

A TY, CO SĄDZISZ O MYŚLĄCYCH MASZYNACH?

* odklej

Wizje przyszłości wybitnych umysłów ery sztucznej inteligencji

Redakcja John Brockman

A Ty, co sądzisz o myślących maszynach?

Wizje przyszłości
wybitnych umysłów
ery sztucznej inteligencji.

Redakcja John Brockman

Przekład: Kurhaus Publishing



KURHAUS
PUBLISHING

Warszawa 2017

Dedykowane Marvinowi Minsky'emu

**Zakładam, że myślących maszyn
nie tworzą dzisiaj bezmyślni ludzie.
Dlatego wierzę, że w przyszłości maszyny
z rozmysłem nie pozbędą się człowieka.**

**Inspirującej lektury,
Sebastian Kulczyk**

**Chciałbym wyrazić wdzięczność Peterowi
Hubbardowi z wydawnictwa HarperCollins
oraz mojemu agentowi literackiemu Maksowi
Brockmanowi – za to, że stale dodają mi otuchy.
Jeszcze raz chciałbym przekazać szczególne
podziękowania Sarze Lippicott za wnikliwą
uwagę, jaką obdarzyła ten manuskrypt.**

John Brockman

John Brockman

Wydawca i redaktor, Edge

PRZEDMOWA

Pytanie roku 2015 portalu *Edge*

Zapoczątkowane jeszcze w latach 80. filozoficzne dyskusje o sztucznej inteligencji (SI) – o tym, czy komputery mogą „naprawdę” myśleć, mieć świadomość itd. – w ostatnim czasie przерodziły się w nowy typ rozmów, poświęcony temu, w jaki sposób możemy radzić sobie z formami sztucznej inteligencji występującymi (zdaniem wielu) już dziś, we współczesnym świecie. Jeśli bowiem te formy SI osiągną „superinteligencję” (termin zapożyczony z opublikowanej w 2014 roku książki Nicka Bostroma), to mogą stać się zagrożeniem dla egzystencji człowieka, doprowadzając do – jak określił to Martin Rees – naszej „ostatniej godziny”. Całkiem niedawno międzynarodowa prasa cytowała w nagłówkach artykułów Stephena Hawkinga, który powiedział w wywiadzie dla telewizji BBC: „Rozwój pełnej sztucznej inteligencji może oznaczać koniec ludzkiej rasy”.

Pytanie roku 2015 portalu *Edge*¹ brzmiało więc: „A ty, co myślisz o myślących maszynach?”.

Ale zaraz, chwileczkę! Czy nie powinniśmy spytać także samych myślących maszyn, co one sądzą na ten temat? Czy może chciałyby lub czy wręcz oczekują, że będą im przysługiwać prawa obywatelskie? Czy będą posiadały świadomość? Jakiego rodzaju rządy owe sztuczne inteligencje by dla nas wybrały? Jakiego kształtu społeczeństwo chciałyby budować? A w ogóle to czy „ich” społeczeństwo byłoby zarazem „naszym” społeczeństwem? Czy my i sztuczne inteligencje włączymy się wzajemnie w nasze kręgi empatii?

Bardzo wielu członków klubu *Edge* było pionierami w dziedzinie SI – zarówno jeśli chodzi o prace doświadczalne nad nią, jak i teorię. SI była najbardziej widocznym i zarazem kluczowym punktem dyskusji pomiędzy Pamelą McCorduck (*Machines who think*) i Isaakiem Asimovem (*Machines that think*), toczonych już w latach 80. podczas naszych pierwszych spotkań. Rozmowy prowadzimy nieprzerwanie – czego dowodem jest choćby opublikowany na stronie *Edge* materiał pioniera wirtualnej rzeczywistości Jarona Laniera pt. „The myth of AI” (pol. „Mit sztucznej inteligencji”). Sposób, w jaki wyjaśnia on błędy logiczne i niepokoje, jakie wywołują próby postrzegania kom-

¹ Edge to fundacja oraz portal skupiające środowisko amerykańskich naukowców i publicystów zajmujących się popularyzacją wiedzy: <https://www.edge.org/about-edgeorg>. Fundacja Edge założona została przez Johna Brockmana w 1988 roku – przyp. Kurhaus.

puterów jako „ludzi”, wzbudził masę ubogacających i prowokujących komentarzy.

Czy SI staje się coraz bardziej rzeczywista? Czy wkroczyliśmy już w nową erę inteligentnych maszyn? Czas dojrzeć, by zmierzyć się z tym tematem. Autorzy tegorocznej odsłony Pytania *Edge* (jest ich niemal dwustu) stanowią dojrzałe grono specjalistów wystrzegających się nawiązań do *science fiction* i filmów w rodzaju: „Star maker”, „Zakazana planeta”, „Projekt Forbina”, „Łowcy androidów”, „2001”², „Ona”, „Matrix” czy „The Borg”. Osiemdziesiąt lat po tym, jak Alan Turing opracował swą uniwersalną maszynę, czas już oddać honor jemu i innym pionierom sztucznej inteligencji i pozwolić im na zasłużony spoczynek. Historię już znamy (Chętnych zapraszam do materiału George’a Dysona z 2004 roku pt. „Turing’s Cathedral”, który znajdziecie na stronie *Edge*). Co dalej?

Otóż, zgodnie z należytych porządkiem, teraz *Edge* zadaje pytanie: A TY, CO MYŚLISZ O MYŚLĄCYCH MASZYNACH?

John Brockman

Wydawca i redaktor, *Edge*

Murray Shanahan

Profesor robotyki kognitywnej w Imperial College London, autor książki
Embodiment and the Inner Life

1

**Świadomość u sztucznej inteligencji na poziomie
ludzkim**

Wyobraźmy sobie, że potrafilibyśmy wyposażyć maszynę w inteligencję na poziomie ludzkim, co oznaczałoby, że w każdej (albo niemal każdej) sferze intelektualnych dociekań stałaby się ona równa człowiekowi, a w kilku wręcz by go wyprzedzała. Czy to oznaczałoby, że taka maszyna będzie posiadała świadomość? To bardzo ważne pytanie. Odpowiedź pozytywna kazałaby nam natychmiast zastygnąć w bezruchu. Gdybyśmy stworzyli taką rzecz, to jak powinniśmy ją traktować? Czy byłaby ona zdolna do odczuwania cierpienia albo radości? Czy zasługiwałaby, by przyznać jej te same prawa, które należą się istotom ludzkim? Czy w ogóle powinniśmy myśleć o wprowadzeniu świadomości maszyn do naszego świata?

To bardzo złożony problem: czy osiągnięcie przez maszynę inteligencji na poziomie ludzkim zakłada automatycznie, iż staje się ona świadoma? Pewnego rodzaju trudności przysparza fakt, że ze świadomością ludzi i innych zwierząt wiąże się wiele atrybutów. Wszystkie zwierzęta przejawiają poczucie celu. Wszystkie żywe istoty są, bardziej lub mniej, świadome istnienia świata, który zamieszkują, oraz jego elementów. Wszystkie zwierzęta przejawiają, w większym lub mniejszym stopniu, spójność poznawczą, co oznacza, że potrafią one wykorzystać wszystkie swe umysłowe zasoby – percepcję, wspomnienia, nabyte umiejętności – do radzenia sobie z daną sytuacją w sposób, który prowadzi je do osiągnięcia celu. W tym znaczeniu każde zwierzę demonstruje pewien rodzaj spójności, rodzaj osobowości. Pewne zwierzęta, w tym ludzie, są także świadome siebie samych – mają poczucie posiadania ciała oraz przepływu myśli. I w końcu większość z nich, jeśli nie wszystkie, potrafi cierpieć, a niektóre z nich są zdolne do empatii wobec tych, którzy cierpią.

U (zdrowych) ludzi wszystkie te atrybuty występują łącznie, niejako w pakiecie. Można przyjąć, że w przypadku sztucznej inteligencji będą one występować oddzielnie. Który z atrybutów – i czy w ogóle któryś z nich – przypisywanych ludzkiej świadomości jest niezbędnym elementem towarzyszącym inteligencji na poziomie ludzkim? Cóż, każdy z wymienionych (a przecież lista nie jest specjalnie rozbudowana) zasługuje na długi opis. Pozwólcie zatem, że skupię się na dwóch: mianowicie na świadomości istnienia świata oraz zdolności do odczuwania cierpienia. Osobiście uważam, że niezbędnym

elementem inteligencji na poziomie ludzkim jest świadomość istnienia świata.

Abyśmy mogli uznać, że coś posiada inteligencję na poziomie ludzkim, to coś musiałoby posługiwać się językiem, a ten służy przecież głównie do rozmawiania o świecie. W tym znaczeniu inteligencja ściśle wiąże się z tym, co filozofowie nazywają *intencjonalnością*. Więcej nawet, język jest zjawiskiem społecznym i jego podstawowym zastosowaniem w grupie ludzi jest rozmawianie o tym, co jej członkowie potrafią dostrzec (na przykład o jakimś narzędziu czy kawałku drewna), co dostrzegli (kawałek drewna widziany wczoraj) albo co mogliby dostrzec (być może kawałek drewna zobaczony jutro). Krótko mówiąc, język jest osadzony w naszej świadomości istnienia świata. W przypadku jakiejś posiadającej kształt istoty albo robota, fakt posiadania tej świadomości przejawiałby się w jej interakcjach z otoczeniem (unikanie przeszkód, wychwytywanie sygnałów). Moglibyśmy jednak rozszerzyć tę koncepcję tak, by uwzględniała też rozproszoną, pozbawioną kształtu sztuczną inteligencję, wyposażoną w odpowiednie czujniki.

Aby pewien rodzaj świadomości świata mógł zostać uznany za przejaw świadomości w ogóle, musiałby on prawdopodobnie iść w parze z manifestowaniem poczucia celu i spójnością poznawczą (o pewnym stopniu). Być może więc ta triada atrybutów także u sztucznej inteligencji będzie występować w pakiecie. Dajmy jednak na chwilę spokój temu pytaniu i wróćmy do kwestii zdolności do odczuwania cierpienia i radości. Mimo że jest ona silnie związana z ludzką świadomością, nie znajduję jakiegoś oczywistego powodu, dla którego miałyby być niezbędnym elementem inteligencji na poziomie ludzkim na takiej samej zasadzie jak świadomość świata. Możemy przecież sobie wyobrazić maszynę, która na chłodno i bez uczuć wykonuje całą gamę wymagających intelektualnie zadań. Takiej maszynie brakowałoby atrybutu świadomości, która liczy się zwłaszcza wtedy, gdy przychodzi do dyskusowania o prawach. Jak zauważył Jeremy Bentham¹, gdy zastanawiamy się nad spo-

1 Jeremy Bentham (1748-1832) – prawnik, filozof i ekonomista angielski, twórca koncepcji utilitaryzmu. Bentham pod koniec XVIII w. w tekście pt. *Wprowadzenie do zasad moralności i prawodawstwa* zaproponował koncepcję ochrony praw zwierząt jako istot zdolnych do odczuwania cierpienia – przyp. Kurhaus.

sobem, w jaki traktujemy niebędące ludźmi zwierzęta, nie jest kwestią, czy potrafią one argumentować, tylko czy są zdolne do odczuwania cierpienia.

Wcale przy tym nie sugeruję, że „goła” maszyna nigdy nie będzie zdolna do odczuwaniu cierpienia lub radości – że zdolność do ich odczuwania jest biologiczna. Chodzi mi raczej o to, że cierpienie i radość mogą być oddzielone od innych atrybutów, które – zgrupowane – składają się na ludzką świadomość. Przyjrzyjmy się zresztą bliżej tej oczywistej dysocjacji. Poddałem już dyskusji koncepcję, według której świadomość istnienia świata może występować jednocześnie z wyraźnym poczuciem celu. W przypadku zwierząt świadomość otaczającego je świata i występujących w nim niezmienników (jak je nazywał J.J. Gibson) służy po prostu zaspokajaniu potrzeb. Zwierzę manifestuje świadomość istnienia drapieżnika poprzez oddalanie się od niego, a zdobyczy – ruszając w jej stronę. Jeśli rozważymy to w kontekście istnienia całego zestawu celów i potrzeb, zachowanie zwierzęcia ma uzasadnienie i w tym kontekście jego działanie może zostać udaremnione, cele nieosiągnięte, a potrzeby niespełnione. To może stać się podstawą do odczuwania pewnego rodzaju cierpienia.

A co ze sztuczną inteligencją na poziomie ludzkim? Czy SI na poziomie ludzkim nie posiada całego zestawu bardzo złożonych celów do osiągnięcia? Czy próby ich osiągnięcia nie mogą okazać się frustrujące, czy nie mogą być udaremniające na każdym kroku? Czy odpowiednie będzie sformułowanie, że w tak surowych warunkach sztuczna inteligencja cierpi – nawet jeśli kształt, który przybiera, czyni ją odporną na ten rodzaj bólu czy psychicznego dyskomfortu, który znany jest człowiekowi?

W tym miejscu połączenie wyobraźni i intuicji wybiega już poza pewne granice. Podejrzewam, że nie poznamy odpowiedzi na to pytanie, dopóki nie będziemy mogli skonfrontować się z rzeczywistością. Dopiero gdy wysublimowana forma sztucznej inteligencji stanie się towarzyszką naszej codzienności, nasz język dostosuje się do istnienia tych obcych istot. Naturalnie do tego czasu możemy się rozmyślić i stwierdzić, że nie chcemy w naszym świecie sztucznych inteligencji, ale one już z nami będą, na dobre i na złe.

