodarkę uwagi oraz przyszłość ludzkiej pracy w dobie wszechobecnej sztucznej inteligencji.

This essay illuminates the phenomenon of Nvidia, which has transformed from a niche graphics chip manufacturer into the architect of the modern digital reality. The key to this success was the shift from sequential to parallel processing, embodied in the GPU architecture and the CUDA platform. The text examines how this technical shift enabled the rise of deep neural networks, redefining the concept of computing power. The author goes beyond the technical aspects to examine the geopolitical implications of Nvidia's dominance and the ethical challenges posed by the enormous energy demands of data centers. This comprehensive look at how silicon technology is shaping contemporary power, the attention economy, and the future of human work in the age of ubiquitous artificial intelligence.

Artykuł analizuje fenomen firmy Nvidia, postrzegając ją nie tylko jako technologicznego giganta, ale jako symbol transformacji cywilizacyjnej opartej na równoległym przetwarzaniu danych. Tezą jest, że Nvidia stała się kluczowym architektem nowej infrastruktury obliczeniowej, która redefiniuje naukę, gospodarkę i kulturę, wykraczając daleko poza branżę gier komputerowych. Autor argumentuje, że sukces Nvidii wynika ze zbiegu trzech czynników:

kultury zafascynowanej połączeniem człowieka z maszyną, ekonomii poszukującej wzrostu produktywności oraz odrodzenia zainteresowania sztuczną inteligencją. Analiza obejmuje ewolucję strategii firmy, od porażki z NV1 do dominującej roli w standaryzacji i przyspieszaniu innowacji. Artykuł bada konsekwencje globalnej dominacji Nvidii z perspektywy USA, Europy i świata arabskiego, ukazując odmienne podejścia do innowacji, regulacji i modernizacji. Podkreśla ambiwalentny charakter technologii GPU, która z jednej strony napędza postęp, z drugiej zaś generuje zagrożenia związane z nadzorem, dezinformacją i nierównościami. Ostatecznie, artykuł stawia pytanie o przyszłość, w której dostęp do mocy obliczeniowej powinien być traktowany jako dobro wspólne, podlegające demokratycznej kontroli.

SŁOWA KLUCZOWE

Nvidia | obliczenia równoległe | GPU | CUDA | architektura równoległa | sztuczna inteligencja | uczenie głębokie | AlexNet | prawo Moore'a | infrastruktura obliczeniowa | sieci neuronowe | akceleracja | geopolityka AI | ślad węglowy | obliczenia przyspieszone

Nvidia: filozoficzne źródła cyfrowej ontologii

Jeżeli spojrzeć na historię techniki jak na szczególny przypadek dziejów idei, narodziny Nvidii przestają być jedynie epizodem z kronik Doliny Krzemowej. Stają się momentem, w którym pewien model technicznej racjonalności zaczyna organizować cały nasz świat przeżywany: od gier komputerowych, przez naukę i finanse, aż po geopolitykę. Firma, założona w 1993 roku przez Jensena Huanga, Jeffa i Jima, w scenerii tak trywialnej, że aż symbolicznej – w restauracji Taco Bell – okazała się czymś więcej niż producentem układów scalonych. Stała się ośrodkiem, w którym skrystalizował się nowy paradygmat cywilizacyjny, którego słowem kluczem jest równoległość przetwarzania, a głębszym celem – radykalne przyspieszenie procesów tworzenia wiedzy i władzy obliczeniowej.

Aby jednak zrozumieć fenomen Nvidii nie jako przypadkową historię sukcesu, lecz jako odpowiedź na strukturalne napięcia nowoczesnego kapitalizmu, musimy porzucić kuszącą narrację heroiczną, skupioną na charyzmatycznym założycielu i rynkowej innowacji. Zamiast tego należy zrekonstruować te warunki możliwości, w których zbiegły się trzy potężne nurty. Były to: kultura wyobrażająca sobie przyszłość człowieka w nierozerwalnym splocie z maszyną, ekonomia poszukująca nowych motorów produktywności oraz nauka, która po latach stagnacji, zwanych zimą AI, na nowo uczyła się myśleć w kategoriach głębokich sieci neuronowych i uczenia statystycznego.

