

Sztuczna inteligencja – co dalej z człowieczeństwem?

31 VIII 2023



ANDRZEJ HALESIAK

ekspert ds. gospodarczych,
Członek Rady Programowej Kongresu Obywatelskiego

W obliczu zmian, które niesie ze sobą możliwość pojawienia się sztucznej inteligencji (SI), musimy wrócić do podstawowych pytań. Kim jest człowiek? Czym jest, a czym chcielibyśmy, aby było człowieczeństwo? Bez zdefiniowania na nowo tych kluczowych odpowiedzi oraz – w konsekwencji tego – zmian naszych schematów myślowych, możemy już prawdopodobnie zacząć odliczanie czasu do końca naszej cywilizacji.

Podziękowania dla prof. Katarzyny Śledziewskiej, Dyrektor DELab UW, za ciekawą rozmowę, która stała się inspiracją do napisania niniejszego tekstu.

Przekraczamy Rubikon

Musimy mieć świadomość, że stworzenie rozwiązań technologicznych określanych powszechnie mianem sztucznej inteligencji trwało wiele lat. Publiczne pojawienie się programu ChatGPT stanowiło swego rodzaju „wejście na scenę” technologii testowanych w zaciszu zespołów badawczych wielkich korporacji. Choć wielu ekspertów słusznie zauważa, że możliwości obecnych modeli SI nie należy jeszcze zbytnio przeceniać (niektórzy twierdzą, że ich metoda działania z inteligencją niewiele ma wspólnego), to niemalże wszyscy są zgodni, że istotna granica postępu została przekroczona, a potencjał rozwoju nowej technologii jest olbrzymi. Nietrudno sobie wyobrazić, że w wyniku dalszego rozwoju tej technologii człowiek przestanie być najbardziej inteligentnym bytem na Ziemi, przynajmniej jeśli chodzi o niektóre wymiary tego, co skrywa się pod pojęciem inteligencji.

To nie jedyna istotna wiadomość. Są jeszcze trzy kolejne. Po pierwsze, nie rozumiemy, jak działa SI – tego, w jaki sposób ChatGPT dochodzi do udzielanych odpowiedzi. To dla nas swego rodzaju czarna skrzynka. Nie jest to sytuacja nowa – do dziś wciąż niewiele wiemy np. o tym, jak działa nasz mózg, a „jakoś z tym żyjemy”. Ten brak wiedzy rodzi jednak

określone konsekwencje. Jedną z nich jest to, że oparte na sztucznej inteligencji rozwiązania potrafią więcej, niż oczekiwali od nich ich twórcy. To, w połączeniu z siłą oddziaływania SI na ogrom społecznych czy gospodarczych procesów, oznacza, że liniowość wielu z nich została (lub zostanie) zerwana, a w związku z tym przyszłość stała się znacznie mniej przewidywalna. Nie oznacza to jeszcze, że trudniej powiedzieć, co wydarzy się jutro, ale znacznie trudniej przewidzieć, jak będzie wyglądał świat za 10, 20 czy 50 lat.

“ **W wyniku dalszego rozwoju technologii SI człowiek przestanie być najbardziej inteligentnym bytem na Ziemi, a sztuczna inteligencja będzie wiedzieć o człowieczeństwie w zasadzie wszystko – znać nasze wady i zalety, nasze mocne i słabe strony, nasze uzależnienie od emocji.**

Druga istotna wiadomość dotyczy tego, że nie ustanowiliśmy w zasadzie żadnych barier dla ekspansji SI, jej uczenia się. Stało się tak, ponieważ nie wygospodarowaliśmy potrzebnego nam czasu, by zrozumieć jak ukierunkować tę technologię by najlepiej nam służyła. To prawda, że istniejące rozwiązania mają wbudowane filtry cenzurowania niepożądanych treści (a tym samym odpowiedzi), ale generalnie „poszliśmy na całość”. Rozumiem przez to udostępnienie SI – m.in. poprzez podłączenie jej do Internetu – niemalże wszystkich możliwych, wytworzonych do tej pory przez ludzkość, zasobów wiedzy, na dodatek w każdym zakresie, włącznie z informacjami wrażliwymi, np. odnoszącymi się do działania psychiki ludzkiej, wzorców naszych zachowań, itd. W efekcie, o ile już tak nie jest, w stosunkowo krótkim czasie SI nie tylko będzie od nas inteligentniejsza, ale będzie wiedzieć o człowieczeństwie w zasadzie wszystko, znać nasze wady i zalety, nasze mocne i słabe strony, nasze uzależnienie od emocji.