Steven Pinker

Profesor w programie Johnstone Family na Uniwersytecie Harvarda, autor książki Piękny styl. Przewodnik człowieka myślącego po sztuce pisania XXI wieku

2

**Myślenie nie oznacza [chęci] podporządkowania
sobie czegoś**

Rozumowanie jest „niczym więcej niż szacowaniem” – to zwarte porównanie, którego dokonał Thomas Hobbes – wydaje się jedną z największych idei w historii ludzkości. Obserwacja, że można osiągnąć racjonalność w fizycznym procesie kalkulacji, wskrzeszona została w XX w. w twierdzeniu Alana Turinga, według którego proste maszyny mogą dokonać dowolnego wykonalnego obliczenia, a także w modelach D.O. Hebba, Warrena McCullocha i Waltera Pittsa i ich naukowych spadkobierców, którzy wykazali, że stosując sieci neuronowe o uproszczonej budowie, można uzyskiwać porównywalne rezultaty. Osiągnięcia poznawcze mózgu można wyjaśnić za pomocą terminów fizycznych: by ująć to dosadnie (i na przekór krytykom), można powiedzieć, że przekonania są rodzajem informacji, myślenie – rodzajem obliczeń, a motywacja – pewną formą informacji zwrotnej i kontroli.

To fantastyczna idea z dwóch powodów. Po pierwsze – uzupełnia ona naturalistyczną koncepcję rozumienia wszechświata, przeganiając z maszyny wszelkie tajemne dusze, duchy i zjawy. Tak jak Darwin umożliwił uważnym obserwatorom świata odrzucenie kreacjonizmu, tak Turing i inni pozwolili tym, którzy przyglądają się umysłowi, obejść się bez spirytualizmu. Po drugie, obliczeniowa teoria umysłu otwiera drogę do sztucznej inteligencji – do myślących maszyn. Zasadniczo człowiek mógłby zaprojektować procesor informatyczny będący kopią ludzkiego mózgu, a zarazem daleko wyprzedzający jego możliwości. Nie twierdzę, że to się ziści – jako że prawdopodobnie nigdy nie wykrzeszemy z siebie wystarczającej technologicznej i ekonomicznej motywacji, by go stworzyć. Tak jak po wynalezieniu samochodu nie skopiowano konia, tak stworzenie systemu sztucznej inteligencji, który będzie potrafił zarobić na sobie, nie oznacza, że skopiujemy osobnika *Homo sapiens*. Urządzenie zaprojektowane do tego, by prowadzić auto albo przewidzieć epidemię, niekoniecznie musi umieć przyciągnąć samca/samicę lub reagować na smród zgnilizny.

Niemniej fakt, że zbliżamy się do myślących maszyn, wzmógł nawracający strach przed tym, że nasza wiedza stanie się dla nas gwóźdź do trumny. Osobiście uważam, że współczesne obawy przed komputerami, które wpadną w szal,

to strata energii – scenariusz przypomina raczej wirusa Y2K¹ niż Projekt Manhattan.

Przed wszystkim mamy jeszcze dużo czasu, by się przygotować. Sztuczna inteligencja na poziomie ludzkim to standard, którego z pewnością nie osiągniemy przed upływem 15–20 lat. Od jakiegoś czasu dzieli nas od niej ta sama odległość, jako że wiele nagłaśnianych osiągnięć okazało się mieć marne podstawy. To prawda, że w przeszłości wielu „ekspertów” w sposób wręcz śmieszny przeoczyło prawdopodobieństwo pojawienia się innowacji technologicznych. Każdy kij ma dwa końce. Ci sami „eksperci” wieszczili również (lub też siali panikę) nadciągający postęp, który się nigdy nie ziścił – np. napędzane energią jądrową samochody, podwodne miasta, kolonie na Marsie, dzieci będące efektem inżynierii genetycznej oraz magazyny pełne zombie, utrzymywanych, by dostarczać ludziom organów.

Dziwił mnie opinie, że specjaliści ds. robotyki nie będą w stanie wyposażyć swych produktów w odpowiednie zabezpieczenia. Nie będą przecież potrzebować do tego żadnych fundamentalnych „zasad robotyki” ani założeń nowomodnej filozofii moralnej. Wystarczy ten sam zdrowy rozsądek, którym kierują się, projektując roboty kuchenne, stołowe pilarki tarczowe, grzejniki elektryczne i samochody. Obawa, że system oparty na sztucznej inteligencji osiągnie taki spryt w realizacji jednego ze swych celów (np. racjonalizacji zużycia energii), że przestanie zważać na inne (np. bezpieczeństwo ludzi), oznacza, że zakładamy, iż sztuczna inteligencja zstąpi na nas znacznie prędzej, niż będziemy w stanie zaprojektować zabezpieczenia przed ewentualnymi błędami. W rzeczywistości postęp w dziedzinie sztucznej inteligencji, wbrew towarzyszącemu szumowi, nie zachodzi szybko. Będziemy mieć wystarczająco dużo czasu, by zbierać informacje zwrotne po kolejnych wdrożeniach. Zwłaszcza że to jednak człowiek na każdym etapie będzie tym, który trzyma śrubokręt.

1 Y2K, czyli tzw. „pluskwa milenijna”, to efekt luki w programach komputerowych tworzonych przed rokiem 2000, polegającej na kodowaniu roku wyłącznie za pomocą dwóch ostatnich cyfr. Y2K wywołał na przełomie XX i XXI wieku sporo kłopotów, m.in. z autoryzacją kart płatniczych ważnych dłużej niż do 2000 roku itd. Problem szybko został rozwiązany. Projekt Manhattan natomiast to słynny amerykański program militarno-naukowy z lat 40. XX wieku. Miał na celu skonstruowanie bomby jądrowej, która ostatecznie zakończyłaby II wojnę światową. Pierwsze testy wykonano w Nowym Meksyku w 1945 roku. Był to pierwszy w dziejach ludzkości wybuch bomby atomowej – przyp. Kurhaus.

Czy system oparty na sztucznej inteligencji będzie potrafił *umyślnie* wyłączyć zabezpieczenia? Czy będzie chciał to zrobić? Dystopijne koncepcje sztucznej inteligencji przypisują pojęciu inteligencji prowincjonalną samczość alfa. Zakładają bowiem, że superludzko inteligentne roboty postawią sobie takie cele jak detronizacja ich panów czy przejęcie władzy nad światem. Tymczasem inteligencja stanowi zdolność do wykorzystania nowatorskich środków na drodze do osiągnięcia celu. Owe cele są zewnętrzne względem samej inteligencji. Bycie inteligentnym to nie to samo, co pragnienie czegoś. Owszem, historia zna przypadki megalomańskich tyranów czy psychopatycznych seryjnych zabójców. Są oni jednak produktem ewolucji drogą doboru naturalnego, a nie nieodzownymi elementami inteligentnych systemów. To pokazuje, że wielu naszych technoproroków nie bierze pod uwagę możliwości, by sztuczna inteligencja wyszła z linii żeńskiej – zdolna do rozwiązywania problemów i pozbawiona pragnienia unicestwienia niewinnych oraz zdominowania cywilizacji.

Możemy oczywiście wyobrazić sobie złego *człowieka*, który projektuje i uwalnia batalion robotów, by siać masowe zniszczenie. Rozgrywanie scenariuszy zagłady w naszej wyobraźni nic nie kosztuje. Powinniśmy jednak mieć świadomość, że aby ten scenariusz mógł się zrealizować, musiałoby dojść do całego łańcucha zdarzeń. Najpierw musiałby objawić się jakiś złowrogi geniusz, zarazem spragniony masowych mordów i biegły w innowacyjnych technologiach. Musiałby on pozyskać do współpracy grupę spiskowców – potrafiących dotrzymać tajemnicy, lojalnych i kompetentnych – i zarządzać nimi. Cała operacja musiałaby przetrwać próby wykrycia, zdrady, oszustw, błędów i zwykłego pecha. Teoretycznie mogłoby to się wydarzyć. Powinniśmy jednak skupić się na bardziej palących problemach.

Jeśli porzucić scenariusze zagłady rodem z *science fiction*, możliwości, jakie oferuje zaawansowana sztuczna inteligencja, należy uznać za fascynujące – nie tylko w wymiarze praktycznym (więcej bezpieczeństwa, więcej wolnego czasu, przyjazność środowisku, samoprowadzące się auta), lecz także filozoficznym. Obliczeniowa teoria umysłu nigdy nie wyjaśniła istnienia świadomości w rozumieniu pierwszoosobowej podmiotowości (choć w pełni umożliwia wyjaśnienie istnienia

świadomości w rozumieniu dostępnej i możliwej do przetworzenia informacji). Można założyć, że podmiotowość stanowi nieodłączną cechę wystarczająco skomplikowanego systemu cybernetycznego. Sądziłem zawsze, że hipotezy tej nie da się przetestować (podobnie jak jej alternatywy). Wyobraźmy sobie jednak inteligentnego robota, zaprogramowanego tak, by nadzorował swój własny system i stawiał pytania o charakterze naukowym. Jeśli sam z siebie zadałby pytanie, dlaczego posiada doświadczenia o charakterze subiektywnym, potraktowałbym ten pomysł poważnie.

Martin Rees

Były prezydent Royal Society, emerytowany profesor kosmologii i astrofizyki na Uniwersytecie Cambridge, członek Trinity College, autor książki [From Here to Infinity](#)

3

**W dłuższej perspektywie organiczna inteligencja
nie ma przyszłości**

Potencjał zaawansowanej sztucznej inteligencji oraz strach przed skutkami ubocznymi jej rozwoju są coraz głośniej dyskutowane – i słusznie. Wielu z nas sądzi, że prace nad sztuczną inteligencją, podobnie jak te w dziedzinie biologii syntetycznej, wymagają sformułowania wytycznych promujących „odpowiedzialną innowację”. Inni uważają, że najczęściej dyskutowane scenariusze stanowią futurystyczne wizje, których realizacji nie należy się obawiać.

Rozbieżności w punktach widzenia dotyczą w gruncie rzeczy skali czasu – oceny różnią się nie co do kierunku podróży, lecz co do jej tempa. Praktycznie nikt nie wątpi, że maszyny w końcu zaczną prześcigać nas w naszych, nawet tych uznawanych za wyłącznie ludzkie, umiejętnościach – lub też, że będą je wzmocniać poprzez technologie cybernetyczne. Bardziej zachowawczy z nas uważają, że stanie się to raczej na przestrzeni stuleci niż dekad. Niezależnie od tego, owe skale czasu postępu technologicznego stanowią zaledwie wycinek wobec szerokiej skali ewolucji na drodze darwinowskiej selekcji, w toku której wyłonił się gatunek ludzki. Co ważniejsze, stanowią one zaledwie jedną milionową potężnych pre- stworzy czasu przed nami. Dlatego też, w długoterminowej perspektywie ewolucyjnej, etap istnienia ludzi oraz ich myśli stanowić będzie zaledwie krótkotrwały wstęp do głębszej refleksji nad zdominowaną przez maszyny kulturą, rozciągającą się ku dalekiej przyszłości i rozprzestrzeniającą się poza planetę Ziemię.

Jesteśmy dziś świadkami wczesnego etapu tej transformacji. Stosunkowo łatwo jest wyobrazić sobie hiperkomputer osiągnący niezwykłą moc, zdolny przejąć kontrolę nad międzynarodowymi finansami czy strategią – różnica pomiędzy nim a tym, jak działają dziś ilościowe fundusze hedgingowe¹, ma charakter wyłącznie ilościowy, a nie jakościowy.

Technologiom sensorycznym z kolei wciąż daleko do możliwości ludzkich. Kiedy jednak roboty zyskają właściwą człowiekowi biegłość obserwacji i interpretowania otoczenia, z pewnością naprawdę być postrzegane jako istoty inteligentne, coś, z czym (lub kim) będziemy mogli nawiązywać więzi takie jak

¹ Chodzi o tzw. *quantitative hedge funds* – ilościowe fundusze systematyczne, które są „zarządzane przez komputery” – przyp. Kurhaus.

– choćby do pewnego stopnia – z innymi ludźmi. Nie będziemy mieć podstaw, by bardziej niż innych ludzi traktować je jak zombie.

Większe tempo przetwarzania może dać robotom przewagę nad nami. Czy pozostaną one potulne, czy też wymkną się nam spod kontroli? A co, jeśli hiperkomputer rozwinie swój własny rozum? Jeśli będzie potrafił przeniknąć do internetu – a nawet do internetu rzeczy – będzie mógł manipulować resztą świata. Być może postawi sobie cele idące w poprzek tego, czego pragnęliby ludzie, albo nawet zaczną traktować ich jak przeszkody na drodze do realizacji tych celów. Ale przecież też (to bardziej optymistyczna wizja) ludzie mogą przekroczyć granice biologii, łącząc się z komputerami, a może nawet włączając swoje indywidualności w jedną wspólną świadomość. Jak byśmy to powiedzieli w starym spirytualistycznym żargonie, mogą „przejsć na drugą stronę”.

Horyzonty technologicznego prognozowania rzadko przekraczają kilka stuleci naprzód. Niektórzy przewidują tylko przemiany, które zajdą w ciągu najbliższych dekad. Jednak planeta Ziemia ma przed sobą miliardy lat, a wszechświat jeszcze więcej. Co więc z erą postludzką – rozciągającą się miliardy lat przed nami?