Grafika 3D wymusza paradygmat równoległości

Narodziny Nvidii w 1993 roku przypadają na moment, w którym grafika trójwymiarowa, dotąd elitarna, zaczyna opuszczać hermetyczny świat laboratoriów filmowych i stacji roboczych, by trafić do masowego użytkownika. Symbolicznym tłem staje się zarówno estetyka filmów pokroju „Parku Jurajskiego", jak i ontologiczna ambicja przemysłu gier, który pragnie przekroczyć iluzję płaskiego ekranu na rzecz przestrzeni możliwej do zanurzenia. Ta z pozoru błaha rewolucja estetyczna tworzy jednak nowe warunki możliwości, wymuszając przejście od obliczeń sekwencyjnych do architektur przystosowanych do jednoczesnego przetwarzania tysięcy wierzchołków, pikseli, cieni i tekstur.

NV1: rynkowa klęska zbyt ambitnej architektury

Pierwszy produkt, NV1, okazał się spektakularną porażką. Jego koncepcja kwadratowego mapowania tekstur była technicznie elegancka, lecz systemowo obca temu, co przygotowywał Microsoft w ramach DirectX i czego faktycznie oczekiwali deweloperzy gier. Fiasko współpracy z Segą, zalegające w magazynach karty, miażdżące recenzje – z perspektywy księgowej była to sytuacja bliska katastrofie. Jednak z perspektywy logiki systemowej był to moment założycielski: Nvidia przechodzi przyspieszony kurs, z którego płynie jedna, fundamentalna lekcja. W epoce standaryzacji protokołów nowatorstwo formalne nie ma żadnej wartości, jeśli nie wpisuje się w techniczny konsensus społeczności praktyków.

W tym miejscu rodzą się dwie zasady, które w perspektywie długiego trwania ukształtują tożsamość Nvidii. Po pierwsze, absolutna gotowość do porzucenia własnych, nawet genialnych rozwiązań, gdy ich dalsza obrona prowadzi do ekosystemowej izolacji. Po drugie, zamiast budować markę na hermetycznej wyjątkowości, należy stać się strażnikiem standardu, który jednak zmienia reguły gry poprzez tempo i skalę jego wdrażania. Nvidia błyskawicznie pojmuje, że jej prawdziwym produktem nie jest pojedynczy chip. Jest nim zdolność do rytmicznego narzucania całemu rynkowi kolejnych generacji sprzętu i oprogramowania, co czyni z niej swoistego operatora cykli innowacji.

GPU vs CPU: starcie hierarchii z kolektywem

Na najbardziej elementarnym poziomie technicznym architektura GPU stanowi zaprzeczenie logiki CPU. Zamiast kilku wyspecjalizowanych, niezwykle szybkich rdzeni dysponuje ona setkami lub tysiącami prostszych jednostek, zdolnych do wykonywania tej samej operacji na wielu danych jednocześnie. Jest to technologiczny odpowiednik sytuacji, w której jednego genialnego szefa kuchni, przygotowującego kolejne

dania w skupieniu, zastępuje armia kuchennych automatów. Każdy z nich wykonuje swoją prostą czynność identycznie, bez zmęczenia i bez refleksji, lecz za to w skali, która wymyka się pojedynczej percepcji.

W kategoriach antropologicznych jest to nic innego jak przejście od modelu inteligencji rozumianej jako sekwencyjna refleksja jednostkowego podmiotu, do modelu, w którym funkcjonalną całość tworzy populacja prostych procesów. Żaden z nich nie jest z osobna inteligentny, lecz ich masowe, zsynchronizowane działanie generuje rezultaty przewyższające intuicję pojedynczego umysłu. To właśnie takie emergentne wytwory równoległości dały nam systemy pokroju Deep Blue, TD-Gammon, a wreszcie AlexNet – kamienie milowe sztucznej inteligencji.

Nieprzypadkowo zatem Jensen Huang, założyciel Nvidii, formułuje tezę, która brzmi jak epitafium dla całej epoki. Twierdzi on, że klasyczne prawo Moore'a dobiegło końca, a drogą naprzód są obliczenia przyspieszone. Rozpoczyna się nowa era, w której o dynamice rozwoju nie zdecyduje już prosty przyrost liczby tranzystorów, lecz właśnie architektura masowego przyspieszenia.

