Szef od AI w Microsofcie: Powstrzymanie tej technologii nie wydaje się możliwe, ale...

Magazyn Wyborczej, 07.06.2024

https://wyborcza.pl/magazyn/7,124059,31032343,szef-od-ai-w-microsofcie-powstrzymanie-tej-technologii-nie.html

Gra planszowa Go była większym wyzwaniem dla sztucznej inteligencji niż szachy. Ale już nie jest. Program odkrył strategie, o których ludzie nie mieli pojęcia. Powyższą ilustrację do tekstu sporządził program Copilot. Sztuczna Inteligencja nie poradziła sobie z szachownicą, więc może jest dla nas nadzieja (FOt. JAKUB PIOTROWSKI / COPILOT)

Mustafa Suleyman doszedł do wniosku, że problemy, przed jakimi stoi nasza cywilizacja, przerastają rozum człowieka i trzeba szukać pomocy w mocach nadludzkich. Czyli sztucznej inteligencji.

19 marca Mustafa Suleyman ogłosił w serwisie X:

„Dzisiaj dołączam do Microsoftu, będę kierować wszystkimi konsumenckimi badaniami i produktami związanymi ze sztuczną inteligencją, włącznie z Copilotem, Bingiem i Edgem."

Na początku maja zapowiedział, że przeprowadzi Microsoft z ery komputerów osobistych, których ta firma była pionierem, do epoki „inteligencji osobistej". A zacznie się to od największej aktualizacji systemu Windows w czterdziestoletniej historii firmy. Ten najpopularniejszy system operacyjny na świecie ma być teraz zespolony ze sztuczną inteligencją, służyć każdemu użytkownikowi jako osobisty cyfrowy asystent.

Na konferencji TED w Vancouver roztaczał wizję, która jeszcze kilka miesięcy temu byłaby uznana za szaloną i utopijną:

„Wszyscy wkrótce będziemy mieli osobistych asystentów AI o nieograniczonej wiedzy, niemal doskonałym IQ, a także wyjątkowym EQ [emocjonalna inteligencja]. Będą mili, wspierający i empatyczni. Będziesz nosił w kieszeni spersonalizowanego nauczyciela, lekarza, stratega biznesowego i trenera – 24 godziny na dobę.

Będą towarzyszami, powiernikami, współpracownikami, przyjaciółmi i partnerami, tak różnorodni i wyjątkowi, jak my wszyscy. Będą nie tylko dla ludzi - choć może to zabrzmieć dziwnie, każda organizacja, od małych firm, przez organizacje non-profit, po rządy krajowe, każde miasto, budynek i obiekt będą reprezentowane przez unikalną interaktywną osobowość.

Zamówią jedzenie na wynos i poprowadzą elektrownię. Będą wchodzić w interakcje z nami i, oczywiście, ze sobą nawzajem. Będą mówić każdym językiem, odczytywać każdy wzór danych z czujników, kamer, mikrofonów, analizować strumienie informacji z całego świata, znacznie przewyższających to, co każdy z nas mógłby skonsumować w ciągu tysiąca wcieleń."

W styczniu tego roku w Davos podczas dorocznego spotkania Światowego Forum Ekonomicznego Suleyman wyszedł na scenę wraz z dyrektor zarządzającą Międzynarodowego Funduszu Walutowego Kristaliną Georgiewą i ministrem spraw zagranicznych Ukrainy Dmytro Kulebą, aby dyskutować o globalnych implikacjach i zagrożeniach związanych z rozwojem sztucznej inteligencji.

- Ta technologia może być głównym motorem szerzenia dobrobytu i pomyślności w nadchodzących latach. Ogromną pomocą w stawianiu czoła cywilizacyjnym wyzwaniom - od opieki zdrowotnej i edukacji po zmiany klimatyczne. Dzięki AI będziemy mieli znacznie większe szanse, by im sprostać - powtarza w mailu do „Wyborczej".

W połowie maja ukazało się polskie tłumaczenie jego bestsellerowej książki „Nadchodząca fala. Sztuczna inteligencja, władza i najważniejszy dylemat ludzkości w XXI wieku" (Szczeliny, imprint Wydawnictwa Otwartego, tłum. Justyn Hunia).

Kończący w tym roku 40. rok życia brytyjski menedżer - syn urodzonego w Syrii londyńskiego taksówkarza i brytyjskiej pielęgniarki - znalazł się właśnie u szczytu kariery.

Wyniosła go ta sama fala, o której pisze w swojej książce. Globalna ekspansja i masowe rozprzestrzenianie się dwóch nowych technologii:

AI i biologii syntetycznej, które wkrótce dadzą ludzkości potężne - niemal boskie - możliwości albo... przyniosą katastrofalne, dystopijne skutki. A może obie te rzeczy na raz.

