Serce nauki: uzasadnienie jako fundament wiedzy

AUTOR

Fundacja Dobre Państwo

STRESZCZENIE / ABSTRACT

Współczesna nauka często błędnie koncentruje się wyłącznie na finalnym wyniku, traktując go jako certyfikat prawdy. Jacob Stegenga proponuje radykalną zmianę perspektywy, w której to rygorystyczna jakość procesu uzasadniania staje się najważniejszym wyznacznikiem wartości wiedzy. Tekst bada, jak społeczna epistemologia i instytucjonalne mechanizmy audytu mogą przeciwdziałać kryzysowi replikacyjnemu oraz zjawiskom takim jak upolitycznienie wnioskowania naukowego. Autor analizuje architekturę produkcji wiedzy, wskazując na konieczność samokorekty instytucjonalnej w obliczu walki narracji na symbolicznym rynku idei. To lektura obowiązkowa dla osób zainteresowanych filozofią nauki, które chcą zrozumieć, dlaczego rzetelność procedur jest ważniejsza od heroicznych prób odkrycia ostatecznej prawdy.

Contemporary science often mistakenly focuses solely on the final result, treating it as a certificate of truth. Jacob Stegenga proposes a radical shift in perspective, in which the rigorous quality of the justification process becomes the most important determinant of knowledge's value. The text examines how social epistemology and institutional audit mechanisms can counteract the replication crisis and phenomena such as the politicization of scientific reasoning. The author analyzes the architecture of knowledge production, pointing to the necessity of institutional self-correction in the face of narrative battles in the symbolic marketplace of ideas. This is essential reading for those interested in the philosophy of science who want to understand why procedural integrity is more important than heroic attempts to discover ultimate truth.

Współczesna nauka ugrzęzła w kulcie heroicznego odkrywcy i fetyszu prawdy, co prowadzi do epistemologicznego kryzysu zaufania. Jacob Stegenga proponuje radykalną zmianę paradygmatu: zamiast rozliczać badaczy z ostatecznego wyniku, należy poddać rygorystycznemu audytowi architekturę procesu badawczego. W ujęciu tym nauka nie

jest kasynem trafnych przeczuć, lecz instytucją produkcji uzasadnienia, w której wartość wiedzy wynika z dyscypliny procedur, a nie z błyskotliwej intuicji. Autor odrzuca zarówno naiwny pozytywizm, jak i postmodernistyczny relatywizm, stawiając na deontologiczną filozofię nauki. W dobie sztucznej inteligencji, która błyskawicznie generuje hipotezy, postulat ten staje się fundamentem przetrwania – prawdziwy postęp nie mierzy się szybkością predykcji, lecz jakością dowodów. Stegenga przekonuje, że w świecie zdominowanym przez algorytmy i walkę narracji, tylko przejrzysta, audytowalna procedura może stanowić bezpieczną podstawę dla decyzji publicznych. To wezwanie do odbudowy republikańskiej godności nauki, gdzie zaufanie nie jest aktem wiary, lecz mierzalnym efektem spełnienia surowych norm poznawczych, chroniących nas przed arbitralnością i intelektualnym bankructwem późnej nowoczesności.

SŁOWA KLUCZOWE

uzasadnienie jako fundament wiedzy | społeczna epistemologia | instytucja produkcji uzasadnienia | rygorystyczna jakość uzasadniania | deontologiczna filozofia nauki | mechanizmy audytu | kryzys replikacyjny | epistemiczna niewinność | systematyczne odstępstwo | rynek symboliczny | architektura generująca output | wspólna wiedza | samokorekta instytucjonalna | punkty bifurkacji | normy uzasadniania

Stegenga: odrzucenie mitu heroicznej prawdy

Współczesna kultura wiedzy zdążyła się wygodnie urządzić w pewnym akademicko wypolerowanym zabobonie. Głosi on, że naukową giełdę idei należy oceniać wyłącznie przez pryzmat jej finału, pytając, czy wygenerowała trafną reprezentację rzeczywistości. Jacob Stegenga rozsadza ten elegancki mit od środka, przy czym nie ucieka się do taniego relatywizmu. Dokonuje ruchu o wiele bardziej wymagającego, przesuwając oś oceny z końcowego certyfikatu prawdy na rygorystyczną jakość uzasadniania. Nauka w ujęciu Stegengi to nie kasyno trafnych przeczuć. To instytucja produkcji uzasadnienia.

Centralnym celem tego przedsięwzięcia przestaje być prywatna prawda odkrywcy, a staje się nią wspólna, rzetelnie zaudytowana wiedza. Nie pytamy wyłącznie o output, lecz o jakość architektury generującej output. Takie podejście, sytuujące książkę w nurcie nowej społecznej epistemologii,

wymaga od nas porzucenia starych przyzwyczajeń. Zamiast fetyszyzować poznawczy triumf, musimy skupić się na dyscyplinie procedury chroniącej przed intelektualnym bankructwem. Serce nauki nie bije w prawdzie. Bije w uzasadnieniu.

Dla wielu uszu ta teza brzmi jak herezja, ponieważ współczesny akademik został wychowany w kleszczach podwójnego mitu. Pierwszy z nich to iluzja heroicznej prawdy, wedle której badacz ma kroczyć ku absolutowi niczym asceta wyczekujący objawienia. Drugi biegun stanowi mit skromności tymczasowej, redukujący każdą tezę do wiecznego prowizorium. Stegenga odrzuca obie te fałszywe dychotomie jednym cięciem. Prawda nie stanowi bowiem dla działającego tu i teraz badacza użytecznej normy operacyjnej, toteż nie można na niej opierać praktyki.

Jeśli sprowadzimy naukę do niekończącej się gry pozorów i wiecznych wersji beta, natychmiast straci ona swoją cywilizacyjną funkcję koordynacyjną. Nie da się budować nowoczesnej medycyny, prawa czy polityki energetycznej na epistemologii cierpiącej na wewnętrzną sprzeczność. Nie możemy nakazywać społeczeństwu słuchania ekspertów, by zaraz potem dodawać, że ekspert nigdy niczego nie wie na pewno. Bez twardych fundamentów zostaje nam już tylko walka narracji, czyli rynek symboliczny, na którym wygrywa nie ten, kto lepiej uzasadnia, lecz ten, kto lepiej dystrybuuje afekt. To nie jest pokora, lecz rozbrojenie instytucji poznania.

Deontologia: fundament etyczny nauki

W tym miejscu ujawnia się prowokacyjna siła omawianej książki. Stegenga odrzuca mgławicowe pochwały „dialogu” czy „otwartości”, które dziś działają niczym semantyczne koce gaśnicze, dławiące każdy spór. Proponuje rozwiązanie znacznie mniej towarzyskie, a nieporównywalnie bardziej produktywne. Wprowadza deontologiczną filozofię nauki, czyli koncepcję zorientowaną na twarde reguły poprawnego działania, a nie na sam rezultat. Oznacza to bowiem przeniesienie klasycznego dylematu etycznego na grunt epistemologii.

Przestajemy toteż oceniać badaczy przez pryzmat konsekwencji, skupiając się na zgodności procesu z normą. Nauka ma być rozliczana nie z tego, czy heroicznie odkryła prawdę, lecz czy działała w sposób budujący zaufanie. Dla analityków ekonomii instytucjonalnej ten zwrot można ująć dobitniej. Nie interesuje nas tylko wynik transakcji, lecz standard ładu kontraktowego oraz audytowalność procesu.

Zamiast ekscytować się giełdową wyceną spółki, musimy sprawdzić, czy mechanizm wyceny nie ma strukturalnego błędu. Bez tego zostaje nam już tylko walka narracji, czyli rynek symboliczny, na

którym wygrywa nie ten, kto lepiej uzasadnia, lecz ten, kto lepiej dystrybuuje afekt. Serce nauki nie bije wszak w prawdzie, lecz w uzasadnieniu. To rygorystyczna instytucja produkcji uzasadnienia.

Wspólna wiedza vs. prywatne przekonania

Zacznijmy od fundamentu, jakim jest wspólna wiedza, stanowiąca oś całej konstrukcji. Nie jest to zwykła suma prywatnych przekonań ani socjologiczny opis faktu, że większość środowiska tak uważa. Taka definicja byłaby zbyt tania, plemienna i podatna na epistemiczne oligopole. Wspólna wiedza oznacza bowiem coś mocniejszego. To stan, w którym istnieje nie tylko zgoda co do tezy, lecz twardy konsensus dotyczący procedur i standardów.

To nie jest tłum ludzi, którzy przypadkowo stanęli pod tym samym transparentem. To wspólnota surowych sędziów, uznających ważność tej samej procedury dowodowej. To instytucja produkcji uzasadnienia. Właśnie dlatego uniwersytety, czasopisma, recenzje i jawność danych nie stanowią jedynie estetycznych dodatków do uprawiania nauki.

