CZŁOWIEK NA ROZDROŻU

SZTUCZNA INTELIGENCJA – 25 PUNKTÓW WIDZENIA

Seth Lloyd
Judea Pearl
Stuart Russell
George Dyson
Daniel C. Dennett
Rodney Brooks
Frank Wilczek
Max Tegmark
Jaan Tallinn
Steven Pinker
David Deutsch
Tom Griffiths
Anca Dragan

Chris Anderson
David Kaiser
Neil Gershenfeld
W. Daniel Hillis
Venki Ramakrishnan
Alex „Sandy" Pentland
Hans Ulrich Obrist

Alison Gopnik
Peter Galison
George M. Church
Caroline A. Jones
Stephen Wolfram

REDAKTOR
JOHN BROCKMAN

Helion

Tytuł oryginału: Possible Minds: Twenty-Five Ways of Looking at AI

Tłumaczenie: Marcin Machnik

ISBN: 978-83-283-6209-3

Drogi Czytelniku!
Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
http://helion.pl/user/opinie/czloro
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Helion SA
ul. Kościuszki 1c, 44-100 Gliwice
tel. 32 231 22 19, 32 230 98 63
e-mail: *helion@helion.pl*
WWW: *http://helion.pl* (księgarnia internetowa, katalog książek)

Printed in Poland.

SPIS TREŚCI

Rozdział 9.

DYSYDENCKIE PRZEKAZY

JAAN TALLINN

Jaan Tallinn, *programista, fizyk teoretyczny i inwestor, jest współzałożycielem Skype'a i Kazaa.*

Jaan Tallinn wychował się w Estonii i był jednym z niewielu estońskich programistów gier, gdy to państwo należało jeszcze do Związku Socjalistycznych Republik Radzieckich. W tym eseju porównuje dysydentów, którzy przyczynili się do zerwania żelaznej kurtyny, do dysydentów, którzy biją na alarm w związku z raptownymi postępami w pracach nad sztuczną inteligencją. Paradoksalnie korzeni współczesnych dysydenckich postaw wobec AI doszukuje się w pracach takich pionierów jak Wiener, Alan Turing i I.J. Good.

Jego głównym zainteresowaniem są zagrożenia egzystencjalne, wśród których AI jest jednym z najpoważniejszych. W 2012 roku wraz z filozofem Huwem Price'em i astronomem królewskim Martinem Reesem założył na Uniwersytecie w Cambridge Centre for the Study of Existential Risk — interdyscyplinarny instytut badawczy, którego celem jest złagodzenie zagrożeń związanych z pojawiającymi się technologiami i działalnością człowieka.

Swego czasu przedstawił mi się jako „świadomy konsekwencjalista" — świadomy do tego stopnia, żeby przekazać większość majątku na instytut Future of Life (którego jest współzałożycielem), instytut Machine Intelligence Research oraz na inne organizacje zajmujące się minimalizacją zagrożeń. Max Tegmark napisał o nim: „Jeśli jesteś inteligentną formą życia i czytając ten tekst za milion lat od teraz, nie możesz wyjść z podziwu nad rozkwitem życia, prawdopodobnie zawdzięczasz swoją egzystencję Jaanowi".

W trakcie ostatniej wizyty w Londynie Jaan uczestniczył razem ze mną w będącym elementem Serpentine Gallery's Marathon panelu AI w londyńskiej City Hall pod egidą Hansa Ulricha Obrista (jednego z autorów zamieszczonych tu esejów). Jak przystało na świat sztuki, tego samego wieczoru zorganizowano wystawne przyjęcie w posiadłości wypełnionej śmietanką towarzyską Londynu — artystami, modelkami, oligarchami i gwiazdami sceny i ekranu. Po zapoznaniu się ze wszystkimi na swój niewzruszony sposób („Cześć, jestem Jaan") Jaan niespodziewanie stwierdził: „Czas na hiphopowy taniec", po czym opuścił się na podłogę i oparty na dłoni zaczął demonstrować zdumionej śmietance towarzyskiej spektakularne ruchy taneczne. A później zniknął w subkulturze klubów tanecznych. Tak najwyraźniej kończył się każdy jego wieczór w trasie. Kto by pomyślał?

