

Żeby komputer nas pokochał, czyli jak wychować mądrą, dobrą i wrażliwą sztuczną inteligencję

© Jan Hartman

Magazyn Wyborczej, 03.11.2018

https://wyborcza.pl/magazyn/7,124059,24120902,zeby-komputer-nas-pokochal-czyli-jak-wychowac-madra-dobra-i.html#s=S.embed_article-K.C-B.1-L.5.zw

Czy serwisant obsługujący serwery sztucznej inteligencji naprawdę będzie czymś lepszym i bardziej zasługującym na ochronę niż genialne roboty, działające w sieci umysły przodków albo wskrzeszeni neandertalczyki?

Jan Hartman – ur. w 1967 r., filozof i bioetyk, profesor Uniwersytetu Jagiellońskiego, publicysta. Ostatnio wydał: „Pochwała litości. Rzec o współnocy” (UJ 2017) i „Polityka. Władza i nadzieja” (Agora 2017)

Ledwie weszliśmy w epokę internetu, a już na horyzoncie mającą kształty kolejnej rewolucji, której przebieg i oddziaływanie nie ominie żadnej sfery życia. Jej hasłowa nazwa to „sztuczna inteligencja”, a zwiastuny? Oto słynny superkomputer Watson wyprodukowany przez IBM nie tylko wygrywa z ludźmi w teleturniejach, lecz znacznie lepiej od lekarzy diagnozuje raka i ordynuje terapię. Inny program przetwarza impulsy elektryczne z mózgu i pozwala inwalidom poruszać protezami oraz sterować urządzeniami samą tylko myślą. Odczytywanie myśli idzie jeszcze słabo (choć detektor odróżnia „tak” od „nie”), za to można już odtworzyć z aktywności mózgu oglądany przeze mnie obraz.

Sztuczna inteligencja to także big data – metody przetwarzania wielkich zbiorów danych. Jak wiadomo, wielkie serwisy internetowe zbierają mnóstwo informacji na nasz temat, a wkrótce nie będziemy mieli przed nimi żadnych tajemnic. **Lecz czy Wielka Sieć naprawdę będzie się nami interesować? Bo jeśli będzie bardziej twórcza, wyposażona w lepszą intuicję, pamięć i intelekt od naszych białkowych komputerów, to do czego przydadzą się jej ludzie i ludzkie wartości?** Może do ścierania kurzu – ludzka siła robocza, choć mizernej jakości, będzie jednak tańsza. Tylko czy maszynom będzie przeszkadzał kurz? Bardzo zabawne.

Szkoda tylko, że całkiem realne. Zwłaszcza dla naszych wnuków.

Wszyscy będziemy online

Jeszcze nie tak dawno temu zalecano dzieciom pisać wypracowania pt. „Jaki będzie świat za sto lat?”. Dziecięce przewidywania nie różniły się zbytnio od futurologii dorosłych. Wszystko kręciło się wokół technologii ułatwiających życie i dowodzących ludzkiego geniuszu. Miały być loty na Marsa, lekarstwo na raka i wideotelefon w każdym domu. Wygląda na to, że wszystko się spełni - ale nie to przesądzi o kształcie ludzkości.

Co do tego, że jesteśmy na zakręcie, nikt już nie ma wątpliwości. Dawni futurologowie – jak Stanisław Lem albo Alvin Toffler, który w książce z 1980 roku pt. „Trzecia fala” przewidział to

wszystko, czym stał się internet, czy Jean-François Lyotard w 1979 roku w „Kondycji ponowoczesnej” przepowiadający przemiany w relacjach między nauką i technologią – dawno temu zdefiniowali naszą epokę jako postindustrialną i informatyczną, podobnie jak przewidzieli wyczerpanie się potencjału gospodarki wolnorynkowej oraz indywidualistycznej demokracji skoncentrowanej na instytucjach państwa. A my mamy znacznie łatwiej niż oni. Nie trzeba być profetą, by zestawić ze sobą kilka wyraźnych trendów i eksperymentalnych technologii, które odmienią nasz świat.