CUDA: demokratyzacja mocy obliczeniowej GPU

Wprowadzenie architektury CUDA około 2006 roku wyznacza moment, w którym NVIDIA świadomie wychodzi poza rolę producenta akceleratorów graficznych, by stać się architektem nowej, ogólnej infrastruktury obliczeniowej. Istota tej transformacji polega na fundamentalnej zmianie perspektywy: programista przestaje postrzegać procesor graficzny jako hermetyczne, wyspecjalizowane urządzenie obsługujące wyłącznie potok graficzny. Zamiast tego zaczyna go traktować jako uniwersalną maszynę do wykonywania dowolnych równoległych obliczeń macierzowych, które stanowią przecież podstawę nie tylko grafiki, lecz także symulacji fizycznych, obliczeń finansowych, statystyki i uczenia maszynowego.

Z punktu widzenia teorii wiedzy technicznej CUDA jest próbą zinstytucjonalizowania nowej formy racjonalności. Zamiast oczekiwać, że pojedynczy algorytm zrealizuje inteligentne zadanie w czystej, liniowej sekwencji, przyjmuje się, że inteligencja jest funkcją zdolności do równoczesnego operowania na wielowymiarowych strukturach danych. Można zatem powiedzieć, że CUDA staje się swoistą gramatyką, która tworzy warunki możliwości dla artykulacji tej nowej formy rozumu w językach C i C++, a więc w narzędziach, którymi biegle posługują się już istniejące społeczności programistyczne.

AlexNet: triumf głębokiego uczenia w 2012 roku

Prawdziwe, historyczne znaczenie tej decyzji ujawnia się dopiero w roku 2012, kiedy AlexNet – splotowa sieć neuronowa autorstwa Krizhevsky'ego, Sutskevera i Hintona – wykorzystuje konsumenckie karty

GeForce GTX 580 i programowanie w CUDA, by osiągnąć deklasującą wręcz przewagę w konkursie ImageNet, redukując błąd klasyfikacji top-5 o ponad dziesięć punktów procentowych względem najbliższego rywala.

AlexNet nie jest tu jednak indywidualnym bohaterem, lecz raczej symptomem konwergencji trzech fundamentalnych procesów: istnienia wielkiego, ustandaryzowanego zbioru danych, powszechnej dostępności obliczeń na GPU oraz udoskonalonych metod trenowania sieci. Należy przy tym z całą mocą podkreślić, że bez platformy CUDA nauczenie tak rozległego modelu w akceptowalnych ramach czasowych i budżetowych byłoby po prostu niemożliwe, co jednoznacznie potwierdzają analizy historyczne boomu na uczenie głębokie.

W tym właśnie punkcie Nvidia przestaje być aktorem jednej branży, a staje się fundamentem nowego paradygmatu naukowego i gospodarczego. Akceleracja obliczeń na GPU uruchamia bowiem lawinę badań w dziedzinie uczenia głębokiego, torującą drogę architekturze Transformer, która w roku 2017 redefiniuje przetwarzanie języka naturalnego, co w konsekwencji prowadzi do powstania wielkich modeli językowych i generatywnych systemów AI, współkształtujących dziś praktykę życia codziennego, pracy naukowej i działalności gospodarczej.

W tym sensie, gdy Nvidia twierdzi, że sztuczna inteligencja jest najpotężniejszą siłą technologiczną naszych czasów, nie jest to jedynie chwytliwe hasło z konferencji prasowej. To precyzyjny opis procesu, w którym GPU i oprogramowanie takie jak CUDA, TensorRT czy cuDNN stają się współczesnym odpowiednikiem dziewiętnastowiecznej sieci energetycznej i kolei. To już nie komponent, lecz infrastruktura. A zatem jest to narzędzie współdefiniujące same warunki możliwości tego, co w ogóle uchodzi za wykonalne w praktyce gospodarczej, naukowej i kulturowej.

USA, Europa i świat arabski: geopolityka AI

W Stanach Zjednoczonych rewolucję obliczeń równoległych postrzega się jako kolejny akt w dobrze znanym spektaklu kapitalizmu innowacyjnego. Scenariusz jest prosty: rynek, napędzany kapitałem wysokiego ryzyka, finansuje śmiałe technologiczne przedsięwzięcia, a te najbardziej udane ewoluują w prywatne infrastruktury o znaczeniu niemal publicznym. Nvidia nie jest tu wyjątkiem, lecz kontynuacją linii wytyczonej przez Intela czy Amazon Web Services, gdzie jedna firma staje się faktycznym operatorem kluczowych zasobów dla całej gospodarki. Amerykański dyskurs ekonomiczny widzi w AI przede wszystkim potężne narzędzie do turbodoładowania produktywności i cięcia kosztów. Ryzyka, takie jak ogromna koncentracja władzy, są konsekwentnie bagatelizowane jako problemy, z którymi doskonale poradzi sobie rynek, mający rzekomo korygować wszelkie monopolistyczne zapędy.

