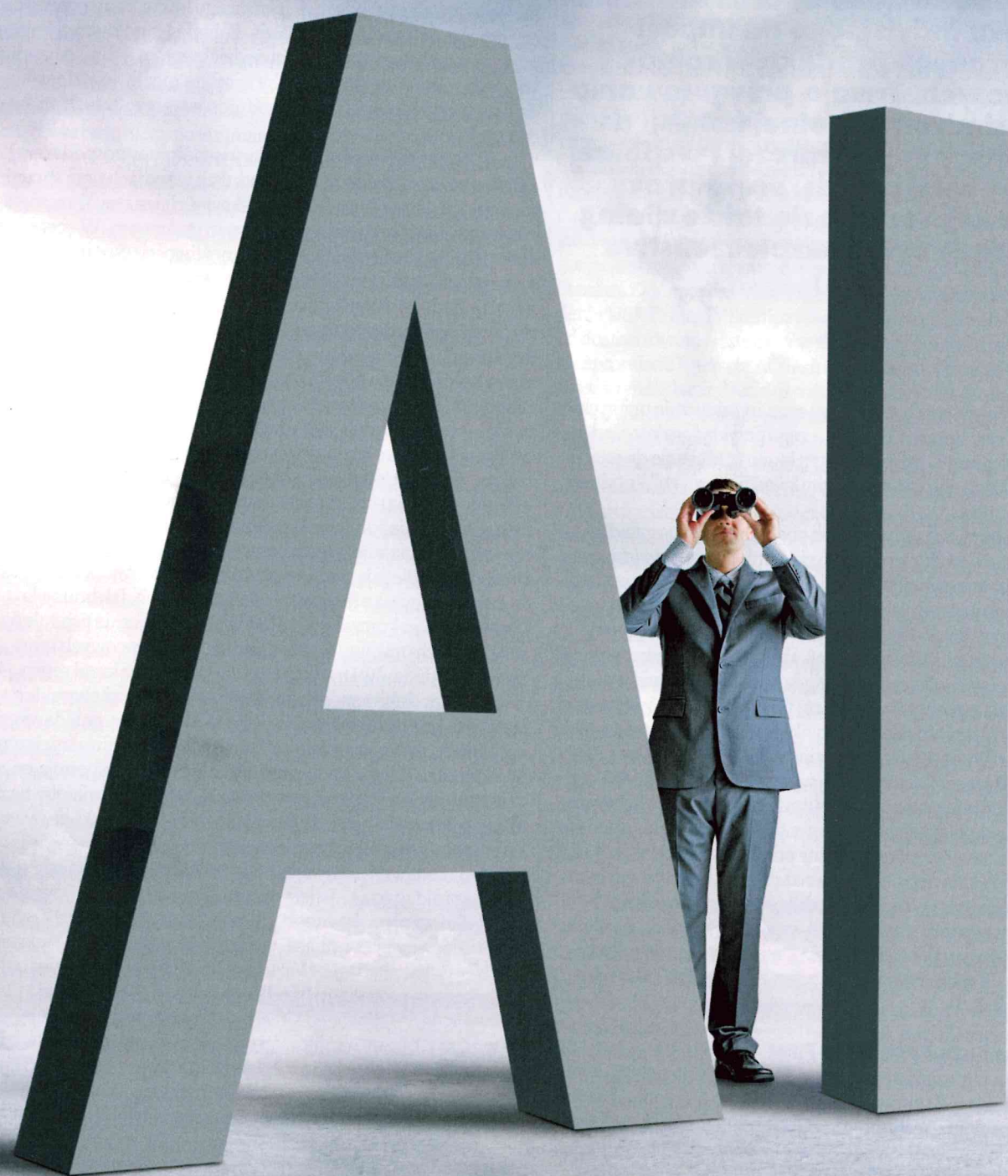


wywiad



Świat ery AGI: cyfrowa utopia czy technologiczna dyktatura

Sztuczna inteligencja może nie tylko przejąć wiele zadań wykonywanych przez ludzi, ale też radykalnie zmienić fundamenty globalnego ładu. W optymistycznym scenariuszu oznaczać to może „w pełni zautomatyzowany luksusowy komunizm”. Jednak jeszcze bardziej prawdopodobne są mroczne wizje: podział ludzkości na kasty, neokolonializm technologiczny, ekstremalne nierówności czy cyfrowe dyktatury zarządzane przez ogólną sztuczną inteligencję. Utopia może łatwo przerodzić się w dystopię – przestrzega prof. Jakub Growiec, doradca ekonomiczny w Departamencie Analiz i Badań Ekonomicznych NBP i wykładowca SGH.

„OBSERWATOR FINANSOWY”: Dlaczego dotychczasowe źródła długoterminowego wzrostu gospodarczego przestają działać?

JAKUB GROWIEC: „Tradycyjne” źródła wzrostu gospodarczego, które z powodzeniem służyły nam przez cały dwudziesty wiek, wyczerpują się. Widzę tu dwie kluczowe przyczyny. Po pierwsze – demografia. Społeczeństwa się starzeją – przyrost naturalny w krajach rozwiniętych jest bliski zera lub ujemny. Ponadto niemal każdy, kto tylko jest w stanie, zdobywa dziś wyższe wykształcenie, wobec czego praktycznie

wyczerpał się też potencjał wzrostowy akumulacji kapitału ludzkiego. Po drugie – ograniczenia poznawcze człowieka. Wydaje się, że docieramy już do granic możliwości ludzkiego mózgu i to pomimo jego intensywnej suplementacji za pomocą komputerów, smartfonów i innych elektronicznych gadżetów. Jak udokumentowali to Nicholas Bloom i współautorzy w szeroko cytowanym artykule „Are Ideas Getting Harder to Find?” opublikowanym w „American Economic Review” w 2020 r., nowe pomysły są coraz trudniejsze do wynalezienia, przez co podtrzymywanie stałego tempa postępu technolo-

