

Filozofowie Doliny Krzemowej i przyszłość ludzkości

Czy longtermiści to dziś najbardziej
wpływowi myśliciele na świecie?
Dlaczego wspiera ich Elon Musk?
Czy ludzkość ma przed sobą
jeszcze tysiące lat rozwoju?

— s. 4-39

**Globalne ocieplenie przyspiesza.
Ale to nie koniec walki — s. 56**

**Roberto Calasso:
Książki nauczyły mnie,
czym jest namiętność — s. 66**

**Małgorzata Mirga-Tas.
Romska artystka
zachwyca świat — s. 96**

Ilustracja: Janusz Jurek

ISSN 0044-488X INDEKS 383716

Nr 9 770044 488232

812

CENA: 26,99 zł | VAT 8% STYCZEŃ (01) 2023



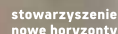
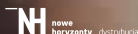
FESTIVAL DE CANNES
UN CERTAIN REGARD
2022 OFFICIAL SELECTION

Godland

REŻYSERIA H. PÁLMASSON

W KINACH

SNOWGLOBE PRESENTS IN ASSOCIATION WITH JOHN HOPKIN PICTURES W SPÓŁPRACY Z MANEHI FILMS FILM I VÁST & GARAGEFILM INTERNATIONAL PRESENTS ELLIOTT CROSSSET HOVE INGYAR SIGURDSSON VÍG CARMEN SONNE JACOB HAUBERG LOHMANN HILMAR GEDDÓNSSON & ÍDA MEGGÍN HLAUSDÓTTIR
DOPRODZIA MARIA VON HAUSWOLFF, DFF DOPRODZIA JULIUS KREBS DANISO REŻYSERKA DOPRODZIA BJÖRN VINDTOSON & KRISTIAN EIDNES ANDERSEN SCENARIUM FROSTI FRIÐRIKSSON GOSCIOLINI NINA GRÖNLUND (KONKURSYJÓKKA) KATRINE TERSGOV REŻYSERKA ALEX ZHANG HUNGTAI
PRODUCENT KOPRODUCENT GILLIAN TROBLE DOPRODUCENT DIDAR BOMEHR ANTHONY MUIR HINNÍ SPÁNG PETER POSSNE & GUÐMUNDUR ARIAR GUÐMUNDSSON PRODUCENT WYKONAWCZA HILNIR PÁLMASSON ELLIOTT CROSSSET HOVE & INGYAR SIGURDSSON
PRODUCENT KATRINE POIRIS ANTON MANN SKANSSON EYLA JARHSEN & MIKKEL JENSEN SCENARIUM I REŻYSERIA HILNIR PÁLMASSON SPRAWDZIŁ SŁUŻBA NEW EUROPE FILM SALES DYSTYBUCJA W POLSCE STOWARZYSZENIE NOWE HORYZONTY





Dominika Kozłowska

Szukać języka, aby wypowiedzieć prawdę

JAKIŚ CZAS TEMU BRAŁAM UDZIAŁ W LETNIEJ SZKOLE w Yad Vashem. Jeden z wykładów wygłosił Yehuda Bauer, wybitny izraelski historyk zajmujący się badaniami nad Holokaustem. Utkwiła mi w pamięci jego polemika z często powtarzaną tezą o niezrozumiałości Zagłady i niewysławialności doświadczenia spotkania z absolutnym złem. Jeśliby ulec temu przekonaniu, oznaczałoby to poddanie rozumu i kapitulację ludzkości wobec zła. A na to nie możemy się zgodzić. Zrozumiałam wtedy sens żmudnych badań naukowych, np. rozpisanego na ponad dekadę projektu rekonstrukcji logistyki Zagłady: organizacji i przebiegu transportów wiozących Żydów z całej niemalże Europy do obozów masowej śmierci. Te badania to prawdziwie drobiazgowe śledztwa. W oparciu o szczątkowe niekiedy fragmenty dokumentów próbuje się zrekonstruować decyzje, za którymi stali konkretni ludzie, oraz losy tych, którzy nie przeżyli.

Myśl o niewysławialności zła powróciła do mnie ze zdwojoną siłą po lekturze znakomitego portretu romskiej artystki Małgorzaty Mirgi-Tas pióra Angeliki Kuźniak. *Porajmos*, zagłada Romów, jest historią nieopowiedzianą; nieznaną szerszej publiczności i nadal nie w pełni opisaną w badaniach naukowych. Po części ma to związek z cechami samej kultury romskiej, która opiera się w dużej mierze na ustnym lub wizualnym przekazie. Przede wszystkim jednak jest wyrazem naszej obojętności na losy społeczności, która została tak tragicznie doświadczona w czasie II wojny światowej. Szacuje się, że w latach 1939–1945 wymordowano nawet 75% europejskich Romów. Antycyganizm jest matrycą, w której możemy widzieć, jak nieznajomość kultury i brak międ-

zykulturowej komunikacji przekładają się na stereotypy, a te na dyskryminujące i nienawistne zachowania. W obliczu wyzwań i odpowiedzialności związanych ze stałym powiększaniem się liczby uchodźców są to dziś sprawy o zasadniczym znaczeniu.

W tym kontekście myślę o postulatach longtermizmu – wpływowego ruchu filozoficznego, który inspirował miliarderów wywodzących się z Doliny Krzemowej, a któremu poświęcamy Temat Miesiąca. Rzecznicy tego ruchu kładą nacisk na moralne zobowiązania wobec przyszłości. Podkreślają, że w naszych decyzjach powinniśmy brać pod uwagę nie tylko doraźne skutki, ale i te dalekosiężne, widziane w perspektywie setek, a nawet tysięcy lat. To podejście jest z pewnością cenne, choć zarazem problematyczne. Zdaję sobie sprawę, że w obliczu katastrof, które dzieją się na naszych oczach, wiara w naukę i postęp techniczny pozwala nam dziś zachować umiarkowany optymizm. Jednak patrząc w daleką przyszłość, nie można rezygnować z dogłębnego poznawania świata, w którym żyjemy tu i teraz. Wciąż bowiem nie jesteśmy w stanie uzyskać konsensusu w sprawach bardziej przyziemnych. Choćby co do tego, aby w ocenie wartości produkowanych dziś towarów brać pod uwagę doraźne, ale niebezpośrednie skutki, takie jak koszty zanieczyszczenia środowiska, ekstensywnej eksploatacji dóbr, a także wpływ na życie i zdrowie człowieka i innych istot żywych. Choć longtermiści rozszerzają horyzonty naszej wyobraźni etycznej o przyszłe pokolenia, nie wiem, czy oferują nam to, czego potrzebujemy równie pilnie – przekonujący język do sprzeciwu wobec niesprawiedliwości, pogardy i zła. — **Z**

SPIS TREŚCI



40 Poszukiwanie prezydenta wszystkich Czechów

Martyna Wasiuta: Wybory nad Wełtawą w cieniu wojny i inflacji



56 Jak uratować Ziemię i zachować duszę

Kamil Fejfer: Nie powstrzymamy zupełnie zmiany klimatycznej, ale możemy uniknąć apokalipsy



96 Romski pałac

Angelika Kuźniak: Małgorzata Mirga-Tas w swojej sztuce zszywa dzieje Romów z ich dzisiejszym życiem



fot. Michał Cizek / AFP / East News / Just Stop Oil / Handout / Anadolu Agency / Getty / Łukasz Gagulski / PAP / Ferdinando Scianna / Magnum / Forum

68

ROBERTO CALASSO

Dla moich dzieci

U dziadków przeczytałem Wichrowe wzgórza. Tamta noc nie miała końca. Chciałem za wszelką cenę czytać dalej. Aż dotąd nie wiedziałem dokładnie, czym jest namiętność – i tamtej nocy się dowiedziałem

TEMAT MIESIĄCA

6 Pojutrze ludzkości

Kacper Kowalczyk: *Longtermiści, najbardziej wpływowi filozofowie świata*

16 Człowiek, gatunek międzyplanetarny

Nick Bostrom w rozmowie z Kacprem Kowalczykiem

22 Przyszłość pod znakiem zapytania

24 Na pokładzie statku „Przyszłość”

Jacek Dukaj w rozmowie z Karolem Kleczką: Czy przestaliśmy marzyć o lepszym jutrze?