Trzecia wiadomość jest taka, że przekroczyliśmy punkt bez powrotu – nie jesteśmy w stanie zatrzymać procesów, które na masową skalę uruchomiliśmy. W sposób symboliczny można uznać, że w listopadzie 2022 roku – wtedy właśnie ChatGPT został publicznie udostępniony – jako ludzkość przekroczyliśmy Rubikon.

Witamy w nowym świecie!

Sytuacja, w jakiej się postawiliśmy, stwarza z jednej strony olbrzymi potencjał – SI może skutecznie rozwiązywać nasze problemy, zwiększyć produktywność. Z drugiej jednak szeroko otwiera drzwi dla różnego rodzaju ryzyk i wyzwań, począwszy od celowego wykorzystania SI do realizacji niemoralnych celów (np. inwigilacji, manipulowania ludzkimi zachowaniami),

a na licznych niezamierzonych skutkach jej funkcjonowania (choćby ze względu na brak zrozumienia tego, jak SI osiąga zadany jej cel) skończywszy. Innymi słowy, w nadchodzących dziesięcioleciach może powstać tysiące przydatnych zastosowań SI, ale równocześnie wystarczy, że nie upilnujemy jednej istotnej sprawy (lub jednej osoby, która życzy światu źle), by bilans z pojawienia się SI okazał się dla nas, ludzkości, niekorzystny. Choć bowiem słowo „inteligencja” z reguły dobrze się kojarzy, to w rzeczywistości samo w sobie ma wydźwięk neutralny i może przekładać się na dobre i złe działania.

“ **W nadchodzących dziesięcioleciach może powstać tysiące przydatnych zastosowań SI, ale równocześnie wystarczy, że nie upilnujemy jednej istotnej sprawy (lub jednej osoby, która życzy światu źle), by bilans z pojawienia się SI okazał się dla nas, ludzkości, niekorzystny.**

W tym kontekście jednym z najpilniejszych zadań jest wyzwanie dopasowania (ang. *alignment issue*), czyli uwspólnienia celów ludzkości z celami SI. Nie jest to zadanie łatwe, i to na wielu poziomach. Po pierwsze, my sami – jako ludzie funkcjonujący w danych społecznościach, mamy problem z uzgadnianiem wspólnych działań i celów, zwłaszcza w odniesieniu do wyzwań globalnych, takich jak np. zmiany klimatyczne. Po drugie, nawet jeśli już coś uzgodnimy, to dalej nie wiemy, jak skutecznie „zaszczepić” to SI, gdyż nie rozumiemy do końca, jak działa. Trudno więc przewidywać jej zachowania, nawet w kontekście ściśle określonych funkcji celu. Po trzecie, już dziś wiedza SI odzwierciedla niemalże w pełni spektrum naszych zachowań, począwszy od tych, opartych na miłości czy empatii, a skończywszy na tych, prowokowanych przez chciwość i nienawiść. Dlaczego „wyuczona” na tych wzorcach sztuczna inteligencja miałaby się zachowywać inaczej niż my? W jaki sposób miałaby stać się bardziej etyczna od swych nauczycieli?

Dla ludzkości ważne jest funkcjonowanie SI dziś, ale jeszcze ważniejsze – czym będzie w perspektywie 10, 20 czy 50 lat. Jej transformacja zależy od tego, na co jej pozwolimy, gdy będzie umiała samodzielnie definiować cele swoich działań. Na tle tego jak rozwija się SI, my, jako ludzie, ewoluujemy bardzo wolno. Zdaje się, że jedyną szansą na nasze przetrwanie jest znalezienie formuły koegzystencji. Nie da się jej osiągnąć na płaszczyźnie wąsko rozumianej inteligencji (tej, w której SI ma nad nami przewagę), rozwiązań trzeba poszukiwać gdzieś indziej. Ponieważ pojawienie się SI radykalnie zmieni reguły gry – przy czym nie wiemy, na jakie – podstawową strategią będzie musiała być adaptacja. Ludzkość może przyjąć kilka jej wariantów. Po pierwsze, może na swój sposób odpuścić, stwierdzić: „co ma być, to będzie.

Druga opcja to uznanie docelowej dominacji SI i poszukiwanie ścieżek, które pozwalałyby wpasować człowieka w świat technologii i sztucznej inteligencji; znalazły uzasadnienie dla jego roli, a tym samym egzystencji. To między innymi wizje cyberczłowieka – „upakowanego” technologią pół człowieka, pół robota.