**Od technologii cyfrowych,
w tym AI, nie ma ucieczki**

Punktem zwrotnym będzie moment, w którym AI przestanie być komplementarna względem pracy umysłowej człowieka, a stanie się substytucyjna

gicznego wymaga nie stałego, lecz przyspieszającego wzrostu nakładów na badania. W długim okresie dynamika taka jest nie do utrzymania.

Czy nadszedł moment, by zakwestionować dominację wzrostu gospodarczego jako głównego celu rozwoju?

Osobiście nie wierzę w postwzrost (koncepcja przyszłości bez lub ze znacznie powolniejszym niż dziś wzrostem gospodarczym – przyp. red.). Oczywiście PKB nie jest miarą doskonałą, w szczególności, gdy chcemy precyzyjnie mierzyć jakość naszego życia. Wystarczy tu wspomnieć niewliczaną do PKB produkcję domową czy darmowe usługi cyfrowe, które podnoszą jakość naszego życia, albo wliczane do PKB wydatki na zbrojenia i ochronę, które jakości życia nie podnoszą. Można też wymienić nieuwzględniane w PKB efekty zewnętrzne wzrostu, jak np. zanieczyszczenie środowiska czy emisja gazów cieplarnianych. Uważam jednak, że na dłuższą metę bez wzrostu PKB nie da się systematycznie podnosić jakości życia. Trudno też się spodziewać, że człowiek z własnej woli zrezygnuje ze wzrostu. Naturą człowieka jest dążenie do rozwoju, postępu. To jest wrodzone i uwarunkowane ewolucyjnie. Nawet jeśli przez tysiące lat ery przedprzemysłowej wzrost gospodarczy był minimalny, to nie dlatego, że tak świadomie wybraliśmy, lecz dlatego że pomimo nieustających wysiłków po prostu nie byliśmy w stanie sprawić, by ten wzrost był szybszy. Brakowało nam wtedy jeszcze niezbędnych ku temu technologii.

