



**D**rogi czytelniku, czy słyszałeś ostatnio o nowym pomysle biznesowym lub innowacji technologicznej, których realizację zatrzymano, bo budziły wątpliwości etyczne? Czy wiesz, nad czym pracują centra badawczo-rozwojowe w firmach? Czy jesteś w stanie określić konsekwencje społeczne i etyczne, jakie niesie ze sobą sztuczna inteligencja lub przemysł 4.0? Ja nie. Czy zatem nie siedzimy na tykającej bombie? →

- Wyzwania polityczne, społeczne i etyczne związane z nowymi technologiami i sztuczną inteligencją (AI) porównuje się dziś do tych, z jakimi mierzono się, gdy pięćdziesiąt lat temu nastąpił przełomowy moment w historii – ludzkość po raz pierwszy osiągnęła możliwość dokonania autodestrukcji. Dziś ponownie znaleźliśmy się w takim momencie w dziejach. A jednak nie prowadzimy na ten temat poważnej debaty.

### Cóż to jest dobro

Nie chcę wyjść na naiwniaka, który toczy bezsensowną walkę z fundamentalną zasadą współczesnego świata, jaką jest rozwój dla rozwoju. Wręcz przeciwnie: jestem optymistą, ale... To „ale” dotyczy po pierwsze postawy każdego z nas wobec etycznych wyzwań związanych z nowymi technologiami. A po drugie – wzięcia odpowiedzialności przez firmy za ich aktywność w zakresie badań i rozwoju.

Przedsiębiorczość to nie tylko kreowanie biznesu, ale również organizowanie społeczeństwa. Wymaga to odpowiedzialności przedsiębiorców. Jednak i my wszyscy – nie tylko politycy i naukowcy – o ile możemy ignorować to, co się dzieje w obszarze nowych technologii, o tyle już skutków tych zdarzeń ignorować nie zdołamy. Dlatego warto mieć chociaż wiedzę, a najlepiej kontrolę, żeby móc uniknąć negatywnych skutków społecznych albo przynajmniej je minimalizować. Tak czy inaczej w centrum rozważań etycznych na temat nowych technologii powinno stać dobro człowieka i Kantowska zasada: „Czyń tak, abyś nigdy nie używał człowieka tylko jako środka, ale zawsze również jako celu”. I tu zaczynają się schody. Konia z rzędem temu, kto jest w stanie odpowiedzieć na pytanie o to, czym to dobro człowieka jest.

### Demon społeczeństwa zmechanizowanego

W czasach błyskawicznego postępu, najprawdopodobniej u progu etapu zaawansowanych rozwiązań w zakresie sztucznej inteligencji, ze zdwojoną siłą wracają obawy o miejsce człowieka w świecie technologii. Erich Fromm już w 1968 roku w książce *Rewolucja nadziei. W stronę ucztowionej technologii* pisał, że pośród nas krąży nowy demon totalnie zmechanizowanego społeczeństwa: oddanego maksymalnej produkcji dóbr materialnych i ich konsumpcji, kierowanego przez komputery. Społeczeństwa, w którym człowiek, chociaż dobrze nakarmiony i zabawiany, pozostaje bierny<sup>1</sup>. A nawet więcej,

<sup>1</sup> E. Fromm, *Rewolucja nadziei. W stronę ucztowionej technologii*, tłum. A. Kochan, Kraków 2017.

**Są trzy elementy, które zdaniem Fromma pozwalają społeczeństwu odpowiedzieć na zagrożenia ze strony technologii: zaktywizowanie jednostki, przywrócenie kontroli człowieka nad systemem społecznym, uczłowieczenie technologii. Aby procesy te się powiodły, potrzebna jest jednak nadzieja. Bez niej nie dojdzie do zmian społecznych, których celem będzie pobudzenie ludzkiej aktywności, rozkwit świadomości i rozumu.**

jest kierowany i manipulowany, o czym w tym samym roku pisał Zbigniew Brzeziński. W świecie zaawansowanych technologii – twierdził Brzeziński – obywatele będą się skupiać wokół magnetycznych i atrakcyjnych osobowości skutecznie wykorzystujących techniki komunikacji, by manipulować emocjami i kontrolować umysły<sup>2</sup>.