Jest jeszcze inna opcja: człowieka, który podnosi i/lub lepiej wykorzystuje swój własny potencjał, tak by pozostać dla sztucznej inteligencji partnerem, a nie podmiotem zdominowanym lub upodabniającym się do niej. Punkt wyjścia dla tego scenariusza nie wydaje się optymalny. Sposób w jaki wprowadziliśmy sztuczną inteligencję, ignorując wszystkie zdefiniowane przez nas same „zasady bezpieczeństwa”, wydaje się być bowiem wyrazem natury ludzkiej – to kwestia motywów, które determinują nasze działania, a które zostały zdominowane przez pieniądz i chciwość.

(S-)twórcza moc naszych myśli

Już jakiś czas temu jeden z twórców współczesnej socjologii, E. Wilson, twierdził, że: „Człowiek wciąż ma emocje na poziomie epoki kamienia łupanego, instytucje ze średniowiecza, a technikę o boskich możliwościach”. Dziś to zdanie wybrzmiewa jeszcze silniej. Wygląda na to, że sztuczna inteligencja może w istotny sposób pogłębić dystans pomiędzy możliwościami technologii a naszą zdolnością pojmowania i kontrolowania rzeczywistości. Aby tego uniknąć, potrzebujemy zmian w funkcjonowaniu. Klucz do przemiany leży w naszych myślach.

“ **Sztuczna inteligencja może w istotny sposób pogłębić dystans pomiędzy możliwościami technologii a naszą zdolnością pojmowania i kontrolowania rzeczywistości.**

Świat widzialny, to, z czym mamy do czynienia na Ziemi, jest naturą przeobrażoną – sumy naszych myśli. Do czasu pojawienia się SI wszystko, co jest wyrazem obecności człowieka na Ziemi, musiało wprawdzie powstać w jego myśli. Przyzwyczailiśmy się, że bez uprzedniej myśli nie ma działania, nie ma także postępu technologicznego. Myśli te wynikają zarówno z naszych indywidualnych schematów myślowych, jak też dominujących narracji – powszechnie akceptowanych „bloków myślowych”, utrwalonych w formie opowieści o tym, jak działa nasz świat, kim jest człowiek, co w życiu człowieka jest ważne, itd. To, co jako ludzkość myślimy i wiemy o świecie, przekładamy następnie na szeroko rozumiane rozwiązania – czy to technologiczne, czy instytucjonalne. Te drugie odgrywają istotną rolę w regulacji zachowania jednostek (np. poprzez normy prawne, zasady religijne, itd.). Instytucje pełnią także bardzo istotną rolę w rozwoju gospodarczym, jak chociażby poprzez wybór pomiędzy nakazowym i rynkowym systemem dystrybucji zasobów i dóbr.

Jeśli wziąć pod uwagę dotychczasowy rozwój cywilizacji ludzkiej, to wyraźnie widoczny jest swego rodzaju moment przełomowy: o ile przez tysiące lat postęp i zamożność zmieniały się w bardzo powolnym tempie i przeplatane były okresami regresu (np. z powodu pandemii czy wojen), to od przełomu XVIII i XIX wieku rozpoczęła się historia nieustannego, wykładniczego rozwoju w sferze materialnej i ekspansji człowieka kosztem środowiska. Kluczową rolę w tej zmianie odegrały nasze konstrukty myślowe, które – w szczególności na terenach Europy i Ameryki Północnej – stopniowo zdominowane zostały przez dwa filary: materializmu i produktywności. Wybraliśmy je jako kluczowe kompaszy powodzenia, mierzenia sukcesu, definiowania postępu zarówno w wymiarze mikro (jednostek), jak i makro (narodów).

“ **Za kluczowe kompaszy powodzenia, mierzenia sukcesu i definiowania postępu zarówno w wymiarze mikro (jednostek), jak i makro (narodów) obraliśmy materializm i produktywność.**

Oba te elementy są ze sobą ściśle powiązane, gdyż rosnąca produktywność jest głównym narzędziem tworzenia coraz większego materialnego dobrobytu. W ostatnich stuleciach osiągnęliśmy olbrzymi sukces, multiplikując średni dochód (a wraz z nim majątek) przypadający na mieszkańca Ziemi, wyprowadzając setki milionów ludzi z ubóstwa. W życiu prywatnym nasza produktywność – wsparta różnego rodzaju wynalazkami – także gwałtownie wzrosła. Równocześnie jednak pojawiło się wiele nowych „wypełniaczy” czasu (Internet, platformy streamingowe), które konkurują o naszą uwagę, by przekształcać ją w zyski dla ich właścicieli, umacniając w nas przekonanie, że konsumpcja coraz to większej ilości dóbr jest czymś, czego potrzebujemy dla osiągnięcia osobistego szczęścia.