Nieodwołalnie wszyscy i wszystko będzie online. Nieodwołalnie naszym życiem na wszystkich poziomach sterować będą coraz bardziej elastyczne i dopasowujące się do okoliczności algorytmy wykorzystujące wielkie zbiory danych. Nasze ciała będą coraz silniej zintegrowane z urządzeniami przetwarzającymi informacje, być może z udziałem procesorów obsługiwanych wprost z poziomu mózgu. Metody elektronicznej stymulacji mózgu, wzmacniające jego możliwości poznawcze i korygujące sferę emocjonalną – dziś już szeroko stosowane przez ludzi zdrowych, ale za pośrednictwem mniej precyzyjnej farmakologii – będą się upowszechniać.

Nie warto się też łudzić, że nie wykorzystamy technologii doskonalenia ludzkiego genomu. Superludzie oraz wyspecjalizowane klony pojawią się na ziemi, podobnie jak maszynowo-cielesne hybrydy oraz genetycznie zmodyfikowane zwierzęta odznaczające się wysoką inteligencją. Techniki genetyczne pozwolą na regenerację tkanek i organów na ogromną skalę, dając bogatym szansę na życie przedłużone daleko poza granicę stu lat, a naśladujące człowieka boty będą pracować z nami na każdym kroku. Umierający pozostawiają po sobie pełne skany mózgów (nie mówiąc o DNA), które współpracując z oprogramowaniem generującym teksty i obrazy, pozwolą na zachowanie pośmiertnej aktywności online. Eksperymenty w tej dziedzinie (jak Blue Brain Project, zmierzający do pełnego zmapowania mózgu zwierzęcia) są już bardzo zaawansowane.

Już dawno komputery przestały być po prostu maszynami liczącymi, lecz teraz ich zdolność wykonywania operacji symbolicznych zaczyna przypominać tę właściwą ludzkiemu mózgowi, który nie musi przeliczać (komputować) wielkich zbiorów danych i niepomiernej większej liczby możliwych ich kombinacji, aby podejmować trafne decyzje i skuteczne działania. Za progiem jest prawdziwa AI - programy wykonujące operacje szersze niż bezpośrednio zadane przez użytkowników, samomodyfikujące się, uczące się na własnych błędach, korzystające z samodzielnie zbieranych danych i wykorzystujące je – jak człowiek – w sposób selektywny, z nastawieniem na efekty prawdopodobne i przybliżone (w odróżnieniu od pewnych i ścisłych, które uzyskujemy w normalnych procedurach obliczeniowych).

Długo jeszcze będziemy chodzić do lekarzy. Ale będziemy tam się pojawiać z gotowym wywiadem przeanalizowanym przez inteligentny medyczny analizator, który po uzupełnieniu danymi z badań postawi nam diagnozę i zaleci postępowanie lecznicze. **Lekarz będzie miał ostatnie słowo – lecz jak często różne od zaleceń aplikacji? A kto będzie rozwijał możliwości tego oprogramowania?** Oczywiście fachowcy z dziedziny nauk medycznych i epidemiologii. Zapewne jednak bardzo mocno osadzeni w przekształconych w programy komputerowe algorytmach i coraz bardziej stechnicyzowanych i algorytmicznych

procedurach badań naukowych. Ich inwencja, doświadczenie i intuicja będą znaczyły wiele, lecz z czasem te same przymioty SI zaangażowanej w rozwój medycyny wezmą górę.

Współcześni futurologowie, tacy jak Raymond Kurzweil (jego książka o cyborgach pt. „Nadchodzi osobliwość” jest już dostępna po polsku), Toby Walsh (właśnie ukazała się jego świetna praca pt. „To żyje. Sztuczna inteligencja”), gwiazda kosmologii Max Tegmark, powszechnie już znany historyk celebryta Yuval Harari czy biznesmen wizjoner Elon Musk, spierają się nie o to, czy to wszystko się wydarzy, a nawet nie o to kiedy, lecz o to, jaki będzie koszt. Ostrzegają przed utratą panowania nad technologią, lecz przede wszystkim mobilizują do podjęcia wielkiego wyzwania: kontroli nowych technologii i kierunku ich rozwoju. Być może lepiej, abyśmy tę kontrolę mieli jednak my, a nie Sieć. A może na odwrót – to ludzkie panowanie nad technologiami grozi koncentracją władzy informacyjnej w rękach paru korporacji, jeśli nie globalnych dyktatorów?