Na tle ostatnich wydarzeń wyborczych na świecie nietrudno nam się odnaleźć w rzeczywistości zarysowanej przez Brzezińskiego. Aż trudno uwierzyć, że te diagnozy są teraz nawet bardziej aktualne niż pół wieku temu, kiedy były formułowane. Przycoczmy jeszcze jedną myśl z książki Fromma. Píše on, że być może najbardziej złowieszcze w sytuacji, w jakiej znalazł się człowiek, jest to, że tracimy kontrolę nad stworzonym przez nas systemem. Czy nie jest to opis świata sztucznej inteligencji? Wykonujemy decyzje, które za nas na podstawie obliczeń podjęły komputery. Fromm zadaje dramatyczne pytanie: jak doszło do tego, że człowiek u szczytu swojej dominacji nad naturą okazał się więźniem własnych wytworów i stanął w obliczu samozagłady?

### Nuda i bierność

Idea ratowania ludzkości jest jedną z tych beznadziejnych, niemożliwych, prowadzących raczej do paraliżu niż do zaangażowania. Wydaje się, że właśnie o zaangażowanie i wspólnotę – lub dokładniej: instytucje wspólnoty – tutaj chodzi. O wyrwanie nas z bierności, apatii, braku zainteresowania, braku poczucia wpływu i siły.

Fromm twierdził, że obecny system społeczny zrozumiemy znacznie lepiej, jeśli przyjrzymy się funkcjonującemu w jego ramach człowiekowi. Dostrzegł niezadowolenie niektórych ludzi z prowadzonego przez nich stylu życia, charakteryzującego się nudą i biernością, brakiem prywatności i depersonalizacją, jak również tęsknotę za radością i wartościowym istnieniem, które odróżniałoby ich z jednej strony od zwierząt, a z drugiej – od komputera. Wieszczył, że skoro niektórzy osiągnęli już pełną satysfakcję materialną i przekonali się, że raj konsumenta nie zapewnia szczęścia (a biedni, chociaż sami nie zaznali tego stanu, również widzą to, obserwując bogatych), to musi dojść do zmiany paradygmatu społecznego. Jednakże przez następne pięć dekad nic takiego się nie stało. Ostatnia duża rewolucja społeczna z roku '68 nie wpłynęła na poprawę zdiagnozowanej przez filozofa sytuacji.

<sup>2</sup> Z. Brzeziński, *The Technetronic Society*, „Encounter” 1968, t. 30, nr 1 (styczeń).

### Nadzieja potrzebna od zaraz

Są trzy elementy, które zdaniem Fromma pozwalają społeczeństwu odpowiedzieć na zagrożenia ze strony technologii: zaktywizowanie jednostki, przywrócenie kontroli człowieka nad systemem społecznym, uczłowieczenie technologii. Aby procesy te się powiodły, potrzebna jest jednak nadzieja. Bez niej nie dojdzie do zmian społecznych, których celem będzie pobudzenie ludzkiej aktywności, rozkwit świadomości i rozumu. Fromm wskazuje na różne rodzaje nadziei. Jedną z nich jest nadzieja bierna, którą autor charakteryzuje jako oczekiwanie na nadejście czasu. W takiej sytuacji ludzie niczego nie spodziewają się osiągnąć teraz, lecz w przyszłości lub nawet w innym świecie. Nie o taką nadzieję nam chodzi. Interesuje nas nadzieja, którą Fromm definiuje jako gotowość na to, co jeszcze się nie narodziło, przy jednoczesnym nieuleganiu rozpacz, gdy przez całe życie nie doczekamy się żadnych narodzin.