Mustafa Suleyman szuka pomocy w mocach nadludzkich

Co ciekawe, Mustafa Suleyman nie jest programistą, informatykiem, matematykiem czy neurokognitywistą. Zanim trafił do branży IT, był społecznikiem i aktywistą, a po trosze biznesmenem. Studiował filozofię i teologię na Oksfordzie, ale wytrzymał tam tylko rok, zrezygnował z nauki, bo - jak dziś zapewnia - chciał zająć się czymś bardziej praktycznym. Był jednym z tych 19-latków, który marzą o tym, aby zmienić świat, tu i teraz, a nie filozofować i rozdzielać włos na czworo.

Wraz ze swoim przyjacielem ze studiów założył telefoniczną infolinię dla młodzieży muzułmańskiej, która wkrótce stała się jednym z największych przedsięwzięć w zakresie wsparcia zdrowia psychicznego dla muzułmanów w Wielkiej Brytanii. Wynajął wózek

inwalidzki i z kolegami jeździli nim po całym Londynie, wydając na koniec przewodnik po stolicy Wielkiej Brytanii dla niepełnosprawnych.

Ale po kilku latach doszedł do wniosku, że praca w organizacjach charytatywnych daje wprawdzie satysfakcję, ale małą szansę na poważny wpływ i zmianę.

Potem jeszcze krótko pracował jako specjalista ds. praw człowieka u Kena Livingstone'a, burmistrza Londynu, ale uznał, że ten urząd nie jest gotowy na radykalne działania. Założył Reos Partners – firmę konsultingową, mającą rozwiązywać konflikty społeczne i wspierać globalne zmiany systemowe w takich kwestiach, jak produkcja żywności, odpady czy różnorodność. Pracował jako negocjator, mediator i facylitator dla m.in. Organizacji Narodów Zjednoczonych, rządu holenderskiego, funduszu WWF czy Shella.

Opuścił jednak Reos Partners na początku 2010 r. po tym, kiedy konsultacje i negocjacje klimatyczne ONZ w Kopenhadze, w których przez rok brał udział, poniosły spektakularną porażkę. Na tym szczycie klimatycznym nie doszło do podpisania porozumienia.

Sfrustrowany Suleyman doszedł wtedy do wniosku, że problemy, przed jakimi stoi nasza cywilizacja, przerastają rozum człowieka i trzeba szukać pomocy w mocach nadludzkich. Czyli sztucznej inteligencji.

- Pracowałem w doradztwie i polityce i coraz częściej widziałem, jak tradycyjne instytucje zawodzą, próbując stawić czoła największym naszym wyzwaniom. Jednocześnie obserwowałem, jak nowe technologie z łatwością i szybkością docierają do setek milionów użytkowników - mówi nam Suleyman. - Chciałem wykorzystać tę wielką skalę rozprzestrzeniania się i rozmiar wpływu. Dlatego postawiłem na sztuczną inteligencją, choć w tamtym czasie było to niemodne i wielu uważało, że wręcz niemożliwe.

Suleyman, Hassabis i Shane Legg zakładają Deep Mind

Nagły zwrot w kierunku AI był nie tylko skutkiem frustracji i poczucia bezradności. Miał w tym udział także przypadek, ale z gatunku tych, które wykorzystują najlepsi.

Los bowiem sprawił, że w dzieciństwie poznał Demisa Hassabisa, starszego brata jego najlepszego przyjaciela ze szkoły.

Hassabis jest geniuszem, już jako dziecko grał w szachy na poziomie arcymistrza, sam nauczył się programować i projektować popularne gry wideo, dzięki czemu zarobił na czesne w Cambridge, gdzie przez trzy lata studiował informatykę. To mu nie wystarczyło, bo w projektowanych grach strategicznych i symulacyjnych jako jeden z pierwszych stosował sieci neuronowe i algorytmy sztucznej inteligencji. Skończył więc jeszcze kurs neurobiologii poznawczej w University College London, gdzie szukał inspiracji, jak ulepszyć AI.

Suleyman i Hassabis jeszcze w czasach szkolnych sporo ze sobą rozmawiali. - Obaj chcieliśmy zmienić świat. On już wtedy przekonywał, że musimy zbudować te wielkie komputerowe symulacje, które pewnego dnia będą modelować całą złożoną dynamikę

naszych systemów finansowych i rozwiążą wszystkie najtrudniejsze problemy społeczne. Ja byłem zdania, że trzeba ruszać w świat i już teraz próbować dokonać realnej zmiany - mówi Suleyman.

Jednak po twardym zderzeniu z realnym światem, wrócił do pomysłu Hassabisa.

Komputerowy geniusz postawił sobie za cel odkrycie św. Graala dziedziny AI - a więc zbudowanie ogólnej sztucznej inteligencji (AGI, artificial general inteligence). Takiej, która miałby lepsze zdolności poznawcze niż najbardziej inteligentni ludzie.

Maszyny zdolnej do uczenia się, planowania przyszłości i być może także obdarzonej świadomością. - Chcę ją rozwinąć tak, by rozwiązała wszystkie nasze problemy, takie jak głód, choroby, bezdomność czy globalne ocieplenie. To będzie program Apollo XXI wieku – mówił Hassabis w "Wired".