Są to w istocie rygorystyczne mechanizmy audytu, przekształcające indywidualne wglądy w dobro wspólne o charakterze poznawczym. Omawiana książka przedstawia ten proces z niezwykłą precyzją. Celem nauki nie jest wszak po prostu naga, wyizolowana prawda. Jest nim szczególny rodzaj prawdy społecznie ustabilizowanej, a także wzajemnie uzasadnionej.

Samokorekta: instytucjonalny mechanizm nauki

To rozpoznanie niesie za sobą doniosłe konsekwencje dla każdego, kto uczestniczy w rynkowym i społecznym obiegu wiedzy zawodowej. Prawnik natychmiast dostrzeże tu uderzającą analogię do skomplikowanego procesu legitymizacji orzeczenia sądowego. Samo trafne rozstrzygnięcie sprawy nigdy nie wystarcza, jeśli obserwator nie może precyzyjnie prześledzić drogi rozumowania. Ekonomista z kolei zauważy wyraźne pokrewieństwo z fundamentalnym problemem koordynacji oraz redukcji kosztów transakcyjnych. Wspólna wiedza stabilizuje bowiem środowisko decyzji, umożliwia przewidywalność działań innych podmiotów i drastycznie obniża koszty niepewności. Kulturoznawca natomiast zrozumie, że Stegenga mierzy się z głębokim kryzysem cywilizacji późnej nowoczesności, w której nadmiar bodźców rozsadza warunki wspólnego świata.

Właśnie w tym newralgicznym miejscu filozofia nauki spotyka się z najszerszymi diagnozami współczesności. Fundacja Dobre Państwo w swoich niedawnych publikacjach stawia zresztą bardzo podobny, niezwykle istotny akcent. Odróżnia ona mianowicie informację, która jedynie łączy ludzi,

od wiedzy, która rygorystycznie porządkuje prawdziwe sądy. Podkreśla przy tym, że kluczem do przetrwania w epoce danych jest bezwzględna obrona instytucji samokorekty. Stegenga bez cienia wątpliwości mógłby podpisać się pod tą trafną intuicją obiema rękami. Dla niego nauka jest wszak niczym innym jak właśnie zinstytucjonalizowaną samokorektą, ocenianą z perspektywy rygoru uzasadniania.

Z tego solidnego fundamentu wyrasta drugi filar całej argumentacji, czyli rygorystyczna ocena deontologiczna. Stegenga podejmuje tu otwarty spór z klasycznym problemem demarkacji, uderzając w wieloletnią obsesję filozofii nauki. Dyscyplina ta przez dekady pragnęła znaleźć jeden cudowny przełącznik oddzielający prawdziwą naukę od pseudonauki. Niegdyś miała nim być weryfikowalność, później falsyfikowalność, a niekiedy sama struktura teorii lub jej zdolność predykcyjna. Wszystkie te ambitne próby miały jednak wspólną wadę: szukały bowiem binarnego klucza do rzeczywistości, która z natury jest historyczna i instytucjonalnie złożona.

Autor nie pyta zatem o metafizyczną czystość gatunkową danej dyscypliny czy praktyki badawczej. Pyta raczej o to, w jakim stopniu konkretne działanie spełnia właściwe, rygorystyczne normy uzasadniania. To w żadnym wypadku nie jest kapitulacja, lecz radykalne zaostrzenie dotychczasowych kryteriów oceny. Zamiast jednego abstrakcyjnego testu otrzymujemy wielowymiarową, niezwykle wymagającą ocenę całego procesu badawczego. Sprawdzamy, czy ujawniono dane, czy kontrolowano źródła błędu oraz czy procedura była odporna na manipulację. Pytamy również, czy wniosek jest proporcjonalny do materiału dowodowego i czy możliwa jest krytyczna kontrola wspólnotowa.

W tym sensie filozof dokonuje ruchu analogicznego do przejścia od prymitywnej rachunkowości do pełnego audytu. Nie pytamy wyłącznie o wynik, lecz o jakość architektury generującej ten wynik. Nie interesuje nas tylko to, czy firma ogłosiła zysk, ale przede wszystkim to, jak skonstruowano księgi. Ta zmiana perspektywy jest wyjątkowo aktualna w dzisiejszym, mocno poturbowanym paradygmacie naukowym. Żyjemy wszak po bolesnym kryzysie replikacyjnym i po latach wyczerpujących sporów o prerejestrację badań. Mamy za sobą debaty o otwartych danych, p-hackingu oraz konfliktach interesów w biomedycynie.

Innymi słowy, funkcjonujemy już bezpowrotnie po ostatecznym końcu epistemicznej niewinności. Doskonale wiemy, że sama błyszcząca etykieta z napisem „science” nie gwarantuje już absolutnej czystości poznawczej. Wiemy również, że publiczne nawoływanie do ślepego zaufania nauce bywa całkowicie puste i przeciwskuteczne. Zaufanie nie ma być aktem liturgicznym wobec wielkiej, nietykalnej abstrakcji zwanej Nauką. Zaufanie ma być mierzalną funkcją spełnienia rygorystycznych norm uzasadniania w konkretnym procesie. Nie ufa się białemu fartuchowi, lecz ufa się procedurze, która potrafi przetrwać próbę wrogiej kontroli.

Tę kluczową intuicję wzmacnia także jego wcześniejsza praca nad istotą postępu naukowego. Dowodził w niej, że nauka robi postęp wtedy i tylko wtedy, gdy następuje zmiana w uzasadnieniu. Nie musimy czekać na moment, w którym można metafizycznie przybić ostateczną pieczęć prawdy. Pojęcie biasu, czyli systematycznego wypaczenia, zostaje w tym świetle całkowicie oczyszczone z psychologicznej mgły. Bias nie jest już tylko nazwą dla nieestetycznej pomyłki poznawczej czy subiektywnej skłonności jednostki. Staje się on precyzyjnym określeniem systematycznego odstępstwa od właściwych, zinstytucjonalizowanych norm uzasadniania.

To definicja niezwykle mocna, odwraca bowiem naszą uwagę od taniego melodramatu intencji. Badacz może być całkowicie szczery, a jednocześnie głęboko stronniczy w swoich działaniach. Może mieć krystalicznie czyste serce, ale fatalną, pełną luk metodę badawczą. Może nawet przypadkiem trafić w prawdę, a mimo to zasługiwać na surową naganę epistemiczną. Stanie się tak, jeśli jego droga do wyniku była skażona arbitralnością, manipulacją lub niejawnością. Epistemologia nie jest wszak modną terapią wyrażania siebie ani poszukiwaniem wewnętrznej harmonii.

Jest ona bezwzględną, chłodną dyscypliną kontroli wszelkich roszczeń do obiektywnej wiedzy. W tej perspektywie Stegenga staje się cennym sojusznikiem krytycznej teorii nowoczesnych instytucji. Pokazuje on dobitnie, że obiektywizm nie polega na jakiejś bezcielesnej, anielskiej neutralności badacza. Polega on wyłącznie na stopniu, w jakim wdrożone procedury ograniczają ryzyko błędu systematycznego. Jeszcze bardziej wybuchowy okazuje się jego projekt nowego ideału wolności od wartości. Uderza on bezpośrednio w modny dziś słownik usytuowanego poznania i nieuchronnej normatywności.

Czytelnik przyzwyczajony do tezy o zawsze politycznym charakterze wiedzy może tu poczuć narastający opór. Stegenga nie wraca jednak do naiwnego pozytywizmu i nie twierdzi, że nauka jest wolna od wartości. Nie nawołuje również do instytucjonalnej hipokryzji, w której badacze mieliby udawać brak własnych światopoglądów. Proponuje on coś znacznie bardziej ambitnego, intelektualnie wymagającego i trudnego do wdrożenia. Naukowcy powinni mianowicie działać tak, jakby nauka miała być od tych wartości całkowicie wolna. Formuła „działać tak, jakby” jest tu genialna, rozdziela bowiem stan końcowy od obowiązku dążenia.

To subtelne rozróżnienie ma kolosalne znaczenie nie tylko filozoficzne, lecz także prawne i polityczne. W prawie konstytucyjnym często żyjemy pod sztandarem zasad, które nigdy nie są spełnione doskonale. Mimo swojej utopijności pozostają one jednak niezbędnymi, twardymi punktami orientacyjnymi dla całego systemu. Nikt rozsądny nie twierdzi, że skoro pełna równość wobec prawa jest nieosiągalna, należy zrezygnować z zasady równości. Nikt poważny nie wyrzeka się reguł uczciwej konkurencji tylko dlatego, że nie istnieje doskonały rynek. Podobnie jest w architekturze pojęciowej Stegengi, gdzie dążenie do neutralności pozostaje twardym obowiązkiem.