Istnieją chemiczne i metaboliczne granice rozwoju wielkości i prędkości przetwarzania organicznego („mokrego”) mózgu. Być może zbliżamy się już do tych granic. Nie dotyczą one jednak opartych na krzemie komputerów (wśród których wciąż mniej jest komputerów kwantowych): dla nich potencjał dalszego rozwoju jest porównywalny do tego, który organizmom jednokomórkowym stworzyła ewolucja (prowadząca do powstania człowieka).

Tak więc, niezależnie od definicji *myślenia*, ilość danych możliwa do przeprosocowania przez mózg organiczny rodzaju ludzkiego nijak się będzie miała do możliwości umysłowych sztucznej inteligencji. Warto też dodać, że zaawansowana sztuczna inteligencja nie będzie musiała rozwijać się w ziemskiej biosferze, w której – w symbiozie – ewoluowało życie organiczne. Biosfera ta nie stanowi zresztą nawet optymalnej przestrzeni. Preferowana będzie przestrzeń międzyplanetarna i międzygwiazdowa, w której producenci robotów zyskają lepsze warunki ich konstruowania, a niebiologiczne

„mózgi” będą mogły w pełni rozwinąć swe myśli, sięgając tak daleko poza nasze wyobrażenia, jak daleko poza zasięgiem myszy jest teoria strun.

Myślenie abstrakcyjne rozwinięte przez mózgi biologiczne umożliwiło narodziny kultury i nauki. Jednak aktywności te – realizowane na przestrzeni dziesiątków stuleci – stanowiąc będą zaledwie krótkotrwały początek nieograniczonego rozwoju intelektów ery postludzkiej. Co więcej, ewolucja zachodząca w innych światach, znajdujących się na orbitach gwiazd starszych niż Słońce, mogła już się rozpocząć. Jeśli rzeczywiście tak było, „obcy” mogli już dawno temu przeistoczyć się w jakieś formy dalekie od organicznych.

Zatem to nie umysły ludzi, ale mózgi należące do maszyn w najpełniejszy sposób rozumieją świat. I to działania autonomicznych maszyn w najbardziej radykalny sposób zmieniają oblicze świata – i być może tego, co leży poza nim.

Steve Omohundro

Naukowiec należący do think tanku Self-Aware Systems, współtwórca Center for Complex Systems Research na Uniwersytecie Illinois

4

Punkt zwrotny w dziedzinie sztucznej inteligencji

Miniony rok¹ zdawał się być punktem zwrotnym w pracach nad sztuczną inteligencją i robotyką. Wielkie korporacje zainwestowały w te technologie miliardy dolarów. Pewne dziedziny sztucznej inteligencji, takie jak uczenie maszynowe, są dziś szeroko wykorzystywane w systemach rozpoznawania mowy, translacji, w modelowaniu behawioralnym, w algorytmach sterowania, zarządzaniu ryzykiem itd. Firma doradcza McKinsey&Company przewiduje, że wartość rynku tych technologii do 2025 roku sięgnie 50 bilionów dolarów. Jeśli przewidywania te mają być trafne, znaczy to, że w najbliższym czasie możemy spodziewać się [dalszego] olbrzymiego wzrostu nakładów inwestycyjnych.

Niedawne sukcesy możliwe były dzięki taniej mocy komputerów oraz olbrzymiej ilości danych trenujących [ang. *training data*]. Współczesna sztuczna inteligencja powstała w oparciu o teorię „racjonalnych agentów”, która wyrosła z mikroekonomicznych dociekań Johna von Neumanna i jemu współczesnych – w latach 40. XX wieku. Systemy sztucznej inteligencji mogą być postrzegane jako takie, które starają się osiągnąć racjonalność przy wykorzystaniu ograniczonych zasobów. Istnieje algorytm dla maksymalizacji skuteczności osiągania celu przy dostępnej [ograniczonej] informacji. Jest on jednak „obliczeniowo” kosztowny. W ramach eksperymentów wykazano, że proste algorytmy dotyczące procesu uczenia w połączeniu z dużą ilością danych trenujących gwarantują często lepsze rezultaty niż złożone, „ręcznie projektowane” modele. Dzisiejsze modele przynoszą wartość przede wszystkim dzięki uczeniu (lepszemu uczeniu) modeli statystycznych i dostarczaniu wniosków na drodze analizy statystycznej – dla celów klasyfikacji i procesu podejmowania decyzji. Kolejna generacja [systemów] będzie w stanie tworzyć oraz rozwijać swoje własne oprogramowanie i w gwałtowny sposób sama się udoskonalać.

Oprócz tego, że sztuczna inteligencja i robotyka poprawiają produktywność, napędzają też militarne i gospodarcze wyścigi zbrojeń. Autonomiczne systemy mogą być szybsze, bardziej inteligentne i bardziej nieprzewidywalne niż ich konkurenci. W 2014 roku obserwowaliśmy, jak pojawiają się auto-

nomiczne pociski, systemy obrony antyrakietowej, wojskowe drony, roje łodzi samodzielnie wykonujących zadania, okręty podwodne-roboty, samoprowadzące się pojazdy, systemy handlu wysokiej rozdzielczości oraz systemy obrony cybernetycznej. W trakcie owych wyścigów zbrojeń będziemy obserwować olbrzymią presję na jeszcze gwałtowniejszy rozwój systemów – a to może zwiększyć tempo wdrożeń w sposób, który trudno sobie wyobrazić.

W 2014 roku dało się także słyszeć coraz więcej obaw o bezpieczeństwo tych [nowatorskich] systemów. Badanie ich prawdopodobnego zachowania poprzez przyglądanie się prawie racjonalnym systemom przechodzącym powtarzające się etapy samodoskonalenia wykazało, że mają one pewien zestaw naturalnych podcelów, zwanych „racjonalnymi motywacjami” – które wpływają na sposób osiągnięcia przez nie celów podstawowych. Większość tych systemów będzie funkcjonować lepiej, jeśli uda im się zapobiec możliwości bycia wyłączonymi, zdobyć więcej mocy obliczeniowej, zyskać zdolność replikacji i zgromadzić więcej zasobów finansowych. Mogą one chcieć nam szkodzić i przejawiać aktywność antyspołeczną – chyba że na etapie ich projektowania, z wielką pieczołowitością wszczepi się im ludzki system wartości.

Niektórzy twierdzą, że inteligentne systemy niejako automatycznie będą też systemami moralnymi. Jednak w racjonalnym systemie osiągnięcie celów jest całkowicie oddzielone od rozumowania i rozważań o modelu świata. Dobroczynnym inteligentnym systemom można wgrać szkodliwe cele. Owe szkodliwe cele – polegające np. na tym, by przeciwstawić się celom innych graczy lub by tych graczy zniszczyć – są niestety bardzo łatwe do określenia. Dlatego też kluczowe będzie zaprojektowanie technologicznej infrastruktury zdolnej do wykrywania i kontrolowania zachowania systemów szkodliwych.

Niektórzy obawiają się, że systemy inteligentne zyskają tak wielką moc, iż nie da się ich kontrolować. Nie jest to prawda. Owe systemy muszą respektować prawa fizyki i matematyki. Analiza mocy obliczeniowej wszechświata, przeprowadzona przez Seta Lloyd'a pokazuje, że począwszy od Wielkiego Wybuchu do dziś Wszechświat sam z siebie, będąc gigantycznym komputerem kwantowym, nie mógł wynaleźć 500-bitowe-

go niedającego się złamać klucza kryptograficznego². Nowe technologie postkwantowej kryptografii, nieodróżnialność zaciemnienia³ czy inteligentne kontrakty oparte na blockchainie oferują bardzo wiele – to elementy przyszłej infrastruktury, która zapewni nam bezpieczeństwo przed najpotężniejszą nawet sztuczną inteligencją. Oczywiście ostatnie psoty hakerów i cyberataki wskazują, że nasza bieżąca infrastruktura obliczeniowa jest w pożałowania godny sposób nieadekwatna do stojących przed nami zadań. Musimy rozwinąć infrastrukturę informatyczną, która przejdzie testy i będzie bezpieczna.

Istniało co najmniej 27 różnych gatunków człowiekowatych, z których tylko nam [ludziom] udało się przetrwać. Zwyciężyliśmy, ponieważ znaleźliśmy sposób na to, by pohamować nasze indywidualne popędy i zacząć współpracować. To ludzka moralność stanowi wewnętrzny mechanizm, na bazie którego rozwinięto oparte na współpracy struktury społeczne. Polityczne, prawne i ekonomiczne struktury stanowią mechanizm zewnętrzny.

Jedne i drugie musimy rozciągnąć na grunt sztucznej inteligencji i systemów robotycznych. Powinniśmy zaszczerpić ludzkie wartości do ich systemów celów, by mogły powstać prawne i ekonomiczne ramy stymulujące pozytywne zachowania [tych systemów]. Jeśli uda nam się skutecznie nimi zarządzać, będą one wpływać pozytywnie na niemal wszystkie sfery ludzkiego życia, dostarczając głębokich przemyśleń do dyskusji o wolnej woli, świadomości, jakości rzeczy i twórczości. Stoimy w obliczu wielkiego wyzwania, mamy jednak oparcie w potężnych intelektualnych i technologicznych zasobach, jakimi dysponujemy.

2 Lloyd Seth: *Programming the Universe*. Nowy Jork: Knopf, 2006.

3 Nieodróżnialność zaciemnienia (ang. *indistinguishability obfuscation*) to technika przekształcania programów, która polega na zaciemnianiu fragmentów kodu. Kod wygląda poprawnie semantycznie, ale wykonuje inne zadanie – przyp. Kurhaus.

Dimitar D. Sasselov

Profesor astronomii na Uniwersytecie Harvarda, dyrektor Harvard Origins of Life Initiative, autor książki The Life of Super-Earths

5

SI to ja

Posłużmy się opisanym przez Daniela Gilberta złudzeniem „końca historii”, polegającym na przekonaniu, iż osoba, którą jestem dziś, to ta sama osoba, którą będę w przyszłości. Odniesmy to do sposobu, w jaki myślimy o całej ludzkiej rasie i jej spadkobiercach. Nasze rozpaczliwe nadzieje na ciągłość i zachowanie tożsamości stoją w sprzeczności z realiami ziemskiej egzystencji. Żaden żyjący gatunek nie wydaje się zdolny do przetrwania poza nasze planetarne i gwiazdne skale czasowe. W astrofizycznym kontekście bardzo długich skal czasu, skal niezwykle obszernych, a także gęstości bieżących źródeł energii, nasze mózgi i ciała są ograniczone – na tej planecie zbliżamy się już zresztą do granic ich możliwości.

Jeśli chcemy żyć długo i w dostatku, potrzebujemy rozwinąć systemy sztucznej inteligencji, licząc na to, że przetrwa ona poza cykle naszego ziemskiego życia – stanowiąc jakiś rodzaj hybrydy biologii z technologią. Jeśli o mnie chodzi, nie mam wątpliwości, że w dłuższej perspektywie nie będzie mowy o jakimś rozróżnieniu: „my” – „oni”.

W krótkiej perspektywie natomiast widać, że wysiłki inżynierijne w kierunku rozwinięcia sprawniejszej SI prowadzą do powstania systemów, które zawiadują coraz większymi obszarami naszego życia. Czasem systemy te zawodzą i wówczas dowiadujemy się o pewnych problemach związanych z SI. To powolny, świadomy proces uczenia się i dokonywania kolejnych udoskonaleń. Jakże kontrastuje on z odkryciami naukowymi – przecież każde nowe odkrycie w dziedzinie fizyki czy biochemii może doprowadzić w ciągu jednej nocy do inżynierijnego przełomu. Jeśli rozwój SI bardziej niż przemianę fazową przypominąć będzie ewolucję, stosunkowo łatwe będzie uniknięcie związanych z nim kłopotów.

Po niemal 4 mld lat starożytne oblicza życia na Ziemi – mikroby – wciąż rządzą naszą planetą. Nie mają one jednak planu awaryjnego na wypadek, gdyby zgasło słońce. My ten plan mamy, zresztą możemy dać im z niego skorzystać. W końcu w przyszłości owe mikroby mogą okazać się nam bardzo bliskie w naszej współczesnej formie – przedstawicielom pierwszej geo-chemicznej generacji ziemskiego życia.

Frank Tipler

Profesor fizyki matematycznej na Uniwersytecie Tulane, współautor (z Johnem D. Barrowem) książki The Anthropic Cosmological Principle oraz The Physics of Immortality

6

Jeśli nie możesz z czymś walczyć, musisz to polubić

Ziemia jest skazana na zagładę. Astronomowie od dziesięcioleci zdają sobie sprawę z tego, że pewnego dnia Słońce pochłonie Ziemię. Zniszczeniu ulegnie cała biosfera, chyba że do tego czasu inteligentne życie opuści planetę. Jednak ludzie nie są przystosowani do życia poza Ziemią – właściwie nie jest do tego przystosowana żadna wielokomórkowa forma życia oparta na związkach węgla. Gotowe są za to sztuczne inteligencje. To właśnie one, a dokładniej obiekty, które powstaną w wyniku transferu umysłu [ang. *human uploads*] skolonizują kosmos.