32 Musk. Móc, chcieć, musieć

Agata Sikora: *Elon Musk jako obiekt kultu*

SPOŁECZEŃSTWO • ŚWIAT

40 Poszukiwanie prezydenta wszystkich Czechów

Martyna Wasiuta

48 Jeden na jeden

Anna Mateja: *Pierwszą chorobą osób w kryzysie bezdomności jest samotność*

PUBLICYSTYKA

56 Jak uratować Ziemię i zachować duszę

Kamil Fejfer

FELIETONY

54 Feminatywy

Olga Gitkiewicz

55 Springer z trasy

Filip Springer

65 Kościół – mój dom?

Janusz Poniewierski

IDEE

68 Dla moich dzieci

Roberto Calasso

76 O czym milczą bogowie

Stefan Klemczak: *Cienka nić łączy z sobą wszystkie książki Roberto Calassa*

STACJA: LITERATURA

86 Przeprawa

Opowiadanie Grzegorza Bogdała

LUDZIE • KSIĄŻKI • ZDARZENIA

96 Romski pałac

Angelika Kuźniak

106 Problem z talentem

Adrian Stachowski: *Czy Kapuściński w końcu zmyślał czy nie?*

109 Plus minus

Jerzy Franczak: *Georges Perec bez „e”*

112 Dwie historie

Klaudia Muca o podróży przez świat osób głuchych

114 Noty o książkach

116 Ale książka!

Zbigniew Rokita: *Czas opowiedzieć polsko-niemiecką historię Brechtawia, Gdańziga i Szczettina*

118 Jeź na autostradzie

Julia Fiedorczyk czyta wiersz Miłosza Biedrzyckiego

120 Kadry – znaki szczególnie

Diana Dąbrowska: *Roztrzępany pan Hulot w świecie przyszłości*

ZNAK

M I E S I Ę C Z N I K

ul. Tadeusza Kościuszki 37, 30-105 Kraków
tel. (12) 61 99 530, fax (12) 61 99 502
e-mail: miesiecznik@znak.com.pl
www.miesiecznik.znak.com.pl

redakcja: Mateusz Burzyk (sekretarz redakcji), Michał Jędrzejek, Karol Kleczka, Ilona Klimek, Dominika Kozłowska (redaktor naczelna), Janusz Poniewierski, Agnieszka Rzonca, Henryk Woźniakowski

zespół: Michał Bardel, Wojciech Bonowicz, Halina Bortnowska, Bohdan Cywiński, Bartłomiej Dobroczyński, Tomasz Fiałkowski, Tadeusz Gadacz, ks. Michał Heller, Jerzy Illg, Piotr Kłotkowski, Maria Makuch, Janina Ochojska-Okońska, abp Grzegorz Ryś, Władysław Stróżewski, Karol Tarnowski, Łukasz Tischner

współpraca: Ewa Bolińska-Gostkowska, Andrzej Brzeziecki, Marta Duch-Dyngos, Magdalena Falińska, Monika Gałka, Ewelina Kaczmarczyk, Krzysztof Kornas, Dobrosław Kot, Elżbieta Kot, Artur Madaliński, Barbara Matecka, Marek Maraszek, Anna Marchewka, Anna Mateja, Andrzej Muszyński, Małgorzata Nocur, Urszula Pieczek, ks. Eligiusz Piotrowski, Martyna Stowik, Marcin Wilk, Edyta Zielińska, Marcin Żyła

projekt graficzny pisma:

Władysław Buchner, Rafał Buchner

automatyzacja składu:

Paweł Żeleznikowicz (Kobugi)

dyrektor artystyczny:

Władysław Buchner

łamanie:

Ilona Klimek

fotodycja:

Marcin Kapica

korekta:

Barbara Gąsiorowska, Magdalena Wołoszyn-Cępa

ilustracje portretowe:

Patrycja Podkościelny

nakład:

2600 egz.

promocja i reklama:

tel. (12) 61 99 530
e-mail: miesiecznik@znak.com.pl

prenumerata:

tel. (12) 61 99 561
e-mail: prenumerata@znak.com.pl

Redakcja nie zwraca tekstów niezamówionych oraz zastrzega sobie prawo ich redagowania i skracania. Redakcja nie odpowiada za treść zamieszczanych ogłoszeń. Rozpowszechnianie redakcyjnych materiałów publicystycznych bez zgody wydawcy jest zabronione.



Ewangelia Doliny Krzemowej

Czy filozofia może zmieniać świat? Tak, zwłaszcza w połączeniu z wielkimi pieniędzmi. Ruch filozofów, zwanych longtermistami, szuka sposobów, by skutecznie rozwiązać globalne problemy. Nie tylko dzisiejsze, lecz przede wszystkim te, które nadejdą. Wspiera walkę z malarią, ale i badania nad pandemią czy sztuczną inteligencją. Do jego sympatyków należą m.in. Elon Musk oraz Dustin Moskovitz, współtwórca Facebooka. Dzięki hojności darczyńców tylko w 2021 r. na takie inicjatywy wydano 600 mln dolarów.

Mimo szlachetnych intencji longtermizm budzi też obawy: czy jego dalekosiężna perspektywa nie przyćmiewa współczesnych wyzwań? I czy nie grozi nam władza garstki technokratów? Warto uważniej przyrzeć się tej „dobrej nowinie”, która zachęca do przemiany naszego myślenia o świecie i etycznego nawrócenia.

Jakie są największe zagrożenia dla naszej cywilizacji? Czy ludzkość ma przed sobą jeszcze tysiące lat rozwoju? Dlaczego powinniśmy oddawać 10% dochodu na organizacje charytatywne? Kiedy Elon Musk uwierzył, że jest superbohaterem?

Odpowiadają:

**Kacper Kowalczyk, Nick Bostrom,
Jacek Dukaj i Agata Sikora**

Pojutrze ludzkości

KACPER KOWALCZYK

Longtermiści głoszą, że powinniśmy już dziś zatroszczyć się o ludzi, którzy przyjdą na świat za setki, tysiące, a nawet miliony lat. Od dawna żaden ruch filozoficzny nie wzbudził tak dużych kontrowersji

LUDZKOŚĆ MOŻE MIEĆ PRZED SOBĄ OGROMNĄ PRZYSZŁOŚĆ. CZŁOWIEK pojawił się na Ziemi co najmniej 200 tys. lat temu. Typowy gatunek ssaka żyje mniej więcej milion lat. *Homo sapiens* może więc mieć przed sobą jeszcze 800 tys. lat. Jak pisze australijski filozof Toby Ord, w skali jednego ludzkiego życia ludzkość miałaby teraz mniej więcej 16 lat, a zatem byłaby „wystarczająco dorosła, żeby się czymś wykazać albo żeby wpakować się w poważne tarapaty”. Burzliwy XX w. to w tym porównaniu zaledwie trzy dni.