W marcu 2009 roku znalazłem się w nijakiej franczyzowej jadłodajni przy ruchliwej kalifornijskiej autostradzie. Umówiłem się w tym miejscu na spotkanie z młodym człowiekiem, którego blog obserwowałem. Abym mógł go rozpoznać, miał naszywkę z napisem: *Mów prawdę, nawet gdy drży ci głos*. Nazywał się Eliezer Yudkowsky i kolejne cztery godziny spędziliśmy na omawianiu jego przekazów dla świata — przekazów, które skłoniły mnie do wizyty w tej knajpie i ostatecznie zdominowały moją następną pracę.

PIERWSZY PRZEKAZ: OKUPACJA SOWIECKA

W *The Human Use of Human Beings* Norbert Wiener patrzał na świat przez pryzmat komunikacji. W jego oczach świat maszerował w takt drugiego prawa termodynamiki w stronę nieuniknionej śmierci z przegrzania. W takim świecie jedynymi (meta)stabilnymi jednostkami były *przekazy* — wzorce informacyjne, które rozchodziły się w czasie jak fale na powierzchni jeziora. Nawet ludzi można postrzegać jako przekazy, ponieważ atomy w naszych ciałach są zbyt ulotne, żeby wiązać z nimi swoją tożsamość. Jesteśmy więc raczej „przekazem" utrzymywanym przez funkcje naszego ciała. Jak wyraził to Wiener: „To struktura utrzymywana przez homeostazę jest kryterium naszej osobistej tożsamości".

Ja bardziej przywykłem do traktowania procesów i przetwarzania jako podstawowych elementów składowych naszego świata. Trzeba jednak przyznać, że perspektywa Wienera wyłuskuje pewne interesujące aspekty świata, które miały

olbrzymi wpływ na moje życie, chociaż z innej perspektywy pozostałyby w tle. Chodzi o dwa przekazy, oba sięgające genezą do drugiej wojny światowej. Początkowo były to ciche przekazy dysydenckie — takie, na które ludzie nie zwracali większej uwagi, nawet jeśli po cichu się z nimi zgadzali. Pierwszy z nich brzmiał: *Związek Radziecki składa się z wielu nielegalnych okupacji. Tym okupacjom trzeba położyć kres.*

Jako Estończyk dorastałem za żelazną kurtyną i widziałem jej upadek z pierwszego rzędu. Słyszałem ten pierwszy przekaz w nostalgicznych wspomnieniach dziadków i w zakłóconym hałasami Głosie Ameryki. Nabrał on mocy w erze Gorbaczowa, gdy państwo bardziej pobłażliwie traktowało dysydentów, a szczyt nastąpił w trakcie śpiewającej rewolucji pod koniec lat osiemdziesiątych ubiegłego wieku.

Jako nastolatek byłem świadkiem rozpowszechniania się tego przekazu wśród ludzi, najpierw wśród aktywnych dysydentów, którzy wyrażali go przez pół wieku wbrew grożącym im konsekwencjom, potem wśród artystów i literatów, wreszcie wśród członków partii i polityków, którzy przeszli na drugą stronę. Ta nowa elita stanowiła dość eklektyczną zbieraninę pierwotnych dysydentów, którym udało się przetrwać represje, znanych intelektualistów oraz nawet (ku wielkiemu rozgoryczeniu dysydentów) byłych komunistów. Zatwardziali dogmatycy — nawet najbardziej prominentni — zostali w końcu zepchnięci na boczny tor, a część z nich zbiegła do Rosji.