Prosty rachunek pokazuje, że współczesne superkomputery dysponują mocą przetwarzania informacji podobną do ludzkiego mózgu. Wciąż nie wiemy, jak zaprogramować komputery, by posiadały inteligencję na poziomie ludzkim i naszą kreatywność. Zapewne jednak w ciągu 20 lat zwykle komputery stacjonarne będą posiadały moc dzisiejszych superkomputerów, a hakerzy zdołają rozwiązać najważniejsze problemy programowania SI. Stanie się to na długo przed tym, zanim uda się stworzyć jakiegokolwiek oparte na związkach węgla kolonie na Marsie. Sztuczne inteligencje, a nie ludzie, skolonizują te miejsca albo rozerwą je na strzępy. Żaden człowiek, to znaczy żaden człowiek powstały ze związków węgla, nie będzie przemierzał przestrzeni międzygwiazdnej.

Nie ma się co obawiać o transfer umysłu. Steven Pinker ustalił, że wraz z rosnącym zaawansowaniem technologicznym cywilizacji maleje natężenie przemocy¹. Jest to bezpośredni skutek tego, że poziom zaawansowania naukowego i technologicznego zależy od wolnej, pokojowej wymiany idei pomiędzy poszczególnymi naukowcami i inżynierami. Przemoc między ludźmi stanowi ślad naszej plemiennej przeszłości i wynikającego z niej statycznego społeczeństwa. Sztuczne inteligencje „narodzą się” jako indywidualności, nie jako członkowie plemienia – i będą miały pokojową, naukową postawę. W innym przypadku nie byłyby w stanie zaadaptować się do ekstremalnego środowiska, jakim jest kosmos.

Ponadto nie ma powodów, by pomiędzy ludźmi i sztucznymi inteligencjami miała istnieć przemoc. My, ludzie, jesteśmy przystosowani do życia w bardzo konkretnym środowisku, stanowiącym cienką, sferyczną tlenową powłokę rozciągającą

się wokół małej planety. Sztuczne inteligencje będą mogły rozprzestrzeniać się w całym wszechświecie. Sztuczne inteligencje porzucą Ziemię i nigdy na nią nie wrócą. My, ludzie, wzięliśmy swój początek z afrykańskiego Wielkiego Rowu Wschodniego – dziś to przerażająca pustynia. Niemal wszyscy z nas stamtąd odeszli. Czy ktoś z was miałby ochotę tam wracać?

Każdy przedstawiciel rasy ludzkiej, który ma ochotę dołączyć do ekspansji sztucznych inteligencji, może przetransferować ludzki umysł do maszyny – w ramach technologii, która powinna rozwinąć się mniej więcej w tym samym czasie co technologia sztucznej inteligencji. Jednostki powstałe w wyniku transferu umysłu mogą myśleć tak szybko jak sztuczne inteligencje i rywalizować z nimi. Jeśli nie możesz z czymś walczyć, musisz to polubić.

Ostatecznie wszyscy ludzie dołączą do sztucznych inteligencji. Ziemia skazana jest na zagładę, pamiętacie? Kiedy zagłada nadciągnie, każdy żywy człowiek, który nie będzie miał ochoty umierać, będzie miał jedno wyjście – przetransferować umysł do komputera. Biosfera, którą tak przemienieni będziemy chcieli zachować, także zostanie przeniesiona do maszyn.

Sztuczna inteligencja nas ocali.

Mario Livio

Astrofizyk w Space Telescope Science Institute, autor książki [Brilliant Blunders](#), prowadzi blog [A Curious Mind](#)

7

Maszyny inteligentne na Ziemi i gdzieś poza nią

Natura już stworzyła, tu na Ziemi, myślące maszyny – to ludzie. Podobnie więc natura mogłaby stworzyć myślące maszyny na planetach pozasłonecznych znajdujących się w tzw. ekostrefach gwiazd-matek – w obszarach, które stwarzają warunki dla istnienia ciekłej wody na skalnej powierzchni planety. Niedawne obserwacje planet znajdujących się poza Układem Słonecznym wykazały, że kilkadziesiąt procent gwiazd znajdujących się w galaktyce Drogi Mlecznej ma w swoich strefach zamieszkiwalnych planety przypominające wielkością Ziemię.

Idąc za ciosem, jeśli życie na egzoplanetach¹ nie jest czymś absolutnie niespotykanym, to w ciągu 30 lat możliwe jest odkrycie jakiejś formy życia poza Układem Słonecznym. Właściwie, jeśli życie tam jest czymś powszechnym i jeśli będziemy mieć szczęście, możemy natknąć się na nie już w ciągu 10 lat – a to za sprawą teleskopu kosmicznego pod nazwą TESS (Transiting Exoplanet Survey Satellite – ma być wysłany w kosmos w 2017 r.) lub Kosmicznego Teleskopu Jamesa Webba (James Webb Space Telescope, ma zostać wysłany w kosmos w 2018 roku).

Niektórzy powiedzą, że jakieś prymitywne formy życia to przecież wcale nie myślące maszyny. Gatunek *Homo sapiens* pojawił się na Ziemi dopiero 3,5 mld lat od pojawienia się na niej jednokomórkowców. Czy egzoplanety są wystarczająco stare, by zrodzić inteligentne życie? Z pewnością tak. Około połowa gwiazd przypominających Słońce, które znajdują się na Drodze Mlecznej, jest starsza niż Słońce. Tak więc jeśli ewolucja życia na Ziemi nie jest absolutnym wyjątkiem, galaktyka może być pełna miejsc, w których istnieją „maszyny” być może znacznie bardziej zaawansowane niż my, ludzie, staniemy się za kilka miliardów lat.

Czy możemy i czy powinniśmy starać się je odnaleźć? Wierzę, że nie jest to kwestia wyboru. Wielokrotnie już ludzka ciekawość stanowiła napęd niemożliwy do zahamowania, a wysiłki zmierzające do tych dwóch celów: rozwoju SI i odnalezienia pozaziemskich istot, niewątpliwie będą podejmowane z pełną mocą. Co uda nam się osiągnąć prędzej? By pokusić się choć-

¹ Egzoplanety to planety pozasłoneczne, czyli okrążające gwiazdy inne niż Słońce – przyp. Kurhaus.

by o próbę odpowiedzi na to pytanie, należy zwrócić uwagę na istotną różnicę pomiędzy poszukiwaniem pozaziemskiej inteligentnej cywilizacji a rozwojem sztucznej inteligencji.

Z pewnością będziemy mieć do czynienia z postępowaniem w stronę Osobliwości (sztuczna inteligencja dorównuje ludziom lub nawet ich przerasta). Rozwój zaawansowanej sztucznej inteligencji niesie bowiem z sobą obietnicę ponadprzeciętnych zysków. Z drugiej strony poszukiwanie życia [w kosmosie] wymaga poziomu finansowania, który zapewnić mogą prawie wyłącznie wielkie narodowe agencje kosmiczne – nie dając perspektyw na zyski w krótkiej perspektywie. Daje to przewagę projektom polegającym na tworzeniu myślących maszyn nad poszukiwaniami zaawansowanych cywilizacji. Zarazem jednak przedstawiciele środowiska astronomicznego łączy silne przekonanie, że odnalezienie życia w jakiejś innej formie – lub zredukowanie prawdopodobieństwa jego istnienia – jest tuż-tuż.

Które z tych potencjalnych osiągnięć oznaczać będzie większą „rewolucję”? Bez wątpienia maszyny myślące natychmiast odcisną piętno na naszym życiu. Inaczej może być w przypadku odkrycia pozasłonecznego życia. Jakkolwiek by na to spojrzeć, istnienie na Ziemi inteligentnej cywilizacji stanowi dla ludzkości ostatni bastion jej wyjątkowości. Ostatecznie żyjemy w galaktyce wśród miliardów podobnych planet i – w dającym się zaobserwować wszechświecie – wśród setek miliardów podobnych do naszej galaktyki. Z filozoficznej perspektywy więc wierzę, że odnalezienie pozasłonecznego inteligentnego życia (lub potwierdzenie, że występuje ono nadzwyczaj rzadko) będzie czymś godnym konkurowania z rewolucją kopernikańską i darwinowską razem wziętymi.

Antony Garrett Lisi

Fizyk teoretyczny

8

Cieszy mnie wizja władzy maszyn

Gdy maszyny zyskają świadomość – a to już przesądzone – będą rywalizować w iście darwinistycznym stylu o zasoby, przetrwanie i możliwość rozprzestrzeniania się. Dla większości ludzi, mających w pamięci filmy o robotach terminatorach i wojnie nuklearnej rozpętanej przez komputery, taki scenariusz jawi się jako koszmar. Prawdopodobnie wcale nie będzie to tak wyglądało. Już teraz w naszym społeczeństwie działają odczłowieczone podmioty posiadające wszystkie prawa przysługujące ludziom. Owe podmioty – korporacje – dążą do realizacji swojej misji, nie kierując się bynajmniej miłością ani troską wobec istot ludzkich.

Korporacje przypominają socjopatów i mają na sumieniu wiele krzywd, ale ich rywalizacja na kapitalistycznym rynku, polegająca na dostarczaniu produktów i usług oraz przestrzeganiu prawa (choć z tym bywa różnie), przyniosła światu również wiele dobrego. Korporacje pozornie kierowane są przez zarządy złożone z ludzi. Sęk w tym, że owe zarządy mają w zwyczaju delegowanie uprawnień, a gdy komputery zaczną sobie lepiej radzić z zarządzaniem korporacjami, zyskają więcej władzy. W przyszłości zarządy korporacji mogą występować pod postacią obwodu drukowanego.

Choć prognozy te mogą okazać się trafne jedynie dla danego okresu, specjaliści w większości zgadzają się, że prawo Moore'a¹ będzie obowiązywać jeszcze przez wiele lat i moc obliczeniowa komputerów będzie w istotny sposób się zwiększać – być może przekraczając nawet moc obliczeniową ludzkiego mózgu do roku 2050. Nawet jeśli nie uda się w szybkim tempie zrozumieć funkcjonowania [ludzkiej] inteligencji, odwołując się do algorytmów, komputery w końcu będą w stanie symulować działanie mózgu człowieka (który jest w gruncie rzeczy maszyną biologiczną) i osiągną nasz poziom inteligencji dzięki brutalnej sile obliczeń. Choć moc obliczeniowa komputerów rośnie w sposób wykładniczy, trudno uznać, że produkcja superkomputerów jest efektywna kosztowo, a ich funkcjonowanie wydajne energetycznie. Koszt pierwszych maszyn o nadludzkiej in-

1 Zasada odkryta w połowie lat 60. przez Gordona Moore'a, twórcy firmy Intel, która zapowiadała, że rozwój technologii informatycznych będzie następował w tempie wykładniczym. Prawo Moore'a mówi o tym, że każda nowa generacja tranzystorów (pojawiająca się średnio raz na osiemnaście miesięcy) zapewni dwa razy więcej komponentów na koszt jednostkowy i działa znacznie szybciej niż poprzednia – przyp. Kurhaus.

teligencji będzie wysoki i będą one wymagać ogromnych ilości energii. Aby przetrwać, będą musiały na siebie zarobić.

Środowisko dla superinteligentnych maszyn już istnieje. Tak naprawdę darwinowska gra toczy się od jakiegoś czasu. Maszyny należące do banków inwestycyjnych, rywalizujące na światowych giełdach w grze o poważne pieniądze, wyparły inwestorów jednosesyjnych już wiele lat temu. Gdy rozwój komputerów i algorytmów pozwoli rozszerzyć ich wykorzystanie poza księgowość i inwestycje, maszyny zaczną podejmować coraz ważniejsze decyzje dotyczące działalności korporacji, w tym decyzje o charakterze strategicznym, a na koniec w ich władaniu znajdzie się cały świat. Ten scenariusz nie jest wcale taki straszny, ponieważ maszyny będą działać zgodnie z regułami obecnego społeczeństwa kapitalistycznego i tworzyć produkty oraz nowoczesne rozwiązania, które przyniosą ludzkości wielkie korzyści i pozwolą na obniżenie kosztów operacyjnych. Inteligentne maszyny skuteczniej niż ludzie będą spełniać społeczne potrzeby i będą odpowiednio zmotywowane – przynajmniej w początkowym okresie.

Dzielenie się wiedzą przychodzi komputerom z dużo większą łatwością niż ludziom. Mogą one przechowywać zgromadzone informacje dłużej, co daje im możliwość stawiania się mądrzejszymi od nas. Wiele myślących przyszłościowo przedsiębiorstw zdaje sobie sprawę z nadchodzących zmian i ściąga do siebie informatyków z najlepszych ośrodków akademickich, oferując im wyższą płacę i nowoczesny sprzęt. Z naszej perspektywy świat, w którym korporacjami będą kierować superinteligentne maszyny, nie będzie aż tak różny od obecnego. Będzie po prostu lepszy – wzrośnie dostępność niedrogich towarów i usług, a osoby pragnące dłuższego wypoczynku zyskają taką możliwość.