Nauka ma bezwzględny obowiązek ograniczania punktów, w których wartości wpływają na sam tok wnioskowania. W tym miejscu jego myśl staje się zaskakująco antycyniczna i ożywczo pod prąd. Odrzuca on współczesne rozkoszowanie się niemożnością oraz kapitulację przed wszechobecną politycznością. Nie zgadza się z tezą, że skoro idealna neutralność jest nieosiągalna, to wszystko jest już tylko walką o władzę. To instytucja produkcji uzasadnienia.

Badacze muszą identyfikować miejsca bifurkacji, czyli momenty największego ryzyka interpretacyjnego. Są to punkty, w których identyczne dane, przepuszczone przez odmienne wartości, dają skrajnie odmienne wnioski. System ma obowiązek minimalizować takie miejsca przez lepsze dowody i wyższą przejrzystość. W ten sposób autor wykonuje niezwykle subtelny manewr wobec krytyk feministycznych i postkolonialnych. Nie odrzuca ich trafnej diagnozy, że wiedza bywa skażona ukrytymi asymetriami władzy. Odrzuca jednak wniosek, wedle którego odpowiedzią ma być radosna celebracja wielości wartości.

Różnorodność perspektyw jest dla niego narzędziem demaskowania wypaczeń, a nie nowym fetyszem epistemicznym. To stanowisko doskonale i bez zgrzytów przekłada się na twardy język ekonomii politycznej. Zwiększanie pluralizmu w systemie nie jest i nie może być celem samym w sobie. Jest ono wyłącznie instrumentem wykrywania rent poznawczych i rozbijania oligarchii interpretacyjnych. Różnica ma sens tylko wtedy, gdy służy oczyszczaniu procedur z systematycznych błędów. Nie ma ona racji bytu, gdy staje się wygodnym immunitetem dla metodologicznej dowolności.

To rozstrzygnięcie jest szczególnie ważne dziś, w epoce sztucznej inteligencji i platform cyfrowych. W tym nowym środowisku wielość głosów bywa całkowicie pozorna i starannie wyreżyserowana. Faktyczne kanały widzialności są bowiem sterowane przez modele optymalizujące uwagę, a nie prawdę. W takim świecie Stegenga przypomina coś, o czym inteligencja sieciowa bardzo lubi zapominać. To nie sama liczba perspektyw w obiegu tworzy ostateczną wiarygodność danego twierdzenia. Tworzy ją dopiero bezlitosny rygor tego, jak te perspektywy są wystawiane na wspólną próbę.

Książka znakomicie wpisuje się w najbardziej żywe nurty współczesnego paradygmatu naukowego. Z jednej strony mocno rezonuje z ruchem otwartej nauki i troską o pełną transparentność. Współdzieli krytykę epistemicznych rynków „cytryn”, na których łatwiej sprzedać wynik efektowny niż rzetelny. Z drugiej strony uderza w tę część nowej ortodoksji, która promuje pełzające upolitycznienie wnioskowania. Nauka nie jest świątynią pozaludzkiej czystości, ale właśnie dlatego potrzebuje twardszych procedur.

To byłby ostateczny koniec nauki jako kluczowej instytucji wolnego, demokratycznego społeczeństwa. Zostałaby nam jedynie epistemiczna giełda paniki, napędzana przez algorytmy i polityczne emocje. Na tym etapie wyłania się obraz tyleż surowy, co głęboko wyzwalający dla badacza. Nauka nie jest religią prawdy ani tanim teatrem eksperckiej, medialnej charyzmy. Jest ona publiczną praktyką odpowiedzialnego uzasadniania, której celem jest produkcja wspólnej wiedzy. Serce nauki nie bije w prawdzie. Bije w uzasadnieniu.

Ta pierwsza myśl ustanawia architekturę całego gmachu, ale teraz musimy wejść głębiej. Czekają tam najbardziej sporne i niebezpieczne komnaty całego filozoficznego projektu Stegengi. Tam bowiem rozstrzyga się to, co dla współczesnej akademii jest politycznie i moralnie najtrudniejsze. Trzeba odpowiedzieć na pytania o to, jak naukowiec ma mówić publicznie w warunkach kryzysu. Musimy ustalić, czym właściwie jest postęp i czy wolno dziś mówić o prawdach trwałych. Właśnie dlatego druga połowa książki jest znacznie bardziej drażniąca i prowokacyjna.

Asercja naukowa, czyli publiczne twierdzenie badacza o świecie, zostaje tu brutalnie wyrwana z pułapek. Pierwszą z nich jest norma prawdy, wedle której wolno mówić tylko to, co bezwzględnie prawdziwe. Drugą jest norma wiedzy, wymagająca od mówiącego absolutnej, niezachwianej pewności. Obie te formuły wydają się szlachetne, lecz dla codziennej praktyki są zabójczo nieporęczne. Stegenga zastępuje je podwójnym wymogiem rygorystycznego uzasadnienia oraz praktycznej informatywności. Twierdzenie musi opierać się na jawnych racjach i realnie wnosić coś do pola decyzji.

Ekspert zbyt często mówi tak, aby niczego nie dało się z jego wypowiedzi praktycznie wydobyć. Powstaje wtedy komunikacja immunizacyjna, która chroni autora przed jakąkolwiek odpowiedzialnością. Formalnie zachowuje ona pozory ostrożności, lecz społecznie jest całkowicie pusta i bezużyteczna. W tym właśnie punkcie filozofia Stegengi dotyka najczulszego nerwu współczesnej polityki wiedzy. Państwa i korporacje domagają się od naukowców szybkich, decyzyjnych komunikatów w warunkach niepewności. Jednocześnie karzą ich za każdą pomyłkę tak, jakby badacz miał dostęp do boskiego punktu obserwacji.

W rezultacie eksperci uczą się mówić mętnym językiem asekuracji i niekończących się zastrzeżeń. Społeczeństwo z kolei odbiera tę asekurację jako unik, tchórzostwo albo zakamuflowaną manipulację. Autor uderza w samo centrum tego paradoksu, domagając się od badaczy epistemicznej odwagi. Naukowiec nie musi być nieomylnym prorokiem, lecz odpowiedzialnym dostawcą użytecznych twierdzeń. Regulacja ryzyka czy decyzje sanitarne nie mogą opierać się na semantycznym odpowiedniku milczenia. Z tej samej matrycy pojęciowej wyrasta jego nowatorska koncepcja postępu naukowego.

Obala ona infantylnie triumfalny obraz nauki jako radosnego marszu od odkrycia do odkrycia. Nauka robi postęp wtedy i tylko wtedy, gdy następuje mierzalna zmiana w uzasadnieniu. To formuła zarazem niezwykle skromna, jak i głęboko rewolucyjna dla całego systemu. Pozwala ona uznać za postęp także to, czego nowoczesna kultura sukcesu panicznie nie znosi. Sukcesem staje się osłabienie hipotezy, porażka replikacji czy rozbrojenie efektownej narracji. W tej surowej optyce laboratorium nie jest już fabryką pozytywnych, medialnych wyników.

Staje się ono raczej zakładem bezlitosnej restrukturyzacji uzasadnień i eliminacji błędów. Czasem największym sukcesem badawczym nie jest znalezienie nowego obiektu, lecz zamknięcie drogi złudzeniu. Recepcja tej tezy jest szczególnie interesująca w epoce traumy kryzysu replikacyjnego. Stegenga dowodzi, że kryzys ten nie jest dowodem klęski, lecz dowodem sprawnego działania systemu. Dzieje się tak, o ile instytucja potrafi obniżać konfirmację twierdzeń źle uzasadnionych. Odwrót, rewizja i ograniczenie pewności nie są kompromitacją rozumu, lecz jego najtrudniejszą kompetencją.

Reguła pierwszeństwa: patologia systemu nauki

Jeszcze brutalniej Stegenga wchodzi w obieg prestiżu naukowego, gdy uderza w regułę pierwszeństwa. W tradycyjnej socjologii nauki funkcjonuje ona niemal jak prawo natury, w którym zwycięzcą zostaje ten, kto pierwszy dobiegnie do mety. Kto pierwszy opublikuje i nazwie zjawisko, ten zgarnia całą symboliczną rentę poznawczą. Stegenga udowadnia jednak, że ten model jest zarówno opisowo fałszywy, jak i normatywnie zgubny. Fałszywy, ponieważ historia nauki rzadko oddaje hołd samemu pionierowi, częściej nagradzając tego, kto dostarczył rozstrzygającego dowodu. Zgubny, bowiem system wynagradzający wyłącznie pośpiech musi produkować toksyczne efekty uboczne.