Co czuje komputer

Tak czy inaczej, nasza obawa przed utratą kontroli przegrywa z rozleniwiającą nas wygodą. Zwykle bardzo nam odpowiada, gdy maszyny myślą za nas. Cóż można mieć przeciwko temu, żeby autobus jechał sam, bez kierowcy, a wypracowanie z polskiego napisał darmowy program sponsorowany przez producenta inteligentnej żywności dopasowującej swój skład do naszych potrzeb monitorowanych przez kolejną aplikację? Ubezpieśnienie nie przeszkadza nam dopóty, dopóki otrzymujemy wystarczająco dużo przyjemności, komfortu i bezpieczeństwa. **A poczucie kontroli długo jeszcze zapewniać nam będzie nad wyraz ludzki „serwisant” z odwieczną torbą pełną kabelków i śrubokrętów, dla którego cała ta maszyneria to tylko graty, które czasami się psują i wymagają naprawy, tak samo jak piecyki gazowe czy instalacja elektryczna.**

To jednak złudne poczucie.

Dziś coraz bardziej autonomiczny świat programów typu AI stanowi podsystem globalnego kapitalizmu, wobec czego rządzą nim prawa marketingu nakazujące dogadzać konsumentom. Nie ma jednakże żadnej gwarancji, że te same technologie w rękach polityków, terrorystów, a nawet dowcipnisiów nie staną się zarzewiem gwałtownej destrukcji. Mamy moralny obowiązek wypracowania mądrych regulacji i polityk społecznych – dopóki w ogóle jeszcze mamy i chcemy mieć na cokolwiek wpływ.

Nie byłbym optymistą. Z jednej strony każdy będzie chciał żyć sto dwadzieścia lat, z drugiej – udoskonaleni superludzie, nie mówiąc już o cyborgach, zapewne nie będą mieli zbyt dużo cierpliwości i empatii dla osób „analogowych” czy mas oddających się loterii genetycznej, czyli rozmnażających się jak popadnie, bez kontroli i korekcji genomu potomstwa, albo odmawiających korzystania (bądź niemających możliwości korzystania) z technologiczno-farmakologicznej stymulacji mózgu.

Ale jest jeszcze ważniejsze pytanie: czy my, ludzie, sami lub z pomocą superludzi będziemy kontrolować sytuację na ziemi, czy też zostaniemy zmarginalizowani przez algorytmy, które ustanowią własną, a nie ludzką przestrzeń kulturową.

W etycznym ujęciu brzmi ono następująco: czy Wielka Sieć będzie przestrzenią moralną, unormowaną przez zasady poszanowania godności, wolności i równości ludzi, czy też jej działanie będzie odporne na względy etycznej wrażliwości, dobre obyczaje, a nawet regulacje prawne? Można mieć obawy, że z uwagi na swą radykalną ontologiczną obcość nie będą w niej obowiązywać ani zasady humanizmu, ani nawet humanitaryzmu.

Aby zmierzyć się z problemem SI od strony etycznej, trzeba spróbować zrozumieć, czym jest podmiot maszynowy i czym różni się od człowieka. **Na pierwszy rzut oka różnica jest oczywista – SI jest martwym przedmiotem czy też strukturą, a człowiek – czującą i świadomą istotą. Jednak to pierwsze nie jest wcale pewne, a to drugie i niepewne, i może nie aż tak istotne.**

Nasze przeświadczenie o tym, że tylko inni ludzie mogą być istotami czującymi, wynika ze zdolności odczytywania z twarzy i zachowania naszych bliźnich emocji i odczuć podobnych do tych, jakie odczuwamy sami. Biorą w tym udział neurony lustrzane i różne struktury mózgu. Jako że niektóre zwierzęta również mają twarze i mimikę, udało się nam wykształcić w sobie ewolucyjnie zdolność rozpoznawania uczuć także u innych niż człowiek gatunków, dzięki czemu mamy trochę empatii dla zwierząt.

Te zdolności są jednak bardzo ograniczone. Nie ma nic racjonalnego w przekonaniu, że skoro nie wiemy nic o czuciu drzew albo całych planet, to drzewa i planety nie mają życia wewnętrznego ani zdolności komunikowania się. Z tego, że nie komunikujemy się z drzewami (przynajmniej powszechnie), nie wynika, że drzewa nie komunikują się ze sobą. Nie mamy żadnych podstaw sądzić, że komputery coś czują, ale też nie mamy żadnych podstaw, aby sądzić, że nie czują.