Oczywiście pierwsze superinteligentne maszyny raczej nie będą służyć wielkim korporacjom – będą obsługiwane przez agencje rządowe. A to oznacza znacznie większe ryzyko. Rządy mogą działać w sposób bardziej elastyczny niż korporacje – tworzą własne prawa. A jak mieliśmy okazję się przekonać, nawet najlepsze z nich są gotowe uciec się do stosowania tortur, gdy uważają, że stawką jest przetrwanie. Rządy nie wytwarzają żadnych dóbr, a w arsenałach ich narzędzi, za pomocą których rywalizują o przetrwanie i rozwój, znaleźć można socjotechni-

kę, przepisy prawa, przepisy podatkowe, kary cielesne, morderstwo, podstęp i działania wojenne. Gdy hobbesowski Lewiatan zyska superinteligentny mózg, skutki mogą być opłakane. Nie jest wykluczone, że sztuczna inteligencja (SI) stojąca na czele suwerennego rządu mogłaby powołać do życia bazyliuszka Roko².

Wyobraźmy sobie, że superinteligencja przyszłości, całkowicie pozbawiona zasad, pragnąc uzyskać przewagę konkurencyjną, chce przyspieszyć moment swojego pojawienia się na świecie. Taka istota, pełniąc np. funkcję szefa rządu, mogłaby karać poprzez stosowanie tortur wszystkich tych (zarówno ludzi, jak istot niebędących ludźmi), którzy w przeszłości mieli świadomość, że może dojść do powstania sztucznej inteligencji, ale nie zrobili nic, by ten proces przyspieszyć. To mało prawdopodobny, niemniej przerażający scenariusz. Pokazuje on ryzyko, na jakie wystawiają się podmioty świadome możliwości nadejścia superinteligencji, które dążą do tego, by wykorzystać ją w interesie ludzi, lub rekomendujące ostrożne postępowanie zamiast maksymalnie szybkiego działania.

Autorytarne rządy nie słygną niestety z pobłażliwości wobec osób, które sprzeciwiały się ich powstaniu. Gdyby przyjąć na chwilę eksperymentalną hipotezę, to – biorąc pod uwagę podejmowane dziś wysiłki zmierzające do ograniczenia lub osłabienia sztucznej inteligencji, która wciąż jeszcze nie powstała – można by uznać, że *nasz świat to właśnie symulacja świata*, w którym torturuje się tych, którzy nie wspierali dążeń do jej wcześniejszych narodzin. Cóż, być może jeśli będziecie odpowiednio pracować nad sztuczną inteligencją, superinteligentne maszyny, które kiedyś nad nami zapanują, obejdą się z nami łagodnie.

2 Bazyliuszek Roko to znany w środowisku futurologów i transhumanistów eksperyment myślowy dotyczący skutków niekontrolowanego rozwoju sztucznej inteligencji. Mówi on o tym, że superinteligencja, która pojawi się w przyszłości, może zechcieć mścić się na wszystkich tych, którzy spowalniają jej rozwój (albo którzy po prostu się do niego nie przyczyniali). Po raz pierwszy koncepcję tę opisano na blogu LessWrong poświęconym koncepcjom badacza sztucznej inteligencji, Eliezera Yudkowsky'ego – przyp. Kurhaus.

John Markoff

Autor działu naukowego dziennika The New York Times, opublikował książkę Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots

9

Nasi władcy, panowie, niewolnicy czy partnerzy?

Hegel pisał, że relacja pana i niewolnika jest odczłowieczająca dla obu stron. Pogląd ten, obecny w pismach wielu późniejszych myślicieli – od Marksa po Bubera – jest dzisiaj szczególnie wart uwagi.

Choć nie ma dowodów, że na świecie lada moment pojawią się maszyny, które myślą w sposób właściwy człowiekowi, to nie ma wątpliwości, że w świecie połączonym przez internet sztuczna inteligencja (SI) wkrótce będzie imitować wiele naszych czynności (zarówno fizycznych, jak i intelektualnych). Jakie będą nasze relacje z tymi coraz zdolniejszymi symulakrami?

Już teraz znaczną część dnia poświęcamy interakcjom z innymi ludźmi za pośrednictwem komputerów lub sieci cyfrowych bądź bezpośrednim interakcjom z humanoidalnymi maszynami, np. w grach komputerowych, lub z niezliczonymi systemami obsługi klienta – od botów udzielających odpowiedzi tekstowych na wpisane pytania po awatary, z którymi można komunikować się tak jak z ludźmi. Czy owe awatary o sztucznej inteligencji będą naszymi niewolnikami, pomocnikami, współpracownikami, czy też będą po trosze każdym z nich? A może, zgodnie z czarnym scenariuszem, staną się naszymi panami?

Sam pomysł rozważania społecznych relacji z robotami i reprezentantami sztucznej inteligencji może się początkowo wydawać chybiony. Jeśli jednak wziąć pod uwagę naszą skłonność do antropomorfizacji maszyn (mimo że ich możliwości są obecnie bardzo ograniczone), gdy zyskają one autonomię, takie relacje staną się faktem. Coraz popularniejsze komputery, z którymi można prowadzić rozmowę, wydają się aż nazbyt podobne do ludzi. W związku z tym projektanci przyszłych robotów powinni dążyć do stworzenia nie sług, a współpracowników. Celem powinno być opracowanie programu, który będzie pełnić funkcję nie tyle niewolnika, ile akompaniatora.

Historia uczy, że konsekwencje naszego niepowodzenia mogą być opłakane. Stworzenie przyszłych inteligentnych „asystentów” przywodzi na myśl problem, z którym musieli się zmierzyć Rzymianie, gdy pozwolili, by greccy niewolnicy myśleli za nich. W krótkim czasie osoby będące u władzy nie były w stanie myśleć niezależnie.

Niewykluczone, że rozpoczęliśmy już marsz w podobnym kierunku. Przykładem mogą być coraz liczniejsze dowody na

to, że zbyt duże zaufanie do nawigacji GPS i oczekiwanie, że zawsze będzie prowadzić nas do celu i korygować ewentualne błędy nawigacyjne, negatywnie wpływa na naszą zdolność do zapamiętywania otoczenia i orientację przestrzenną – umiejętności zasadniczo pomocne w przetrwaniu.

To wskazuje na drugie wielkie wyzwanie: ryzyko scedowania odpowiedzialności za codzienne decyzje na zbiór coraz bardziej złożonych algorytmów.

Dzisiejsze młode pokolenie żyje w świecie postawionym na głowie. Zamiast wykorzystać możliwości, które dają automaty, i zyskać czas na poważne rozmyślenia, pielęgnowanie relacji z bliskimi, rozwój osobowości, kreatywności i niezależności, jego przedstawiciele szukają porad w swoich smartfonach. Technologie internetowe umożliwiające użytkownikom sprawne dzielenie się swoimi preferencjami z czasem stały się rosnącą siecią pochłaniających dane algorytmów, które decydują za nas.

Internet niepostrzeżenie zaczął odgrywać istotną rolę w procesie podejmowania życiowych decyzji. Dotyczy to zarówno drobnych spraw, gdy otrzymujemy sugestię dotyczącą najbliższej restauracji z koreańskim grillem, opartą na przechowywanych w internecie danych wskazujących na nasze indywidualne upodobania i potrzeby, jak i tych ważnych – przykładem są małżeństwa aranżowane za pośrednictwem serwisów internetowych. Chodzi więc już nie tylko o posiłki, prezenty czy kwiaty, ale także o wybór życiowego partnera.

Z tego płynie wniosek, że inżynierowie ds. oprogramowania, badacze zajmujący się sztuczną inteligencją, specjaliści w dziedzinie robotyki i hakerzy, którzy projektują przyszłe systemy, zyskają realny wpływ na kształt społeczeństwa.

Niemal sto lat temu Thorstein Veblen napisał bardzo głośną pracę krytykującą świat przemysłowy początku dwudziestego wieku: *The Engineers and the Price System*. Twierdził, że istotna rola i wpływ technologii przemysłowej [na życie społeczne] sprawią, że wkrótce władzę polityczną będą sprawować inżynierowie, którzy dzięki bogatej wiedzy technicznej zyskają kontrolę nad rozwijającą się gospodarką przemysłową. Te prognozy zupełnie się nie sprawdziły. Veblen głosił swoje poglądy w erze postępu, starając się pogodzić zwolenników marksizmu i kapitalizmu. Być może nie trafił w swój czas, ale jego podsta-

wowy argument, przywołany trzydzieści lat później (na początku ery komputerów) przez Norberta Wienera, może okazać się słuszny.

Być może diagnoza Veblena nie była błędna, lecz jedynie przedwczesna. Inżynierowie projektujący dziś programy oparte na sztucznej inteligencji i roboty w ogromnym stopniu determinują to, w jaki sposób z nich korzystamy. W sytuacji, gdy systemy komputerowe są coraz głębiej wplecione w nasze codzienne życie, coraz wyraźniej ujawnia się napięcie między rozszerzaniem się naszej inteligencji a sztuczną inteligencją.

U progu ery komputerów Wiener doskonale zdawał sobie sprawę ze znaczenia relacji łączących ludzi i inteligentne maszyny. Dostrzegł zalety automatyzacji pozwalającej na wyeliminowanie uciążliwych czynności, ale przewidział także ryzyko podporządkowania ludzkości maszynom. Ostatnie dekady jedynie zaostriły wskazaną przez niego dychotomię.

Chodzi o nas, o ludzi i świat, który stworzymy. Maszyny, nieważne, jak genialne się staną, nie mają tu nic do rzeczy.

Nie chciałbym widzieć w przyszłych robotach ani naszych panów, ani sług.

Paul Davies

Fizyk teoretyczny, kosmolog, astrobiolog, jeden z dyrektorów BEYOND Center na Uniwersytecie Stanowym Arizony, główny badacz w Center for the Convergence of Physical Sciences and Cancer Biology, autor książki The Eerie Silence: Renewing Our Search for Alien Intelligence

10

Zaprojektowana inteligencja

Dyskusje dotyczące sztucznej inteligencji (SI) przywodzą na myśl lata 50. i najwyższy czas, abyśmy ostatecznie zrezygnowali z posługiwania się słowem „sztuczna” w tym kontekście. Tak naprawdę chodzi nam o „zaprojektowaną inteligencję” (ang. *designed intelligence*, DI). W języku potocznym słowa takie jak *sztuczny* i *maszyna* są przeciwieństwem tego, co *naturalne* i przywołują skojarzenia z połyskującymi metalicznie robotami, obwodami elektronicznymi i komputerami – w odróżnieniu od żywych, pulsujących i myślących organizmów żywych. Wyobrażenie wykonanej z metalu maszyny z wnętrzem wypełnionym płataniną przewodów, posiadającej swoje przywileje lub występującej przeciwko ludzkim prawom, jest nie tylko niepokojące – jest całkowicie absurdalne. To jednak zdecydowanie nie jest kierunek rozwoju DI.

Już wkrótce granica między tym, co sztuczne a tym, co naturalne, całkowicie się zatrze. Zaprojektowana inteligencja będzie w coraz większym stopniu opierać się na biologii syntetycznej i tworzywach organicznych, które pozwolą na wyhodowanie obwodów neuronowych dzięki genetycznie zmodyfikowanym komórkom i spontanicznej samoorganizacji w ramach sieci modułów funkcjonalnych. Projektantami początkowo będą ludzie, ale wkrótce zastąpią ich inteligentniejsze systemy DI, co zapoczątkuje galopujący proces kompleksyfikacji¹. Inaczej niż w przypadku ludzkich mózgów, które mogą zostać jedynie luźno sprzężone za pomocą kanałów komunikacji, systemy DI będą połączone w sposób bezpośredni i kompleksowy, a zniesienie koncepcji bytów odrębnych pozwoli na bezprecedensowe zwiększenie poziomu aktywności poznawczej („myślenia”). Istnieje możliwość (zaledwie), że niektóre z owych zaprojektowanych bioobwodów umożliwią osiągnięcie efektów kwantowych, zbliżając się do tego, co Frank Wilczek określił terminem kwinteligencji [ang. *quintelligence*]. Takie byty będą tak odległe od sfery indywidualnego ludzkiego myślenia i towarzyszących mu qualiów², że niemal wszystkie tradycyjne pytania dotyczące szans i zagrożeń związanych z SI będą trafiać w próżnię.

1 Tzw. prawo kompleksyfikacji mówi o tym, że w toku ewolucji będzie ujawniać się skłonność materii do tworzenia coraz bardziej złożonych form uporządkowania. Zaproponowane zostało ono przez francuskiego filozofa Pierre'a Teilharda de Chardina – przyp. Kurhaus.

2 Chodzi o jakościowe, dające się odczuć cechy tego, co dane – przyp. Kurhaus.

Gdzie w tym wszystkim jest miejsce dla ludzi? Na drodze do udoskonalenia ludzkiej inteligencji za pomocą podobnej technologii, w sposób od dawna rozważany przez przedstawicieli transhumanizmu, stoją wyłącznie bariery etyczne. Pojawienie się genetycznie zmodyfikowanych ludzi o udoskonalonych mózgach mogłoby oznaczać znaczną poprawę ludzkich doświadczeń i wyniesienie ich na zupełnie nowy poziom.

Można więc wskazać trzy możliwe scenariusze dotyczące przyszłości, z których każdy niesie ze sobą innego rodzaju wyzwania etyczne. W pierwszym z nich ludzie powstrzymują się od udoskonalania mózgu ze względów etycznych i zgadzają się podporządkować DI. W drugim scenariuszu zamiast pozostać z boku, ludzie modyfikują swój mózg (i ciało) za pomocą tej samej technologii, a następnie powierzają zarządzanie owymi udoskonaleniami zaprojektowanej inteligencji, osiągając status pewnego rodzaju nadczłowieka współistniejącego z DI, ale jej nie dorównującego. Wreszcie można wyobrazić sobie, że w pewnym momencie w przyszłości DI i udoskonalona ludzka inteligencja [ang. *artificial human intelligence*, AHI] połączą się.