Gospodarka światowa w XX w. i pierwszych dekadach XXI w. rosła w średnim tempie ok. 2–3 proc. rocznie i nie mogła rosnąć szybciej, gdyż jej wzrost ograniczała efektywna podaż pracy, w szczególności umysłowej, per capita. W związku z tym w ubiegłym wieku kluczowym motorem wzrostu gospodarczego

Klasyczne modele wzrostu gospodarczego są świetne, ale opisują gospodarki, które już nie istnieją

go był postęp technologiczny zwiększający produktywność pracy. Nie bez znaczenia był też stopniowy wzrost wykształcenia społeczeństw – czyli oba te motory, których moce dziś się stopniowo wyczerpują. Jednocześnie współczesne gospodarki, a w szczególności rynki finansowe czy systemy ubezpieczeń społecznych, budowane są przy założeniu, że wzrost gospodarczy będzie w przyszłości kontynuowany.

Jeśli wzrost gospodarczy wciąż jest konieczny, to co może go napędzać?

W tym miejscu może wkroczyć sztuczna inteligencja. Mam na myśli algorytmy AI na tyle zaawansowane, by mogły w pełni zastępować pracę umysłową ludzi. Technologie takie jeszcze nie istnieją, ale prawdopodobnie wkrótce powstaną. Z punktu widzenia wzrostu gospodarczego będzie to przełom, ponieważ w przeciwieństwie do naszych mózgów, a każdy człowiek ma jeden mózg i nie może mieć ich więcej, moce obliczeniowe AI i jej kod mogą być bez ograniczeń akumulowane per capita.

Na razie nie widać dużego wpływu AI na wzrost. Czy przyszłość przyniesie zmianę?

Jak dotąd przełożenie dynamicznego rozwoju AI na wzrost gospodarczy było rzeczywiście bardzo niewielkie. Wynika to przede wszystkim z faktu, że wielu zadań póki co nie potrafimy zautomatyzować. AI, czy w szczególności narzędzia generatywnej AI, takie jak czatboty czy copiloty, wspomagają nas w pracy, ale na razie nie potrafią nas w niej całkowicie wyręczyć. Praca umysłowa człowieka pozostaje więc komplementarna względem AI, a to oznacza, że jest ona nadal niezbędna i nadal ogranicza tempo wzrostu gospodarczego.

Ekspertów zajmujących się prognozowaniem wpływu AI na globalny wzrost gospodarczy można dziś podzielić na dwa obozy, „konserwatywny” i „radykalny”. Obóz „konserwatywny” uwzględnia w swoich predykcjach wyłącznie technologie i zjawiska, które widoczne są już teraz, które można zidentyfikować w danych historycznych. W efekcie przewidują oni bardzo niewielkie, czasem wręcz pomijalnie małe efekty AI. Takie wnioski wysuwa m.in. noblista Daron Acemoglu. Obóz „radykalny” podkreśla natomiast szaloną dynamikę rozwoju kompetencji AI w ostatnich latach i stara się ekstrapolować ją w przyszłość. W rezultacie otrzymujemy prognozy, że AI może w najbliższym czasie – według Daniela Kokotajlo i współautorów scenariusza „AI 2027” już w 2027 r. – przekroczyć poziom ogólnej inteligencji człowieka, a w dalszej kolejności całkowicie go zastąpić we wszystkich rodzajach pracy umysłowej.

A z którym z tych dwóch podejść pan się utożsamia?

Osobiście znacznie bliżej mi jest do tego drugiego. Uważam, że punktem zwrotnym, który może zaistnieć w potencjalnie całkiem nieodległej przyszłości – nawet jeśli nie w 2027 r., to może np. w roku 2032, jak przewiduje społeczność prognostyczna Metaculus – będzie moment, w którym AI przestanie być komplementarna względem pracy umysłowej człowieka, a stanie się substytucyjna. Od tego momentu głównym motorem wzrostu gospodarczego przestanie być postęp technologiczny zwiększający produktywność pracy człowieka, a stanie się nim akumulacja mocy obliczeniowych zdolnych do wykonywania

kodu AI. A także robotów i innych urządzeń, którymi zaawansowana AI będzie mogła sterować.