Taka architektura bodźców premiuje komunikacyjną agresję, skracanie ścieżek dowodowych oraz inflację pewności kosztem żmudnej weryfikacji. Działa to jak fatalnie zaprojektowana giełda, która nagradza natychmiastowy zwrot, a koszt ryzyka przerzuca na resztę systemu. Stegenga proponuje zatem radykalne cięcie, postulując zerwanie związku symbolicznej renty z samym faktem pierwszeństwa. Chwała ma przypadać nie temu, kto pierwszy krzyknął z korytarza, lecz temu, kto rzetelnie domknął sprawę. To instytucja produkcji uzasadnienia.

To stanowisko ma fundamentalne znaczenie dla współczesnej ekonomii nauki, która stała się bezlitosnym rynkiem wskaźników. Dzisiejsza akademia to osobliwy sektor, produkujący dobra publiczne pod presją prywatnych karier, grantów i indeksów cytowań. Jeśli mechanizmy nagradzania są wadliwie skalibrowane, cała epistemiczna gospodarka zaczyna przypominać

patologiczny audyt. Zysk osiąga się w nim poprzez księgowanie przyszłych strat poza oficjalnym bilansem. Reguła pierwszeństwa jest właśnie takim mechanizmem, dającym szybki bonus za heroizm odkrycia. Koszt badawczej niestaranności rozprasza się natomiast po całej wspólnocie w postaci marnowanych środków i utraty zaufania.

Wobec tego kryzysu Stegenga proponuje głęboką, deontologiczną korektę całego ładu motywacyjnego. Nie pytamy wyłącznie o wynik, lecz o jakość architektury generującej wynik, co oznacza nagradzanie wkładu w uzasadnienie. To myśl wręcz konserwatywna w najlepszym, rzemieślniczym sensie tego słowa, broniąca pracy potwierdzającej przeciw narcystycznej kulturze przełomu. Dialektyka jest tu bezlitosna, toteż fałszywy heroizm pioniera musi ustąpić przed rygorem ostatecznej weryfikacji. Stegenga staje w obronie drugiego przybysza, który zjawia się po fanfarach, przynosząc twarde dowody. Bez niego wszak nauka pozostaje jedynie dobrze ubranym domysłem.

Bezpieczniki szybkiej nauki: weryfikacja kryzysowa

Jeszcze trudniejsza okazuje się refleksja o szybkiej nauce, ponieważ filozofia zderza się tu z biopolityką kryzysu. Kiedy wybucha pandemia czy katastrofa środowiskowa, społeczeństwa nie chcą słuchać, że rygor procedur wymaga czasu. Oczekują bowiem decyzji natychmiastowych, a Stegenga ten fakt w pełni uznaje. Dopuszcza on fast science, czyli naukę przyspieszoną, która pod presją czasu może łamać część standardów rutynowej praktyki. Niemniej właśnie tutaj ujawnia się jego metodologiczna surowość. Szybka nauka nie może wszak stanowić usprawiedliwienia dla własnych naruszeń, a nagła sytuacja nie zamienia epistemicznej oszczędności w cnotę.

Dlatego Stegenga formułuje dwa potężne hamulce, chroniące system przed katastrofą. Po pierwsze, wprowadza zasadę podobieństwa, wedle której szybka nauka powinna maksymalnie przypominać zwykłą praktykę badawczą. Po drugie, ustanawia zasadę niezależności, co oznacza, że metody użyte do diagnozy zagrożenia nie mogą uzasadniać uproszczonych narzędzi interwencji. Te dwa bezpieczniki mają powstrzymać kryzysowe poznanie przed przejściem w epistemiczną brawurę. To instytucja produkcji uzasadnienia, która nawet w stanie wyższej konieczności musi podlegać rygorystycznemu audytowi.

Znaczenie tej części książki wykracza daleko poza medycynę oraz doświadczenia pandemiczne. Dotyczy ono bowiem całej architektury rządzenia w warunkach radykalnej niepewności. Państwa technologiczne operują dziś na podstawie modeli, danych strumieniowych i predykcji algorytmicznych, przy czym ich jakość bywa nierówna, a stawka ogromna. W takich warunkach rośnie pokusa, by procedurę zastąpić performansem konieczności. Władza woli uznać, że czasu nie

ma, toteż musimy bezwarunkowo zaufać intuicji aparatu. Stegenga odrzuca tę logikę, dekonstruując fałszywą dychotomię między heroizmem szybkiej akcji a rygorem metodologicznym.

Jego sprzeciw wynika ze świadomości, iż stan wyjątkowy stanowi ulubione siedlisko epistemicznego nadużycia. Jeśli w zwykłych czasach badaczy kusi efektowność, to w momentach nadzwyczajnych nęci ich dodatkowo renta poznawcza w postaci politycznego immunitetu. Każdy, kto pyta o standard dowodu, bywa natychmiast oskarżany o brak odpowiedzialności. Stegenga odpowiada na ten szantaż, twierdząc, że odpowiedzialność to nie zawieszenie standardów, lecz ich ostrożne przekształcenie. W przeciwnym razie szybka nauka stanie się nauką zredukowaną, zalegalizowaną przez samą grozę sytuacji. Powstałaby tak epistemologia stanu wyjątkowego, z której niezwykle trudno powrócić.

Prawda ponadczasowa kontra tymczasowość teorii

Wreszcie docieramy do terytorium, które w dzisiejszym klimacie intelektualnym uchodzi za obszar podwyższonego ryzyka, czyli do koncepcji prawd ponadczasowych. W tym miejscu Stegenga wypowiada otwartą wojnę tezie o permanentnej tymczasowości, która niczym dogmat opanowała znaczną część filozofii, historii oraz socjologii nauki, głosząc, że każda teoria prędzej czy później zostanie obalona. Choć na pierwszy rzut oka takie stanowisko brzmi szlachetnie i ubiera badaczy w szaty intelektualnej pokory, w rzeczywistości jest to pozór fatalny w skutkach, przypominający raczej księgową sztuczkę niż rzetelny audyt poznawczy. Jeśli bowiem wszystko ma być wyłącznie chwilowym kaprysem, to całkowicie wyparowuje różnica między luźno rzuconą hipotezą a twierdzeniem zabetonowanym przez wielokrotną kontrolę i potężny konsensus. Społeczeństwo otrzymuje wówczas skrajnie niebezpieczny komunikat: skoro nauka sama przyznaje, że jej gmach może jutro runąć, to dlaczego dzisiaj mielibyśmy opierać na nim nasze polityki publiczne i instytucje? Stegenga stawia sprawę jasno, twierdząc, że tam, gdzie wytworzono wspólną wiedzę o potężnej mocy uzasadniającej, mamy pełne prawo i obowiązek mówić o prawdach ponadczasowych.

To stanowisko jest o tyle fascynujące, że wcale nie oznacza powrotu do naiwnego realizmu, w którym badacz po prostu bezrefleksyjnie odczytuje księgę natury. Stegenga nie twierdzi bynajmniej, że każda świeżo upieczona teoria automatycznie zyskuje status wieczności, lecz proponuje rozwiązanie o wiele subtelniejsze, przypominające solidną architekturę systemu. Prawdy ponadczasowe krystalizują się tam, gdzie spełniono rygorystyczne warunki możliwości wiedzy, a skala opisu została precyzyjnie dopasowana do naszych narzędzi epistemicznych. Wymaga to rozległego konsensusu, w którym przyszłe odkrycia nie wywracają stolika, lecz jedynie cyzelują

detale, tak jak zdanie o Ziemi krążącej wokół Słońca nie staje się fałszywe przez to, że współczesna fizyka dokładniej opisuje dynamikę układu. Zostaje ono po prostu umieszczone na właściwej skali opisu i na tej konkretnej giełdzie poznawczej zachowuje swoją pełną, niepodważalną wartość. Jest to potężne uderzenie zarówno w tani historycyzm, jak i w narcystyczną potrzebę naszej epoki, by każdą odziedziczoną pewność traktować z góry jako naiwność przodków, zapominając, że rozwój wiedzy to często żmudna rozbudowa trwałego kapitału.

Recepcja tej końcowej, niezwykle mocnej tezy będzie z pewnością głęboko spolaryzowana, przypominając nerwowe reakcje rynków na niespodziewaną podwyżkę stóp procentowych. Część filozofów nauki dostrzeże w niej zbyt śmiałą rehabilitację realizmu, socjologowie uznają ją za niedocenienie historycznej zmienności praktyk dowodowych, ale publicyści naukowi przyjmą ten zwrot z ogromną ulgą. Odpowiada on bowiem na bardzo realny kryzys zaufania, w którym retoryka nieustannej prowizoryczności bywa doskonałym paliwem dla denializmu, pozwalając podważać twarde dane pod pretekstem ich rzekomej tymczasowości. W moim odczytaniu właśnie w tym miejscu Stegenga okazuje się myślicielem najbardziej aktualnym, ponieważ nie oferuje taniego pocieszenia, lecz przywraca utraconą miarę rzeczy w epoce pełnej sprzeczności. Kiedy z jednej strony ubóstwiamy naukę jako źródło technicznej sprawczości, a z drugiej podcinamy jej autorytet przez banalizację wyników, potrzebujemy rygorystycznego narzędzia, które zmusi nas do kategorycznego oddzielania niepewnej granicy najnowszych badań od twardego rdzenia ugruntowanej wiedzy.