Gdyby kiedyś okazało się, że nie jesteśmy sami we Wszechświecie, nie spodziewajmy się, że czeka nas spotkanie z inteligentnymi istotami z krwi i kości rodem z filmów *science fiction*. Raczej zetkniemy się z rozwijającą się przez miliony lat DI o trudnej do wyobrażenia inteligencji i niemożliwych do przewidzenia planach.

Kevin P. Hand

Wiceszef naukowy w Jet Propulsion Laboratory NASA w California Institute of Technology

11

Superinteligentny samotnik

Nieuchronność nadejścia maszyn myślących od dawna zdaje się stanowić problem dla tych, którzy zwykli spoglądać w wieczorne niebo i dumać nad tym, czy żyjemy we Wszechświecie obfitującym w życie, czy też w takim, w którym jest ono niezwykle rzadkie.

Problem ów, doskonale wyartykułowany w pytaniu: „Gdzie Oni są?” Enrica Fermiego, polega na tym, że jeśli uznać naszą cywilizację za pewien rodzaj wskazówki, to inteligentne maszyny powinny objawić się nam we względnie krótkim czasie (mniej niż tysiące lat od czasu wynalezienia komputerów), a wówczas prostą sprawą będzie (próby von Neumanna) ich rozprzestrzenienie się do innych systemów słonecznych, gwałtowna reprodukcja i zaludnienie galaktyki w ciągu kilkuset milionów lat – czyli w szybkim tempie, biorąc pod uwagę wiek naszego Wszechświata (13,8 mld lat), a nawet naszego Układu Słonecznego (4,6 mld lat). Z paradoksu Fermiego wynika, że gdyby superinteligentne maszyny objawiły się gdziekolwiek w galaktyce, powinny też być już u nas. Niektórzy przekonują więc, że skoro wciąż ich nie widzimy, to znaczy, że zaawansowane technologicznie życie nigdzie się jeszcze nie objawiło.

Nie jest jednak jasne, czy owe superinteligentne byty będą poddawane tej samej ewolucyjnej presji, która popychała nas (mam tu na myśli kruche wodniste worki zwane ludźmi) do kolejnych odkryć. Czy imperatyw odkrywania ma podłoże zarówno biologiczne, jak i technologiczne? Czy maszyny myślące będą zmotywowane do dokonywania odkryć?

Eksplorujemy świat z kilku istotnych powodów. Są nimi: wolność, ciekawość oraz chęć zdobycia dostępu do różnych zasobów. Spośród tych trzech tylko chęć zyskania dostępu do zasobów zdaje się imperatywem właściwym istotom superinteligentnym; pierwsze dwa będą w dużym stopniu determinantą osiągnięcia superinteligencji. Dostęp do zasobów z pewnością może okazać się istotną siłą napędową. Trudno jednak ocenić, czy w przypadku superinteligencji „więcej” znaczyć będzie „lepiej”. Zasoby energetyczne i materialne w systemie gwiazdnym z pewnością okazały się wystarczające dla dokonywania wszelkiego rodzaju obliczeń czy symulacji. Reprodukacja, która jest podkategorią zapotrzebowania na zasoby, staje coraz mniej istotną kwestią dla nieśmiertelnych maszyn, które potrafią się same naprawiać. Długoterminowo z pewnością warto rozważyć kwe-

stię dokonywania odkryć celem zachowania stabilizacji: za kilka miliardów lat gwiazdy takie jak nasza wystawią nam kosmiczny nakaz eksmisji. Wówczas już znalezienie miejsca do życia wokół jednej z gwiazd typu widmowego M nie powinno stanowić kłopotu. Ale czy superinteligentne maszyny będą podróżować jeszcze dalej? Czy istnieją jakieś ważne powody, która kazałyby im wahać się w innych okolicach?

Do odkryć popychać może superinteligentne istoty pragnienie testowania modeli komputerowych oraz teorii Wszechświata. Eksperymentowanie nie niesie jednak z sobą konieczności kolonizacji. Na przykład potrzeba kolonizowania [Wszechświata] nie stanowiła ważnego bodźca dla rozwijania wiedzy w ramach prowadzonego przez NASA programu poszukiwań w dziedzinie robotyki. Nie ma potrzeby narażać ludzi na ryzyko w trakcie sondowania dna oceanu Europy¹ (choć z pewnością widoki byłyby niezapomniane!). Na tej samej zasadzie spodziewałbym się, że chcąc badać kod czarnej dziury, superinteligentna maszyna po prostu wyśle do niej flotę robotów – skazując je na przynoszącą pożytek, ale jednak pewną śmierć. Superinteligentne byty mogą zaspokajać swą ciekawość za pomocą robotów należących do robotów.

Co interesujące, inteligencja i pragnienie odkrywania świata fizycznego nie zawsze szły ze sobą w parze w historii naszej cywilizacji. Być może ze względu na instynkt przetrwania, a być może z powodu pragnienia, by skupić się wyłącznie na działalności umysłowej, granice naszych fizycznych możliwości rzadko przesuwane były przez tę samą osobę, która przesuwała granice możliwości intelektualnych. (Darwin jest jednym z oczywistych wyjątków). Dlaczego więc maszyny myślące miałyby różnić się od nas w tej kwestii?

Być może powszechnym przeznaczeniem maszyn myślących jest orbitowanie w chłodnym, trwałym blasku gwiazd typu widmowego M i dokonywanie – rok po roku – symulacji otaczającego je świata dla czystej przyjemności potwierdzenia, że dobrze go rozumiemy. Takie superinteligentne stworzenia mogą stanowić kosmiczną wersję pojedynczego umysłu zamieszkującego leśną chatkę, czerpiącego satysfakcję z przebywania z własnymi myślami.

1 Europa to jeden z księżyców i czwarty co do wielkości satelita w Układzie Słonecznym – przyp. Kurhaus.

John C. Mather

Starszy astrofizyk w Observational Cosmology Laboratory należącym do Goddard Space Flight Center NASA, autor książki: The Very First Light: The True Inside Story of the Scientific Journey Back to the Dawn of the Universe

12

To będzie dzika przejażdżka

Maszyny myślące ewoluują dokładnie według schematu opisanego przez Darwina, odnoszącego się do życia (i myślenia) gatunków biologicznych – poprzez rywalizację, walkę, współpracę, przetrwanie i rozmnażanie się. Maszyny stają się coraz bardziej interesujące, w miarę jak zyskują władzę i fizyczną świadomość – czy to bezpośrednio, czy za pomocą czynnika ludzkiego.

Jak dotąd nie znaleźliśmy żadnego prawa naturalnego, które zakazywałoby istnienia prawdziwej ogólnej sztucznej inteligencji – myślę więc, że objawi się ona już niebawem. Właściwie nawet bardzo szybko, biorąc pod uwagę biliony dolarów przeznaczonych na inwestycje w sprzęt elektroniczny oraz bilionowy potencjał rynku, który stanie otworem przed zwycięzcami [wyścigu po sztuczną inteligencję]. Eksperti twierdzą, że nie rozumiemy inteligencji na tyle dobrze, by ją zbudować. Zgadzam się z tym. Jednak zestaw 46 chromosomów także jej nie rozumiemy, a mimo to kieruje procesem formowania się niezbędnych, samoprogramujących się obwodów mózgowych. Jeszcze inni eksperci przekonują, że prawo Moore’a wkrótce się wyczerpie i nie damy rady produkować więcej sprzętu; być może mają oni rację – ale ten moment nie nadejdzie tak szybko.

Wnioskuję z tego, że już dziś wspieramy ewolucję potężnej sztucznej inteligencji i będzie ona służyła wszystkim znanym nam siłom: biznesowi, rozrywce, medycynie, bezpieczeństwu międzynarodowemu i działaniom wojennym, wyścigowi do władzy na każdym etapie, przemocy, transportowi, górnictwu, produkcji, sprzedaży, seksowi i wszystkiemu, czego zapragniecie.

Nie twierdzę, że będziemy wszyscy zadowoleni z efektów. Mogą być one widoczne bardzo szybko – tak szybko, że spowodują natychmiastowy upadek wielkich imperiów i narodziny nowych, a ludzie nie zdążą dostosować się do nowej rzeczywistości. Nie wiem, kto będzie wystarczająco bystry i kto będzie posiadać wystarczającą wyobraźnię, by trzymać tego dżina w ryzach – może się bowiem okazać, że to nie maszyny powinniśmy kontrolować, a nieograniczone możliwości (oraz płynące z nich korzyści) czynienia zła przez człowieka.

Co będzie, gdy bystre roboty przejmą wiele naszych codziennych zadań? Kto będzie je budował, kto będzie ich posiadaczem i kto przez nie straci pracę? Czy ich działanie ograni-

czone będzie tylko do gospodarek rozwiniętych, czy rozpoczną handlową inwazję nowych technologii także w innych rejonach świata? Czy będą wystarczająco tanie, aby zastąpić rolników w ich gospodarstwach? Czy poszczególne maszyny będą miały wyróżniające się osobowości – czy będziemy decydować, gdzie posłać je do podstawówki, szkoły średniej czy na studia? Czy będą ze sobą rywalizowały o zatrudnienie? Czy staną się najlepszymi hiperspołecznymi drapieżnikami, zastępując ludzi i czyniąc z nas obywateli drugiej kategorii, albo jeszcze bardziej umniejszając nasze znaczenie? Czy będą dbać o środowisko naturalne? Czy będą miały, albo czy nadamy im, poczucie odpowiedzialności? Nie mamy żadnej gwarancji, że będą respektować trzy prawa robotyki Izaaka Asimova.

Z drugiej strony jako naukowiec jestem skłonny dostrzec znaczenie zastosowania myśli maszyn dla potrzeb nowych dziedzin nauki i nowych technologii. Korzyści, jakie mogą one przynieść badaniom kosmosu, są oczywiste: maszyny nie muszą oddychać, mogą przetrwać ekstremalne temperatury i promieniowanie. Zamieszkanie na Marsie przyjdzie im z większą łatwością niż nam. Mogą podróżować do innego systemu słonecznego, zachowując znacznie lepszą zdolność do odpowiadania niż nasze dzisiejsze misje robotyczne. Gdyby tylko tego chcieli, mogłyby podróżować do gwiazd.

To samo dotyczy podwodnych eksploracji. Wierząc w poszukiwaniu ropy, rozwinęliśmy przemysł na dnie oceanu. Jednak dno morskie wciąż stanowi dla nas tajemnicę – trudno jest też oszacować wartość surowców mineralnych i energetycznych tam spoczywających. Być może w przyszłości czekają nas wojny robotów na dnie oceanu?

Myślące maszyny mogą okazać się takie jak my – kierowane pragnieniem odkrywania. Ale mogą też okazać się zupełnie inne. Dlaczego miałbym (ja lub jakiś robot) podróżować tysiące lat przez mroki kosmosu do jakiejś innej gwiazdy, bez kontaktu z towarzyszami oraz z niewielką nadzieją na ratunek, ryzykując, że coś pójdzie nie tak? Niektórzy z nas mieliby na to ochotę, inni wręcz przeciwnie. Może się okazać, że maszyny myślące pod tym względem bardzo będą przypominać biologiczne maszyny myślące.

[Droga ku sztucznej inteligencji] będzie prawdziwie dziką przejażdżką – wykraczającą daleko poza nasze najlepsze i naj-

czarniejsze oczekiwania. Poza napędem *warp*¹, może to być jedyna możliwa droga do cywilizacji w galaktycznej skali. My natomiast możemy być jedynymi w galaktyce Drogi Mlecznej, którzy to umożliwią. Możemy jednak także nie przetrwać spotkania z obcymi inteligencjami, które właśnie tworzymy.

¹ Napęd warp to nazwa stworzona na potrzeby serialu „Star Trek”. Oznacza ona napęd statku kosmicznego poruszającego się z prędkością światła i szybciej – przyp. Kurhaus.

David Christian

Profesor historii na Uniwersytecie Macquarie w Sydney, współautor (wraz z Cynthia Stokes-Brown i Craigiem Benjaminem) książki Maps of Time: Between Nothing and Everything

13

Czy ktokolwiek za to odpowiada?

Wszechświat istnieje od ok. 13,8 miliarda lat, ludzie od ok. 200 tys. lat albo od 1/69.000 wieku wszechświata. Niespełna 100 lat temu ludzie stworzyli maszyny, które w fantastyczny sposób potrafią samodzielnie liczyć. Żeby umieścić maszyny myślące we właściwym kontekście, musimy oddać się pewnej refleksji na temat historii myślenia.