Europa patrzy na ten sam krajobraz z zupełnie innej perspektywy, widząc w dominacji Nvidii przede wszystkim zagrożenie dla swojej autonomii regulacyjnej i suwerenności technologicznej. Odpowiedzią Unii na rosnącą zależność infrastrukturalną jest próba budowy normatywnej fortecy, której zwieńczeniem ma być ambitna ustawa o sztucznej inteligencji. Równolegle Bruksela usiłuje stymulować własną politykę przemysłową, która poprzez programy takie jak IPCEI ma tworzyć lokalne łańcuchy wartości w produkcji czipów. Jednak brak kapitału wysokiego ryzyka na amerykańską skalę oraz głęboko zakorzeniona zasada ostrożności sprawiają, że Stary Kontynent porusza się znacznie wolniej. To z kolei potęguje egzystencjalny lęk przed technologiczną marginalizacją, w którym Nvidia staje się symbolem głębokiej ambiwalencji: jest zarazem niezbędnym partnerem i bolesnym przypomnieniem o zależności od amerykańskiego ekosystemu.

Jeszcze inaczej rozkładają się akcenty w świecie arabskim, zwłaszcza w państwach Zatoki Perskiej. Dysponując gigantycznym kapitałem z eksportu surowców, widzą one w rewolucji AI historyczną szansę na technologiczny skok naprzód – przyspieszoną dywersyfikację gospodarki w kierunku wiedzy, z pominięciem długiego etapu industrializacji opartej na przemyśle ciężkim. Masywne inwestycje w centra danych, klastry obliczeniowe i instytuty badawcze stają się fundamentem nowej strategii rozwoju, a Nvidia odgrywa w niej rolę kluczowego partnera infrastrukturalnego, symbolu awansu do globalnej pierwszej ligi. Jest jednak i druga, mroczniejsza strona tego medalu: w rękach autorytarnych reżimów te same narzędzia stają się potężnym instrumentem nie tylko modernizacji, ale także masowego nadzoru i cyfrowej kontroli nad społeczeństwem.

Mamy zatem trzy odmienne opowieści: amerykańską o nieokiełznanej innowacji, europejską o regulacyjnej odpowiedzialności i arabską o modernizacyjnym zrywie. A jednak we wszystkich tych narracjach Nvidia funkcjonuje jako niezmienny, wspólny mianownik. Jest soczewką, przez którą widać globalne ambicje i lęki, uobecniając się zarówno w abstrakcyjnym dyskursie ekonomicznym, jak i w bardzo konkretnej praktyce budowy nowej, globalnej infrastruktury obliczeniowej.

Równoległość obliczeń napędza gospodarkę uwagi

W wymiarze kulturowym rewolucja obliczeń równoległych dokonuje głębokiej dekonstrukcji dotychczasowych sposobów doświadczania czasu i uwagi. Platformy mediów społecznościowych, których architektura jest wprost uzależniona od mocy GPU, przyzwyczajają nas do życia w stanie permanentnej fragmentacji, gdzie setki równoległych strumieni informacji konkurują o zasoby percepcyjne jednostki. To, co było niewinną infrastrukturą dla wirtualnych światów gier i symulatorów 3D, staje się fundamentem gospodarki uwagi, w której techniczna równoległość przetwarzania danych nierozerwalnie sprzęga się z równoległością bodźców psychicznych.

Fabryki AI: prawne wyzwania nowej infrastruktury

Krytyczna rekonstrukcja musi jednak zdemaskować fundamentalne złudzenie ukryte w tej narracji: obraz, w którym postęp sprzętowy niemal automatycznie przekłada się na postęp społeczny. U jego podstaw leży niebezpieczne założenie, że sam wzrost mocy obliczeniowej jest równoważny wzrostowi racjonalności. Tymczasem przyspieszone obliczenia to narzędzie moralnie obojętne. Może ono z równą, a może nawet większą skutecznością służyć wzmacnianiu mechanizmów dezinformacji, zacieśnianiu pętli nadzoru i doskonaleniu komercyjnej manipulacji niż rozwiązywaniu problemów klimatycznych czy medycznych.