Człowiek nie jest jednak typowym ssakiem. Ma język, narzędzia, zamieszkuje wszystkie kontynenty. Ludzkość mogłaby więc przetrwać kolejny miliard lat, do momentu kiedy rosnące Słońce wyjąłowi Ziemię. Dla porównania, pisze Ord, szacuje się, że niektóre typy morskich gąbek są już z nami pół miliarda

lat. A jeśli ludzie – albo ich ewolucyjna spuścizna – staną się kiedyś gatunkiem międzygwiazdowym, to mogliby przetrwać kolejne 100 mln lat, do czasu kiedy zgasną ostatnie gwiazdy. Jak pisze kolega Orda szkocki filozof William MacAskill, na każdego człowieka żyjącego dzisiaj może przypadać tysiące, a nawet dziesiątki tysięcy ludzi przyszłych. Istnieje też oczywiście ryzyko, że XXI w. jest naszym ostatnim.

Ta niecodzienna perspektywa stanowi punkt wyjścia longtermizmu (lub, inaczej, długoperpektywizmu), nowego ruchu społeczno-filozoficznego, który głosi, że kluczowym zadaniem naszych czasów jest pozytywny wpływ na przyszłość ludzkości, przy czym chodzi o przyszłość ekstremalnie długofalową: o kolejne tysiące, miliony, może nawet miliardy lat. Do obecnej fali longtermizmu można zaliczyć oprócz wspomnianych MacAskilla i Orda również Nicka Bostroma, dyrektora Future of Humanity Institute, który zatrudnia Orda, i Hilary Greaves, do niedawna szefową Global Priorities Institute, gdzie z kolei pracuje MacAskill. Obie instytucje są częścią Uniwersytetu Oksfordzkiego. Dodajmy od razu, że longtermizm nie jest jedynie poglądem garstki akademików. Ruch staje się coraz głośniejszy: zyskuje finansowe wsparcie w Dolinie Krzemowej, duże materiały poświęciły mu ostatnio „Time”, „The Economist”, „The New York Times” i „New Yorker”, a jednym z jego najbardziej znanych sympatyków stał się niedawno Elon Musk.

Longtermizm i efektywny altruizm

ELEMENTÓW LONGTERMIZMU można doszukiwać się już dawniej, np. w tradycji Irokezów, którzy podtrzymują „zasadę siódmego pokolenia”: zawsze bierz pod uwagę skutki swojej decyzji na siódme pokolenie do przodu. (Chociaż siedem pokoleń to krótki horyzont czasowy dla oksfordzkich longtermistów). Polski czytelnik mógł z kolei zetknąć się z podobnymi tematami w pismach niemieckiego filozofa Hansa Jonasa na temat etyki odpowiedzialności.

Dzisiejszy longtermizm wyrasta w dużej mierze z ruchu efektywnego altruizmu, który „używa faktów i logiki, żeby dowiedzieć się, jak najlepiej pomóc innym i wprowadzać tę wiedzę w życie”. MacAskill i Ord to jego współzałożyciele. Inspiracją dla ruchu pozostaje także Peter Singer, australijski filozof, autor ważnego artykułu *Famine, Affluence and Morality* (Głód, zamożność i moralność) z 1972 r. o moralnych obowiązkach bogatych w świecie biedy. Istotną częścią efektywnego altruizmu jest charytatywne rzeczoznawstwo: porównywanie skuteczności różnych organizacji charytatywnych. Efektywni altruści oceniają każdy problem pod względem skali, zaniedbania i trudności: ilu ludzi dotyczy?, jak dużo poświęca mu się teraz środków?, jak ciężko będzie go rozwiązać? Im problem większy, bardziej zaniedbany i łatwiejszy, tym lepszy jako cel dla efektywnych altruistów. Przez lata wielu efektywnych altruistów wspierało zatem pomoc dla ludzi skrajnie ubogich, których pozostają na świecie setki milionów – np. poprzez organizację GiveDirectly, która nasze pieniądze bezpośrednio przeleje na konta ubogich rodzin w Kenii i Ugandzie – czy też walkę z malarią, która rocznie zabija ok. 600 tys. ludzi – m.in. poprzez Against Malaria Foundation, zajmującą się zakupem i dystrybucją owadobójczych siatek na łóżka.

Celem efektywnych altruistów jest pomagać nie tylko efektywnie, ale także jak najwięcej. Wielu zobowiązuje się przeto do przekazywania przynajmniej części swojego dochodu – zazwyczaj jest to 10%, w świadomym nawiązaniu do praktyki dziesięciny – na rzetelnie wybrane cele charytatywne. Inni zwracają się do organizacji 80 000 Hours – tak nazwanej, bo tyle godzin w życiu przeciętny człowiek spędza w pracy – która doradza, jak nasza kariera zawodowa może zmienić świat na lepsze: czy lepiej zostać lekarzem, naukowcem, działaczem NGO, a może dobrze zarabiającym menedżerem, który swój dochód przekaże na „efektywny altruizm”.

Cały ruch doświadczył gwałtownego wzrostu w ciągu ostatniej dekady. Szacuje się, że w 2021 r. 600 mln dolarów zostało przekazanych na cele związane z efektywnym altruizmem. Głównym powodem była popularność ruchu wśród miliarderów, takich jak Dustin Moskovitz, współzałożyciel Facebooka i współtwórca organizacji Open Philanthropy, oraz Sam Bankman-Fried, były stażysta w jednej z efektywno-altruistycznych organizacji i handlarz kryptowalut, którego niedawne bankructwo z towarzyszącymi mu zarzutami oszustwa i niegospodarności skłoniło licznych efektywnych altruistów do refleksji na temat zagrożeń polegania na szczodrości bogatych kapitalistów. Ostatnio wielu efektywnych altruistów przeszło na longtermizm, przekazując swoją dziesięcinę na rzecz poprawy ludzkiej przyszłości zamiast na skrajną biedę czy siatki antymalaryczne. Uwagę mediów zwróciła pozytywna ocena, którą wspomniany już Musk wystawił na Twitterze książce MacAskilla *What We Owe The Future* (Co jesteśmy winni przyszłości), pisząc, że jest ona „bliska jego filozofii”. MacAskill szybko dostrzegł, że to wpływowy, ale i kontrowersyjny sojusznik. Podkreślił więc w odpowiedzi, że choć łączy ich poważne potraktowanie moralnych zobowiązań wobec przyszłości, to dzielą ich też istotne różnice zdań dotyczące oceny skali zagrożeń. Przykładowo MacAskill – inaczej niż Musk – nie sądzi, iż kolonizacja kosmosu powinna być



□ 35-letni filozof z Oksfordu William MacAskill to lider ruchu efektywnych altruistów i longtermistów. Mówi, że inspirują go trzy osoby: Mozi, starożytny myśliciel chiński wzywający do bezstronnej troski o wszystkich ludzi, Benjamin Lay, kwakier i abolicjonista z XVIII w. oraz Irena Sendlerowa ratująca Żydów w czasie II wojny światowej
fot. Matt Crockett

naszym kluczowym celem w najbliższym czasie; albo że spadek dzietności stanowi istotne zagrożenie dla naszej cywilizacji. Kłopotliwa etykieta „filozofa Muska” jednak zaczęła być przyklejana MacAskillowi. Jakże są zatem w rzeczywistości wady i zalety longtermizmu?