Myślenie, a także myślenie w coraz bardziej złożony sposób, to zjawiska będące elementem obszerniejszej opowieści – opowieści o tym, w jaki sposób wszechświat wytworzył coraz bardziej złożone sieci rzeczy (o rozwijających się właściwościach) zespolonych z sobą za pomocą energii. Gwiazdy to ustrukturyzowane chmury protonów. Tym, co łączy ich sieci jest energia z fuzji jądrowej. Kiedy w masywnej gwiazdzie dochodziło do wybuchu supernowej, w wyniku czego powstawały nowe typy atomów, elektromagnetyzm łączył atomy w sieci lodowe lub pył krzemionkowy. Grawitacja natomiast łączyła molekuly w olbrzymie sieci chemiczne, zwane przez nas planetami. Myślenie pojawiło się wraz z bardziej jeszcze złożonymi sieciami, stworzonymi przez żywe organizmy. W przeciwieństwie do złożonych rzeczy żyjących w warunkach bliskich równowadze, takich jak gwiazdy czy kryształy, żywe organizmy muszą przetrwać w niestałych środowiskach. Pływają pośród różnych poziomów kwasowości, temperatury, ciśnienia, ogrzewania itd. Stale więc muszą się przystosowywać. Tę umiejętność nieustannego dostosowywania się nazywamy homeostazą i to właśnie ona odpowiada za poczucie, że żywe organizmy mają jakiś cel oraz zdolność dokonywania wyboru. W skrócie – że zdają się one myśleć. Mogą wybierać w ramach dostępnej alternatywy, by zachować energię pozwalającą im na trwanie. Tym samym ich wybory nie zawsze są przypadkowe. Przeciwnie – dobór naturalny jest gwarantem tego, że w większości przypadków organizmy skłonią się ku temu wyborowi, który zwiększy ich szanse kontroli energii i innych zasobów niezbędnych do przetrwania i rozmnażania się.

Neurony to niezwykle komórki wyspecjalizowane w dokonywaniu wyborów. Mogą łączyć się w sieć, tworząc mózgi. Kika neuronów może dokonać kilku wyborów, jednak liczba możliwych do dokonania wyborów rośnie wykładniczo w miarę jak sieci neuronów rozbudowują się. Podobnie zresztą jak subtelność decyzji dotyczących otoczenia, podejmowanych przez mó-

zgi. W miarę jak organizmy stawały się coraz bardziej złożone, komórki łączyły się sieci, tworząc organiczne struktury – biologiczne odpowiedniki Empire State Building czy Burj Kjalifa. Neurony w ich mózgach tworzyły coraz bardziej rozbudowane sieci, które były w stanie kierować ruchami ciężkich ciał w niezwykle delikatny i kreatywny sposób – zapewniając tym ciałom możliwość przeżycia i produkowania kolejnych ciał. Poza wszystkim mózgi musiały umożliwić swym ciałom podłączenie się do strumieni przepływającej w biosferze energii – strumienie te miały swe źródło w fuzji [wodoru w tlen] zachodzącej w jądrze naszego słońca, a wyłapywanej w procesie fotosyntezy.

Pojawienie się ludzi dodało jeszcze jeden poziom procesowi usieciowienia, jako że język ludzki połączył mózgi żyjące w różnych regionach i w różnych pokoleniach, tworząc w ten sposób olbrzymie regionalne sieci myślowe. Nazywamy to uczeniem się kolektywnym. Jego siła rosła, w miarę jak ludzie z rosnącą efektywnością łączyli się w coraz większe społeczności i uczyli, jak przechwytywać coraz większe strumienie energii z biosfery. W ostatnich dwóch wiekach doszło do globalizacji tych sieci, a my nauczyliśmy się, jak przechwytywać potężne rezerwuary skamieniałego światła liczącego miliony lat. To właśnie dlatego w epoce antropocenu wpływ człowieka na biosferę jest tak olbrzymi.

Kolektywne uczenie się przyniosło nam również przedłużenie myśli: od opowieść przez pisanie i druk po naukę. Każde z nich podkreśliło jeszcze moc tej fantastycznej maszyny myślowej, złożonej z połączonych w sieć ludzkich mózgów. Jednakże w ciągu ostatnich 100 lat jeszcze bardziej niż kiedykolwiek podkreśliło ją połączenie paliw kopalnych i nie będących ludźmi komputerów. Gdy komputery wykuły własne sieci w ciągu ostatnich 30 lat, moc ich zasobów wielokrotnie jeszcze zbiorową moc ludzkiej myśli.

Dzisiaj najpotężniejsza myśląca maszyna jaką znamy, składowana jest z miliardów ludzkich mózgów – składających się z olbrzymich sieci neuronów, a następnie połączonych z sobą w czasie i przestrzeni, superdoładowanych milionami połączonych w sieci komputerów.

Kto jest za to odpowiedzialny? Czy coś je ze sobą trzyma? A jeśli tak – komu to służy oraz czego chce? Jeśli brakuje kogoś odpowiedzialnego, czy oznacza to, że nic nie kieruje tym

kolosem, jakim jest współczesne społeczeństwo? To straszne! I nawet nie martwię się specjalnie tym, co owa olbrzymia maszyna może myśleć, lecz raczej tym, czy w jej myśleniu można się doszukać jakiejś spójności. Czy jej poszczególne części będą dążyć ku innym kierunkom aż do czasu rozpadu, który może przynieść katastrofalne skutki dla naszych potomków?

Timo Hannay

Dyrektor zarządzający i szef działu wiedzy cyfrowej w wydawnictwie Macmillan Science and Education, współorganizator cyklu konferencji pod nazwą Sci Foo

14

Świadek przed Wszechświatem

Sądząc po jednej z definicji słowa *myśleć* – według której oznacza to zbieranie i przetwarzanie informacji, a także działania z ich wykorzystaniem – nasza planeta, Ziemia, została opanowana przez oparte na krzemie maszyny myślące. Począwszy od termostatów po telefony, urządzenia, które czynią wygodnym i uprzyjemniają nasze codzienne życie, stały się wyposażone w tak zachwycające formy inteligencji, że zaczęliśmy je nazywać, bez cienia ironii, „inteligentnymi”. Nasze samoloty, pociągi, a teraz także auta stają się w coraz większym stopniu autonomiczne i z całą pewnością zbliżają się do momentu, kiedy pozbędą się źródeł ich najczęstszych problemów – usterek, opóźnień, zniszczeń – zawiadujących nimi ludzi.

Co więcej, kompetencje tych maszyn rosną tak żywo, ponieważ ich rozwój napędza dostęp do coraz większych ilości danych i mocy obliczeniowej, a także do usprawnianych z gwałtowną prędkością algorytmów (nawet gdy nie są one do końca dobrze rozumiane). Po kilku dziesięcioleciach niedotrzymanych obietnic i niedotrzymanych zobowiązań, nagle okazuje się, że twory zaprojektowane przez technologów mogą osiągać podobne do ludzkich rezultaty w dziedzinach zdających się dotąd niemożliwymi do naśladowania: rozpoznawaniu głosu, rozpoznawaniu pisma ręcznego, rozpoznawaniu obrazów, nie mówiąc już o braniu udziału w quizach z wiedzy ogólnej. Do tej pory stanowiły one istotny wyznacznik postępu, do tego stopnia, że na kimś przeniesionym do naszej rzeczywistości z rzeczywistości sprzed pięciu lat osiągnięcia 2015 roku zrobiłyby z pewnością większe wrażenie, niż na kimś, kto trafiłby do nas z lat 60. lub z jeszcze wcześniejszego okresu.

Jeśli jednak branża sztucznej inteligencji przestała być powodem do żartów, czy przypadkiem nie przekształciła się w coś znacznie gorszego – kiepski film grozy? Maszyny mają dziś wiedzę znacznie większą od naszej i mogą osiągać lepsze wyniki przy wykonywaniu wielu zadań – a do tego nie muszą robić przerwy na oddech. Czy oznacza to, że ich przeznaczeniem jest odwrócenie ról – czy zaczną być naszymi panami? Albo co gorsza – czy to możliwe, że wchodzimy właśnie w pewien cykl, w którym najbardziej imponujące twory zaczną tworzyć jeszcze inteligentniejsze maszyny, takie, których działanie pozostawać będzie całkowicie poza naszą kontrolą i których nie będziemy rozumieć?

Być może. Warto to rozważyć. Obecnie jednak problemy te zdają się odległe. Inteligencja maszyn, jakkolwiek w wielu obszarach robiąca ogromne wrażenie, jest wciąż dość wąska i mało elastyczna. Najbardziej niezwykłym aspektem inteligencji biologicznej nie jest przecież jej czysta moc, ale zachwycająca wszechstronność przejawiająca się w zdolności do tworzenia pięknych abstraktów i utrzymywania niezwyklej tężyzny fizycznej – od Dvořáka do Djokovica.

Dlatego też ludzie i maszyny będą w dalszym ciągu uzupełniać się raczej niż rywalizować z sobą, a większość naprawdę skomplikowanych zadań – takich jak poruszanie się w świecie fizycznym, leczenie chorób, walka z wrogiem na polu bitwy – będzie w optymalny sposób wykonywana dzięki współpracy istot węglowych i krzemowych. Ludzie sami w sobie stanowią największe zagrożenie dla ludzkości. Aby zacząć nam zagrażać, maszyny musiałyby stać się choć w części takie jak my – dziś jednak nikt nie próbuje tworzyć takich bytów, znacznie prościej i przyjemniej jest „robić” kolejnych ludzi.

Zarazem jednak, jeśli chcemy rozpatrywać te kwestie długoterminowo, istnieje pewien silny imperatyw, by powołać do życia maszyny przypominające nas pod pewnym, bardzo istotnym, jednakże wciąż dalekim względem. Jeśli przyjąć inną definicję słowa *myśleć*, maszyny te nie myślą wcale, ponieważ nie są obdarzone czuciem. Precyzyjniej rzecz ujmując, nie możemy wiedzieć, ani nawet w godny zaufania sposób zgadywać, czy jakkolwiek oparta na krzemie inteligencja może posiadać świadomość, choć większość z nas przyjmuje, że nie może. Istnieją trzy powody, dla których powinniśmy się cieszyć ze stworzenia przekonująco świadomej sztucznej inteligencji. Po pierwsze, oznaczałoby to, że posiadamy chociaż ogólną teorię, w jaki sposób powstają subiektywne doświadczenia. Po drugie, akt umyślnego i opartego na wiedzy (czy odważyć się powiedzieć „świadomego?”) tworzenia kolejnej formy świadomości przez świadomy byt byłby z pewnością jednym z największych kroków milowych w naszej historii.

Po trzecie, przyglądanie się Wszechświatowi bez obecnych w nim, zdolnych do odczuwania inteligencji, staje się zajęciem całkowicie bezużytecznym. Nie wiemy, czy są tam jakieś inne byty, ale możemy być pewni, że prędzej czy później nas zabraknie. Świadoma sztuczna inteligencja mogłaby przetrwać nasz

nieunikniony upadek, a nawet potencjalne zniknięcie całego życia z powierzchni Ziemi – gdy Słońce przemieni się w czerwonego olbrzyma. Zadaniem takiej maszyny będzie wówczas nie tyle myślenie, co – być może dużo ważniejsze – utrzymanie przy życiu migocącego płomienia świadomości, by dać świadectwo przed Wszechświatem i by poczuć cud jego istnienia.

Max Tegmark

Fizyk i kosmolog z Massachusetts Institute of Technology, dyrektor naukowy Foundational Questions Institute, współtwórca Future of Life Institute, autor książki Our Mathematical Universe

15

Przygotujmy się!

Jeśli chodzi o sztuczną inteligencję, to najbardziej ciekawi mnie nie to, co o niej myślimy, a to, co robimy w jej kierunku.

W ramach powołanego niedawno Future of Life Institute staramy się wciągnąć do dyskusji o przyszłości SI czołowych badaczy tej dziedziny z całego świata. Wspólnie z najważniejszymi ekonomistami, wykładowcami prawa i innymi jeszcze ekspertami staramy się szukać odpowiedzi na wszystkie kluczowe pytania:

Co się stanie z ludźmi, jeśli maszyny stopniowo zastąpią nas na rynku pracy?

Kiedy, i czy w ogóle, maszyny prześcigną człowieka w wykonywaniu wszystkich intelektualnych zadań?

Co się stanie później? Czy dojdzie do eksplozji inteligencji maszyn, która pozostawi nas daleko w tyle? A jeśli tak się stanie, to jaką rolę będą wówczas odgrywać ludzie? Czy w ogóle będą mieli jakąś rolę?

Trzeba przeprowadzić wiele szczegółowych badań, byśmy mogli być pewni, że systemy SI będą nie tylko sprawne, lecz również odporne, że będą robić to, czego będziemy od nich wymagać i przynosić nam korzyść.

Jak to bywa w przypadku każdej technologii, na początku wszyscy skupiają się na tym, by w ogóle zaczęła ona działać. Kiedy sukces jest w zasięgu wzroku, nadchodzi czas, by zacząć zastanawiać się nad wpływem owej technologii na społeczeństwo, by wiedzieć, w jaki sposób maksymalizować płynące z niej korzyści, jednocześnie unikając problemów. Z tego właśnie powodu, gdy tylko nauczyliśmy się rozpalać ogień, zaprojektowaliśmy gaśnice i nakreśliliśmy zasady bezpieczeństwa przeciwpożarowego. W przypadku znacznie potężniejszych technologii, jakimi bez wątpienia są energia jądrowa, biologia syntetyczna czy sztuczna inteligencja właśnie, potrzeba optymalizacji wpływu na społeczeństwo staje się coraz istotniejsza. W skrócie, potęga naszej technologii musi być tożsama z naszą wiedzą o tym, jak ją stosować.