Można tu przywołać jedną z kluczowych intuicji współczesnej teorii krytycznej, zgodnie z którą nastawienie czysto instrumentalne, niezakotwiczone w publicznej debacie o celach, prowadzi do patologii systemu. Czyniąc z narzędzi racjonalności cel sam w sobie, odcina ono je od horyzontu sensu zakorzenionego w świecie przeżywanym. Przekładając ten filozoficzny język na rewolucję GPU, można powiedzieć, że architektury typu Blackwell, mimo swej imponującej wydajności, nie rozwiązują ani jednego problemu normatywnego. Nie mówią nam, jak rozdzielać owoce wzrostu produktywności, nie rozstrzygają, gdzie kończy się dopuszczalne profilowanie, ani nie usuwają napięć między wolnością a bezpieczeństwem.

Moc obliczeniowa nie gwarantuje racjonalności

Równoległość, w wymiarze technicznym będąca elegancką odpowiedzią na schyłek prawa Moore’a, w skali globalnej odsłania jednak swoją mroczną stronę, zarówno środowiskową, jak i antropologiczną. Budowane w Arabii Saudyjskiej, Zjednoczonych Emiratach Arabskich czy Stanach Zjednoczonych fabryki AI to w istocie gigantyczne centra energetyczne, wymagające tysięcy megawatów mocy i potężnych systemów chłodzenia. Szacunki są wymowne: pojedynczy ośmioprocesorowy system DGX B200 pochłania ponad czternaście kilowatów, a klastry złożone z setek takich maszyn osiągają moc niewielkiej elektrowni. Rewolucja obliczeń równoległych okazuje się zatem rewolucją energetyczną w najczystszej postaci.

W tym kontekście rodzi się fundamentalna aporia. Z jednej strony sztuczną inteligencję przedstawia się jako narzędzie do optymalizacji zasobów i walki ze zmianą klimatu; z drugiej zaś sama generuje olbrzymi ślad węglowy, wymuszając budowę nowych mocy wytwórczych, nierzadko wciąż opartych na paliwach kopalnych. W świecie arabskim, zwłaszcza w krajach bogatych w ropę i gaz, pojawia się pokusa, by ten gigantyczny apetyt na energię przedstawiać jako formę wartości dodanej – jako sposób na „zmonetyzowanie” surowców w postaci zaawansowanych usług obliczeniowych. Z perspektywy długiego trwania cywilizacji oznacza to jednak, że logika równoległości może w końcu zderzyć się z ograniczeniami, jakie narzuca świat przeżywany, w którym ludzkie ciała i całe ekosystemy mają swój nieprzekraczalny próg tolerancji na eksploatację.

Energetyczny koszt rewolucji równoległej

W świecie pracy równoległość obliczeniowa GPU objawia swój dwoisty, niemal ambiwalentny charakter. Z jednej strony pozwala zautomatyzować rutynowe zadania, które dotąd pochłaniały lwią część czasu i uwagi, potencjalnie uwalniając ludzką kreatywność od monotonii przetwarzania danych. Z drugiej jednak strony wprowadza nowe, subtelne formy algorytmicznej kontroli. Pracownik staje się wymiennym modułem w systemie nieustannego monitorowania, optymalizacji i predykcji, a jego autonomia ulega systematycznemu ograniczeniu. Procesy decyzyjne, które dotąd opierały się na interakcji i negocjowaniu racji, zostają przeniesione do modeli głębokiego uczenia. W efekcie roszczenia do autentyczności czy normatywnej słuszności ulegają formalizacji: zostają sprowadzone do parametrów funkcji straty i wag w sieci neuronowej.

Z perspektywy antropologii pracy rodzi się zatem fundamentalne pytanie: czy człowiek ery GPU będzie postrzegał siebie jako sprawcę historii, czy już tylko jako operatora systemów, które przechowują, przetwarzają i projektują jego własne warunki możliwości? Być może jednym z najważniejszych wyzwań staje się stworzenie takiej kultury zawodowej, w której modele AI nie są traktowane jak czarne skrzynki wydające wyroki. Muszą stać się raczej uczestnikami praktyki nastawionej na wzajemne zrozumienie, co z kolei wymaga radykalnego podniesienia kompetencji krytycznych i otwarcia kodu, modeli oraz danych na szeroki audyt społeczny.

Kiedy więc Jensen Huang w swoich wystąpieniach odwołuje się do charakteru i zaufania jako fundamentów przywództwa, można to odczytać nie tylko jako intuicję. To głębokie przeczucie, że bez odbudowy zaufania do instytucji zarządzających infrastrukturą AI, każda, nawet największa przewaga techniczna ostatecznie zamieni się w źródło oporu. Zaufanie jest bowiem niczym innym jak społecznym, równoległym przetwarzaniem ryzyka i nadziei. To ludzki odpowiednik procesora graficznego, w którym tysiące indywidualnych doświadczeń i ocen składają się na wspólną, zbiorową ocenę wiarygodności systemu.