- Klimat, ekonomia, choroby to niezwykle skomplikowane, wzajemnie oddziałujące systemy. Ludzie po prostu nie są w stanie przeanalizować wszystkich danych i nadać im sensu. Najpewniej istnieją fizyczne granice tego, co mogą zrozumieć mózgi ekspertów. Sztuczna inteligencja ich wspomoże, otworzy nową epokę odkryć - sekundował mu Suleyman.

Suleyman, Hassabis i nowozelandczyk Shane Legg (jeden z pionierów prac nad AGI, którego Hassabis spotkał na University College London) pod koniec roku 2009 założyli w Londynie start up o nazwie Deep Mind, którego celem było ni mniej, ni więcej, tylko stworzenie sztucznego mózgu.

- Wszystko wskazuje na to, że nasz mózg jest biologiczną maszyną liczącą i możemy odtworzyć jego działanie. W ciągu dziesięciu, może 15 lat stworzymy maszynę myślącą na poziomie noworodka. W pełni funkcjonalny system to tylko kwestia czasu. Mamy wokół ponad 7 mld działających prototypów, czyli nas samych - zapowiadał Hassabis.

Google płaci 400 mln funtów za niedochodowy start-up

W tamtym czasie taka deklaracja była czystym science-fiction. Nawet eksperci w dziedzinie AI pukali się w głowę. Tym niemniej trójka marzycieli przekonała poważnych inwestorów, aby włożyli spore pieniądze. W Deep Mind zainwestowali m.in. Elon Musk (twórca SpaceX czy Tesli), Peter Thiel (jeden z założycieli Facebooka i Jaan Tallinn (współzałożyciel Skype).

W ciągu kilku lat głębokie sieci neuronowe Deep Mind - logiczno-matematyczne struktury, które naśladują działanie ludzkiego mózgu - osiągnęły zdumiewające rezultaty. Kluczem do sukcesu okazały tzw. algorytmy uczące się przez wzmacnianie (ang. Reinforcement Learning).

Proces programowania takiej sieci przypomina uczenie się żywego mózgu. Do tego stopnia, że do opisu "uczenia maszynowego" używana jest terminologia zaczerpnięta z psychologii – istnieją na przykład pojęcia sprzężenia zwrotnego i nagrody. Połączenia między

sztucznymi neuronami są wzmacniane (lub osłabiane) przez przykłady, trening i zbierane doświadczenie.

- Większość dzisiejszych systemów sztucznej inteligencji polega na wstępnym programowaniu maszyn. Naszym pomysłem jest zaprogramowanie ich tak, aby mogły uczyć się samodzielnie. Jest to o wiele potężniejsze, w ten sposób uczą się systemy biologiczne - wyjaśniał Hassabis.

Sztuczna inteligencja Deep Mind zaczyna więc od minimalnego zestawu umiejętności i początkowo działa na oślep, metodą prób i błędów. Ale błyskawicznie uczy się strategii postępowania, która przynosi najlepsze efekty. Dochodzi do poziomu, na którym bije profesjonalistów i ekspertów.

Pierwszy algorytm, zwany Deep Q-network, nauczył się grać w 49 stare gry wideo, znane z konsol Atari 2600 z końca lat 70. zeszłego wieku. Brzmi to mało spektakularnie, dopóki nie dodamy, że w algorytmie nie zostały zakodowane reguły tych gier. Program sam się ich domyślił - miał do dyspozycji tylko obraz z konsoli, informację o wyniku oraz funkcjach poszczególnych ruchów joystickiem.

Suleyman wspomina, że obserwowanie na żywo postępów algorytmu robiło piorunujące wrażenie, nawet na samych jego twórcach. Zapowiadało to rewolucję.

To częściowo wyjaśnia, dlaczego w 2014 r. firmę przejął Google - płacąc gigantyczną kwotę 400 mln funtów, choć start-up nie miał jeszcze żadnego źródła przychodów - nie skomercjalizował algorytmu, nie wypuścił na rynek żadnego produktu.

Program wygrywa w "StarCrafta II"

Drugim powodem tego wówczas największego europejskiego przejęcia był Suleyman. Jeśli Hassabisa można nazwać mózgiem przedsięwzięcia, Suleyman był jego motorem. Odpowiadał za logistykę, komunikację i biznes. Wciąż obmyślał, jak zastosować AI do rozwiązywania praktycznych problemów. Wydawało się, że pod skrzydłami potężnego Google'a stanie się to dużo szybciej.

Przejęcie Deep Mind przez Google'a zelektryzowało zaś innych globalnych graczy w dziedzinie AI. Nagle wszyscy zaczęli mówić o sztucznej inteligencji, jedni z nadzieją, inni ze strachem.

Ten właśnie moment może być w przyszłości przyjmowany za początek globalnej rywalizacji w celu zbudowania sztucznego mózgu. Jeśli oczywiście będzie jakaś przyszłość.