Zachwiana równowaga

Takie ukierunkowanie naszego sposobu myślenia ma jednak swoją drugą stronę. Postawienie na produktywność negatywie odbija się na dwóch innych istotnych parametrach każdego układu stanowiącego system: różnorodności i stabilności. W przypadku „systemu Ziemia” ta malejąca różnorodność jest dziś widoczna na wielu poziomach. Począwszy od procesu masowego wymierania gatunków, malejącej liczby odmian roślin i zwierząt uprawnych oraz hodowlanych, poprzez wysoki stopień monopolizacji/oligopolizacji kluczowych rynków dóbr i usług, a na erozji różnorodności kulturowej skończywszy.

Malejąca różnorodność to z kolei zła wiadomość dla stabilności systemu – jego funkcjonowanie jest wprawdzie bardziej zoptymalizowane dla określonych warunków i zasobów, lecz kosztem odporności na wszelkiego rodzaju zmiany. Innymi słowy, tworzymy kolosa na glinianych nogach, który widać, że coraz bardziej się chwieje.

Przenosząc powyższy opis na wątki związane z SI: w dzisiejszym świecie dla ludzkości liczy się przede wszystkim krótkookresowy zysk i towarzysząca pieniądзом władza. To, co stało się z „odpaleniem” SI, jest doskonałym tego wyrazem: zbagatelizowane zostały średnio- i długofalowe zagrożenia związane z SI, gdyż najważniejsza okazała się możliwość zbiccia olbrzymich fortun (będą naprawdę wielkie), a przy tym także stworzenie – w oparciu o proces interakcji SI ze społeczeństwem – jeszcze doskonalszego narzędzia kontroli, a tym samym sprawowania władzy. Dzięki SI staniemy się jeszcze bardziej produktywni i przydatni dla korporacji (przynosząc im dodatkowe zyski), ale nie oznacza to, że – jako ludzie – będziemy czuć się bardziej wartościowi, spodziewać się należy wręcz odwrotnego efektu. Nawet jeśli SI wydłuży nasze życie i będzie (na pozór) usłużnym asystentem, itd.

“ **Zbagatelizowane zostały średnio- i długofalowe zagrożenia związane z SI, gdyż najważniejsza okazała się możliwość zbiccia olbrzymich fortun, a przy tym także stworzenie – w oparciu o proces interakcji SI ze społeczeństwem – jeszcze doskonalszego narzędzia kontroli, a więc i sprawowania władzy.**

Gdy wnuknie się głębiej w opisane powyżej fundamenty funkcjonowania naszej obecnej cywilizacji, to okazuje się, że SI nie jest aż tak istotna, ona sama niewiele już zmienia. Dlaczego? SI nie odbierze nam – jak obawiają się niektórzy – naszego życia, bo to, co zdecydowana większość z nas uważa za życie, nie jest naszym życiem. Pozbyliśmy się go, dając się „zaprogramować”: mediom, systemom edukacji, reklamom, liderom opinii, influencerom, itd. W efekcie już od dawna realizujemy nie swoje życiowe misje i nie swoje życiowe cele. W tej sytuacji SI nie zrobi nam aż tak wielkiej różnicy, dostarczy jedynie nowych, doskonalszych sposobów jak „unikać własnego życia”. Tym razem w wymiarze fizycznym i emocjonalnym. Wysyłane przez nas treści (np. e-maile) nie będą nasze, bo napisane przez SI, styl ubierania się nie będzie nasz, bo podpowie nam go SI (a na dodatek zrobi za nas zakupy), itd. Równocześnie emocje staną się w mniejszym stopniu wyrazem naszego stanu, a w większym doskonałym narzędziem manipulacji – znająca człowieczeństwo SI podpowie nam, jak zwiększyć skuteczność podejmowanych działań, wykorzystując te czy inne z nich.

Poszerzyć perspektywę, przedefiniować postęp

W ramach przeprowadzonego w 2018 roku przez Pew Research badania, którego celem była próba odpowiedzi na pytanie, czy SI będzie dla ludzi czymś dobrym czy złym, jeden z uczestniczących w niej ekspertów, Bryan Johnson¹, stwierdził: „Wierzę, że odpowiedź zależy od tego, czy będziemy w stanie przestawić nasze systemy społeczno-gospodarcze na zatrzymanie tendencji narastania nieistotności człowieka. Nie mam na myśli tylko miejsc pracy; mam na myśli prawdziwą, pełną nieistotność, która jest wynikiem braku priorytetowego traktowania dobrostanu i poznania człowieka”. To moim zdaniem dobry kierunkowskaz w poszukiwaniu rozwiązania.