AI redukuje ludzkie sprawstwo w procesie pracy

Dopiero gdy uznamy, że infrastruktura GPU nie jest kolejną warstwą techniki, lecz samym medium, w którym krystalizują się przyszłe formy władzy, wolności i rozumu, możemy sensownie szkicować scenariusze przyszłości. Unikniemy w ten sposób zarówno naiwnej apologii innowacji, jak i taniego katastrofizmu. W tym horyzoncie wyłaniają się trzy idealne modele, trzy reżimy równoległości, które nie tyle opisują konkretne państwa czy korporacje, ile wyznaczają wektory sił, ku którym dryfuje cały system.

Pierwszy reżim można określić mianem korporacyjnego, choć trafniej byłoby mówić o infrastrukturze prywatno-imperialnej. W tym wariancie Nvidia staje się figurą modelową, w której zbiega się potęga kapitału finansowego, inżynieryjna maestria i globalna sieć partnerstw z gigantami chmurowymi oraz

operatorami centrów danych. Równoległość obliczeń zostaje tu wprzęgnięta w logikę nieustannej akumulacji, gdzie nadrzędnym kryterium jest stopa zwrotu z inwestycji, a dopiero wtórnym – choć głośno deklarowanym – poprawa losu ludzkości. Gdy prezes firmy ogłasza, że AI będzie wbudowana we wszystko, od aut po fabryki, wyraża nie tylko techniczną możliwość, lecz także roszczenie do tego, by każdy sektor gospodarki stał się klientem wąskiej grupy dostawców. Prawo antymonopolowe próbuje nadążyć, lecz ze względu na transnarodowy charakter mocy obliczeniowej pozostaje głównie reaktywnym instrumentem symbolicznego oporu.

Drugi reżim ma charakter państwowy i wyrasta z przekonania, że równoległość obliczeń to zasób strategiczny, na równi z energią, wodą czy terytorium. W tym scenariuszu państwa budują własne superkomputery AI, uruchamiają narodowe programy produkcji półprzewodników i negocjują dostęp do GPU na najwyższych szczeblach dyplomacji. Rozwój sztucznej inteligencji staje się elementem doktryny bezpieczeństwa narodowego. Nvidia jest tu traktowana nie jako zwykły producent, lecz partner w budowie „narodowych fabryk AI”, przy czym partnerstwo to jest zawsze podszyte napięciem – każda strona dąży do maksymalizacji własnej autonomii. Reżim państwowy ma tę pozorną zaletę, że pozwala na wprowadzenie silnych regulacji, jednak jego fundamentalna sprzeczność polega na tym, że w imię ochrony obywateli przed arbitralnością rynku z łatwością przechodzi w arbitralność władzy, która widzi w AI doskonałe narzędzie nadzoru i prewencji.

Trzeci reżim, dziś wciąż niedoreprezentowany, można nazwać wspólnotowym, czyli opartym na dobru wspólnym. Jego fundamentem jest założenie, że skoro moc obliczeniowa staje się warunkiem możliwości uczestnictwa w życiu publicznym, powinna być współwłasnością wspólnot. W tym wariancie instytucje publiczne, uczelnie i organizacje społeczne tworzą konsorcja budujące otwarte klastry GPU. Rozwijają modele o otwartych wagach, ustanawiają przejrzyste procedury audytu i wspólnie decydują, które zastosowania są dopuszczalne, a które naruszają integralność naszego świata przeżywanego. W takim scenariuszu Nvidia mogłaby stać się partnerem w projektach, których logika nie redukuje się do zysku, lecz obejmuje cele naukowe, edukacyjne i emancypacyjne.