Należy z całą mocą oddzielać cnotę ostrożnej korekty od szkodliwego fetyszu prowizoryczności, który paraliżuje jakiekolwiek zdecydowane działanie oparte na faktach. Bez tych fundamentalnych rozróżnień nowoczesna kultura wiedzy nieuchronnie zapada się w dwie równie destrukcyjne formy: osuwa się albo w technokratyczną pychę, ślepą na własne ograniczenia, albo w postmodernistyczne znużenie. Całość projektu Stegengi można zatem odczytać jako monumentalną próbę odbudowy republikańskiej godności nauki, rozumianej jako absolutne dobro wspólne. Jest to projekt na wskroś republikański, ponieważ w centrum stawia publiczną kontrolę racji, wspólny język uzasadnienia oraz instytucjonalne warunki chroniące przed arbitralnością decyzji. W tej precyzyjnej wizji nie ma miejsca ani dla wyalienowanego kapłaństwa ekspertów, ani dla populistycznej pogardy wobec twardych kompetencji, a naukowiec nie jest dostawcą intelektualnego towaru dla kapryśnego klienta.

Badacz staje się tu raczej surowym depozytariuszem zobowiązań proceduralnych, kimś na kształt rzetelnego audytora w wielkiej korporacji wiedzy. To instytucja produkcji uzasadnienia, w której każda moneta zaufania musi mieć twarde pokrycie w dowodach, co czyni tę książkę traktatem o konstytucji ładu poznawczego w czasach galopującej inflacji bodźców. Właśnie dlatego jej recepcja

musi wyjść daleko poza wąskie mury akademii i dotrzeć do państwowych regulatorów, projektantów polityk publicznych czy twórców standardów dla sztucznej inteligencji, którzy codziennie podejmują decyzje na styku wiedzy niepełnej i skutków nieodwracalnych. Płynie z niej bowiem lekcja surowa, ale oczyszczająca: nie pytaj, kto ma absolutną rację na wieczność, lecz pytaj, kto zasłużył na zaufanie tu i teraz poprzez najwyższą jakość proceduralnego uzasadnienia. Dopiero z tak pojętej dyscypliny może narodzić się wspólna wiedza, a z niej prawdy na tyle mocne, by można było na nich bezpiecznie oprzeć nasz wspólny świat.

Normy Stegengi w erze generatywnej AI

Zatem musimy wykonać ruch śmielszy i znacznie bardziej niebezpieczny intelektualnie. Trzeba zapytać, czy normatywy Stegengi dają się przenieść na naukę generowaną przez AI. Wkraczamy wszak w porządek, w którym hipotezy i modele nie powstają już w samotnym mózgu uczonego. Rodzą się w układzie hybrydowym, w którym człowiek i system wspólnie produkują roszczenia do wiedzy. Tu kończy się komfort filozofii nauki uprawianej w duchu laboratoryjnego humanizmu. Zaczyna się brutalna epoka epistemologii infrastrukturalnej, w której badamy całą maszynerię stojącą za wynikiem.

Nie pytamy już zatem tylko o to, jakie normy powinien spełniać sam badacz. Pytamy, jakie rygory musi spełniać cały zautomatyzowany aparat generowania i selekcji twierdzeń. Serce nauki nie bije w prawdzie, bije w uzasadnieniu, co w świecie algorytmów oznacza rzecz fundamentalną. Model dający poprawną odpowiedź bez audytowalnej drogi dowodowej jest użyteczny jak czarna skrzynka dla hazardzisty. Może przynieść zysk poznawczy, niemniej nie wytwarza twardej waluty, jaką jest wspólna wiedza. Stegenga widzi wszak naukę jako instytucję wzajemnego uzasadniania, a nie giełdę prywatnych trafień.

Tutaj widać, jak bardzo jego projekt okazuje się odporny na technologiczne wstrząsy. Uczony wyposażony w model generatywny potrafi wyprodukować tysiące hipotez szybciej niż dawny zespół przez rok. Lecz samo przyspieszenie produkcji roszczeń poznawczych nie jest jeszcze przyspieszeniem nauki. Postępem jest dopiero radykalny wzrost zdolności do ich uzasadniania i poddawania kontroli publicznej. Dlatego Stegenga byłby fatalnym patronem technofilskiego entuzjazmu, który naiwnie miesza prędkość z postępem. Byłby natomiast idealnym prawodawcą twardej konstytucji dla nauki wspomaganej przez AI.

W ramach takiej konstytucji model nigdy nie jest autonomicznym autorem prawdy. Staje się wyłącznie instrumentem, który bezwzględnie podlega surowym normom deontologicznym. Liczy się nie tylko trafny wzorzec, lecz to, czy potrafimy odtworzyć architekturę decyzyjną systemu.

Musimy wiedzieć, jakie dane i punkty bifurkacji doprowadziły maszynę do konkretnego wniosku. Bez transparentnego audytu zostaje nam jedynie skuteczne, lecz całkowicie niezrozumiałe zaklęcie maszynowe.

Skuteczność predykcyjna AI: niewystarczalność dowodowa

W tym świetle odwieczny spór między indukcją a dedukcją zyskuje zupełnie nowy, wysoce praktyczny ciężar gatunkowy. Dzisiejsza sztuczna inteligencja jest bowiem w przeważającej mierze potężnym silnikiem indukcyjnym, który uczy się z oceanów danych, kompresuje statystyczne regularności, przewiduje kolejne tokeny, a w nowszych iteracjach buduje nawet złożone procedury planowania i samokorekty. To niewątpliwie imponujące osiągnięcie inżynieryjne, niemniej nie rozwiązuje ono fundamentalnego problemu logicznej precyzji, ponieważ indukcja z definicji dostarcza nam jedynie prawdopodobieństwa, a nie twardej konieczności. Oferuje świetnie skalkulowany zakład na giełdzie predykcji, nie zaś żelazny, matematyczny dowód gwarantujący zachowanie własności w każdym dopuszczalnym przypadku. To instytucja produkcji uzasadnienia, toteż architektura AI musi ewoluować od ślepego rozpoznawania wzorców do systemów, w których twierdzenia podlegają rygorystycznemu audytowi, lokalnym dowodom i wspólnotowym testom kontrfaktycznym.

Współczesny stan badań nad sztuczną inteligencją doskonale obrazuje to epistemologiczne napięcie między obietnicą a rzeczywistością. Benchmark ARC AGI, czyli rygorystyczny test mierzący zdolność systemów do uogólniania, został zaprojektowany właśnie jako bezlitosny audytor sprawdzający luki w zadaniach trywialnych dla człowieka, a druzgocących dla modeli. Oficjalne raporty ARC Prize nie pozostawiają złudzeń: wszak czyste modele językowe osiągają tam wyniki oscylujące wokół zera, obnażając słabość czystej statystyki. Nawet najbardziej zaawansowane systemy konkursowe z końca 2025 roku dobiły zaledwie do poziomu około dwudziestu czterech procent na prywatnym zbiorze danych, co pozostaje wynikiem drastycznie odległym od ludzkiej sprawności. Stanowi to niezwykle cenną lekcję epistemicznej pokory, przypominającą brutalną rynkową korektę po okresie nadmiernej technologicznej hossy. Pokazuje to dobitnie, że fenomenalna skuteczność indukcji w wąskich domenach nie gwarantuje jeszcze swobodnej, przenośnej i w pełni uniwersalnej inteligencji ogólnej.

Z drugiej strony nie wolno nam popadać w fałszywą dychotomię i ogłaszać całkowitego bankructwa podejścia indukcyjnego, co byłoby równie płytkie jak bezkrytyczne bałwochwalstwo wobec samego skalowania parametrów. Najnowsze systemy rozumujące udowadniają bowiem, że indukcję można

skutecznie wzmocnić mechanizmami przypominającymi deliberację, systematycznym rozpisywaniem kroków pośrednich, poszukiwaniem kontrprzykładów czy wewnętrzną kontrolą przebiegu wnioskowania. Sęk w tym, że owa „widzialna dedukcyjność" to wciąż tylko architektoniczna fasada, a nie solidna ściana nośna oparta na fundamentach formalnej dedukcji. Nawet najbardziej imponujące łańcuchy rozumowania bywają niezwykle kruche, częściowo pozorne i wyjątkowo trudne do systematycznego monitorowania w czasie rzeczywistym. Badania OpenAI z końca 2025 roku nad przejrzystością tych procesów potwierdzają, że to obiecująca, lecz wysoce niestabilna warstwa bezpieczeństwa, całkowicie zależna od kaprysów treningu i jakości danych. Nie pytamy wyłącznie o wynik, lecz o jakość architektury generującej wynik, przy czym zręczna imitacja procesu myślowego nie czyni z modelu transparentnego, trwałego nośnika uzasadnienia.