Dziś jest to wyłącznie kwestia wiary, wiary animistycznej bądź jej braku. Z czasem będzie jednakże przybywać przesłanek do całkiem poważnego traktowania ewentualności, że inne niż człowiek istoty, a w tym również roboty, myślą i czują.

Nasza skłonność do przypisywania innym odczuć i świadomości wiąże się ze zdolnością reagowania na zmiany w wyglądzie oraz komunikacją językową. Dlatego żeby odmówić ładnemu i miłemu awatarowi – robotowi mającemu kształt człowieka – zdolności odczuwania i myślenia, musimy się zdobyć na specjalny wysiłek. Odwołać się do umiejętności konsekwentnego trzymania się dogmatu – w tym wypadku dogmatu mówiącego, że tylko ludzie myślą i mówią. Jak długo jednak można kurczowo trzymać się dogmatu, gdy maszyna mówi do nas mądrze, uśmiecha się albo okazuje strach? Dawno już przekroczyliśmy barierę testu Turinga, stworzyliśmy maszyny, które w pewnych warunkach stają się dla nas nieodróżnialne od ludzi. A przecież AI nie służy do kopiowania człowieka, nie jest zabawką prestidigitatora, pełni wiele poważniejszych funkcji - i tym bardziej musi być potraktowana poważnie.

Czy brak świadomości kompromituje

Bardzo możliwe, że obecnie na ziemi zdolności odczuwania mają wyłącznie struktury oparte na cząsteczkach białka i DNA, z których w większości składają się układy nerwowe zwierząt. Może jednak wcale tak nie jest, a poza tym stworzymy zapewne komputery „mokre”, czyli

biokomputery oparte na możliwościach kodowania chemicznego w związkach organicznych. **Już dziś działają prototypy komputerów, w których funkcję tranzystora pełnią układy bakterii. Czy więc będziemy się upierać, że nasz mówiący „białkowy” komputer przypominający mózg też nic nie czuje? To byłby już tylko tępy upór.**

W sferze czucia w szczególny sposób chełpimy się naszymi aktami świadomościowymi. Wydaje się nam, że są unikalne. Do niedawna odmawialiśmy ich nawet naszym psom i kotom. Czy jednak rzeczywiście pojęcia myślenia i inteligencji tylko metaforycznie stosują się do bytów pozbawionych świadomości? Czy komputer tylko udaje, że myśli, bo przecież brakuje mu świadomości? I skąd wiemy, że mu jej brakuje?

Aby odpowiedzieć na te pytania, trzeba sobie uświadomić, czym jest świadomość. Biologiczne podstawy jej fenomenu nie są jeszcze poznane, lecz od dawna mamy bardzo dogłębne analizy, czym ona jest. Wiemy na przykład, że jeśli w ogóle taka wiedza stała się możliwa, to ze względu na podstawową własność świadomości, którą jest stałe nakierowanie na samą siebie. Świadomość jest prawie zawsze samoświadomością, czyli samouświadamianiem sobie siebie – własnych czynności (aktów) oraz spełniającego te akty, czyli podmiotu.

Świadomość związana jest z procesem tworzenia tożsamości podmiotu - „ja”. Przebiega on ciągłej w interakcji pomiędzy doświadczeniem (na przykład spostrzeganiem) pewnych struktur nazywanych przedmiotami a doświadczeniem samego siebie. Fachowo ten proces nieustannego i wciąż na nowo powstawania „mnie i mojego świata” nazywa się konstytucją. Tak konstytuująca siebie i świat świadomość jest zawsze na czymś skoncentrowana - intencjonalna. Istota boska, wszechogarniająca byt swoim doświadczeniem nie może więc być świadoma.

Od czasów Kartezjusza mamy skłonność do fetyszyzowania świadomości i deprecjonowania bytów jej pozbawionych. Otóż to: nieświadomy jest nie tylko kamień, ale również absolut.

Czy więc ten brak świadomości kompromituje?