Troska o przyszłe pokolenia

LONGTERMIŚCI TRAKTUJĄ POWAŻNIE POGLĄD, ŻE SAMA W SOBIE ODLEGŁOŚĆ w czasie powinna mieć równie małe znaczenie jak odległość w przestrzeni. Brytyjski filozof Derek Parfit wspólnie z amerykańskim ekonomistą Tylerem Cowenem, pytają, co byśmy powiedzieli, gdyby okazało się, że wkrótce musimy umrzeć na raka, bo 2000 lat temu Kleopatra zażyczyła sobie dodatkę deseru. Byłoby to skandaliczne, uważa Parfit, przy czym podobnie skandaliczne jest nasze lekceważenie ludzi, którzy za 2000 lat będą tak samo realni jak my teraz.

Nie jest to jednak pogląd powszechny. Przykładowo: ekonomiczne analizy zmiany klimatycznej na ogół pomniejszają przyszłe korzyści o stały procent rocznie. Przy typowym założeniu 3-procentowa korzyść z ocalenia jednego życia dzisiaj jest traktowana jako równoważna korzyści z ocalenia mniej więcej 20 ludzi za 100 lat. Taka procedura ma sens w przypadku rozważania korzyści pieniężnych, które da się zainwestować i za 100 lat odebrać z odsetkami, ale nie jeśli chodzi o ludzkie życie. Istnieje też oczywiście ryzyko, że żadnej ludzkiej przyszłości nie będzie. Nie jest to jednak dobry powód, żeby ryzyko zignorować, zamiast je wyeliminować. Lepszy powód jest taki, że

z masami przyszłych ludzi nie łączą nas żadne specjalne więzy. Ludzie żyjący dziś to w końcu nasze rodziny, znajomi, rodacy. Jednak to, że osoba nam znajoma jest dla nas bliższa, nie znaczy chyba, że przyszły nieznajomy nie liczy się wcale. A takich nieznajomych może być całkiem sporo.

Przekonanie o moralnym znaczeniu przyszłych ludzi i o ich potencjalnie astronomicznej liczbie prowadzi longtermistów do wniosku, że bardzo ważne jest, jaki mamy na nich wpływ. Tym bardziej że długofalowa przyszłość stanowi obecnie kwestię zaniedbaną. Przyszli ludzie nie głosują w naszych wyborach ani nie posiadają udziałów w naszych firmach, zaś większość naszych instytucji posługuje się bardzo krótkoterminową perspektywą.

Egzystencjalne ryzyko

CZY JEDNAK W OGÓLE DA SIĘ ZNACZĄCO WPŁYNAĆ NA DŁUGOFALOWĄ PRZYSZŁOŚĆ? Wyginięcie ptaka dodo stanowi przykład długotrwałych, może nawet nieodwracalnych, skutków ludzkich działań. Podobnie byłoby z wyginięciem naszego gatunku. Dla longtermistów ważnym problemem jest więc ryzyko egzystencjalne, czyli – w definicji Bostroma – „ryzyko przedwczesnego wyginięcia inteligentnego życia na Ziemi lub drastycznego i trwałego



Elementów longtermizmu można doszukiwać się w tradycji Irokezów, którzy podtrzymują „zasadę siódmego pokolenia”: zawsze bierz pod uwagę skutki swojej decyzji na siódme pokolenie do przodu

zniszczenia jego potencjału na dalszy pozytywny rozwój”. Zdaniem Orda ryzyko egzystencjalnej katastrofy w ciągu następnych 100 lat wynosi jeden do sześciu, czyli tyle co w rosyjskiej ruletce. Bostrom ocenia je na jeden do pięciu, a bardziej optymistyczny MacAskill na jeden do stu.

Nie chodzi tutaj o zaobserwowaną częstotliwość egzystencjalnych katastrof, bo wygnać możemy tylko raz, ale o miarę pewności, jaką uzasadnia nasza baza dowodowa. Chodzi zatem o taki rodzaj prawdopodobieństwa, z którym ma do czynienia sędzia, kiedy musi ocenić ryzyko, że oskarżony jest jednak niewinny.

Jednym czynnikiem ryzyka egzystencjalnego jest zmiana klimatu. Dwutlenek węgla uwolniony przez rewolucję przemysłową ma krążyć w atmosferze przez tysiące lat i już powoduje poważne problemy. Orda martwi ponadto możliwość niekontrolowanego efektu cieplarnianego, który mógłby zmienić Ziemię w drugą Wenus. Niepewność ma tutaj efekt niesymetryczny: możemy wykluczyć, że zmiana klimatu będzie bardzo dobra, lecz nie że będzie bardzo zła. MacAskilla z kolei martwi, że węgiel i ropa wydają się niezbędne do rozwoju zaawansowanej cywilizacji. Spalenie naszego zapasu nie tylko może doprowadzić do katastrofy, ale pozbawi też ludzkość drugiej szansy, jeśli udałoby jej się przetrwać.

Innym czynnikiem ryzyka egzystencjalnego są pandemie, szczególnie ludzkiego autorstwa. Jak podaje Ord, w 2003 r. sekwencjonowanie ludzkiego DNA wymagało międzynarodowej współpracy i setek milionów dolarów; w 2022 r. jest ogólnie dostępne i kosztuje dolarów kilkaset. Podobnie jest z modyfikacją i syntezą DNA. W takim tempie niedługo tanio i łatwo będzie można stworzyć wirusa łączącego śmiertelność eboli z zaraźliwością odry. Zagrożeniem okazuje się tutaj zarówno zła wola, jak i błąd ludzki. Jak przypomina Ord, mimo że ospa prawdziwa została wytępiona w przyrodzie w 1977 r., to ostatnia ofiara zmarła w 1978 r., po tym jak wirus wydostał się z nieszczelnego laboratorium w brytyjskim Birmingham.

Longtermiści przestrzegają więc, że wkrótce możemy znaleźć się w sytuacji, w której broń masowego rażenia jest też masowo dostępna, inaczej niż to było w wypadku broni nuklearnej, która dalej zresztą stanowi istotny czynnik ryzyka egzystencjalnego. Według Orda w XX w. ludzkość wiele razy stanęła na krawędzi egzystencjalnej katastrofy, np. w czasie kryzysu kubańskiego, kiedy to John F. Kennedy oceniał ryzyko nuklearnej konfrontacji z ZSRR na co najmniej jeden do trzech. Na świecie wciąż jest ok. 14 tys. głowic nuklearnych, wiele gotowych do startu w ciągu kilku minut. Longtermiści podkreślają jednak, że dramatyczna redukcja nuklearnych arsenałów w latach 90. – w 1986 r. było na świecie blisko 70 tys. głowic – pokazuje, że postęp jest możliwy, zarówno w tej kwestii, jak w przypadku innych globalnych problemów, takich jak pandemie, wojny i zmiana klimatu.

Etyczne aspekty końca świata

GROŹBA WOJNY NUKLEARNEJ BYŁA PUNKTEM WYJŚCIA DLA PROTOLOG-termistów, takich jak astronom Carl Sagan, pisarz Jonathan Schell, filozof Derek Parfit, Bertrand Russell i Albert Einstein. Ci dwaj ostatni w 1955 r.