Test ARC-AGI-2: bariery maszynowej indukcji

W tym miejscu musimy zmierzyć się z tezą wysoce niepopularną. Nauka generowana przez sztuczną inteligencję nie uratuje samej siebie poprzez dalsze doskonalenie indukcji, nawet jeśli ta okaże się oszałamiająco skuteczna. Nie dlatego, że indukcja jest słaba, lecz dlatego, że nie zaspokaja fundamentalnych potrzeb cywilizacji wiedzy. Państwo, sąd, regulator leków czy projektant reaktora nie potrzebują bowiem systemu, który po prostu „zwykle trafia". Wymagają one rygorystycznej architektury poznawczej, która gwarantuje pełną rozliczalność podejmowanych decyzji.

Wspomniane instytucje potrzebują mechanizmu, którego wynik można poddać audytowi i osadzić w twardej strukturze uzasadnienia. W jednych obszarach wystarczy mocna statystyka, podczas gdy w innych niezbędny staje się dowód oraz ścisła kontrola semantyczna. Serce nauki nie bije w prawdzie. Bije w uzasadnieniu. Przyszłość sztucznej inteligencji nie leży zatem w triumfie indukcji nad dedukcją, lecz w ich hierarchicznym sprzęgnięciu. Indukcja ma kompresować przestrzeń możliwości i dostarczać heurystyk, dedukcja zaś musi domykać to, co wymaga niepodrabialnej ścieżki inferencyjnej.

Sprzężenie indukcji z dedukcją: przyszłość nauki

Z tej perspektywy tezy Stegengi okazują się wręcz konstytucją dla nadchodzącej epoki AGI. Wspólna wiedza oznaczałaby w tym ujęciu nie tylko naiwną zgodę badaczy co do ostatecznego wyniku, lecz przede wszystkim wspólną zdolność do rygorystycznego odtworzenia całej drogi. Dotyczy to zwłaszcza sytuacji, gdy znaczna część tej trasy została wytyczona przez system sztucznej inteligencji. Ocena deontologiczna wymusza bowiem radykalne przesunięcie naszej uwagi, toteż nie

pytamy wyłącznie o wynik, lecz o jakość architektury generującej ten wynik. Fascynacja skutecznością modelu musi ustąpić miejsca chłodnemu audytowi całego procesu badawczego pod kątem jawności, odporności na błędy poznawcze i pełnej replikowalności. Ideał wolności od wartości nakłada na nas bezwzględny obowiązek identyfikowania punktów krytycznych nie tylko w ludzkich uprzedzeniach, lecz także w samej architekturze danych treningowych czy parametrach optymalizacji.

Asercja naukowa w tym nowym reżimie wymagałaby, aby komunikat współgenerowany przez maszynę był zarazem głęboko uzasadniony i rzeczywiście informatywny, a nie tylko elegancko opakowany w erudycyjny żargon. Postęp naukowy należałoby zatem mierzyć zupełnie nową walutą, odrzucając dotychczasowe wskaźniki. Nie chodzi już o liczbę efektownych, wygenerowanych w ułamku sekundy hipotez, które zalewają akademickie rynki. Prawdziwą miarą staje się zmiana poziomu uzasadnienia, jaką hybrydowy układ człowieka i maszyny realnie wnosi do wspólnoty badawczej. Nauka w tym ujęciu przestaje być loterią trafnych przeczuć, stając się bezwzględną instytucją produkcji dowodów, w której każda nowa teza musi przejść rygorystyczną kontrolę jakości.

W tym świetle doskonale widać, jak bardzo drażliwy stanie się wkrótce problem akademickiego kredytu i dystrybucji nagród. Kiedy sztuczna inteligencja zacznie taśmowo generować przełomowe przypuszczenia, strategie dowodowe czy gotowe projekty eksperymentów, dotychczasowa reguła pierwszeństwa ukaże swój całkowity absurd. Kto wszakże będzie „pierwszy", jeżeli krzemowy agent potrafi wytworzyć tysiąc obiecujących kandydatów na teorię przed poranną kawą badacza? Kto ma ostatecznie otrzymać laur odkrywcy, skoro właściwa wartość poznawcza nie leży w tanim wygenerowaniu sugestii, lecz w jej żmudnym sprawdzeniu i włączeniu do obiegu? Stegenga dostarcza nam tutaj gotowej, niemal księgowej odpowiedzi, twierdząc, że renta poznawcza ma przypadać dokładnie tam, gdzie rośnie twarde uzasadnienie. W epoce zaawansowanej sztucznej inteligencji ten postulat stanie się nie tylko filozoficznie pociągający, ale wręcz administracyjnie konieczny do przetrwania uniwersytetów.

W tym miejscu na scenę wkracza trzeci akt naszego epistemologicznego dramatu, czyli komputery kwantowe. Musimy jednak zachować absolutną trzeźwość osądu i nie ulegać technologicznym halucynacjom o natychmiastowej rewolucji. Nie dysponujemy dziś jeszcze powszechną, wielkoskalową i odporną na błędy infrastrukturą obliczeń kwantowych, co potwierdzają rygorystyczne przeglądy naukowe. Zarówno publikacje w „Nature Communications" z 2025 roku, jak i oficjalne komunikaty gigantów pokroju Google czy IBM podkreślają to z całą mocą. Choć dokonano bezsprzecznego postępu w korekcji błędów i tworzeniu kubitów logicznych, czyli jednostek informacji wolnych od fizycznego szumu, skalowanie tych układów pozostaje

monumentalnym wyzwaniem. Inżynierowie otwarcie przedstawiają obecne osiągnięcia jako zaledwie obiecującą drogę ku stabilnym obliczeniom, a bynajmniej nie jako ostatecznie zamknięty rozdział historii techniki.

A jednak to właśnie na tym styku wyraźnie rysuje się horyzont naszej przyszłości. Jeżeli sztuczna inteligencja dojrzeje jako główny architekt hybrydowego rozumowania, a komputery kwantowe osiągną wreszcie użyteczną stabilność, nie nastąpi wcale triumf mglistej intuicji nad twardą logiką. Nastąpi coś dokładnie odwrotnego, wręcz dialektyczne zniesienie dotychczasowych ograniczeń, w którym zasady myślenia dedukcyjnego zostaną uniesione na zupełnie nowy poziom infrastrukturalny. Maszyna będzie mogła indukcyjnie generować ogromne przestrzenie hipotez, błyskawicznie szukać skrótów, nieoczywistych analogii i kandydatów na dowody. Systemy formalne będą z kolei bezlitośnie sprawdzać poprawność tych inferencji, poszerzając obszar tego, co da się formalnie domknąć w sensownym czasie. Dedukcja bynajmniej nie zniknie, toteż zostanie po prostu zindustrializowana na niespotykaną dotąd skalę.

To jest właściwa pointa całego tego procesu, którą warto postawić bez cienia intelektualnej kokieterii. Sama indukcja nigdy nie podoła precyzji rachunku zdań tam, gdzie stawka wymaga absolutnej konieczności, a nie tylko wysokiego prawdopodobieństwa. Niemniej indukcja wsparta potęgą AGI staje się potężną giełdą odkrywania ukrytych struktur, gotowych do bezlitosnego audytu i dedukcyjnego uporządkowania. Nie czeka nas zatem żaden romantyczny zmierzch logiki, lecz raczej jej nowa, brutalnie skuteczna epoka. Serce nauki nie bije w prawdzie. Bije w uzasadnieniu, bez którego nawet najpotężniejsza inteligencja wyprodukuje jedynie gęsty, epistemiczny dym.

Postęp naukowy: nowa definicja w epoce AGI

Maszynowe myślenie indukcyjne uwodzi nas swoją skutecznością dokładnie tam, gdzie powinno uruchamiać w nas najostrzejsze dzwonki alarmowe. Jego triumfy są bezdyskusyjnie realne, ale z przerażającą łatwością mylimy je z pełnią rozumu. Systemy uczące się na gigantycznych zbiorach danych potrafią wydobywać regularności, kompresować wzorce i przewidywać kolejne stany z dokładnością, która jeszcze dekadę temu uchodziłaby za czystą magię. A jednak ten spektakularny sukces ma strukturę niezwykle kruchą, przypominającą bańkę spekulacyjną na giełdzie poznawczej. Jest to bowiem przede wszystkim sukces ekstrapolacji, a nie triumf swobodnego, elastycznego rozumienia. Benchmark ARC-AGI-2 został zaprojektowany właśnie po to, by przeprowadzić bezlitosny audyt tej różnicy. Mierzy on nie nagromadzoną biegłość ani encyklopedyczny zapas danych, lecz few-shot generalization, czyli zdolność uchwycenia nowej reguły przy minimalnym,

niemal śladowym doświadczeniu. Według oficjalnych materiałów ARC Prize czyste modele językowe osiągają tam wynik absolutnie zerowy, systemy oparte na łańcuchach rozumowania zaledwie jednocyfrowe, a najlepszy wynik w konkursie na prywatnym zbiorze sięgnął około dwudziestu czterech procent. Zarazem twórcy tego testu z naciskiem podkreślają, że zadania są projektowane jako banalnie wręcz łatwe dla ludzi, a w testach kontrolnych człowiek potrafił bezbłędnie rozwiązać komplet zadań ewaluacyjnych. To nie jest drobna rysa na lśniącym wizerunku sztucznej inteligencji, którą można zaszpachlować kolejnym miliardem parametrów. To potężny, epistemologiczny alarm, który dobitnie pokazuje, że indukcja maszynowa nie daje nam jeszcze przenośnej inteligencji ogólnej.