Zgodnie z obserwowanym od lat 80. XX w. prawem Moore’a, całkowita moc obliczeniowa komputerów na świecie podwaja się w okresie krótszym niż dwa lata. Nie wykluczam, że w świecie z nadludzką ogólną sztuczną inteligencją (*artificial general intelligence* – AGI), w takim tempie rosnąć będzie światowy PKB. Oznaczałoby to przyspieszenie wzrostu gospodarczego o cały rząd wielkości. Przy czym oczywiście praca ludzka miałaby w tym wzroście coraz mniejszy wkład. Jednocześnie w scenariuszu tym ludzkość, pozbawiona statusu najinteligentniejszego gatunku na świecie, najprawdopodobniej utraci kontrolę nad światem i swoją przyszłością.

Czy w takim scenariuszu demografia przestanie być problemem?

Kiedy ludzkość utraci kontrolę nad swoją przyszłością, odda AGI także możliwość decydowania o własnym życiu i śmierci. Demografia przestanie być wówczas naszym problemem, a stanie się parametrem podlegającym optymalizacji przez AGI. Zresztą problemy demograficzne nie będą wtedy miały tego znaczenia, które mają teraz – dziś dywagujemy nad starzeniem się społeczeństw przede wszystkim ze względu na rynki pracy czy systemy emerytalne.

Domyślam się, że pani pytanie ma jednak wydźwięk bardziej krótkookresowy, dotyczy bardziej świata, w którym AI pozostaje narzędziem w rękach ludzi, a praca ludzka pozostaje niezbędna w wytwarzaniu produktu. W odniesieniu do takiego świata odpowiedziałbym: nie. Sądzę, że dopóki AI nie może w pełni zastąpić człowieka, dopóty tendencje demograficzne będą oddziaływały na tempo wzrostu gospodarczego.

prof. Jakub Growiec,
doradca ekonomiczny
w Departamencie Analiz i Badań
Ekonomicznych NBP i wykładowca
SGH, w pracy naukowej zajmuje
się m.in. długookresowymi
konsekwencjami automatyzacji
i rozwoju AI



Czy dysponujemy modelami ekonomicznymi, które mogą pomóc przewidywać i zarządzać skutkami rozwoju AI?

Tak, oczywiście. Ekonomisci stale nad tym pracują. Nie ulega jednak wątpliwości, że tak, jak rozwój AI jest dziś bardzo dynamiczny, tak samo dynamicznie zmienia się też nasze pojmowanie jego kluczowych cech i kanałów oddziaływania na gospodarkę. Dlatego też poszczególne modele ekonomiczne mogą mieć skrajnie różne implikacje w zależności od przyjętych założeń. Czy AI jest traktowane w modelu jako narzędzie zwiększające produktywność pracy człowieka czy jako niezależny agent, który może człowieka zastąpić? Czy preferencje konsumentów lub bariery instytucjonalne wymagają, by pewne zawody nie zostały nigdy zautomatyzowane? Czy AI

może wносить istotny wkład w działalność badawczo-rozwojową, w tym w szczególności w rozwój modeli AI – tworząc tym samym pętlę dodatniego sprzężenia zwrotnego? Kto będzie uprawniony do pobierania wynagrodzenia z tytułu pracy wykonywanej przez algorytmy AI? I czy AGI może przejąć kontrolę nad zasobami i zmarginalizować, a nawet całkowicie wyeliminować ludzkość? Każde z tych pytań to nadal otwarty temat badawczy.

Czy zatem klasyczne modele wzrostu gospodarczego straciły już swoją aktualność?