Obecnie, szacuje
Toby Ord, więcej niż
na zapobieganie
egzystencjalnej
katastrofie świat
wydaje na lody –
60 mld dolarów
w 2018 r.

podpisali list otwarty, dający ludzkości wybór między wojną a przetrwaniem gatunku. Na tym tle rozwinął się – może nieco dziwaczny na pierwszy rzut oka – nurt rozważań nad etycznymi aspektami końca świata. Parfit twierdził np., że wbrew pozorom wojna nuklearna, w której zginęłoby 100% ludzkości, byłaby dużo gorsza niż wojna nuklearna, w której zginęłoby 99% ludzkości. W pierwszym wypadku zginęliby nie tylko ludzie, ale też ludzkość. Byłaby to nie tylko utrata potencjalnych szczęśliwych ludzi w przyszłości, ale także koniec postępu w nauce, sztuce i walce o sprawiedliwszy świat.

Są też inne powody, żeby przemawiać się przedwczesnym końcem ludzkości. Nowozelandzka filozofka Annette Baier dołącza do długiej konserwatywnej tradycji, kiedy pisze, że jesteśmy tylko niedawnymi przybyszami do międzypokoleniowej wspólnoty i nie możemy pozwolić, aby się rozpadła. W długofalowej perspektywie obecne pokolenie jest nie tylko znikomym ułamkiem przyszłości, ale też niewielkim ułamkiem przeszłości. Szacuje się, że na świecie żyje dziś niecałe 7% wszystkich ludzi, którzy kiedykolwiek się urodzili.

Z kolei amerykański filozof Samuel Scheffler sugeruje, że przetrwanie ludzkości leży w interesie obecnego pokolenia. Powołuje się na dystopijną powieść *Ludzkie dzieci* angielskiej autorki P.D. James – zaadaptowanej później na hollywoodzki hit w reżyserii Alfonsa Cuarona – w której ludzkość stopniowo wymiera w wyniku epidemii bezpłodności. Depresja i desperacja bohaterów powieści pokazują, zdaniem Schefflera, że kluczowe dla naszego dobrostanu są projekty

– w zakresie nauki, sztuki i medycyny – których owoce mogą zebrać dopiero przyszłe pokolenia.

Wreszcie, jak pisze brytyjski astronom Martin Rees, „nasz los może mieć wręcz kosmiczne znaczenie”, bo „szanse powstania i rozwoju złożonego życia wydają się tak nikłe, że Ziemia może okazać się jedyną siedzibą świadomej inteligencji w naszej Galaktyce”. Rees odpowiada w ten sposób na paradoks postawiony przez Enrica Fermiego, jednego z ojców broni atomowej, oparty na kontraście między rozmiarem i wiekiem wszechświata a brakiem wiarygodnych śladów pozaziemskiej inteligencji.

Okazuje się, że istnieją szkoły w filozofii, którym koniec ludzkości jest obojętny. Dobrowolne wyginięcie byłoby przecież tzw. przestępstwem bez ofiar. Nie ucierpieliby na nim ani ludzie przyszli – którzy nigdy by się nie narodzili – ani ludzie obecni – którzy z własnej woli pozostaliby bezdzietni. Derek Parfit zauważa słusznie, że to kiepski argument. Działania nakierowane na przyszłość, nawet te moralnie wątpliwe, rzadko mają bowiem jakiegokolwiek ofiary. Nasze decyzje wywierają wpływ nie tylko na standard życia przyszłych ludzi, ale też na to, kim w ogóle będą. Prowadzą do kaskady zmian, przesądzających o tym, kto spotyka kogo i kiedy, a zatem: kim są ich dzieci i dzieci ich dzieci, itd. Jak wielu z nas może powiedzieć, pyta Parfit, że urodzilibyśmy się, nawet gdyby nigdy nie wynaleziono lokomotywy? Tak więc kiedy spalamy węgiel i ropę, szkodząc pokoleniu naszych prawnuków tak naprawdę sprowadzamy na świat jedną grupę ludzi zamiast innej, być może bardziej szczęśliwej. Jeżeli mają oni zadowalający standard życia, trudno ich uznać za poszkodowanych własnym istnieniem. Parfit nazywa to problemem nietożsamości. Jego wniosek jest taki, że – wbrew pozorom – możliwe są przestępstwa bez ofiar i koniec ludzkości byłby jednym z nich.

Pesymiści uważają z kolei, że wyginięcie ludzkości byłoby czymś dobrym, bo ludzka przyszłość będzie zapewne zawierać więcej cierpienia niż szczęścia. W kwestii tak delikatnej jak jakość życia dobrze zapytać samych zainteresowanych. Według psychologów prawie wszędzie na świecie zdecydowana większość pytanym odpowiada, że są szczęśliwi (niemal 93% w Polsce w 2014 r.). I ten procent będzie zapewne rósł, bo, jak przypominają longtermiści, świat staje się zdrowszy i bogatszy: w 1820 r. w skrajnej biedzie żyło 90% ludzi, dziś jest to mniej niż 10%; w 1800 r. 40% dzieci nie dożywało piątych urodzin, dziś mniej niż 4%. Oczywiście nawet te niższe wskaźniki kryją niewyobrażalną tragedię. MacAskill twierdzi jednak, że będąc na łożu śmierci, wolałby ostatecznie reinkarnację jako losowy człowiek żyjący dziś na świecie, niż nigdy się już nie obudzić.

Sprawa ma się gorzej, jeśli chodzi o inne zwierzęta. W 2018 r. ludzkość zabiła np. 69 mld kurczaków. W wyniku nowoczesnych metod hodowli ich życie było krótkie i bolesne. Oczywiście, mówi MacAskill, zwierzęta nie są sobie równe. Przykładowo kurze mózgi są mniejsze niż nasze: jeden ludzki mózg ma tyle neuronów co 400 kurczaków. Ponieważ to neurony odpowiadają za odczuwanie bólu, wygląda na to, że przeciętna kura ma mniejszy potencjał na złe stany świadomości niż przeciętny człowiek. Mimo liczebnej przewagi zwierząt hodowlanych ich dobrostan może mieć zatem mniejsze znaczenie niż dobrostan ludzi. Pozostają jeszcze jednak



❑ Kadr z filmu *Ludzkie dzieci* w reż. Alfonso Cuaróna. Ludzkość powoli wymiera w nim z powodu epidemii niepłodności. Desperacja bohaterów pokazuje, że kluczowe dla naszego dobrostanu są projekty, których owoce mogą zebrać dopiero przyszłe pokolenia fot. BEGW

zwierzęta dzikie, które przewyższają ludzkość nawet pod względem liczby neuronów. Świat przyrody potrafi być brutalny, a zdaniem MacAskilla nie ma znaczenia, czy ryba dusi się w plastikowej siatce z supermarketu czy w szczękach rekina żarłacza. Z uwagi na cierpienie zwierząt dzikich sam MacAskill nie jest pewien, czy jakość życia przeciętnej żywej istoty na Ziemi jest faktycznie pozytywna. Nie wie zatem, czy dzisiejszy świat jako całość ma wartość pozytywną ani jak ta wartość może zmienić się w przyszłości.

Korekta kursu

MOŻEMY BYĆ LONGTERMISTAMI, NAWET JEŚLI ŚWIAT WYDAJE NAM SIĘ ZŁY, a koniec ludzkości obojętny. Tylko że wtedy naszym priorytetem powinna

być nie tyle dłuższa, ile lepsza przyszłość ludzkości; jakość, nie zaś ilość. Wielki wpływ mają tutaj według MacAskilla normy społeczne i wartości moralne. Bardziej niż egzystencjalne bezpieczeństwo interesuje go więc cywilizacyjna korekta kursu.