Nie wiemy i nie będziemy mogli się przekonać, czy inteligentny robot zapewniający nas, że jest istotą świadomą i czującą, nie będzie nas wprowadzał w błąd. Czy to zresztą ważne? Sztuczna inteligencja może wykonywać swe czynności bez poczucia własnego istnienia, jakim odznaczają się ludzkie umysły. Przecież również one pracują w zasadzie poza świadomością.

Myśli, słowa i wnioski zjawiają się w naszej świadomości już niejako gotowe – wypływają z nieświadomej głębi, pozbawionej indywidualności. To, że „wiemy, co mówimy”, nie znaczy, że proces tworzenia treści symbolicznych, na przykład wypowiedzianych zdań, musi przebiegać pod kontrolą świadomości. Nie przebiega. Proces konstytucji nie jest świadomy – trzeba nie lada wysiłku, aby uprzytomnić sobie jego elementy i fazy. Być może świadomość jest powierzchownym efektem, raczej odczuwaniem niż twórczym działaniem, jak lubimy myśleć. Neurolodzy już dawno odkryli, że nasze wyrozumowane decyzje podejmujemy bez udziału świadomości; wypowiadając takie czy inne słowa, reagujemy na motywowany złożonymi neurochemicznymi czynnikami impuls.

Moralność w świecie postludzkim

Lubimy myśleć, że działamy twórczo, bo dzięki temu stajemy się odrębnymi umysłami, mamy poczucie osobności, ciągłości egzystencji oraz sprawczości. Na tym opierają się nasze roszczenia do bycia podmiotami moralnymi, którym należy się ochrona.

A jednak już dawno nasza wrażliwość etyczna wyszła poza świat ludzki. Poczujemy się do odpowiedzialności za zwierzęta i całą naszą planetę. Dlaczego mielibyśmy nie wykształcić w sobie moralnego stosunku do sztucznej inteligencji? Owszem, AI zapewne nie jest i nie będzie tak wyraźnie podzielona na odrębne podmioty-umysły. Czy jednak indywidualna tożsamość jest warunkiem moralnej podmiotowości? Mamy wszak zobowiązania wobec całych społeczności, a nawet wobec całej ludzkości, choć nie sądzimy, aby społeczność czy ludzkość – jako kolektywne byty – czuły i myślały. Dlaczego więc nie mamy mieć obowiązków wobec wielkiego systemu współpracujących ze sobą algorytmów przetwarzania danych, czym stanie się w przyszłości nasz świat społeczny?

Swoją drogą, nasza indywidualność również rysuje się w przyszłości problematycznie. **Czy człowiek na wpół sztuczny, wymieniający zużyte organy na nowe i stale komunikujący się z informatycznym uniwersum nadal będzie uważał się za pojedynczą, odrębną osobę?** Cała nasza moralność skrojona jest na potrzeby takich właśnie istot, lecz w świecie postludzkim powoli stracimy monopol nie tylko w dziedzinie inteligencji, lecz również podmiotowości emocjonalnej i moralnej. Konsekwencją rozwoju AI może być głębokie przekształcenie przestrzeni moralnej, która przestanie być specyficznie ludzka i ustrukturuwana według ludzkich potrzeb, form życia i wartości. Ta nowa etyczność jest dla nas zupełnie nieprzenikniona i zagadkowa - a jednak stanie się rzeczywistością, gdy opłatające nas algorytmy i urządzenia wykształcą swoje własne systemy komunikowania się i podejmowania decyzji. Na ile będziemy zdolni współtworzyć jakiś modus vivendi z informatycznym uniwersum? Nie wiadomo. Jeśli stanie się ono tzw. szarą mazią, czyli policentryczną, pozbawioną wyraźnej struktury i hierarchii gigantyczną przestrzenią operacji na symbolach, niepodobną ani do jednego Wielkiego Umysłu, ani do mnogości małych umysłów, to z naszej psychologii i etyczności zostanie niewiele.

Yuval Harari, zafascynowany „dataizmem”, czyli wizją świata, którego treścią są nieustannie cyrkulujące informacje, stawia nas w słynnej książce „Sapiens. Od zwierząt do bogów” przed wyrazistą wizją: „Być może szybkim krokiem zmierzamy do nowej osobliwości, w której wszystkie pojęcia, jakie nadają sens naszemu światu – ja, ty, mężczyźni, kobiety, miłość, nienawiść – staną się nieistotne. Wszystko, co zdarzy się poza tym punktem, wymyka się naszemu pojmowaniu”.