Co interesujące, gdy przekaz rozchodził się wśród ludzi, ewoluował. Najpierw był wyrażany w czystej i bezkompromisowej postaci („Okupacja musi się skończyć!") przez dysydentów, dla których prawda była ważniejsza od wolności osobistej. Mainstreamowe grupy miały więcej do stracenia, więc początkowo modyfikowały i osłabiały przekaz, przyjmując stanowiska w rodzaju: „W długoterminowej perspektywie sensownie byłoby delegować kontrolę nad sprawami lokalnymi". (Istniały rzecz jasna wyjątki: niektórzy intelektualiści wyrażali wprost oryginalny przekaz dysydentów). Ostatecznie ten oryginalny przekaz — który był po prostu prawdą — zwyciężył nad swoimi osłabionymi wersjami. W 1991 roku Estonia odzyskała niepodległość, a trzy lata później ostatnie oddziały sowieckie opuściły kraj.

Ludzie, którzy podjęli ryzyko i głośno wyrażali prawdę w Estonii i pozostałych krajach bloku wschodniego, odegrali olbrzymią rolę w doprowadzeniu do ostatecznego wyniku, który zmienił życie setek milionów osób, w tym moje. Oni mówili prawdę, mimo że drżały im głosy.

DRUGI PRZEKAZ: ZAGROŻENIA AI

Drugi rewolucyjny przekaz poznałem dzięki wspomnianemu na początku blogowi, który skłonił mnie do nawiązania kontaktu z autorem i umówienia się z nim w Kalifornii. Przekaz brzmiał: *Ciągły postęp w AI może doprowadzić do zmiany kosmicznych proporcji. Ten niekontrolowany proces przypuszczalnie doprowadzi do zabicia wszystkich ludzi. Musimy włożyć sporo wysiłku w to, by uniknąć takiego skutku.*

Po spotkaniu z Yudkowskim próbowałem zainteresować tym przekazem swoich kolegów i współpracowników ze Skype'a. Nie udało mi się. Przekaz był zbyt szalony, zbyt dysydencki. Jego czas jeszcze nie nadszedł.

Dopiero później dowiedziałem się, że Yudkowsky nie był pierwszym dysydentem, który wypowiedział tę prawdę. W kwietniu 2000 roku w „Wired" ukazał się długi felieton *Why the Future Doesn't Need Us* Billa Joya, współzałożyciela i głównego badacza Sun Microsystems. Joy ostrzegał:

> Przyzwyczajeni do życia w towarzystwie niemal rutynowych przełomów naukowych będziemy musieli jakoś uporać się z faktem, że najbardziej interesujące technologie XXI wieku — robotyka, inżynieria genetyczna i nanotechnologia — przedstawiają zupełnie inne zagrożenia niż wcześniejsze przełomy technologiczne. W szczególności roboty, sztuczne organizmy i nanoboty mają wspólną, coraz istotniejszą cechę: samoreplikację (…). Jeden bot może stać się wieloma i szybko wyrwać spod kontroli.

Artykuł Joya wywołał mnóstwo gniewu, lecz niewiele działań.

Bardziej zaskoczyło mnie jednak to, że przekaz o zagrożeniach AI niemal symultanicznie wyłonił się z nauki o komputerach. Na wykładzie z 1951 roku Alan Turing stwierdził: „Wydaje się prawdopodobne, że skoro pojawiła się metoda myślenia maszynowego, przerośnięcie przez nią naszych kiepskich możliwości jest tylko kwestią czasu. (…) Oznacza to, że powinniśmy oczekiwać, że w pewnym momencie maszyny przejmą kontrolę"[1]. Mniej więcej dekadę później jego kolega z Bletchley Park I.J. Good napisał: „Pierwsza ultrainteligentna maszyna jest *ostatnim* wynalazkiem, jaki człowiek będzie musiał wynaleźć, pod warunkiem że okaże się na tyle miła, że powie nam, jak nad nią panować"[2]. W *The Human Use of Human Beings* naliczyłem pół tuzina miejsc, w których Wiener

[1] Wykład ukazał się pośmiertnie w „Philosophia Mathematica" 4, nr 3, 1966, s. 256 – 260.