MacAskill sądzi, że normy i wartości są nie tylko wpływowe, ale też długotrwałe i w dużym



Organizacja 80 000 Hours – tyle godzin w życiu przeciętny człowiek spędza w pracy – doradza, jak nasza kariera zawodowa może zmienić świat na lepsze: czy lepiej zostać lekarzem, działaczem NGO, a może menedżerem, który swój dochód przekaze na „efektywny altruizm”

stopniu przygodne, to znaczy łatwo mogłyby być inne. Podobnie uważał John Adams, jeden z ojców założycieli Stanów Zjednoczonych, który pisał, że „instytucje teraz budowane w Ameryce nie zatrą się całkowicie przez tysiące lat. Jest więc sprawą najwyższej wagi, żeby nadać im prawidłowy kierunek. Jeśli wyruszą w złą stronę, to nigdy, chyba że przypadkiem, nie powrócą już do

słusznej ścieżki”. Jak na razie miał rację: amerykańska konstytucja dalej obowiązuje i trudniej ją zmienić niż w 1800 r. MacAskill przygląda się też różnym ruchom religijnym i społecznym, np. kampanii, która doprowadziła do zniesienia niewolnictwa w imperium brytyjskim.

Mimo wielu sukcesów historia jest przynębiającym pasmem moralnych horrorów. Małe są szanse, że jesteśmy pierwszym pokoleniem, które dotarło do moralnej prawdy. MacAskilla martwi przeto ryzyko „moralnego zamrożenia”, kiedy jedno pokolenie narzuca swoje wartości kolejnym, zatrzymując postęp moralny.

Są staroświeckie sposoby na moralne zamrożenie, jak np. globalny totalitaryzm, ale także bardziej subtelne, jak choćby stopniowa homo-



genizacja światowej kultury. Futurystyczne technologie również mogą zwiększyć ryzyko zamrożenia. MacAskill jako jej przykład podaje zaawansowaną sztuczną inteligencję, która znacznie przewyższa człowieka w prawie wszystkich zadaniach intelektualnych. W ankiecie z 2016 r. eksperci w tej dziedzinie widzieli 10% szansy na zaawansowaną sztuczną inteligencję przed rokiem 2025 i 50% szansy przed rokiem 2061. Inteligentne maszyny byłyby szybsze i tańsze w utrzymaniu niż biologiczne maszyny, takie jak my, a także łatwiejsze w powielaniu, niepodatne na zapomnienie i biologiczny rozkład. Ekonomisci szacują więc, że nadejście inteligentnych maszyn może gwałtownie przyspieszyć tempo rozwoju technologicznego, a zatem rozwoju gospodarczego. Kto pierwszy zbuduje inteligentne maszyny, może mieć decydującą przewagę gospodarczą nad resztą świata i zarazem opcję zamrożenia swoich wartości na długo.

Inni longtermiści, przykład wśród nich Bostrom i Ord, skupiają się na możliwości przejścia władzy przez inteligentne maszyny. Nie trzeba do tego maszyn wrogo nastawionych. Ludzkie umysły zajmują zapewne tylko mały wycinek przestrzeni wszystkich możliwych umysłów. Inteligentne maszyny mogą więc być obce, nieludzkie. Wraz z postępem sztucznej inteligencji coraz więcej obszarów świata może być kształtowanych w sposób, który z naszego punktu widzenia wydaje się niezrozumiały i bezwartościowy, podobnie jak świat ludzi może wyglądać z punktu widzenia orangutanów.

Wobec ryzyka moralnego zamrożenia longtermiści chcieliby, żeby po obecnej epoce egzystencjalnych zagrożeń nadeszła epoka Długiej Refleksji, tzn. według MacAskilla „stabilny stan świata, w którym ludzkość jest chroniona od katastrofy i ma możliwość refleksji i debaty nad naturą dobrego życia, a także wypracowania zasad najlepszego społeczeństwa”. Byłoby to coś w rodzaju Oświecenia czy Renesansu, niekoniecznie transformacja globu w wydział filozofii. Tak jak warto przeznaczyć pięć minut, żeby zaplanować dwugodzinny obiad, tak też, sugeruje MacAskill, warto byłoby debatować tysiące lat, co ludzkość powinna zrobić ze swoją astro-nomicznie wielką przyszłością. Wynikiem Długiej Refleksji mogłaby być według filozofa konkretna utopia, zbiorowe samobójstwo gatunku albo kompromis między moralnymi światopoglądami, których różnic nie da się rozstrzygnąć racjonalnie.

Oczywiście epoka Długiej Refleksji w rzeczywistości pewnie nigdy nie nadejdzie. Jako bardziej realistyczne środki MacAskill proponuje więc zmniejszanie ryzyka egzystencjalnej katastrofy i odwołanie okazji do moralnego zamrożenia. Sugeruje też bardziej ambitną wizję społeczeństwa promującego polityczne, kulturowe i osobiste eksperymenty, np. w postaci specjalnych stref politycznych, gdzie obywatele mieliby wybór, czy zamieszkać w marksistowskiej komunie czy w kapitalistycznej metropolii.

Antylongtermizm

U PODSTAW LONGTERMIZMU LEŻĄ ZATEM WIARYGODNE EMPIRYCZNE PRZESŁANKI dotyczące potencjalnej skali przyszłości i przekonujące moralne idee dotyczące bezstronności w odniesieniu do czasu i przestrzeni. Małe są więc szanse, że longtermizm okaże się tylko przelotną intelektualną modą.

Prowadzi on jednak do nowych problemów w teorii i poważnych obaw w praktyce.

Niektórym longtermizm wydaje się niebezpiecznie podobny do tragicznych ruchów utopijnych XX w., których zagrożenia ilustruje przemowa starego bolszewika z powieści *Ciemność w południe* Arthura Koestlera: „Nigdy w historii tak nieliczne ręce nie miały tak wielkiej władzy nad przyszłymi losami ludzkości. Każda fałszywa koncepcja, jaką się posługujemy, jest zbrodnią popełnianą na przyszłych pokoleniach. Dlatego musimy fałszywe koncepcje karać, jak inni karzą zbrodnie: śmiercią” (tłum. T. Terlecki). Niektórzy krytycy longtermizmu, jak Peter Singer i Émile Torres, sugerują, że astronomiczne stawki mogą w końcu skusić longtermistów, żeby uciec się do oszustwa czy nawet przemocy w zmierzaniu do osiągnięcia swoich celów. Są jednak powody do optymizmu. Inaczej niż np. marksiści longtermiści nie oferują szablonu idealnego społeczeństwa, podkreślając raczej ryzyko szablunowych rozwiązań.

Inni obawiają się natomiast, że popularność longtermizmu sprawi, iż naglące problemy teraźniejszości zostaną pozostawione same sobie. Longtermiści uspokajają jednak, że kształt ludzkiej przyszłości jest problemem ogromnym i zaniebzanym, ale też bardzo trudnym. Nie będzie możliwe, by na egzystencjalne ryzyko produktywnie wydać więcej niż parę procent światowego dochodu. Obecnie, szacuje Toby Ord, więcej niż na zapobieganie egzystencjalnej katastrofie świat wydaje na lody – 60 mld dolarów w 2018 r.