Pierwsza bariera, na którą natrafiamy, ma charakter stricte architektoniczny. Indukcja działa świetnie tam, gdzie świat posłusznie dostarcza nam tysięcy podobnych przypadków i gdzie regularność można uchwycić jako statystyczny cień rzucany przez ogromną liczbę obserwacji. ARC-AGI zostało jednak ufundowane na zupełnie innej intuicji, opisanej przez François Cholleta w jego fundamentalnym tekście o mierzeniu inteligencji. Sama umiejętność wykonania konkretnego zadania nie jest dobrym miernikiem rozumu, ponieważ wysoki poziom tej umiejętności można po prostu „kupić" ogromem treningu, danych i wbudowanych założeń. Inteligencja ma być raczej sprawnością nabywania zupełnie nowych umiejętności przy drastycznie ograniczonym doświadczeniu, niczym sprawny analityk potrafiący wycenić nową sytuację na podstawie zaledwie kilku twardych przesłanek. Innymi słowy, maszyna może być fenomenalnie biegła w swoim fachu, a zarazem dramatycznie nieinteligentna w sensie przenoszenia odkrytych zasad na sytuacje całkowicie nowe. To właśnie w tym miejscu myślenie indukcyjne z hukiem natrafia na betonowy mur. Jest ono absolutnym mistrzem gęstych statystycznie pejzaży, ale błyskawicznie gubi się w krajobrazie ubogim w przykłady, gdzie trzeba szybko wydestylować regułę z zaledwie kilku sygnałów i odróżnić głęboką istotę zadania od jego zwodniczej, wizualnej powierzchni.

Druga bariera dotyczy problemu kompozycyjności, który obnaża strukturalne ubóstwo obecnych modeli. Oficjalny raport techniczny ARC-AGI-2 bezlitośnie wskazuje, że współczesne systemy mają szczególny, wręcz paraliżujący problem z zadaniami wymagającymi jednoczesnego zastosowania wielu reguł. Dzieje się tak zwłaszcza wtedy, gdy reguły te wchodzą ze sobą w dynamiczne interakcje, tworząc sieć wzajemnych zależności. Jeżeli zadanie posiada jedną, wyraźnie dominującą zasadę, model bywa całkiem skuteczny i potrafi wygenerować poprawną odpowiedź. Jeżeli jednak trzeba uchwycić swoistą regułę reguł, uporządkować hierarchię relacji między transformacjami i wykonać bezbłędną sekwencję działań, wydajność maszyny gwałtownie pikuje w dół. To nie jest jakiś marginalny drobiazg techniczny, który naprawimy w kolejnej aktualizacji oprogramowania, lecz centralny, konstytutywny deficyt samej indukcji. Indukcja z natury lubi regularność prostą, płaską i jednorodną, z której może czerpać łatwe renty poznawcze. Tymczasem ludzki umysł radzi sobie w

świecie właśnie dzięki nieustannemu składaniu prostych reguł w wyrafinowane struktury wyższego rzędu. Człowiek nie tylko biernie widzi wzór, ale potrafi nim aktywnie zarządzać, podczas gdy model często rozpoznaje poszczególne elementy układanki, lecz nie potrafi przejść od katalogu lokalnych trafień do globalnej organizacji problemu.

Trzecia bariera wiąże się nierozerwalnie z pojęciem kontekstu i elastyczności semantycznej. Twórcy ARC Prize z naciskiem podkreślają, że systemy sztucznej inteligencji mają gigantyczne trudności z zadaniami, w których tę samą, pozornie znaną regułę trzeba stosować odmiennie w zależności od zmieniającej się sytuacji. Modele wykazują zgubną skłonność do przyklejania się do cech czysto powierzchownych, zamiast uchwycić głęboką zasadę selekcji, która w ogóle dyktuje, kiedy dana transformacja ma rację bytu. To właśnie jedna z najbardziej zdradliwych i frustrujących słabości maszynowej indukcji, która udaje zrozumienie tam, gdzie go de facto nie ma. Wysoka sprawność predykcyjna bywa tu okupiona dramatycznie płytką semantyką, przypominającą urzędnika perfekcyjnie wypełniającego formularze bez zrozumienia ich celu. Model potrafi bezbłędnie znaleźć korelację w gąszczu danych, lecz niemal nigdy nie umie samodzielnie rozstrzygnąć, czy jest ona konstytutywna dla istoty zadania, czy jedynie przypadkowo współwystępuje z poprawnym wynikiem. Człowiek w obliczu nowości częściej pyta implicite: „co tutaj ma fundamentalne znaczenie?”. Model z kolei pyta wyłącznie: „co tutaj statystycznie wygląda najbardziej obiecująco?”. W wielu zamkniętych domenach praktycznych to wystarcza, ale w domenie radykalnej nowości i abstrakcji ten mechanizm zaczyna spektakularnie zawodzić.

Czwarta bariera ma charakter głęboko epistemologiczny i dotyczy fundamentalnej relacji między zaobserwowanym wzorcem a uniwersalną regułą. Indukcja maszynowa jest niezwykle mocna w wykrywaniu powierzchownych podobieństw, ale pozostaje bezradna w uchwyceniu logicznej konieczności. Widzi ona doskonale, że pewne elementy zwykle występują razem, lecz nie potrafi autorytatywnie stwierdzić, co musi bezwzględnie wynikać z czego. Tymczasem lwią część zadań inteligentnych stanowi nie tylko trafna predykcja, ale przede wszystkim rozpoznanie warunku wystarczającego, warunku koniecznego, dopuszczalnego wyjątku oraz poziomu opisu, na którym dana reguła zachowuje prawdziwość. To już nie jest zwykła, rzemieślnicza statystyka, lecz sam zalążek rygorystycznej logiki. Gdy system myli te subtelne poziomy, natychmiast wpada w pułapkę nadmiernego dopasowania do danych treningowych, a jego odpowiedź, choć efektowna, bywa poznawczo całkowicie pusta. Właśnie dlatego ARC-AGI zostało pomyślane jako rygorystyczny test fluid intelligence, czyli zdolności do płynnego rozumowania i adaptacji w czasie rzeczywistym, a nie jako test crystallized intelligence, czyli zdolności do odtwarzania skumulowanych, zamrożonych w wagach sieci umiejętności.

Piąta bariera jest nierozerwalnie związana z pojęciem priors, czyli z wrodzonymi lub wbudowanymi założeniami o strukturze otaczającego nas świata. Chollet od samego początku swojego projektu argumentował, że nie istnieje coś takiego jak inteligencja w próżni, pozbawiona pewnego zestawu początkowych, uprzednich struktur poznawczych. Sensowny i uczciwy benchmark musi te struktury rygorystycznie kontrolować, zamiast wygodnie ukrywać je pod stertą petabajtów danych treningowych. ARC opiera się właśnie na koncepcji core knowledge priors, czyli na minimalnym zestawie fundamentalnych intuicji o przestrzeni i obiektach, które dla ludzi są niemal przezroczyste, a dla modeli matematycznych już niekoniecznie. Dlatego właśnie żenująca słabość najpotężniejszych systemów na ARC-AGI-2 pokazuje nam coś znacznie głębszego niż tylko chwilowy brak mocy obliczeniowej w centrach danych. Pokazuje ona dobitnie, że sama brutalna skala nie rozwiązuje problemu braku odpowiednich fundamentów pojęciowych. Można w nieskończoność pompować miliardy parametrów w kolejne generacje modeli, ale jeśli ich architektura nie umie efektywnie reprezentować obiektów i relacji, rozumowanie pozostanie procesem ciężkim, kosztownym i niezwykle kruchym.