[2] Irving John Good, *Speculations Concerning the First Ultraintelligent Machine*, w: „Advances in Computers" 6, MA: Academic Press, Cambridge 1965, s. 31 – 88.

wspomina o jakimś aspekcie Problemu Kontroli („Maszyna w rodzaju dżinna, która potrafi się uczyć i podejmować decyzje na podstawie zdobytej wiedzy, nie będzie w żaden sposób zobligowana do podejmowania takich decyzji, które sami byśmy podjęli lub które byłyby dla nas do przyjęcia"). Najwyraźniej pierwszymi dysydentami rozpowszechniającymi przekaz o zagrożeniach AI byli sami pionierzy AI!

FATALNY BŁĄD EWOLUCJI

Podawano wiele argumentów — część wyrafinowanych, a część nie za bardzo — za tym, że Problem Kontroli jest realny i nie należy go traktować jako fantastycznonaukowej fantazji. Pozwolę sobie przedstawić jeden z argumentów, który ilustruje znaczenie tego problemu.

Przez ostatnie sto tysięcy lat świat (co oznacza Ziemię, ale argument obejmuje Układ Słoneczny i prawdopodobnie cały Wszechświat) znajdował się pod panowaniem ludzkiego mózgu. W trakcie tych rządów mózgi *homo sapiens* były najbardziej wyrafinowanymi mechanizmami kształtowania przyszłości (niektórzy nawet nazywali je najbardziej skomplikowanymi obiektami we Wszechświecie). Początkowo nie robiliśmy z nich większego użytku poza rozpatrywaniem kwestii przetrwania i prowadzeniem polityki plemiennej w grupie poszukiwaczy pożywienia, lecz dzisiaj wpływ tych instrumentów wykroczył poza samą ewolucję naturalną. Planeta przeszła z produkcji lasów do produkcji miast.

Zgodnie z prognozami Turinga skonstruowanie superludzkiego AI („metodą myślenia maszynowego") zakończy rządy ludzkiego mózgu. Rozejrzyj się wokół siebie — jesteś świadkiem ostatnich dekad rządów, które trwały sto tysięcy lat. Sama ta myśl powinna skłonić ludzi do refleksji, zanim wrzucą AI do szufladki z napisem „zwykłe narzędzia". Jeden z najwybitniejszych badaczy AI na świecie wyznał mi ostatnio, że odetchnąłby z ulgą, gdyby okazało się, że jednak nie jesteśmy w stanie stworzyć sztucznej inteligencji na poziomie ludzkim.

Oczywiście realizacja tego projektu może jeszcze potrwać. Mamy jednak podstawy, by sądzić, że wcale tak nie jest. W końcu ewolucja — ślepy i niezręczny proces optymalizacyjny — stworzyła ludzki poziom inteligencji względnie szybko po pojawieniu się zwierząt. Czy raczej po pojawieniu się organizmów wielokomórkowych: sprawienie, by komórki trzymały się razem, było dla ewolucji znacznie trudniejszym przedsięwzięciem niż stworzenie ludzi po pojawieniu się

tych organizmów. Nie wspominając o tym, że nasz poziom inteligencji był ograniczony przez tak groteskowe czynniki jak szerokość kanału rodnego. Wyobraź sobie, że inżynier AI nie może kontynuować pracy, gdyż nie jest w stanie dostosować rozmiaru czcionki w komputerze!

Tkwi w tym interesująca symetria: formując ludzi, ewolucja stworzyła system, który pod wieloma istotnymi względami lepiej sprawdza się w planowaniu i optymalizacji niż sama ewolucja. Jesteśmy pierwszym gatunkiem, który wie, że powstał w drodze ewolucji. Mało tego, stworzyliśmy wiele artefaktów (radia, broń palną, rakiety kosmiczne), których ewolucja raczej by nie stworzyła. Nasza przyszłość nie zależy więc już od ewolucji biologicznej, lecz od naszych decyzji. W tym sensie ewolucja padła ofiarą własnego Problemu Kontroli.