Jeszcze inni zarzucają longtermistom fanatyzm i moralne ryzy-

kanctwo, bo są gotowi poświęcić konkretny sukces na rzecz astronomicznie małych szans gigantycznej wygranej. MacAskill i Hilary Greaves wyliczają np., że opłacałoby się wydać 20 mld dolarów, żeby ryzyko katastrofalnego zderzenia z asteroidą zmniejszyć o jedną milionową, bo nawet niewielka szansa sukcesu pomnożona przez ogromną wartość ocalenia ludzkości daje olbrzymią spodziewaną wartość dodatnią. Longtermiści odpowiadają, że całkowite ryzyko egzystencjalnych katastrof nie jest wcale małe: według MacAskilla wynosi jeden do stu. Jest to więcej niż roczne ryzyko śmierci w wypadku samochodowym, a więc skala ryzyka, którą bierzemy poważnie, gdy zapinamy pasy.

Przy inwestycjach takich jak system obrony planetarnej opłacalność nie będzie znana przez lata, może nawet miliony lat. Poza tym probabilistyczne prognozy ciężko zweryfikować. Jak stwierdzić po fakcie, czy ryzyko katastrofy wynosiło jeden do trzech czy jeden do dwóch? W związku z tym longtermiści muszą na ogół opierać się na informacjach spekulatywnych i niepewnych. Jest to odejście od praktyki efektywnego altruizmu, gdzie dotychczas ideałem były eksperymenty kliniczno-kontrolne, testujące skuteczność charytatywnych interwencji, podobnie jak testuje się skuteczność nowych leków. Niektórzy nadal wolą działać na podstawie solidniejszych dowodów, z możliwością uczenia się na błędach.

Są też tacy, którym longtermizm wydaje się przytłaczający i wyobcowany. Trudno o inną emocjonalną reakcję na perspektywę eonów historii i bilionów ludzi,

szczególnie jeśli – z powodu ewolucji czy bioinżynierii – będą się oni od nas dużo różnili. Niektórzy longtermiści, jak choćby Bostrom, związani są z transhumanizmem, który postuluje, że człowiek nie jest ostatnim przystankiem ewolucji i może, a nawet powinien, być zastąpiony lepszą formą życia. Inni popierają kosmiczny ekspansjonizm, który głosi, że światowym priorytetem powinno być zasiedlenie kosmosu. Nie są to jednak elementy dla longtermizmu niezbędne. MacAskill przestrzega np., że przedwczesne wyruszenie w kosmos może zamrozić niedoskonałe normy społeczne. Ord podkreśla natomiast, że naszym priorytetem powinno być egzystencjalne bezpieczeństwo, a kwestię transformacji naszego gatunku możemy zostawić na później.

Nasze miejsce w historii może też być powodem do swego rodzaju egzystencjalnej nadziei, podobnej do tej, którą brytyjski filozof John Stuart Mill wyraża w swoim dzienniku z 1854 r.: „Lekarstwa na wszystkie nasze choroby zostaną odkryte długo po naszej śmierci, a świat stanie się miejscem zdatnym do zamieszkania dzięki staraniom tych, których większość umrze, zanim to nastąpi. Miejmy nadzieję, że ludzie, którzy będą kiedyś na świecie, ze współczuciem wspomną swoich znanych i nieznanych dobroczyńców”. — 

Korzystałem m.in. z: N. Bostrom, *Existential Risk Prevention as Global Priority*, „Global Policy” 2013, vol. 4, issue 1, s. 15–31; W. MacAskill, *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference*, London 2016; W. MacAskill, *What We Owe the Future: A Million-Year View*, London 2022; T. Ord, *The Precipice: Existential Risk and the Future of Humanity*, London 2020; P. Singer, *The Hinge of History*, „Project Syndicate”, <https://www.project-syndicate.org/commentary/ethical-implications-of-focusing-on-extinction-risk-by-peter-singer-2021-10> (dostęp: 5 grudnia 2022 r.)

Kacper Kowalczyk

Dr filozofii, specjalizujący się w etyce, absolwent Cambridge i Oksfordu, pracownik naukowy na University College London, obecnie wspierany przez grant Longview Philanthropy

Człowiek, gatunek międzyplanetarny

Nick Bostrom
w rozmowie
z Kacprem Kowalczykiem

Nie jesteśmy skazani na wieczną stagnację życia na Ziemi. Możemy wykorzystać kosmiczne zasoby, by w pełni zrealizować ludzki potencjał

Z *Pisał Pan kiedyś, że we wszechświecie marnuje się mnóstwo energii, bo wygląda na to, że żadna cywilizacja w pełni nie eksploatuje dostępnych gwiazd. Nazywa Pan to „astronomicznym marnotrawstwem” i sugeruje, że ludzkość mogłaby w przyszłości temu zaradzić.*

Zgadza się. Co sekundę marnują się astronomiczne ilości energii potencjalnej i innych kosmicznych zasobów. Uważam, że warto byłoby zrobić z tym coś dobrego.

Z *Twierdzi Pan jednak, że wyruszenie w gwiazdy i kolonizacja innych planet nie powinny być dla nas priorytetem. Dlaczego?*

Mamy na razie jeszcze ważniejsze zadanie, a mianowicie zagwarantować, że faktycznie do tych gwiazd dotrzemy. Ze względu na samą długowieczność takich astronomicznych obiektów mniejsze ma znaczenie, czy stanie się to za 100 czy za 1000 lat, niż fakt, że w końcu się to uda. Uważam więc, że o ile wychodzimy od przyjętej wizji wszechświata i pewnych popularnych przesłanek moralnych, to powinniśmy przyjąć zasadę Maxipok, która głosi,

że trzeba maksymalizować prawdopodobieństwo OK rezultatu, a więc uniknąć egzystencjalnej katastrofy, w końcu skolonizować wszechświat i wykorzystać dostępną energię w dobrym celu.

Z *A więc co myśli Pan o ludziach takich jak Elon Musk, którzy chcieliby szybszej kolonizacji kosmosu, a przynajmniej naszego Układu Słonecznego?*

Elon uważa pewnie, że jest to sposób na obniżenie ryzyka egzystencjalnego. Jeśli staniemy się gatunkiem międzyplanetarnym, będzie mniej prawdopodobne, że jakaś katastrofa położy kres całej ludzkiej cywilizacji. A więc nie chodzi mu o to, że powierzchnia Marsa się marnuje, ale że obecnie jako ludzkość stawiamy wszystko na jedną kartę, a byłoby dobrze postawić chociaż na dwie.

Z *Na przykład poprzez budowę kolonii na Marsie.*

Właśnie. Uważam jednak, że mamy pilniejsze zadania. Przypuszczam, że kluczowe egzystencjalne zagrożenia pojawiają się, zanim będziemy mieli okazję na założenie samo-

wystarczalnych kolonii na Marsie. Co więcej, nie sądzę, żeby taka kolonia miała pomóc nam uporać się np. z zagrożeniami, które niesie sztuczna inteligencja. Inteligentne roboty mogłyby polecieć na Marsa równie dobrze jak my. Kolonia na innej planecie dałaby nam większe szanse na przetrwanie niektórych egzystencjalnych zagrożeń, np. związanych z asteroidami i kometami, ale wielu z nich łatwiej można by zaradzić tutaj, na Ziemi, przez tańsze rozwiązania, choćby budowanie większej liczby umocnionych schronów.

Z *Ostatnio sugerował Pan, że postęp technologiczny jest podobny do losowania z urny, zawierającej korzystne kule białe, szare kule o mieszanych skutkach, ale także apokaliptyczne kule czarne. Czemu mielibyśmy przyjmując taką hipotezę?*

Czarna kula to technologia lub idea, której domyślnym skutkiem jest zniszczenie cywilizacji, jaka ją odkryje. Nie wydaje się wykluczone, że wśród wszystkich możliwych idei i technologii jest przynajmniej jedna czarna kula.