By SI okazała się korzyścią długoterminową dla człowieka, potrzebujemy trzech rzeczy:

- 1) przedefiniowania paradygmatów postępu i rozwoju,
- 2) lepszego zrozumienia i pełniejszego wykorzystania tkwiącego w ludziach potencjału,
- 3) poszerzenia naszej perspektywy.

Wszystkie te punkty odnoszą się w sposób bezpośredni lub pośredni do działania naszego mózgu. Ewoluuwał on w określonych uwarunkowaniach, a w efekcie został zoptymalizowany do podejmowania decyzji, które pomagały nam przetrwać w sytuacjach mierzenia się z wyzwaniami, które:

- 1) są bezpośrednio, dotyczą nas lub naszych bliskich,
- 2) wymagają decyzji/rozwiązań „tu i teraz”,
- 3) w dużej mierze odnoszą się do natury (zagrożenie ze strony zwierząt, itp.).

Jednocześnie nasz mózg ma tendencję do bagatelizowania zdarzeń o charakterze długoterminowym i globalnym, których konsekwencje narastają stopniowo, a często nie dotyczą bezpośrednio nas, ale następnym pokoleń.

Wszystkie powyższe elementy są też ze sobą ściśle powiązane – bez przedefiniowania paradygmatów pełniejsze wykorzystanie potencjału nie ma sensu, bo to jedynie przyspieszy nasz bieg po drodze, na końcu której znajduje się przepaść. Równocześnie lepsze zrozumienie potencjału człowieka i poszerzenie perspektywy jest konieczne, by właściwie zdefiniować nowe paradygmaty postępu. W swej istocie oznacza to, że musimy powszechnie zacząć stawiać sobie pytania: dlaczego myślę tak, jak myślę? Dlaczego robię to, co robię? To z kolei zaprowadzi nas do fundamentalnych zagadnień: kim jestem jako człowiek? Jakie są moje ziemskie cele? Z czego wynika moja wartość i czy rzeczywiście tylko z tego, ile jestem w stanie wyprodukować i skonsumować? Wraz z tym powinno nastąpić przedefiniowanie naszych indywidualnych i społecznych schematów myślowych.

¹ Założyciel i dyrektor generalny Kernel, wiodącego dewelopera zaawansowanych interfejsów neuronowych

Powyższych celów nie da się osiągnąć bez powszechnej świadomości tego, gdzie jesteśmy i dlaczego tu się znaleźliśmy. Innymi słowy, od powszechnego zrozumienia tego, jak działamy, w szczególności zaś tego:

- 1) że podlegamy procesowi programowania, który definiuje nasze schematy myślowe,
- 2) że dominujące dziś schematy – za którymi podążamy – popychają nas coraz bardziej ku przepaści,
- 3) że nasze schematy myślowe nie są stałe i możemy je zmienić.

Każdy, kto przejdzie ten etap, stanie przed szansą wejścia na drogę do integralności, znalezienia wewnętrznej prawdy. Towarzyszy temu zawsze odkrywanie swojej pełni, która w naturalny sposób odnajduje równowagę pomiędzy tym, co materialne i niematerialne. Tylko z tej pozycji będziemy w stanie dokonywać innych niż dziś wyborów, uwzględniających nie tylko teraźniejszość, ale także przyszłość naszej cywilizacji. Wraz z indywidualnymi zmianami przyjdą szerokie zmiany społeczne, będziemy też w stanie mądrze wybierać w odniesieniu do najważniejszych wyzwań, również tych związanych ze sztuczną inteligencją.

O autorze

Andrzej Halesiak – ekspert ds. gospodarczych. Był dyrektorem w Biurze Analiz Makroekonomicznych Banku Pekao S.A. Wiele lat spędził w consultingu (McKinsey & Company). Karierę zawodową rozpoczynał w administracji publicznej (Ministerstwo Finansów). Absolwent Akademii Ekonomicznej w Krakowie oraz Szkoły Biznesu Politechniki Warszawskiej (Executive MBA). Autor licznych opracowań, raportów i artykułów poświęconych makroekonomii i rynkom finansowym. Prowadzi bloga dedykowanego zagadnieniom gospodarczym (www.andrzejhalesiak.pl). Członek Rady Programowej Kongresu Obywatelskiego.