Możemy jedynie mieć nadzieję, że jesteśmy w tej kwestii mądrzejsi od ewolucji. Rzecz jasna *jesteśmy*, ale czy wystarczająco? Wkrótce się przekonamy.

AKTUALNA SYTUACJA

Tak więc wyglądamy ponad pół wieku po pierwszych ostrzeżeniach Turinga, Wienera i Gooda oraz dekadę po tym, jak ludzie mojego pokroju zaczęli zwracać uwagę na przekaz dotyczący niebezpieczeństw AI. Cieszę się, że poczyniliśmy olbrzymie postępy w podejściu do tej sprawy, to jednak ciągle zdecydowanie za mało. Zagrożenie ze strony AI przestało co prawda być tematem tabu, lecz badacze AI nadal nie w pełni je zauważają. Zagrożenie to nie jest też jeszcze wiedzą powszechną. W chronologii poprzedniego dysydenckiego przekazu jesteśmy mniej więcej na poziomie roku 1988, gdy podniesienie tematu okupacji sowieckiej nie było już równoznaczne z końcem kariery, wciąż jednak trzeba się było jakoś asekurować. Dzisiaj mamy do czynienia z podobnym asekuranctwem — stwierdzeniami typu: „Nie przejmuję się superinteligentnym AI, ale rosnąca automatyzacja rodzi pewne konkretne problemy etyczne", „Dobrze, że niektórzy badają temat zagrożeń związanych z AI, ale nie jest to palący problem" czy wręcz bardzo rozsądnie brzmiące: „To mało prawdopodobne scenariusze, ale ich potencjalnie poważne skutki uzasadniają poświęcenie im uwagi".

Tak czy inaczej w kwestii rozchodzenia się przekazu zbliżamy się do punktu krytycznego. Z ostatniej ankiety wśród badaczy AI, którzy opublikowali swoje prace na dwóch głównych międzynarodowych konferencjach AI w 2015 roku, wynika, że zdaniem 40% z nich ryzyko ze strony wysoce zaawansowanej

sztucznej inteligencji to „istotny problem" lub „jeden z najważniejszych problemów tej dziedziny"[3].

Oczywiście tak jak część dogmatycznych komunistów nigdy nie zmieniła zdania, tak i na pewno znajdą się tacy, którzy nigdy nie uznają AI za potencjalnie niebezpieczne. Wielu „zaprzeczaczy" pierwszego typu miało swoje korzenie w sowieckiej nomenklaturze i podobnie ci, którzy negują zagrożenie ze strony AI, często kierują się kwestiami finansowymi lub innymi praktycznymi pobudkami. Jedną z najczęstszych jest zysk korporacyjny. Sztuczna inteligencja przynosi zyski, a nawet jeśli nie, jest modnym, przyszłościowym przedsięwzięciem, z którym warto być kojarzonym. Dlatego negatywna postawa jest często produktem korporacyjnego PR i machinerii prawniczej. Z pewnej bardzo realnej perspektywy korporacje to nieludzkie maszyny, które realizują własne cele, niekoniecznie tożsame z celami jakiegokolwiek pracującego w nich człowieka. Jak zauważył Wiener w *The Human Use of Human Beings*: „Gdy ludzkie atomy współtworzą organizację, która je wykorzystuje, ale nie w pełni ich praw jako odpowiedzialne istoty ludzkie, lecz jako trybiki, zębatki i przekładnie, nie ma większego znaczenia, że są zbudowane z krwi i kości".

Innym silnym bodźcem skłaniającym do zlekceważenia ryzyka związanego z AI jest (bardzo ludzka) ciekawość, która nie zna granic. „Gdy widzisz coś technicznie kuszącego, sięgasz po to, wykonujesz to, a dopiero po sukcesie technologicznym zastanawiasz się, co o tym myśleć. Tak właśnie było z bombą atomową", stwierdził J. Robert Oppenheimer. Echo jego słów dało się słyszeć niedawno w wypowiedzi Geoffreya Hintona, prawdopodobnie wynalazcę głębokiego uczenia, który w kontekście zagrożeń związanych z AI powiedział: „Mógłbym podać zazwyczaj przytaczane argumenty, ale prawda jest taka, że perspektywie dokonania odkrycia nie sposób się oprzeć".