W przeszłości mogło się wydawać, że czarną kulą okaże się broń jądrowa. Szczęśliwym zbiegiem okoliczności brakowało jej jednej szczególnej cechy – łatwości wytwarzania. Broń jądrowa wymaga wysoko wzbogaconego uranu lub plutonu. Nie ma łatwego sposobu, żeby uwolnić energię atomu na dużą skalę. Możemy sobie wyobrazić, co by było, gdyby broń jądrową można było zrobić tak łatwo jak czekoladowy pudding z mikrofalówki. Demokratyzacja dostępności do broni masowej

zagłady byłaby najpewniej końcem ludzkiej cywilizacji. Przez większość ludzkiej historii nie mieliśmy o takim zagrożeniu żadnego pojęcia. Czy był to ostatni taki przypadek? Wygląda na to, że sięgając do urny biotechnologii, nanotechnologii i sztucznej inteligencji, możemy mieć pecha i trafić na technologię, która sprawi, że narzędzia masowej zagłady staną się łatwo dostępne.

Z *W Pana książce na ten temat sztuczna inteligencja okazuje się potencjalną czarną kulą, bo istnieje ryzyko, że maszyny w końcu osiągną superinteligencję – a więc poziom rozwoju dużo wyższy od naszego – lecz nie uda nam się im przekazać ludzkiego systemu wartości. Czy dalej Pan tak myśli?*

Podstawowe idee w tej książce wciąż wydają mi się słuszne. Przy czym oprócz scenariuszy, w których pojedyncza sztuczna inteligencja staje się bardzo potężna, ale pozostaje nieskalibrowana z naszymi wartościami, pisałem też o scenariuszach wielobiegunowych, gdzie złe skutki wynikają z konkurencji między różnymi sztucznymi inteligencjami lub inteligentnymi systemami.

Gdy wydałem *Superinteligencję* w 2014 r., właściwie nikt nie pracował nad problemem skalibrowania sztucznej inteligencji. Od tego czasu nastąpił duży wzrost liczby badaczy i ośrodków. Kluczowe laboratoria pracujące nad sztuczną inteligencją, takie jak DeepMind i OpenAI, stworzyły zespoły ds. bezpieczeństwa sztucznej inteligencji, podobnie jak niektóre ośrodki akademickie, takie jak nasz oksfordzki Future of Humanity Institute.

Z *Był to też czas dużego postępu technicznego w dziedzinie samej AI.*

Oczywiście. Postęp jest dużo szybszy, niż się tego spodziewaliśmy, a rewolucja uczenia głębokiego idzie pełną parą do przodu. Daje nam to jaśniejszy obraz potencjalnych ścieżek do superinteligencji, a zarazem jaśniejszy obraz architektury, którą przyjmą wczesne superinteligentne systemy. Żeby lepiej je sobie wyobrazić, możemy pomyśleć o większych wersjach systemów obecnych. Daje to też konkretny cel badaczom, którzy pracują nad problemem kalibracji sztucznej inteligencji.

Z *Czy ogólnie rzecz biorąc, jest Pan bardziej optymistyczny, jeśli chodzi o sztuczną inteligencję niż kiedyś?*

Raczej nie. Jestem bardziej optymistyczny, jeśli chodzi o ilość pracy intelektualnej włożonej w rozwiązywanie tego problemu. Jest jej więcej, niż było, i więcej, niż można się było spodziewać w tym okresie. Uważam też, że przejście do świata sztucznej inteligencji nastąpi prędzej czy później. Byłoby błędem starać się temu zapobiec albo nawet za długo z tym zwlekać. Powinniśmy wziąć się w garść, dopóki możemy, jednak w którymś momencie trzeba będzie po prostu trzymać kciuki i liczyć na dobry obrót sprawy. Wiąże się z tym pewne ryzyko, którego skalę trudno jest dziś oszacować.

Z *Przyjmijmy, że jest w urnie technologii czarna kula. Wtedy albo ulegamy zagładzie, albo tworzymy instytucje, które – jeśli będą miały wystarczającą*



władzę, żeby kontrolować proces losowania – niosą ze sobą inne zagrożenia, np. ryzyko permanentnego totalitaryzmu. Wygląda więc na to, że nasz los jest przesądzony. Przyszłość ludzkości jest albo krótka, albo negatywna.

Dopóki jest niewiedza, jest też nadzieja. Nie wiemy, czy czarna kula istnieje.

Co więcej, różne środki stabilizacyjne, jak powszechny nadzór w czasie rzeczywistym lub silniejszy porządek międzynarodowy, niosą zarówno zagrożenia, jak i możliwe korzyści, jak np. dostępność bardziej niezawodnych środków do rozwiązywania problemów z koordynacją w globalnie istotnych sprawach. Można sobie wyobrazić, że takie środki stabilizacyjne mogłyby doprowadzić do utopii raczej niż do dystopii; że dałyby nam nie tylko ochronę przed zagrożeniami, ale także światowy pokój. Być może ten proces byłby w pewnym sensie podobny do nastania epoki państw. To prawda, że istnieją dystopijne państwa totalitarne, lecz przecież nie wszystkie państwa są całko-

wicie zepsute. W ostatecznym rachunku rozpowszechnienie form państwowości działało być może więcej dobrego niż złego poprzez tłumienie lokalnych konfliktów, opamiętanie społecznego nieładu i przestępczości, spisanie praw, umożliwienie wzrostu gospodarczego.

Mimo że system nadzoru mógłby zostać wykorzystany przez dyktatora lub totalitarny reżim do eliminacji wszelkich form sprzeciwu, to mógłby też służyć bardziej dobroczynnemu rządowi do eliminacji wielu form przestępczości bez konieczności potężnej formacji policyjnej i długich wyroków więziennych. Gdyby istniała gwarancja, że przestępca zostanie zawsze szybko złapany i skazany, dałoby się też ograniczyć inne wydatki związane z walką z przestępczością.

■ *Jak z długofalowej perspektywy wyglądają bieżące konflikty na świecie?*

Myślę, że naszym ogólnym celem powinno być zwiększenie szansy na pokojowy, kooperatywny porządek na świecie, szczególnie jeśli chodzi o największe i najważniejsze punkty sporne między światowymi mocarstwami. Brak światowej koordynacji – czy to w postaci aktywnego konfliktu, czy zbiorowej nieumiejętności zapobiegania różnego rodzaju wypadkom, wyścigom technologicznym i zbrojeniowym – stanowi źródło innych bezpośrednich zagrożeń egzystencjalnych.

Należy więc postawić sobie pytanie, jak najlepiej promować światowy pokój i harmonię; trudniej jednak oczywiście o dobrą odpowiedź. Wiemy przecież, że



przyszłość wyglądałaby dużo lepiej, gdyby światowe mocarstwa potrafiły się dogadać, rozwiązując światowe problemy, czy będzie to globalne ocieplenie, rozprzestrzenianie się broni nuklearnej czy wyzwania związane ze sztuczną inteligencją. Byłoby dużo łatwiej sobie z nimi poradzić, gdyby np. Stany Zjednoczone i Chiny pozostawały w przyjacielskich stosunkach. Niektórzy krótkowzroczni politycy woleliby jednak zapomnieć o nadrzędnym znaczeniu budowania dobrych relacji na rzecz krótkoterminowych korzyści.

Przejście do
świata sztucznej
inteligencji nastąpi
prędzej czy później.
Wiąże się z tym
ryzyko, którego
skalę trudno jest
dziś oszacować