To są świetne modele, ale opisywane przez nie gospodarki już nie istnieją. Kluczowe regularności, które modele te objaśniają, takie jak np. sławne „stylizowane fakty” Nicholasa Kaldora, odnosiły się do gospodarki ery przemysłowej i przestały obowiązywać mniej więcej w latach 80. XX w. Od tego czasu ze względu na rozwój technologii cyfrowych obserwujemy nowe trendy, takie jak np. stopniowy spadek udziału wynagrodzenia pracy w PKB, postęp technologiczny wspierający pracowników wykwalifikowanych (*skill-biased technological change*) czy automatyzacja prac rutynowych, a wraz z rozwojem generatywnej AI także niektórych prac nierutynowych. Od lat 80. XX w. widzimy też stopniowy wzrost nierówności dochodowych w krajach rozwiniętych.

Warto zresztą pamiętać, że neoklasyczne modele wzrostu, jak np. model Roberta Solowa z 1956 r., czy też modele endogenicznego wzrostu, jak np. model Paula Romera z 1990 r. czy model „kreatywnej destrukcji” Philippe’a Aghiona i Petera Howitta z 1992 r., nigdy nie były w pełni uniwersalne. Jeszcze zanim zdaliśmy sobie sprawę, że automatyzacja i technologie cyfrowe podważają ich podstawy, już wcześniej zauważono, że równie słabo radzą sobie one z objaśnianiem wzrostu gospodarczego w czasach przedprzemysłowych. Dlatego też, aby połączyć w jednym modelu te dwa odmienne światy i tym samym zrozumieć kluczowe mechanizmy rewolucji przemysłowej, zaczęto formułować tzw. jednolite, zuniifikowane teorie wzrostu. Najdonioślejszy wkład w ten nurt literatury wniósł jego założyciel Oded Galor, badacz z Uniwersytetu Browna w USA.

Dzisiaj potrzebujemy jednak teorii „jeszcze bardziej jednolitej” niż modele Galora – pozwalającej zrozumieć w ramach jednej spójnej struktury nie tylko przejście od gospodarki rolniczej do przemysłowej, ale także od gospodarki przemysłowej do cyfrowej.

Patrząc zatem w przyszłość – które sektory gospodarki mogą najbardziej skorzystać na rozwoju zaawansowanej AI?

Jak pokazują badania empiryczne Tyny Eloundou z OpenAI i jej współautorów, opublikowane w prestiżowym „Science”, generatywna AI jest technologią powszechnego zastosowania. W związku z tym może być i jest aplikowana w każdym sektorze gospodarki, choć oczywiście z różną intensywnością. Generatywna AI jest użyteczna nie tylko w zadaniach rutynowych, powtarzalnych i łatwych do skodyfikowania, ale także w pracy kreatywnej czy analitycznej, którą dotąd wykonywali wyłącznie wysoko wykwalifikowani i dobrze opłacani specjaliści. AI może też być wykorzystywana w procesach zarządczych i przy podejmowaniu decyzji.

A czy rozwój AI może doprowadzić do powstania zupełnie nowych sektorów gospodarki?

Bardzo trudno to przewidzieć. Zgadywałbym, że w pierwszej kolejności usprawni istniejące sektory, a następnie utworzy też nowe. Wyobrażam sobie, że w przyszłości mogą zacząć powstawać na przykład zupełnie nowe podobszary robotyki, produkujące różnego rodzaju zdalnie sterowane urządzenia precyzyjne, których ludzie nie wymyśliли, gdyż jak dotąd nie były im potrzebne.

Jakie bariery uniemożliwiają dziś pełne wykorzystanie AI jako motoru wzrostu gospodarczego?