Niewątpliwie to przedsiębiorczości i naukowej ciekawości zawdzięczamy niemal wszystkie fajne rzeczy, które w dzisiejszych czasach uznajemy za oczywiste. Trzeba sobie jednak uświadomić, że postęp *nie jest zobligowany* do poprawiania naszego losu. Ujmując to słowami Wienera: „Można wierzyć w postęp, nie uznając go za zasadę etyczną".

Koniec końców nie możemy sobie pozwolić na czekanie, aż wszystkie korporacyjne głowy i eksperci AI potwierdzą ryzyko ze strony AI. Wyobraź sobie, że

[3] Katja Grace i in., *When Will AI Exceed Human Performance? Evidence form AI Experts*, *https://arxiv.org/pdf/1705.08807.pdf*.

siedzisz w samolocie, który za chwilę ma wystartować. Nagle słyszysz komunikat, że według 40% ekspertów na pokładzie znajduje się bomba. Wiadomo, co należy zrobić w takiej sytuacji, i na pewno nie jest to siedzenie i czekanie, aż pozostałe 60% ekspertów zmieni pogląd.

KALIBRACJA PRZEKAZU O ZAGROŻENIACH ZE STRONY AI

Przekaz pierwszych dysydentów o zagrożeniach AI był niewątpliwie proroczy, ale miał poważną skazę — podobnie jak wersja, która zdominowała aktualny dyskurs publiczny. W obu przypadkach lekceważy się zarówno powagę problemu, jak i potencjalne zalety AI. Innymi słowy, przekaz niedokładnie wyraża stawki w tej grze.

Wiener przede wszystkim przestrzegał przed zagrożeniami *społecznymi*, wyrastającymi z bezmyślnej integracji generowanych maszynowo decyzji w procesy zarządzania i niewłaściwego użytkowania (przez ludzi) tak zautomatyzowanego procesu decyzyjnego. Podobnie dzisiaj „poważne" debaty o zagrożeniach AI skupiają się głównie na kwestiach w rodzaju bezrobocia technologicznego lub uprzedzeniach w uczeniu maszynowym. Takie dyskusje bywają wartościowe, gdyż zajmują się krótkoterminowymi problemami, jednocześnie są jednak zdumiewająco zaściankowe. Przypomina mi się wpis Yudkowskiego na blogu: „Pytanie o wpływ *maszynowej superinteligencji* na konwencjonalny rynek pracy jest jak pytanie o to, jak na handel amerykańsko-chiński wpłynęłoby rozbicie się Księżyca o Ziemię. Oczywiście, że by wpłynęło, ale takie postawienie problemu oznacza, że nie dostrzegasz istoty rzeczy".

Moim zdaniem istotą rzeczy jest to, że *superinteligentne AI stanowi zagrożenie środowiskowe*. Pozwólcie, że to wyjaśnię.

Douglas Adams opisał w jednej z książek przypowieść o kałuży, która po przebudzeniu się zauważa, że znajduje się w dziurze. Co więcej, wydaje się, że dziura jest „całkiem zgrabnie dopasowana". Na podstawie tej obserwacji kałuża wysnuwa wniosek, że świat został stworzony specjalnie dla niej. Z tego powodu, jak pisze Adams, „chwila, z którą znika, raczej ją zaskakuje". Podobnym błędem jest założenie, że zagrożenia AI ograniczają się do niekorzystnych zmian społecznych. Brutalna rzeczywistość jest taka, że Wszechświat nie został stworzony dla nas. Wręcz przeciwnie, w drodze ewolucji zostaliśmy przystosowani do bardzo wąskiego zakresu parametrów środowiska. Między innymi potrzebujemy

temperatury na poziomie gruntu mniej więcej równej temperaturze pokojowej, ciśnienia około 100 kPa i odpowiedniego stężenia tlenu. Każda zmiana tej bezcennej równowagi — nawet tymczasowa — w ciągu kilku minut skończyłaby się naszą śmiercią.