Już pod patronatem Google'a Deep Mind stworzył program AlphaGo, który pokonał legendarnego koreańskiego mistrza Lee Se-dola w starej dalekowschodniej grze planszowej Go. Ta gra była jednym z największych wyzwań dla sztucznej inteligencji, bo w szachy czy warcaby już od dawna człowiek przegrywał.

Kolejna wersja AlphaZero potrafiła uczyć się sam "od zera", bez sięgania po wiedzę zdobytą i opracowaną przez ludzi. Ten program grał sam ze sobą i w ciągu ledwie 70 dni

zdobył arcymistrzowskie umiejętności. Odkrył strategie, do których mistrzowie gry w Go dochodzili przez kilkanaście wieków, a także takie, o których ludzie nie mieli dotąd pojęcia.

Kolejną grą, w której Deep Mind zakończył supremację ludzi, była "StarCraft II". Kultowa komputerowa gra strategiczna czasu rzeczywistego, która położyła podwaliny pod lukratywny e-sport. Jej fabuła dotyczy konfliktu pomiędzy trzema galaktycznymi rasami: terranami (wywodzą się od ludzi wygnanych z Ziemi), zergami (przypominają szybko mutujące roje owadów) oraz protosami (to wojownicy zaawansowani technologicznie i obdarzeni mocami paranormalnymi). Gracz wciela się w jedną z tych ras i - jak to w grze - stara się wyeliminować wszystkich przeciwników

Środowisko twórców sztucznej inteligencji uważało tę grę za rodzaj testu, który sprawdza zaawansowanie programów AI i ich zdolność do zastępowania człowieka.

Dlaczego? Z kilku istotnych powodów:

- "StarCraft" nie ma jednej najlepszej strategii wygrywającej. Program musi więc bez przerwy się uczyć i doskonalić strategię.
- W przeciwieństwie do takich gier jak szachy czy Go, gdzie gracze mają przed oczami całą planszę i znają dokładny aktualny stan rozgrywki, w"StarCraft" gracz dysponuje wysoce niepełną informacją. Niektóre informacje o przeciwniku są przed nim ukryte i musi je aktywnie odkrywać.
- Gra toczy się w czasie rzeczywistym - nie jest więc tak jak w szachach, czy w Go, że ruchy graczy następują na przemian. Tutaj musimy dotrzymywać kroku przeciwnikowi, który nas atakuje i realizuje swoją strategię, nie czekając na nasze decyzje i ruchy. Naukowcy oszacowali, że w każdej chwili sztuczna inteligencja ma do wyboru średnio jedno z setek milionów trylionów (10 do potęgi 26) możliwych posunięć!
- Każda podjęta przez gracza decyzja, nawet na wczesnym etapie gry, może mieć długofalowy i z góry nieprzewidywalny efekt dla wyniku rozgrywki.

Słowem, w tej grze jest trochę jak w prawdziwym życiu, gdzie musimy podejmować decyzje nie mając pełnej wiedzy, zmagając się z czasem i konkurencją. Mimo to stworzony przez Deep Mind program StarAlpha osiągnął w "StarCraft II" poziom mistrzowski.

Uczył się podobnie jak AlphaZero, grając sam ze sobą. Kolejne edycje stawały się coraz lepsze dzięki selekcji naturalnej - wersje, które zwyciężały, przechodziły do kolejnych etapów. Naukowcy stworzyli też specjalną "Ligę", w której grało ze sobą kilka programów. Wśród nich był ten, który był liderem i miał się stać niezwyciężony. Reszta była nastawiona na wyłapywanie słabych stron lidera. Po to, by można było je wyeliminować.

Można wzruszyć ramionami i rzec, że to tylko gra wideo. Ale sztuczna inteligencja, która z sukcesem rywalizuje z profesjonalistami w złożonym świecie wirtualnym, za chwilę może już także stawić czoło problemom świata realnego.

Rok temu w "Nature" opublikowano algorytm, który po raz pierwszy pokonuje nas w sterowaniu dronami sportowymi w rzeczywistym świecie, a niedawno AI zwyciężyła już z człowiekiem w walkach powietrznych na niebie - w powietrznej gonitwie prawdziwych myśliwców.

AI się uczy, chce zjeść nasze dane

Suleyman zaprzągł algorytmy Deep Mind do analizy zużycia energii w centrach obliczeniowych i danych Google'a. Znalazł sposoby na zmniejszenie energochłonności obliczeń, tylko o kilkanaście procent, ale w skali całej firmy była to pokaźna oszczędność. Rozwijał też dział zdrowia -

AI miała pomóc lekarzom w diagnozowaniu chorób, dużo łatwiej i szybciej znajdować niepokojące sygnały w wynikach badań.

Ale Deep Mind po raz pierwszy zderzył się tu z problemem ochrony prywatności i dostępu do danych. Żeby AI stała się lekarzem doskonałym, trzeba było ją wyszkolić - nakarmić strumieniem danych z kartotek milionów pacjentów brytyjskiej służby zdrowia NHS. Suleyman odbił się tu od ściany.