Można wymienić dwie główne bariery szerokiego wdrożenia technologii AI we współczesnym świecie. Po pierwsze, reorganizacja procesów produkcyjnych czy biznesowych, tak by w większym stopniu bazować na rozwiązaniach AI, wymaga poniesienia kosztów. Po drugie, i póki co chyba najważniejsze, nadal brakuje nam zaufania do wyników pracy AI. Narzędzia AI, choć rozwijają się niezwykle dynamicznie, nie są jeszcze wystarczająco niezawodne, by móc im w pełni zaufać. Brakuje im też autonomii. Technologia agentów AI jest obecnie na dość wstępnym etapie: potrafią oni działać autonomicznie w dość ograniczonym zakresie i względnie często napotykają na proste problemy techniczne.

Jestem przekonany, że te bariery technologiczne zostaną jednak wkrótce przełamane, a wtedy agenci AI zaczną być wdrażani bardzo szeroko. Zapewne w pierwszej kolejności tam, gdzie wykonywane są zadania realizowane dziś przez pracowników siedzących przed komputerami. Mówimy zatem o pracy biurowej, analitycznej, kreatywnej czy menedżerskiej. Automatyzacja z wykorzystaniem AI będzie natomiast wolniejsza tam, gdzie kluczowe pozostają nierutynowe prace manualne. Tam automatyzacja będzie ograniczana – przynajmniej przez jakiś czas – przez niewystarczającą dostępność stosownych robotów.

Czy AI zdestabilizuje rynek pracy?

Tak długo, jak AI jest wyłącznie narzędziem w rękach człowieka, nawet bardzo zaawansowanym, zmiany będą stopniowe i będzie można nimi zarządzać przez punktowe regulacje. Natomiast autonomiczni agenci AI, a zwłaszcza ci obdarzeni nadludzką inteligencją i zdolni w pełni automatyzować całe pro-

cesy gospodarcze, to już zupełnie inna historia. Kiedy powstaną tacy agenci, AI zacznie być substytucyjna względem pracy umysłowej człowieka i będzie mogła go bezpośrednio zastępować. Wtedy człowiek będzie w stanie konkurować z maszyną wyłącznie ceną – a to jest konkurencja, w której ludzkość szybko przegra. Bądź co bądź, w przeciwieństwie do biologicznego, analogowego człowieka, cyfrowa AI może się łatwo skalować, tworząc swoje doskonałe kopie lub zależne od siebie procesy pomocnicze. W takim świecie praca umysłowa człowieka (a w dalszej perspektywie także praca manualna) stanie się zbędna dla gospodarki i będzie wykonywana tylko wtedy, gdy człowiek będzie gotów ją wykonywać nawet bez wynagrodzenia.

Czy możliwe jest, aby praca ludzi była opcjonalna, a społeczeństwo nadal prosperowało?

Tak, w pozytywnym scenariuszu jest to jak najbardziej możliwe. Wyobraźmy sobie w pełni zautomatyzowaną gospodarkę, wytwarzającą ogromny, co dwa lata podwajający się PKB. To nieprzebrane bogactwo, które można na różne sposoby podzielić. Tylko że podział dochodu to już kwestia instytucjonalna, a nie technologiczna.

Można sobie wyobrazić wiele sposobów podziału dochodu, kiedy nie będzie to już możliwe do zrobienia poprzez wynagrodzenie za pracę. Przykładowo AGI mogłaby wprowadzić – jak spekuluje Aaron Bastani – „w pełni zautomatyzowany luksusowy komunizm”, zapewniający wszystkim ludziom na świecie sprawiedliwy dostęp do wszystkich dóbr, także tych luksusowych, zarezerwowanych dotąd tylko dla najbogatszych. Oczywiście możliwy jest też podział ludzi na rozmaite klasy czy kasty. W grę może też wchodzić technologiczny neokolonializm, nawet bowiem względnie przyjazna AGI nie musi przecież działać dla dobra wszystkich ludzi, może np. preferować obywateli jednego kraju albo członków jakiejś specyficznej społeczności. Do wyobrażenia jest nawet ekstremalna nierówność, gdzie niemal całe przychody z działalności AI byłyby zgarniane przez skrajnie wąską grupę np. akcjonariuszy firm AGI. Można sobie też wyobrazić skrajnie restrykcyjne cyfrowe dyktatury zarządzane przez scentralizowaną władzę AGI. Jak widać, nawet utopia może łatwo przerodzić się w dystopię, a postęp technologiczny bynajmniej nie gwarantuje sprawiedliwego podziału dochodu.