Krzemowa inteligencja nie musi aż tak przejmować się środowiskiem. Dlatego znacznie taniej jest eksplorować kosmos za pomocą sond maszynowych niż „puszek z mięsem". Mało tego, aktualne warunki na Ziemi są niemal na pewno nieoptymalne dla czegoś, co dla superinteligentnego AI jest najważniejsze: *efektywnego przetwarzania danych*. Może się więc okazać, że nagle z antropogenicznego globalnego ocieplenia przechodzimy do maszynogenicznego globalnego ochłodzenia. To jedno z najpoważniejszych wyzwań w badaniach bezpieczeństwa AI: jak powstrzymać potencjalnie superinteligentne AI — AI o znacznie silniejszej pozycji od naszej — przed doprowadzeniem środowiska do postaci nienadającej się do zamieszkania przez biologiczne formy życia.

Co interesujące, ponieważ najpotężniejszymi źródłami zarówno badań nad AI, jak i zaprzeczeń dotyczących ryzyka są wielkie korporacje, to gdy odpowiednio mocno przymkniemy oko, przekaz „AI jako zagrożenie środowiskowe" zacznie przypominać chroniczne obawy o wykręcanie się korporacji od odpowiedzialności za środowisko.

Z drugiej strony, snując obawy przed społecznymi skutkami AI, przegapia się większość zalet. Trudno ująć w słowa, jak zaściankowo i mikroskopijnie prezentuje się przyszłość naszej planety w porównaniu z pełnym potencjałem ludzkości. W skali astronomicznej nasza planeta wkrótce zniknie (chyba że okiełznamy Słońce, co jest zupełnie inną kwestią), a niemal wszystkie zasoby — atomy i darmowa energia — potrzebne do podtrzymania cywilizacji na dłuższą metę znajdują się w dalekim kosmosie.

Eric Drexler, wynalazca nanotechnologii, popularyzuje ostatnio koncepcję „paretotopii". Rzecz w tym, że właściwie skonstruowane AI może dać nam przyszłość, w której życie *każdego* ulega diametralnej poprawie, w której nie ma przegranych. Trzeba tu zrozumieć, że głównym czynnikiem powstrzymującym ludzkość przed osiągnięciem pełnego potencjału jest zapewne nasze instynktowne przeczucie, że uczestniczymy w grze o sumie zerowej, w której gracze z trudem wyszarpują dla siebie drobne wygrane kosztem innych graczy. Taki instynkt jest poważnie chybiony i destrukcyjny w „grze", w której gra się o wszystko, a wygrane są dosłownie astronomiczne. Tylko w naszej galaktyce istnieje więcej układów gwiezdnych niż ludzi na Ziemi.

NADZIEJA

Piszę to z ostrożną nadzieją, że przekaz o zagrożeniach AI ocali ludzkość przed zagładą, podobnie jak przekaz o okupacji sowieckiej doprowadził do wyzwolenia setek milionów osób. W 2015 roku przekonał 40% badaczy AI. Nie zdziwiłbym się, gdyby nowa ankieta wykazała, że większość badaczy AI uważa kwestię jego bezpieczeństwa za istotną.

Cieszy mnie pojawianie się pierwszych technicznych prac na temat bezpieczeństwa AI sygnowanych przez DeepMind, OpenAI oraz Google Brain, a także to, że jeśli chodzi o zespoły pracujące nad bezpieczeństwem AI, w tych skądinąd konkurujących ze sobą korporacjach kwitnie współpraca.