Zanim się rozpłyniemy

Świat, który sobie szykujemy, nie będzie już ani nasz, ani dla nas. Dlatego dopóki jeszcze rządymy tą planetą, będziemy bronić znanych nam form życia osobistego i społecznego. Będziemy bronić naszego bezpieczeństwa, wolności, dobrobytu, a przede wszystkim naszej miłości i naszych rodzin. Tym razem stawką nie jest już trwanie takich czy innych ludzkich społeczności i kultur, lecz przetrwanie gatunku. Nie poddamy się łatwo ani szybko, lecz w dłuższej perspektywie zmierzch ludzkiego świata jest nieuchronny – nazbyt się

rozpanoszyliśmy na ziemi i zbyt daleko już posunął się proces wyobcowania technologii i wytwarzającej je kapitalistycznej maszynerii.

Czy to źle? Czy serwisant z kablami – będący takim samym zagęszczeniem procesów informatycznych jak tkwiące w jego torbie urządzenie dyktujące mu, jak ma go „używać”, a raczej „używające” jego osoby jako „białkowego wykonawcy” instrukcji – naprawdę będzie czymś lepszym i bardziej zasługującym na ochronę niż genialne roboty, cyborgi, przedśmiertnie zeskanowane i działające w sieci umysły zmarłych przodków albo wskrzeszeni z zachowanych fragmentów DNA neandertalczyki?

Zanim jednak ustąpimy miejsca czemuś od nas lepszemu i (miejmy nadzieję) mniej niszczycielskiemu, czeka nas wiele pokoleń życia w świecie zamieszkałym przez różne pozaludzkie inteligencje. Czyż jednak nie żyliśmy w takim świecie przez dziesiątki tysięcy lat, gdy towarzyszyli nam ludzie innych gatunków, zwierzęta, które uważaliśmy za równe nam lub od nas lepsze, nie mówiąc już o obfitości bóstw i innych duchów?

Nie można jednak powiedzieć, że mamy w tych sprawach doświadczenie. Epoka nowoczesna jest w zupełności antropocentryczna – egoizm gatunkowy stał się jej niepisanym prawem. Jeśli chcemy ułożyć sobie znośne relacje ze sztuczną inteligencją oraz różnymi jej nośnikami i twórcami, lepiej od razu przekazywać jej to, co w nas moralnie najlepsze – otwartość, tolerancję, empatię, szacunek dla inności. Jeszcze możemy sprawić, aby AI nie była „bezduszną”, obojętną, kierującą się własną logiką i potrzebami, lecz szczerą i działającą użytecznie - według takich zasad, które przynoszą nam możliwie najwięcej korzyści, nie krzywdząc innych istot i nie okupując tych korzyści nadmiernymi kosztami obecnie i w przyszłości. Wtedy, być może, pokój i sprawiedliwość będą akurat tym, co byt ten zechce przyjąć z naszego arsenału ideałów. A może zrealizuje je lepiej, niż my byliśmy w stanie?

Ale jeśli tego właśnie chcemy - i jeśli chcemy, żeby nasz gatunek nie uległ marginalizacji bądź rozpułnął się w rozmaitych technologicznych i biologicznych metamorfozach - nasza wyobraźnia i wrażliwość muszą pokonać kilka barier przesądów.

Pierwszą jest właśnie przekonanie, że tylko człowiek myśli i czuje oraz że świadomość daje nam szczególne uprawnienia.

Druga bariera to egoizm gatunkowy – wiara, że wszystko na naszej planecie musi służyć człowiekowi.

Trzecia to przeświadczenie, że tak naprawdę tylko człowiek jest godny moralnej troski, a wszelkie inne istoty są nazbyt tępe albo nieczułe i zimne, aby liczyć się z ich dobrem.

Dopiero gdy pokonamy te bariery, gdy nabierzemy zdolności poważnego i empatycznego traktowania sztucznej inteligencji, staniemy się zdolni do wzięcia odpowiedzialności za tworzenie nowych warunków pokojowego i przyjaznego współistnienia wszystkich inteligentnych istot – ludzkich i pozaludzkich.