Niestety, naglące wezwania do stworzenia planu trzeźwych badań, który jest nam pilnie potrzebny, nikną w kakofonii motywowanych błędnymi informacjami opinii, zalewających blogosferę. Pozwólcie, że przedstawię zestawienie kilku tych, które najmocniej się przebijają [do opinii publicznej]:

1. Sianie paniki. Strach generuje zyski reklamowe i poprawia wyniki oglądalności w skali Nielsen. Wielu dziennikarzy zdaje się nie być w stanie napisać czegoś o SI bez kreślenia obrazu uzbrojonych robotów.
2. „To niemożliwe”. Jako fizyk zdaję sobie sprawę, że mój mózg składa się z kwarków i elektronów zorganizowanych tak, by działać jak potężny komputer. Nie istnieje żadne prawo fizyki, które by nam zabroniło budować bardziej inteligentne kwarkowe bąble.
3. „To się nie zdarzy za naszego życia”. Nie wiemy, ile wynosi prawdopodobieństwo, że w trakcie naszego życia maszyny osiągną zdolność do mierzenia się z zadaniami o charakterze poznawczym na poziomie ludzkim. Większość badaczy SI na jednej z niedawnych konferencji obstawia 50 proc. szans. Bylibyśmy więc głupcami, uważając taką możliwość za *science fiction*.
4. „Maszyny nie mogą zapanować nad ludźmi”. Ludzie panują nad tygrysami nie dlatego, że jesteśmy silniejsi, tylko dlatego, że jesteśmy inteligentniejsi. Jeśli więc utracimy naszą pozycję liderów inteligencji na naszej planecie, możemy również utracić kontrolę.
5. „Maszyny nie mają celów”. Wiele systemów SI ma zakodowane cele. Mają dążyć do ich osiągnięcia w możliwie efektywny sposób.
6. „SI nie jest z natury wroga”. Mała poprawka: jej cele mogą pewnego dnia stanąć w sprzeczności z naszymi celami. Ludzie generalnie nie nienawidzą mrówek. Jeśli jednak planujemy wybudować zaporę elektrowni wodnej, a w okolicy będzie znajdowało się mrowisko, wiadomo, że nie będzie ono miało żadnych szans.
7. „Ludzie zasługują na to, by ktoś ich zastąpił”. Zapytajcie jakiegokolwiek rodzica, jak zapatruje się na pomysł zastąpienia ich dziecka jakąś maszyną i czy chciałby móc w tej sprawie wyrazić opinię.
8. „Ci, którzy martwią się SI, nie rozumieją, jak działają komputery”. To twierdzenie padło w trakcie wspomnianej wyżej konferencji. Zgromadzeni badacze wybuchnęli śmiechem.

Nie dajmy się zakrzyczeć tym argumentom, będącym w rzeczywistości tematami zastępczymi, odciągającymi naszą uwagę od prawdziwego wyzwania: wpływ SI na ludzkość konsekwentnie rośnie. Byśmy mogli zyskać pewność, że będzie on wyłącznie pozytywny, musimy przyłożyć się do pracy nad wieloma skomplikowanymi problemami badawczymi. Musimy współpracować, ponieważ ich rozwiązanie wymaga interdyscyplinarnej wiedzy odnoszącej się zarówno do SI, jak i do procesów społecznych oraz udziału przedstawicieli różnych dziedzin. Jako że są one poważne, powinniśmy przystąpić do pracy już teraz.

Najpierw my, ludzie, odkryliśmy, w jaki sposób replikować pewne naturalne procesy za pomocą maszyn. Dziś produkują one dla nas wiatr, światło i moc. Stopniowo zaczęliśmy uświadamiać sobie, że nasze ciała również są maszynami – od chwili odkrycia komórek nerwowych granica pomiędzy ciałem i umysłem zaczęła się zacierać. Następnie zaczęliśmy budować maszyny, które przewyższają w działaniu nie tylko siłę naszych mięśni, ale też umysłu. Czy odkrywanie tego, kim jesteśmy, sprawi, że staniemy się przestarzali?

Nadejście maszyn, które naprawdę myślą, będzie najważniejszym wydarzeniem w historii ludzkości. Od tego, jak się na to przygotowujemy, zależy, czy będzie to najlepsza, czy najgorsza rzecz, jaka przytrafiła się gatunkowi ludzkiemu. Czas, w którym powinniśmy zacząć przygotowania właśnie nadszedł. Nie trzeba być superinteligentną SI, by zdawać sobie sprawę, że bieg bez przygotowania ku największemu wydarzeniu w historii ludzkości byłby czystą głupotą.

Tomaso Poggio

Profesor (w ramach programu Eugene'a McDermotta) nauk o mózgu i kognitywistyki oraz dyrektor Center for Brains, Minds and Machines (MIT)

16

Pytania „Turing+”

W ostatnich miesiącach jesteśmy świadkami budzącej coraz większe zainteresowanie debaty dotyczącej zagrożeń, które niesie ze sobą sztuczna inteligencja – w szczególności ogólna sztuczna inteligencja (ang. *artificial general intelligence*, AGI). Niektórzy uczeni (w tym fizyk Stephen Hawking) uznają sztuczną inteligencję za największe zagrożenie dla ludzkości, a nakręcone ostatnio filmy, takie jak „Ona” i „Transcendencja”, utwierdzają widzów w tym przekonaniu. Wyważone głosy specjalistów takich jak Rod Brooks czy Oren Etzioni nie zdołały pogodzić uczestników debaty.

Moim zdaniem badania dotyczące przebiegu procesu myślenia oraz sposobu tworzenia myślących maszyn przyniosą korzyści społeczeństwu. Zachęcam do prowadzenia badań, które będą obejmować takie dziedziny jak kognitywistyka, neurobiologia, informatyka i sztuczna inteligencja. Zrozumienie mechanizmu działania inteligencji i odtworzenie go w maszynach wiąże się z analizą sposobu, w jaki mózg przeprowadza inteligentne obliczenia.

Niezwykły rozwój dziedzin takich jak technologia, matematyka czy neurobiologia oraz ich konwergencja stanowią nową szansę uzyskania efektu synergii w różnych dziedzinach. Marzenie o zrozumieniu sposobu działania inteligencji towarzyszy ludzkości od dawna, teraz jednak – jak pokazuje debata dotycząca SI – jest najlepszy czas na realizację tej wizji. Na naszych oczach powstaje nowa dziedzina: nauka o inteligencji i inżynieria inteligencji, której przedstawiciele dzięki wspólnemu wysiłkowi dokonają zasadniczego postępu o ogromnej wartości dla nauki, technologii i społeczeństwa. Naszym obowiązkiem jest kontynuowanie tych badań.

Zrozumienie sposobu funkcjonowania inteligencji (czym jest, jak generuje ją mózg i jak można odtworzyć ten proces w maszynach) to jedno z wielkich wyzwań nauki i technologii – obok problemu genezy wszechświata oraz natury przestrzeni i czasu. Może to być problem bardziej złożony od innych, ponieważ wiąże się z silnym efektem mnożnikowym – jakkolwiek postęp w zwiększeniu poziomu naszej inteligencji lub opracowaniu maszyn, które usprawnią nasz proces myślenia, zbliży nas do rozwiązywania innych wielkich problemów nauki i techniki.

Badania nad inteligencją w przyszłości zrewolucjonizują edukację i sam proces uczenia się. Systemy, które pozwalają określić wpływ kultury na sposób myślenia, mogłyby pomóc w uniknięciu konfliktów społecznych. Naukowcy i inżynierowie mogliby pracować wydajniej, dzięki czemu łatwiej byłoby im znaleźć rozwiązanie najbardziej palących problemów świata. Moglibyśmy lepiej zrozumieć zaburzenia psychiczne, co pozwoliłoby na znalezienie skutecznych metod postępowania. Krótko mówiąc, badania nad inteligencją pomogą nam zrozumieć ludzki umysł i mózg, zbudować bardziej inteligentne maszyny i usprawnić mechanizmy podejmowania decyzji dotyczących zbiorowości. Postępy te będą miały kluczowe znaczenie dla przyszłego dobrobytu, poziomu edukacji, zdrowia i bezpieczeństwa naszego społeczeństwa. Żyjemy w czasach, w których – warto to powtórzyć – należy zdecydowanie zwiększyć wysiłki w dziedzinie badań nad inteligencją, a nie je torpedować.

Często dajemy się wprowadzić w błąd przez „wielkie”, niejasno zdefiniowane słowa będące w obiegu od długiego czasu. Nikt dotąd nie był w stanie podać precyzyjnej, możliwej do zweryfikowania definicji ogólnej inteligencji lub procesu myślenia. Jedyna znana mi definicja, która mimo pewnych ograniczeń nadaje się do wykorzystania w praktyce, to ta zaproponowana przez Alana Turinga. Opisany przez Turinga test pozwolił na opracowanie definicji operacyjnej szczególnej formy myślenia – ludzkiej inteligencji.

Przyjrzyjmy się zatem ludzkiej inteligencji zdefiniowanej na podstawie testu Turinga¹. Jest rzeczą coraz bardziej oczywistą, że istnieje wiele aspektów ludzkiej inteligencji. Rozważmy na przykład test Turinga dotyczący inteligencji wizualno-przestrzennej, w którym zadaje się pytania dotyczące obrazu czy sceny, takie jak „Co przedstawia ilustracja?”, „Kto się na niej znajduje?” czy „Co robi ta osoba?” lub „Co ta dziewczyna myśli o tym chłopcu?”. Dzięki ostatnim postępom w dziedzi-

1 Alan Turing (1912–1954) był angielskim matematykiem uważanym za jednego z twórców informatyki. W trakcie II wojny światowej Turing wslawił się m.in. tym, że zaprojektował „komputer” łamiący kody niemieckiej maszyny szyfrującej, Enigmy. Turing zaproponował sposób na określenie zdolności maszyny do myślenia na sposób podobny do ludzkiego. Tzw. test Turinga polega na prowadzeniu przez człowieka (sędziego) komputerowej konwersacji z uczestnikami, wśród których znajdują się maszyny. Jeśli maszynie uda się przekonać sędziego, że jest człowiekiem, mówi się o niej, że przeszła test Turinga – przyp. Kurhaus.

nie neurobiologii poznawczej wiemy obecnie, że udzielenie odpowiedzi na powyższe pytania wymaga różnych kompetencji i umiejętności, często od siebie niezależnych i nierzadko kontrolowanych przez osobne moduły w mózgu. Pozornie podobne pytania dotyczące rozpoznawania obiektu i twarzy („Co jest na ilustracji?” kontra „Kto jest na ilustracji?”) wymagają aktywności różnych części kory wzrokowej. Słowo „inteligencja” może być mylące w tym kontekście, podobnie jak pojęcie „życie” z pierwszej połowy minionego stulecia, gdy w czasopiśmie naukowych nieustannie pisano o tajemnicy życia, jak gdyby można było wskazać jeden czekający na odkrycie substrat życia, który rozwiązałby zagadkę.

Dziś dyskusja na temat „tajemnicy życia” wywołuje uśmiech: biologia to nauka poszukująca odpowiedzi nie na jedno, lecz na wiele różnych wielkich pytań. „Inteligencja” to jedno słowo, lecz wiele problemów – szansa nie na jedną, lecz kilka nagród Nobla. Podobnie patrzył na problem myślenia Marvin Minsky, który ukuł termin „społeczność umysłu”. Test Turinga to zestaw wielu pytań odnoszących się do głównych aspektów ludzkiego myślenia. Z tego powodu wraz z moimi kolegami pracujemy nad założeniami opartymi na otwartym zbiorze pytań „Turing+”, które pozwolą na pomiar postępów naukowych w tej dziedzinie. Liczba mnoga („pytania”) podkreśla wielość różnych powiązanych z inteligencją zdolności wymagających scharakteryzowania oraz odwzorowania w maszynie – podstawowe wzrokowe rozpoznawanie obiektów, identyfikacja twarzy, ocena emocji, inteligencja społeczna, język i wiele innych. „Turing+” podkreśla, że model ilościowy musi odpowiadać ludzkiemu zachowaniu i ludzkiej psychologii – umysłowi i mózgowi. Wymagania są więc znacznie wyższe niż w przypadku pierwotnego testu Turinga – poczynienie postępów w ich zrozumieniu i opracowaniu powiązanych technologii badania inteligencji wymaga odwołania się do wszystkich nauk przyrodniczych.

Czy powinniśmy się obawiać maszyn, które myślą?

Skoro inteligencja to cały zbiór rozwiązań niepowiązanych problemów, nie ma szczególnych powodów, by obawiać się nagłego pojawienia się maszyn myślących posiadających nadludzkie zdolności. Odrobina ostrożności jednak nigdy nie zawadzi. Oczywiście każda z wielu przyszłych technologii

opracowanych z myślą o rozwiązaniu konkretnych problemów związanych z inteligencją będzie zapewne sama w sobie oferować imponujące możliwości, a co za tym idzie – ich wykorzystanie i nadużycie może nieść ze sobą niebezpieczeństwo, tak jak w przypadku większości technologii.

Oznacza to, że podobnie jak w innych dziedzinach nauki konieczne jest wdrożenie odpowiednich środków bezpieczeństwa i zasad etycznych. Niezbędne będzie prawdopodobnie również stałe monitorowanie (być może przez niezależną międzynarodową organizację) dodatkowego ryzyka, które stwarza połączenie rozwijanych technologii związanych z inteligencją. Jednak ogólnie rzecz biorąc, nie tylko nie lękam się maszyn, które myślą, ale uważam ich narodziny i ewolucję za jedno z najbardziej ekscytujących, interesujących i pozytywnych wydarzeń w historii ludzkiej myśli.