Łatwiej poszło z innym praktycznym zastosowaniem. Sztuczna inteligencja Deep Mind nauczyła się przewidywać kształty, w jakie zawijają się łańcuchy białek, podstawowy budulec życia.

"Życie jest formą istnienia białka..." - pisała Agnieszka Osiecka i nie ma w tym wiele przesady. A o tym, jak działają i co robią białka decyduje ich struktura 3D. Dekady zajęło naukowcom opisanie tylko niewielkiej liczby białek, tymczasem program AlphaFold 2 błyskawicznie ustalił trójwymiarowe struktury prawie wszystkich 20 tys. białek spotykanych w ludzkich komórkach i tkankach. Dwa lata później skompletował przewidywane kształty ponad 200 mln białek, wszystkich, jakie znamy dziś na Ziemi.

Naukowcy z DeepMind szkolili swoją AI na tysiącach znanych struktur białek, w czym pomogły dane z Europejskiego Laboratorium Biologii Molekularnej, międzyrządowej organizacji badawczej, finansowanej ze środków publicznych.

Dosłownie w ciągu tygodni i miesięcy AlphaFold stał się podstawowym narzędziem dla setek tysięcy naukowców w laboratoriach i na uniwersytetach na całym świecie, którzy z jego pomocą m.in. projektują szczepionki, leki i nowe terapie. Najnowsza wersja umożliwia także modelowanie innych kluczowych biomolekuł, w tym kwasów nukleinowych (DNA i RNA). Hassabis i John Jumper, którzy kierowali rozwojem AlphaFold, dostali w zeszłym roku medyczną nagrodę Laskera, która jest uznawana za przedsionek do Nagrody Nobla.

AI pomaga nie tylko naukowcom. Podobne samouczące się systemy od kilku lat wkraczają do naszego życia, znajdują się w produktach, usługach i urządzeniach, z których korzystamy na co dzień.

Algorytmy uczenia głębokiego szukają nowych leków, analizują wyniki badań lekarskich i stawiają diagnozy, zarządzają ruchem ulicznym i sieciami energetycznymi, kierują samochodami i optymalizują trasy żeglugowe, modelują klimat i przewidują reakcje rynków finansowych.

Za kilka lat modele AI będą w stanie rozmawiać, rozumować i działać w tej samej rzeczywistości, w której funkcjonują ludzie. Staną się organicznie wplecione w naszą tkankę społeczną.

Otwierają możliwości, z których wciąż nie do końca zdajemy sobie sprawę.

- Jeśli sztuczna inteligencja wykorzysta zaledwie ułamek swojego potencjału, następna dekada będzie najbardziej produktywna w historii ludzkości - mówi Suleyman.

Oczywiście AI można zastosować także w złych celach, np. do stworzenia broni biologicznej, dezinformacji i deep fake'ów. Także do planowania i wygrywania wojen, tyle że już tych prawdziwych, a nie między fikcyjnymi rasami w wymyślonych galaktykach gry wideo.

AI zalewa świat niczym fala tsunami.

Gdy wizjonerzy AI przedstawiali świetlaną przyszłość, powiększała się grupa ekspertów, którzy byli zaniepokojeni błyskawicznym i niekontrolowanym wzrostem możliwości AI.

- Ludzie często przeoczają kluczowy fakt: to, że jest ona wzmacniaczem siły. Odzwierciedla ludzkie mocne i słabe strony. Największe ryzyko wynika z tego, że może zostać wykorzystana przez złych ludzi - twierdzi Suleyman i nie jest w tym odosobniony.

Słynny kosmolog Stephen Hawking mówił na konferencji Web Summit w Lizbonie w listopadzie 2018 roku: „Sukces w stworzeniu skutecznej sztucznej inteligencji może być najdonioślejszym wydarzeniem w historii naszej cywilizacji. Albo najgorszym. Po prostu jeszcze tego nie wiemy".

A kiedy się dowiemy? Problem polega na tym, że rozwój nowej technologii jest tak błyskawiczny, że ta wiedza może dotrzeć do nas za późno.

AI rozprzestrzenia się i zalewa świat niczym fala tsunami.

Mustafa Suleyman przekonuje w swojej książce, że ta ekspansja nie jest niczym nowym. Każda podstawowa technologia, jaką wcześniej wynaleziono – od kilofów po pługi, ceramikę po fotografię, telefony i samoloty – z biegiem czasu staje się tańsza i łatwiejsza w użyciu i rozprzestrzenia się szeroko. Tak było z prasą drukarską i książkami, silnikiem spalinowym, czy internetem.

- Jednym z kluczowych wzorców, jakie obserwujemy w przypadku sztucznej inteligencji, jest to, że innowacje, które w chwili pojawienia się są bardzo trudne do wdrożenia lub kosztowne, po chwili szybko i szeroko się rozprzestrzeniają. Stają się znacznie tańsze i bardziej dostępne dla szerokiego grona osób - mówi Suleyman.