Światowe elity polityczne i biznesowe także powoli się budzą: bezpieczeństwo AI zostało uwzględnione w raportach i prezentacjach Institute of Electrical and Electronics Engineers (IEEE), World Economic Forum oraz Organizacji Współpracy Ekonomicznej i Rozwoju (OECD). Nawet niedawny (z lipca 2017 roku) manifest Chinese AI zawierał sekcje poświęcone nadzorowi nad bezpieczeństwem AI oraz opracowaniu praw, regulacji i norm etycznych i stworzeniu systemu oceny bezpieczeństwa AI, żeby między innymi zwiększyć świadomość potencjalnych zagrożeń. Mam olbrzymią nadzieję, że nowe pokolenie przywódców, które rozumie Problem Kontroli oraz zagrożenie środowiskowe AI, będzie potrafiło wznieść się ponad typowe plemienne gry o sumie zerowej i pomoże ludzkości przebyć te niebezpieczne wody, po jakich aktualnie żeglujemy, a tym samym otworzy nam drogę do gwiazd, które czekają na nas od miliardów lat.

Toast za następne sto tysięcy lat! Nie wahaj się mówić prawdy, nawet gdy drży Ci głos.

CZYM JEST AI?
OTO 25 PUNKTÓW WIDZENIA 25 BŁYSKOTLIWYCH UMYSŁÓW!

Rosnące możliwości różnych form sztucznej inteligencji niepokoją ludzi od kilkudziesięciu lat. Stopniowo uzależniamy się od ciągłej asysty nowoczesnych technologii, jednak coraz doskonalsze metody uczenia maszynowego, dostępna i potężna moc obliczeniowa korzystająca z niewyobrażalnie wielkich zasobów danych każą zadawać sobie pytania: do czego doprowadzi ten szalony rozwój techniki? Jak będzie wyglądało społeczeństwo przyszłości? Czy ludzie wyginą? Czy grozi nam zniewolenie ze strony maszyn albo garstki polityków pragnących jeszcze większej władzy? Coraz wyraźniej widać, że AI staje się powoli narracją współczesności. Świat, jaki znamy, przestanie istnieć. Przed jakimi wyborami przyjdzie stanąć ludzkości już wkrótce i czy zdołamy wybrać dobrze?

Ponad 60 lat po ukazaniu się słynnej *Cybernetyki* Norberta Wienera 25 wyjątkowych umysłów naszych czasów postanowiło ponownie przyjrzeć się koncepcjom tego matematyka i filozofa. Efektem kilkuletnich dyskusji na temat umysłu, myślenia, inteligencji i sensu człowieczeństwa jest ten zbiór 25 esejów — znakomite, wyjątkowo wszechstronne wprowadzenie w krajobraz kluczowych problemów związanych z AI. O wartości tego dzieła stanowi zawarte w nim zderzenie różnych poglądów i punktów widzenia: autorami esejów są wybitni informatycy, robotycy, fizycy, ale również osoby związane z kulturą, filozofią i psychologią. To rzecz, która zmusza do szerszego spojrzenia na jedno z najważniejszych wyzwań naszych czasów.

W książce znalazły się poglądy dotyczące AI w wielu kontekstach:

- ograniczeń uczących się maszyn
- egzystencjalnego ryzyka i przyszłości ludzkiej inteligencji
- możliwych zagrożeń dla społeczeństwa, demokracji, wolności
- idei, kreatywności i sztuki
- przyszłości naszej cywilizacji

JOHN BROCKMAN

(redaktor publikacji) jest amerykańskim pisarzem i agentem literackim. Specjalizuje się w tematach związanych z nauką i najnowocześniejszymi technologiami. Założył Edge Foundation, organizację mającą na celu inicjowanie i ułatwianie współpracy ekspertów z wielu dziedzin. Jego pasją jest rozpowszechnianie najbardziej nośnych idei naukowych. Jest założycielem i dyrektorem Brockman Inc., agencji literackiej i programistycznej. Mieszka w Nowym Jorku.

Helion
helion.pl
HELION SA
ul. Kościuszki 1c
44-100 Gliwice
tel.: 32 230 98 63
helion@helion.pl

ISBN 978-83-283-2578-4

Cena: 49,00 zł