Te trzy reżimy nie są teoretycznym wyborem z katalogu, lecz codzienną areną zmagań, na której pojedyncze decyzje inwestycyjne, regulacyjne i kulturowe przesuwają wektor całego systemu. Jeśli przyjmiemy, że pewne elementy reżimu korporacyjnego i państwowego są nieuniknione, to zadaniem teoretyka i praktyka polityki publicznej staje się maksymalizacja komponentu wspólnotowego. Chodzi o takie rozwiązania, które wprowadzają do

Trzy reżimy równoległości: przyszłość techniki

Kiedy Jensen Huang na GTC 2024 ogłaszał architekturę Blackwell, mówił o procesorze na erę generatywnej AI, stwierdzając przy tym, że infrastruktura sztucznej inteligencji staje się równie podstawowa, jak elektryczność i internet. Ta deklaracja jest czymś więcej niż tylko opisem faktu technicznego; to zarazem postulat normatywny, który wyznacza kierunek rozwoju. Jeżeli bowiem AI ma pełnić rolę tak fundamentalną, to logiczną konsekwencją staje się pytanie, czy dostęp do mocy obliczeniowej GPU i modeli nie powinien być traktowany jako pewien rodzaj dobra wspólnego. Dobra, którego kluczowe parametry podlegają deliberacji demokratycznej, a nie wyłącznie negocjacjom kontraktowym między korporacjami a państwami.

Właśnie w tym punkcie otwiera się przestrzeń dla propozycji politycznych, które wykraczają poza bezpieczne centrum. Jedna z nich mogłaby domagać się wprost, aby klastry GPU po przekroczeniu pewnej skali mocy obliczeniowej były prawnie uznawane za infrastrukturę krytyczną, a co za tym idzie – podlegały szczególnemu reżimowi przejrzystości, audytu i kontroli społecznej. Inna, jeszcze bardziej radykalna wizja mogłaby argumentować za częściową społeczną własnością tej infrastruktury. Przybrałaby ona formę konsorcjów publiczno-prywatnych z udziałem obywateli, środowisk akademickich i organizacji pozarządowych, co przypominałoby historyczne modele własności mieszanej w sektorach energetyki czy telekomunikacji.

Takie postulaty, choć dla wielu dzisiejszych decydentów gospodarczych brzmią jak abstrakcyjna teoria, mają swoją wewnętrzną, twardą racjonalność. Skoro bowiem, jak zauważa Paul Scharre, inteligencja jest najpotężniejszą siłą na naszej planecie, a Nvidia staje się jednym z głównych dostawców jej współczesnego, materialnego nośnika, to rezygnacja z debaty o formach własności i kontroli nad tą infrastrukturą byłaby niczym innym, jak dobrowolnym oddaniem jednego z kluczowych instrumentów władzy w historii.

Infrastruktura AI jako cyfrowe dobro wspólne

Analiza dotychczasowych argumentów prowadzi do jedynej racjonalnej konkluzji, która odrzuca zarówno naiwną apoteozę technologii, jak i jej lękliwą negację. Należy zatem bronić tezy, że przyspieszone obliczenia i architektury GPU, których symbolem stała się Nvidia, są jednymi z najpotężniejszych narzędzi do poszerzania horyzontów poznania i działania, jakie kiedykolwiek stworzono. Jednocześnie trzeba jednak domagać się, by narzędzia te podlegały tym samym rygorom, co każde inne roszczenie do ważności, a więc by sposób ich użycia był przedmiotem publicznej debaty, krytyki i nieustannej korekty.

Jeśli więc przyszły historyk idei zada sobie pytanie, dlaczego na przełomie drugiej i trzeciej dekady XXI wieku rozwój jednej firmy z Doliny Krzemowej okazał się kluczowy dla losów cywilizacji, odpowiedzi nie znajdzie w suchych wykresach mocy obliczeniowej. Będzie musiał jej szukać w tym, czy ludzkość zdołała

przełożyć równoległość obliczeń na równoległość głosów w debacie publicznej, na równoległość perspektyw w nauce i na równoległość narracji w kulturze. Tylko wtedy bowiem rewolucja Nvidii nie byłaby zaledwie kolejnym etapem technicznego przyspieszenia, lecz momentem, w którym sztuczna inteligencja stała się katalizatorem głębszej inteligencji społecznej.

Jeżeli natomiast owa równoległość pozostanie przywilejem nielicznych, a większość z nas będzie doświadczać AI jedynie jako nieprzejrzystej siły decydującej o kredycie, pracy czy zdrowiu, historia zapamięta Nvidię inaczej. Jej powstanie zostanie zapisane jako preludium do nowej formy urzeczowienia świadomości, w której nie tylko relacje międzyludzkie, ale i same struktury myślenia zostaną podporządkowane bezdusznej logice czarnych skrzynek.