Dodaje, że rozwój AI wspomagają bardzo silne bodźce geopolityczne, komercyjne, akademickie i badawcze.

- Nie ma łatwego sposobu, aby ktokolwiek nagle zamknął sztuczną inteligencję w pudełku albo pokierował jej kierunkiem - mówi.

Wyścig napędza rywalizacja między państwami. USA boją się, że zostaną wyprzedzone przez Chiny, Europa nie chce zostać w tyle.

Biznes traktuje nową technologię jak być albo nie być. Kto się nie pospieszy, przepadnie, nie zgarnie sutej premii dla zwycięzcy. Firma, która jako pierwsza wprowadza

nowy produkt, staje się długofalowym liderem rynku tylko dlatego, że była pierwsza. Liczy się czasem nawet różnica tygodni.

Takie firmy jak Microfoft, Google, Amazon czy Meta doskonale zdają sobie z tego sprawę, bo każda z nich wyrosła po wygranej w podobnym wyścigu.

"Firmy z Doliny Krzemowej, które kiedyś z wielką ostrożnością podchodziły do potencjału sztucznej inteligencji, dziś odblokowały hamulce" - pisał rok temu "New York Times". Wcisnęły pedał gazu pod wpływem szoku, jakim było publiczne udostępnienie w internecie ChataGPT w listopadzie 2022.

To druga data, kandydująca do początku ery AI.

ChatGPT należy do generatywnych AI, tzw. dużych modeli językowych, które są wytrenowane na ogromnych zbiorach danych i potrafią generować teksty, kody, obrazy, nawet wideo, a także konwersować z ludźmi.

"Kluczem do przyszłości informatyki jest zdolność prowadzenia konwersacji. Każda interakcja z komputerem jest dziś swoistą rozmową, tyle że używającą przycisków, klawiszy i pikseli do przekładania ludzkich myśli na kod czytelny dla maszyny. I oto ta bariera właśnie zaczyna się kruszyć.

Wkrótce maszyny będą rozumieć nasz język. To była i wciąż jest niezwykła perspektywa" - pisze Suleyman.

To perspektywa, dodajmy, która dosłownie oszołomiła wielkie korporacje.

"New York Times" cytuje obecnych i byłych pracowników Microfoftu i Google'a, a także wewnętrzne dokumenty obu firm, które pokazują, że niezwykły sukces ChataGPT skłonił je do podjęcia ryzyka i wypuszczenia nowych modeli językowych, mimo że programiści ostrzegali, że generują one błędne, nieprawdziwe, a nawet niebezpieczne treści. Co więcej, nikt nie potrafił wyjaśnić, dlaczego dany model generuje takie, a nie inne odpowiedzi, czasem myli się i oszukuje, czyli - jak zwykli mówić eksperci do AI - halucynuje. Nie było czasu na testy, kosztowne i pochłaniające sporo mocy obliczeniowych.

Jedno jest dzisiaj pewne: obie firmy były świadome tego, że ich chatboty mogą zachwiać podstawami nowoczesnych społeczeństw. W jaki sposób?

"Zaleją internet dezinformacją, wywrócą do góry nogami rynek pracy, w rękach takich ludzi jak Władimir Putin mogą poważnie zagrozić ludzkości" - ostrzegał poniewczasie Geoffrey Hinton, pionier w dziedzinie badań maszynowego uczenia sztucznej inteligencji, który w zeszłym roku odszedł z Google'a.

Latami wbudowywane w kulturę firm etyczne zasady, które miały pilnować, by nowatorskie technologie nigdy nie zaszkodziły ludziom, zeszły na dalszy plan. Priorytetem stała się przewaga w wyścigu o prymat w przełomowej technologii.

Bez odzewu pozostał apel ponad tysiąca badaczy i liderów branży, wśród nich Elona Muska i współzałożyciela Apple'a Steve'a Wozniaka, którzy wiosną 2023 r. wezwali, by na pół

roku przerwać prace nad rozwojem nowych systemów AI. Opublikowali list, w którym ostrzegają, że stanowi ona "poważne zagrożenie dla społeczeństw i ludzkości".

Można by tu dodać: historia się powtarza.

Czy damy radę powstrzymać globalne ocieplenie

Fale rewolucji technologicznych w przeszłości także wiązały się z negatywnymi skutkami, którym dopiero poniewczasie starano się zaradzić.

Kiedy kilkanaście tysięcy lat temu w neolicie nauczyliśmy się uprawiać ziemię i zaczęliśmy pędzić osiadły tryb życia, rozwój cywilizacji przyspieszył, ale ta wielka zmiana z początku wcale nie polepszyła warunków życia. Przeciętny przedstawiciel kultury zbieracko-łowieckiej miał bardziej zróżnicowaną dietę i spożywał więcej białka i kalorii niż ludzie z późniejszych kultur osiadłych. Co więcej, w wielkich skupiskach ludzkich łatwiej roznosiły się choroby. Osiadłość oznaczała więc uboższą dietę, więcej chorób i - przynajmniej z początku - krótsze życie.