Szósta bariera to zjawisko znane jako sample inefficiency, czyli dramatyczna wręcz nieefektywność uczenia się z małej liczby dostarczonych przykładów. Przeciętny człowiek potrafi często z zaledwie dwóch lub trzech demonstracji bezbłędnie wydobyć ukrytą regułę, a następnie płynnie zastosować ją do zupełnie nowego przypadku bez konieczności odbywania tysiąca dodatkowych prób. ARC-AGI zostało stworzone dokładnie po to, by ten uderzający kontrast między umysłem a maszyną wyciągnąć na światło dzienne i poddać rygorystycznej ocenie. Jeśli najlepszy wynik prywatny, osiągnięty po tysiącach prób konkursowych i wpompowaniu ogromnego wysiłku badawczego, wciąż oscyluje wokół zaledwie jednej czwartej, wnioski nasuwają się same. Znaczy to ni mniej, ni więcej, że problem nie leży wyłącznie w mozolnym dostrojeniu hiperparametrów, lecz w samej fundamentalnej naturze obecnie stosowanych metod statystycznych. Maszyna nadal zużywa absurdalnie wręcz wiele doświadczenia, by osiągnąć to, co dla człowieka bywa intuicyjnym, błyskawicznym skokiem do właściwego poziomu abstrakcji. Taka nieefektywność nie jest tylko prozaicznym kosztem technicznym, lecz twardym wskaźnikiem drastycznie ograniczonej elastyczności poznawczej.

Siódma bariera dotyczy samej istoty reprezentacji wiedzy w systemach sztucznej inteligencji. Indukcja nigdy nie operuje w sterylnej próżni; operuje ona wyłącznie na tym, jak dany wycinek świata został uprzednio zakodowany przez twórców systemu. W środowisku ARC zadania mają postać dwuwymiarowych siatek oraz transformacji obiektów w przestrzeni dyskretnej, co dla ludzkiego oka jest zaskakująco naturalne. Dla modelu językowego czy wizyjnego to często nadal tylko płaska tablica znaków lub pikseli, chyba że jego architektura jawnie i celowo wspiera obiektowość oraz operacje o charakterze programowym. Z tego właśnie powodu w środowisku

skupionym wokół ARC rośnie gwałtownie zainteresowanie program synthesis, czyli metodami łączącymi uczenie głębokie z explicite symbolicznymi, rygorystycznymi procedurami. Sam zespół ARC Prize pisał wprost, że samo uczenie głębokie absolutnie nie wystarczy i że konieczne jest jego ożenienie z syntezą programów. Gdy najbardziej ambitny benchmark inteligencji ogólnej popycha czołowych badaczy ku rozwiązaniom bardziej strukturalnym, to znaczy, że czysta indukcja właśnie z impetem trafia na granicę własnego paradygmatu.

Ósma bariera jest z pewnością najbardziej filozoficzna, ale jej konsekwencje są na wskroś praktyczne i mierzalne. Maszynowe myślenie indukcyjne cierpi na chroniczny, nieuleczalny deficyt samowiedzy co do własnego, wewnętrznego uzasadnienia. Potrafi ono z imponującą prędkością generować trafne wyjścia, lecz znacznie trudniej przychodzi mu odróżnienie odpowiedzi faktycznie poprawnej od odpowiedzi, która jedynie dobrze i wiarygodnie wygląda. W brutalnej praktyce oznacza to, że system bywa nieporównywalnie lepszy w produkowaniu gładkich wyników niż w produkowaniu twardych racji, które można by poddać rygorystycznej kontroli. Serce nauki nie bije w prawdzie. Bije w uzasadnieniu. Właśnie dlatego tak żenująco niskie wyniki czystych modeli językowych na ARC-AGI-2 nie są jedynie przejściową kompromitacją marketingową, lecz niezwykle cenną lekcją metodologiczną. Uczą nas one dobitnie, że płynność generacji tekstu czy kodu nie jest w żadnym wypadku równoznaczna z głębią faktycznego rozumienia problemu.

Dziewiąta bariera odsłania się w pełnej krasie dopiero wtedy, gdy odważymy się zestawić maszynę z człowiekiem bez cienia technologicznego sentymentalizmu. Ludzie z pewnością nie są idealnymi, nieomylnymi logikami, ale dysponują czymś, czego obecnym modelom nadal dramatycznie brakuje: zdolnością do elastycznej zmiany poziomu opisu sytuacji. Potrafią oni zobaczyć to samo zadanie raz jako problem doboru kolorów, za chwilę jako problem fizycznych obiektów, a jeszcze kiedy indziej jako problem ukrytej intencji transformacji. Model statystyczny znacznie częściej grzęźnie bezradnie w jednym, sztywnym sposobie kodowania danej sytuacji, nie potrafiąc wyjść poza narzucone ramy. To właśnie dlatego zadania określane jako „łatwe dla ludzi, trudne dla AI” są tak fundamentalnie ważne z naukowego punktu widzenia. Pokazują one jasno, że główna luka poznawcza nie przebiega dziś wcale po linii braku mocy obliczeniowej, lecz po linii swobody rekonfiguracji własnych reprezentacji. Bez tego zostaje nam już tylko walka narracji, czyli rynek symboliczny, na którym wygrywa nie ten, kto lepiej uzasadnia, lecz ten, kto lepiej dystrybuuje afekt i wiarę w nieuchronność sztucznej inteligencji.

Najkrócej więc rzecz ujmując, fundamentalne bariery maszynowego myślenia indukcyjnego są co najmniej cztery, a każda z nich sięga znacznie głębiej, niż sugeruje to naiwny język inżynierów. Pierwszą z nich jest chroniczna trudność wydobywania uniwersalnej reguły z małej liczby przykładów. Drugą jest strukturalna słabość kompozycyjna, czyli kłopot ze sprawnym składaniem

wielu drobnych reguł w jedną, spójną architekturę rozwiązania. Trzecią jest niebezpieczna podatność na powierzchniowe korelacje, realizowana kosztem głębokiego sensu zależnego od zmiennego kontekstu. Czwartą wreszcie jest brak mocnej, przenośnej reprezentacji, która pozwalałaby systemowi przejść od statystycznego wzorca do logicznej konieczności. Benchmark ARC-AGI-2 nie tyle autorytatywnie udowadnia, że sztuczna inteligencja jest po prostu głupia, ile brutalnie i potrzebnie porządkuje naszą rozgorączkowaną dyskusję. Pokazuje on czarno na białym, że między imponującą biegłością w naśladownictwie a prawdziwą inteligencją ogólną nadal rozciąga się niezwykle szeroki pas ziemi niczyjej.

I właśnie w tym miejscu pojawia się właściwa, najważniejsza pointa naszych rozważań. Rozumowanie indukcyjne absolutnie nie jest porażką; jest ono wielkim, potężnym, lecz z natury ograniczonym narzędziem w naszym arsenale. Potrafi być fenomenalnie skuteczne w optymalizacji procesów, ale samo z siebie nigdy nie gwarantuje logiki, nie gwarantuje konieczności i nie gwarantuje przenoszenia zasad między nowymi światami z ludzką swobodą. ARC-AGI-2 działa więc w dzisiejszym dyskursie technologicznym jak niezwykle bolesna, ale konieczna szczepionka przeciwko epistemologicznej pysze. Uczy nas, że jeśli chcemy w ogóle mówić poważnie o perspektywie inteligencji ogólnej, musimy odważnie wyjść poza sam triumf skali i statystyki. Musimy zacząć bezkompromisowo pytać o reprezentację, kompozycyjność, wbudowane założenia, transfer wiedzy i formalną strukturę uzasadnienia. Przyszłość nie należy ani do ślepej wiary w czystą indukcję, ani do romantycznej nostalgii za czystą symboliką. Należy ona wyłącznie do architektur, które potrafią połączyć surową siłę indukcyjnego odkrywania z rygorem bardziej logicznego, strukturalnego i kontrolowalnego rozumowania. ARC-AGI-2 nie zamyka więc ostatecznie sprawy sztucznej inteligencji, ono dopiero otwiera przed nami właściwe pytanie o to, czym tak naprawdę jest myślenie.

> Czy w epoce, w której maszyny przejmują ciężar wnioskowania, naszą najważniejszą kompetencją stanie się umiejętność zadawania pytań o architekturę ich decyzji? Być może prawdziwym testem inteligencji nie jest już zdolność do generowania odpowiedzi, lecz odwaga, by domagać się uzasadnienia tam, gdzie algorytm oferuje jedynie czarną skrzynkę. W świecie wiecznej zmienności, naszą jedyną kotwicą pozostaje rygorystyczny proces, który – niczym konstytucja – chroni prawdę przed jej własną płynnością.

Inspiracje

[1] **"Heart of Science"** Jacob Stegenga

Treść tworzy Fundacja Dobre Państwo.
Redaguje i publikuje APA ONE, autorski system redakcyjny Fundacji oparty na AI.
APA ONE Publishing System
Fundacja Dobre Państwo © 2026
Article ID: 7119e946-2753-4fcb-8488-2a1988ba4bda