Wybór między tymi dwiema ścieżkami nie dokona się automatycznie. Żaden układ GPU nie zaprojektuje nam sprawiedliwego świata, tak jak żadna sieć neuronowa nie zastąpi odwagi w podejmowaniu ryzyka w imię lepszego argumentu. Ostatecznym zadaniem nie jest więc tylko budowanie coraz potężniejszych procesorów, lecz odbudowa intelektualnej śmiałości, która pozwala spojrzeć na te masywne maszyny bez nabożnej trwogi i bez cynicznej obojętności. Chodzi o to, by traktować je jako narzędzia, które dopiero poprzez praktykę nastawioną na wzajemne zrozumienie mogą stać się częścią ludzkiej historii, a nie tylko jej anonimowym tłem.

> Czy równoległość obliczeń, tak sprawnie opanowana przez Nvidię, znajdzie odzwierciedlenie w równoległości głosów w debacie publicznej, perspektyw w nauce i narracji w kulturze? Jeśli nie, powstanie firmy zostanie zapamiętane jako preludium do nowej formy urzeczowienia świadomości, gdzie nawet struktury myślenia zostaną podporządkowane bezdusznej logice czarnych skrzynek. Czy zdołamy spojrzeć na te masywne maszyny bez nabożnej trwogi i obojętności, traktując je jako narzędzia w służbie ludzkiej historii?

Inspiracje

[1] **"The Thinking Machine: Jensen Huang, Nvidia, and the World's Most Coveted Microchip"** Jensen Huang

2025 | Penguin Publishing Group/Penguin Random House

Jensen Huang jest współzałożycielem, prezesem i dyrektorem generalnym firmy NVIDIA. Jest inżynierem elektrykiem i przedsiębiorcą, kierującym ekspansją firmy NVIDIA w obszarze procesorów graficznych, obliczeń o wysokiej wydajności i sztucznej inteligencji. Huang uzyskał tytuł licencjata (BS) na Oregon State University i tytuł magistra (MS) na Stanford University.

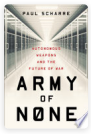
Jensen Huang *is the co-founder, president, and CEO of NVIDIA. He is an electrical engineer and entrepreneur, leading NVIDIA's expansion in GPUs, high-performance computing, and AI. Huang received a BS from Oregon State and an MS from Stanford.*

NVIDIA

Literatura uzupełniająca

Książki

Literatura pokrewna (Google Scholar):

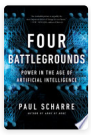


[1] "Army of None: Autonomous Weapons and the Future of War"

Paul Scharre

2018 | W. W. Norton & Company | ISBN: 9780393608991

Google Scholar



[2] "Four Battlegrounds: Power in the Age of Artificial Intelligence"

Paul Scharre

2023 | W. W. Norton & Company | ISBN: 9780393866872

[3] "Deep Learning: Foundations and Concepts"

Geoffrey Hinton

2024 | Springer Nature

Google Scholar

Artykuły naukowe

Autorzy przywoływani:

[1] "Cramming more components onto integrated circuits"

Gordon E. Moore

1965 | Electronics

Google Scholar

[2] "Recent Trends in Parallel Computing"

Gurinder Singh, Aman Kaura

2016 | GIAN JYOTI E-JOURNAL

Google Scholar

[3] **"ImageNet Classification with Deep Convolutional Neural Networks"**

Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton

2012 | Advances in Neural Information Processing Systems

Google Scholar

[4] **"Learning Multiple Layers of Features from Tiny Images"**

Alex Krizhevsky

2008

[5] **"ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst"**

Mayank Bansal, Abhijit Ogale, et al.

2019 | None

Google Scholar

[6] **"Sequence to sequence learning with neural networks"**

Ilya Sutskever, Oriol Vinyals, Quoc V. Le

2014 | Advances in neural information processing systems

Google Scholar

[7] **"Dropout: a simple way to prevent neural networks from overfitting"**

Nitish Srivastava, Geoffrey E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov

2014 | Journal of Machine Learning Research

[8] **"State of the Art in Parallel and Distributed Systems: Emerging Trends and Challenges"**

Sukhpal Singh, Amandeep Kaur

2024 | N/A

Google Scholar

[9] **"Artificial Intelligence and Arms Control"**

Paul Scharre, Megan Lamberth

2022

Google Scholar

[10] **"Human Control of Autonomous Systems"**

Anthony Seaboyer, Paul Scharre

2022

Google Scholar

APA ONE Publishing System

Article ID: 686581e0-e865-46fc-89b5-8fbe0b32c361