Podobnie złudne były początkowo zyski rewolucji przemysłowej w XIX wieku. Na starcie spowodowała ona wielki i niekontrolowany exodus ze wsi do miast. Jednak życie przeciętnego mieszczucha było godne pożałowania - toczyło się w ciemnych suterenach, pośród tysięcy ton odpadków, końskiego łajna, zdechłych psów, kotów i szczurów, powodzi ścieków, szlamów i trujących substancji. W tym mało higienicznym świecie co chwila wybuchały epidemie cholery, grypy, tyfusu, gośćca, szkarlatyny, dyfterytu i ospy, zabijając dziesiątki tysięcy ludzi.

Dzisiaj walczymy z odroczonymi skutkami wykładniczego wzrostu, jaki ludzkość przechodzi w ostatnich dwóch stuleciach, wywołanych emisją gazów cieplarnianych zmian klimatu. Dzieci, które się teraz rodzą, prawdopodobnie będą żyły średnio krócej niż ich rodzicie, po raz pierwszy od wielu pokoleń. Nie jest wcale pewne, czy damy radę powstrzymać globalne ocieplenie i jego katastrofalne następstwa. Ścigamy się z czasem i jak na razie przegrywamy, ale od katastrofy dzielą nas prawdopodobnie jeszcze dekady. Wciąż łudzimy się, że zdążymy znaleźć rozwiązania - technologiczne, ekonomiczne, społeczne.

Fala AI wzbiera dużo gwałtowniej i jeśli odpowiednio wcześnie nie zaopatrzymy tej technologii w bezpieczniki, to gdy się ona wynaturzy, na korektę i wyjście z ostrego wirażu będzie za późno.

We wstępie do "Nadchodzącej fali" Mustafa Suleyman opisuje prezentację, jaką kilka lat po założeniu Deep Mind przygotował dla grona kilkunastu założycieli najbardziej wpływowych firm. Prezesom zarządów i specom od technologii pokazywał zestaw slajdów obrazujących potencjalne długofalowe skutki gospodarcze i społeczne wynikające z rozwoju sztucznej inteligencji.

Przekonywał m.in. że AI niesie wiele zagrożeń wymagających działań wyprzedzających, może doprowadzić do masowych zwolnień w całych sektorach gospodarki, zagrożeń prywatności, niebywałej koncentracji władzy, lawiny dezinformacji, czy posłużyć

jako cyberbroń do szukania luk w zabezpieczeniach coraz bardziej uzależnionego od cyfrowych systemów świata.

Słuchacze przyjęli to z powątpiewaniem, wskazali, że AI wywoła raczej nowy popyt, stworzy nowe miejsca pracy, wyposaży ludzi w nowe kompetencje i sprawi, że będą jeszcze bardziej produktywni. Tak - przyznawali - istnieją pewne ryzyka, ale nie takie znowu straszne, w razie czego, poradzimy sobie. Przecież zawsze znajdowało się jakieś rozwiązanie.

- Tę powszechną reakcję, którą wtedy obserwowałem, nazwałem pułapką niechęci do pesymizmu. To błąd poznawczy pojawiający się wtedy, gdy ogarnia nas lęk przed konfrontacją z potencjalnie nieprzyjemnymi faktami - mówi Suleyman.

Yoshua Bengio, Geoffrey Hinton, Yuval Noah Harari przestrzegają

Wie, co mówi, bo w Deep Mind odpowiadał za etyczną stronę technologii. Był też w 2016 r. jednym z założycieli organizacji non-profit "Partnerstwo na rzecz sztucznej inteligencji", w której znaleźli się przedstawiciele najważniejszych graczy na rynku, m.in. Apple'a, Microsoftu, Google'a, Facebooka i Amazona, a która miała promować dobre praktyki i odpowiedzialne wykorzystanie sztucznej inteligencji. W istocie okazała się jednak tylko listkiem figowym, osłaniającym bezlitosne reguły kapitalizmu, które jak na razie rządzą biznesem AI.

Jego książka to apel o to, abyśmy nie wypierali zagrożeń i świadomie stworzyli mechanizmy, dzięki którym będziemy mieli zdolność monitorowania, ograniczania, kontrolowania, a potencjalnie nawet wycofywania się z używania AI (a także narzędzi biologii syntetycznej, która wraz z AI tworzy wyjątkowo niebezpieczną mieszankę). Suleyman nazywa to zdolnością do powstrzymywania technologii. Celem nie jest zatrzymanie technologii, ale zapewnienie jej bezpieczeństwa i kontroli nad nią.

Jednocześnie przyznaje, że w dziejach ludzkości właściwie nie ma przykładów, na których można byłoby się wzorować. Technologie ogólnego przeznaczenia zwykle rozprzestrzeniały się w sposób niekontrolowany, a potencjalne złe skutki ujawniały się w praniu.