O tym, czy ktoś jest człowiekiem nadziei, nie dowiemy się z jego słów, lecz z wyrazu twarzy, sposobu chodzenia, zdolności okazania zainteresowania innym osobom, z braku fanatyzmu i z umiejętności wysłuchania rozsądnych argumentów. Dlaczego nie ze słów? Fromm tłumaczy, że konwencja społeczna narzuca na człowieka sukcesu, którym każdy pragnie być, konieczność odrzucenia lęku, uczucia znudzenia, zwątpienia, przygnębienia, nudy, rozpacz i samotności. Dodaje, że nadzieja to stan wewnętrznej gotowości do intensywnej, ale wciąż jeszcze potencjalnej aktywności.

### Działać, nie tylko być zajęтым

W tej definicji należy zwrócić uwagę na element działania, z którym zdaniem Fromma wiąże się jedno z najbardziej rozpowszechnionych złudzeń człowieka żyjącego w nowoczesnym społeczeństwie przemysłowym. Cała nasza kultura jest nastawiona na działanie rozumiane jako bycie zajęтым. Problem w tym, że większość ludzi uważających się za bardzo aktywnych nie jest świadoma, iż pomimo swojego zachowania pozostaje bardzo bierna. Ich aktywność jest często przejawem ucieczki przed niepokojem, który rodzi się, gdy ludzie ci zostają sami ze sobą.

Fromm wymienia kilka przejawów braku nadziei. Dziś wydają się one równie aktualne. Znużone twarze zwykłych ludzi, zanik kontaktów międzyludzkich, niemożność podjęcia skutecznych działań w zakresie przeciwdziałania zanieczyszczeniu powietrza i wody, głód w krajach ubogich. Przede

**Nie musimy dokładnie rozumieć, jak funkcjonują techniki związane ze sztuczną inteligencją, ale powinniśmy choćby w zarysie wiedzieć, czego się po nich spodziewać. To wyzwanie wydaje się szczególnie istotne z uwagi na to, że technologie oparte na AI pozwalają wywierać zdecydowanie bardziej rozległy i głęboki wpływ na nasze życie niż technologie analogowe i tradycyjne technologie cyfrowe.**

wszystkim zaś – nasze zaniechania i niemożność kontrolowania własnego życia. Nie jest łatwo zachować nadzieję, będąc świadomym, jaki wpływ na prawa podstawowe mają nowe technologie, w tym sztuczna inteligencja.

Jestem przekonany, że jednym z najistotniejszych zagadnień związanych z procesami rozwijania i wykorzystywania AI jest właśnie zapewnienie efektywnej ochrony praw podstawowych. Możliwość przetwarzania olbrzymiej ilości danych dotyczących obywateli sprawia, że podmioty i instytucje, które będą nią dysponować, uzyskują ogromną przewagę nad osobami, które na coraz to nowych obszarach stają się „dostarczycielami” danych. Autorzy dokumentu *Założenia do strategii rozwoju sztucznej inteligencji. Zagadnienia prawne i etyczne*<sup>3</sup> przygotowanego przez grupę roboczą w Ministerstwie Cyfryzacji zwrócili uwagę na liczne implikacje, jakie niesie sztuczna inteligencja w odniesieniu do podstawowych praw człowieka.

### **Wolność: rozumieć i decydować**

Najpierw ograniczenie prawa do wolności. Jednym z zasadniczych warunków wolności jest możliwość zrozumienia zdarzeń i procesów, z którymi mamy do czynienia, co pozwala na swobodne i niezależne podejmowanie decyzji. Chodzi o zrozumienie sytuacyjne, a nie techniczne. W wielu dziedzinach życia przywykliśmy już, że co do szczegółów sposób działania urządzeń technicznych jest dla nas niezrozumiały. Autorzy dokumentu podkreślają, że znakomita większość ludzi nie rozumie zasad działania pilota do telewizora, kuchenki indukcyjnej czy nawigacji satelitarnej, a mimo to potrafi z nich korzystać bez narażania się na szczególne ryzyko. Taki sam poziom zrozumienia powinniśmy mieć w odniesieniu do AI, z którą możemy się zetknąć w konkretnych sytuacjach życiowych. Nie musimy dokładnie rozumieć, jak funkcjonują techniki związane ze sztuczną inteligencją, ale powinniśmy choćby w zarysie wiedzieć, czego się po nich spodziewać. To wyzwanie wydaje się szczególnie istotne z uwagi na to, że technologie oparte na AI pozwalają wywierać zdecydowanie bardziej rozległy i głęboki wpływ na nasze życie niż technologie analogowe i tradycyjne technologie cyfrowe.