Warto bowiem zdać sobie sprawę, że rynek pracy to tylko wierzchołek góry lodowej. Kiedy mówimy o pełnej automatyzacji produkcji i działalności badawczo-rozwojowej, to oznacza to również przekazanie na rzecz AI kontroli nad podejmowanymi decyzjami. A to już scenariusz utraty kontroli ludzkości nad własną przyszłością – AI takeover.

Trzeba to sobie otwarcie powiedzieć: budując AGI, a więc nadludzką inteligentną, autonomiczną sztuczną inteligencję ogólnego zastosowania, tworzymy byt, który przejmie kontrolę nad światem. Wizja, że mniej inteligentny i wolniej myślący człowiek będzie mógł trwale utrzymać kontrolę nad bardziej inteligentną od niego i szybciej myślącą AI, jest mrzonką.

A stąd widać już rozwidlenie dróg. Jedna zmierza w kierunku katastrofy, a druga – utopii. Kluczowe pytanie brzmi, czy uda nam się zapewnić, że nadludzka AGI będzie się o nas bezinteresownie troszczyć. Jest to problem AI alignment – zgodności celów AGI z długofalowym dobrostanem ludzkości. Pro-

blem ten powinien zostać rozwiązany, zanim utworzymy AGI i oddamy jej kontrolę nad światem.

Załóżmy, że to się uda. Co się wtedy wydarzy?

Pójdziemy wówczas w kierunku utopii. Ludzkość będzie czerpać korzyści z bajecznego wzrostu gospodarczego wypracowywanego przez AGI, co dwa czy trzy lata podwajającego światowy PKB. Będziemy też korzystać z dobrodziejstw znacząco przyspieszonego postępu technologicznego. Wynalezione zostaną lekarstwa na raka i inne trapiące nas choroby, ludzie będą żyli dłużej i zdrowiej, itd.

A jeśli się nie uda?

Konsekwencje będą tragiczne. Całkiem realny jest m.in. scenariusz całkowitej zagłady ludzkości. Mówię to całkowicie serio: AGI może nas wszystkich zabić. I co gorsza, nie musi nas przy tym jakoś aktywnie nienawidzić; nie musi nawet posiadać świadomości. Wystarczy czysta, bezwzględna moc optymalizacyjna. Wystarczy, że powstanie konflikt o zasoby, takie jak energia, surowce czy woda. Konflikt taki ludzkość niechybnie przegra i w konsekwencji albo całkowicie wyginie, albo pozostali przy życiu nieliczni ocaleni będą wegetować, zdani na łaskę i niełaskę AGI.

Niestety, wydaje się, że AGI jest już blisko, a rozwiązania problemu AI alignment zupełnie nie widać. To bardzo trudny problem i póki co nie wiemy, czy w ogóle ma jakiegokolwiek rozwiązanie.

A zmieniając nieco temat: czy Polska może dalej rozwijać się bez intensywnego wdrażania nowych technologii?

Nie widzę innych możliwości. Przez ostatnią dekadę jednym z istotnych i względnie niedocenianych źródeł wzrostu w Polsce, obok inwestycji w kapitał fizyczny i absorpcji nowych technologii, był napływ imigrantów, przede wszystkim z Ukrainy i ich integracja na krajowym rynku pracy. Trudno jednak długofalowo budować swoją gospodarkę na niekończącym się napływie imigrantów – zwłaszcza w świecie, w którym jednocześnie trwa dynamiczny rozwój technologii cyfrowych oraz eksplozja kompetencji generatywnej AI. Od technologii cyfrowych, w tym AI, nie ma ucieczki.