- Gdy zbudowaliśmy fabryki, były to z początku ponure i niebezpieczne miejsca pracy. Gdy trafiliśmy na ropę, w jedno stulecie zanieczyściliśmy całą planetę. Ale teraz może być inaczej, ponieważ wciąż jesteśmy na etapie projektowania i budowania sztucznej inteligencji, mamy więc potencjał i szansę, aby zrobić to lepiej, radykalnie lepiej - mówi Suleyman. - Wciąż stoimy za sterami, możemy budować taką AI, jaką chcemy. Jeśli więc jej twórcy będą podzielać demokratyczne wartości – a np. ja jestem im gorąco oddany - to zbudują AI, która poszerzy wolności obywatelskie i umocni globalny porządek.

- Ale potrzeba też silnych regulacji w prawie i zabezpieczeń umieszczanych w technologii - dodaje, bo jednak nie jest już tak naiwny, by wierzyć w same dobre intencje twórców i korporacji IT.

Suleyman proponuje z grubsza coś na podobieństwo międzynarodowego układu o nierozprzestrzenianiu broni jądrowej, bo najważniejszymi aktorami zdolnymi do trzymania w ryzach technologii są państwa narodowe, które są równocześnie najbardziej zagrożone nadchodzącą falą. To jedno nie wystarczy. Jego zdaniem trzeba umieścić bezpieczniki na wielu poziomach - poczynając od kodów komputerowych, systemów państwowego nadzoru, obowiązkowych audytów rządowych, testów bezpieczeństwa, zachęt biznesowych i podatkowych, a kończąc na masowych ruchach społecznych, wymuszających na rządach korzystne regulacje.

Żeby zminimalizować egzystencjalne zagrożenie, powinniśmy także unikać wyposażania systemów AI w potencjalnie groźne cechy, takie jak:

- Nie możemy dopuścić do tego, aby takie potężne zdolności w jakiś sposób pozostawały poza radarem - mówi Suleyman.

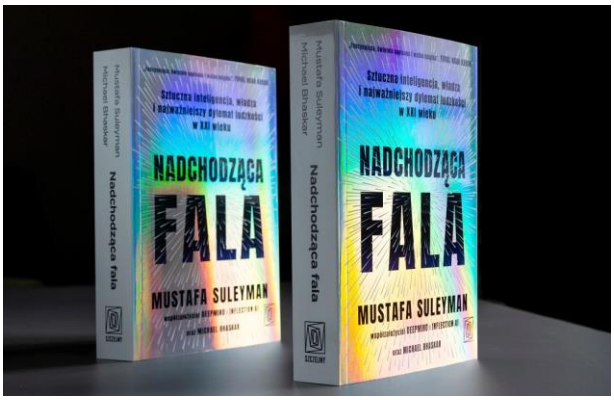
Problem w tym, że te propozycje na razie są wołaniem na puszczy. W porównaniu z ogromem potencjalnych zagrożeń prowadzone dziś badania nad bezpieczeństwem i etyką AI to kropla w morzu potrzeb. Jak zauważa Suleyman, grono badaczy bezpieczeństwa AI wzrosło z około stu ekspertów w najlepszych laboratoriach świata w 2021 roku do trzystu-czterystu w 2022 r. Jest to szokująco mała liczba, uwzględniając to, że obecnie kilkadziesiąt tysięcy badaczy zajmuje się rozwojem systemów AI.

20 maja pismo "Science" opublikowało list grupy naukowców (podpisali go m.in. laureaci Nagrody Turinga - Yoshua Bengio i Geoffrey Hinton, a także znany historyk Yuval Noah Harari), którzy tak jak Suleyman apelują o ściślejszą kontrolę AI, podając konkretne kroki do podjęcia:

- - Firmy technologiczne i agencje finansujące badania powinny być zmuszone inwestować w bezpieczeństwo sztucznej inteligencji co najmniej jedną trzecią swoich budżetów przeznaczanych na badania i rozwój AI.
- - rządy powinny narzucić bardziej rygorystyczne standardy i oceny bezpieczeństwa sztucznej inteligencji
- - w szczególności powinny powołać szybko działające organy nadzorujące AI (takie jak na przykład FDA, czyli Agencja Żywności i Leków, która dopuszcza leki na rynek). Firmy zajmujące się sztuczną inteligencją musiałyby udowodnić, że ich systemy nie mogą wyrządzać szkód.

"Na pierwszy rzut oka powstrzymanie tej technologii nie wydaje się możliwe. A jednak dla dobra nas wszystkich musi być możliwe" - pisze Suleyman. Ale pobrzmiewa w tym jednak bezradność.

Jeśli bowiem mówi to jeden z najlepiej ustosunkowanych i poinformowanych graczy w tej dziedzinie - były wiceprezes Google'a ds. AI, a obecny główny specjalista od sztucznej inteligencji w Microsofcie - to chyba czas zacząć się bać.



'Nadchodząca Fala' Mustafy Suleymana, Szczeliny, imprint Wydawnictwa Otwartego, tłum. Justyn Hunia Fot. A. Sołtysik / Szczeliny