Za to ucieczka w AI w obawie przed stagnacją napędza globalny wyścig technologiczny. Czy to konkurencja dla nas?

W kontekście AI już teraz zaledwie kilka krajów kontroluje globalne innowacje. Bezspornie dominują tu USA i Chiny; udział nielicznych innych krajów, jak np. krajów UE, a także Wielkiej Brytanii, Japonii czy Korei Południowej, jest minimalny, a pozostałych krajów świata – zerowy. To chyba jednak nieuniknione, gdyż wynika wprost z charakteru technologii cyfrowych. Komputerowy software jest czystą informacją, a przez to może być szybko i bezkosztowo przenoszony i kopiowany, bez problemu skaluje się do dostępnego hardware'u. W dobie Internetu jest też jeden globalny rynek na dobra cyfrowe. Wszystko to oznacza silnie rosnące przychody względem skali, skutkujące monopolizacją rynków wskutek dynamiki „zwycięzca bierze wszystko”. W takim krajobrazie kraje konwergujące, niebędące liderami innowacji,

Polska dysponuje wystarczającą infrastrukturą i kapitałem ludzkim, by aktywnie budować ekosystem technologii aplikujących rozwiązania AI

mogą co najwyżej tworzyć swoje specyficzne nisze rynkowe i tam szukać lokalnych przewag.

Czy właśnie taką drogą powinna podążać Polska?

Sądzę, że powinniśmy w kontekście AI rozróżnić dwie kwestie: rozwój najbardziej zaawansowanych modeli (tzw. frontier research) oraz rozwój szerokiego ekosystemu narzędzi wykorzystujących AI. Jeśli chodzi o frontier research, to tam aktywne są dziś wyłącznie ośrodki amerykańskie i chińskie, a bariery wejścia związane z dostępem do gigantycznych mocy obliczeniowych uniemożliwiają innym włączyć się do tego wyścigu.

Tym bardziej, jeśli – jak pan sugeruje – ten wyścig może skończyć się katastrofą.

Tak, moim zdaniem wyścig ten prowadzi z dużym prawdopodobieństwem do katastrofy, w związku z czym najpilniejszym zadaniem dziś powinno być jego zahamowanie, tak byśmy zdążyli na czas rozwiązać problem AI alignment.

A ta druga kwestia?

Jeśli chodzi o szeroko pojęty ekosystem technologii aplikujących rozwiązania AI, to jak najbardziej warto go rozbudowywać. Sądzę, że Polska dysponuje zarówno wystarczającą infrastrukturą, jak i kapitałem ludzkim, by być w tej branży aktywna.

Przed wszystkim jednak chciałbym podkreślić, że wyścig w kierunku nadludzkiej AGI bez uprzedniego rozwiązania problemu zgodności celów AGI z długofalowym dobrostanem ludzkości jest wyścigiem samobójczym. Dlatego uważam, że zamiast próbować się do niego na siłę włączyć, powinniśmy lobbować, by wyścig ten został czym prędzej przerwany. Służą temu działania np. takich środowisk, jak Pause AI. Obecna dynamika rozwoju w branży najbardziej zaawansowanych modeli AI to czyste szaleństwo, napędzane krótkowzrocznością, presją konkurencyjną i wizjami utopii. Jak powiedział Stuart Russell, jeden z najbardziej utytułowanych badaczy AI na świecie: „Liderzy branży AI grają dziś w rosyjską ruletkę z całą ludzkością. Nie zamierzam się na to milcząco godzić”.

Rozmówca wyraża własne opinie, a nie oficjalne stanowisko NBP.

Rozmawiała: Anna Słabosz-Pawłowska

Dziś najpilniejsze jest zahamowanie wyścigu technologicznego, abyśmy zdążyli rozwiązać problem AI alignment. Jeśli długofalowy dobrostan ludzkości nie będzie wzięty pod uwagę, będzie to wyścig samobójczy