Innym ważnym elementem wolności jest swoboda decydowania o tym, z kim wchodzimy w interakcje, co wymaga istnienia wiarygodnego mechanizmu uwierzytelniania potencjalnych partnerów. Sposobem na zagwarantowanie takiego wiarygodnego mechanizmu jest wprowadzenie obowiązku ujawniania w interakcji z ludźmi, że drugą stroną komunikacji jest AI.

**3** S. Szymański, R. Sroka i in., *Założenia do strategii rozwoju sztucznej inteligencji. Zagadnienia prawne i etyczne*, nieopublikowany raport.



### Równość: bez uprzedzeń i manipulacji

Inną wartością etyczną, na którą negatywnie może wpłynąć sztuczna inteligencja, jest równość. W odniesieniu do systemów automatycznego podejmowania decyzji powszechnie dostrzega się dobrze już udokumentowany problem tzw. *algorithmic bias* (stronniczości algorytmów) i *data bias* (stronniczości danych). Polega on na powielaniu przez systemy decyzyjne oparte na nadzorowanym uczeniu maszynowym uprzedzeń autorów algorytmów lub osób dobierających zestaw danych. Aby zapobiec szerzeniu i utrwalaniu poglądów o charakterze dyskryminującym lub wykluczającym, należy chociażby wprowadzić przejrzyste mechanizmy kontroli stosowanych rozwiązań technicznych. Równocześnie systemy sztucznej inteligencji mogłyby być pomocne w bezstronnym wykrywaniu uprzedzeń w decyzjach podejmowanych przez ludzi.

Troska o równość wymaga również ochrony osób narażonych na uleganie manipulacji. Algorytmy wpływające na rodzaj odbieranych treści, dzięki mechanizmowi dostosowywania przekazu do indywidualnego profilu preferencji, mogą szczególnie silnie wpływać na osoby podatne na sugestie i manipulację na przykład ze względu na wiek (dzieci, osoby starsze) lub stan emocjonalny (osoby cierpiące na depresję lub inne zaburzenia nastroju). Profilowanie ze względu na preferencje może w przypadku takich osób pogłębiać stany depresyjne, prowadzić do uzależnień i dostarczać niepożądanych bodźców stanowiących zagrożenie dla ich zdrowia psychicznego i dobrostanu.

Nowe technologie, w tym sztuczna inteligencja, wpływają na realizację bądź ograniczenie realizacji wielu innych zasad: sprawiedliwości, odpowiedzialności, przejrzystości. A w centrum systemu wartości, o który trzeba zadbać, bo nikt tego za nas nie zrobi, jest ochrona godności ludzkiej.

### Uczłowieczyć technologię

Nie wiemy, jakie prace nad innowacjami prowadzą dziś firmy, i tym samym trudno jest nam chronić podstawowe wartości i zasady etyczne. Pozostaje nam jednak apelować do tych, którzy prowadzą te prace, czyli do świata biznesu, o wzięcie za nie odpowiedzialności etycznej. Na etapie rozpoczęcia prac nad nowymi technologiami często nie wiadomo, jaki będzie ich ostateczny kształt i jakie konsekwencje społeczne przyniosą. Proponuję więc, aby firmy wypracowały i wdrożyły skuteczny system etycznej oceny rozwijanych u siebie nowych technologii.

**Troska o równość wymaga również ochrony osób narażonych na uleganie manipulacji. Algorytmy wpływające na rodzaj odbieranych treści, dzięki mechanizmowi dostosowywania przekazu do indywidualnego profilu preferencji, mogą szczególnie silnie wpływać na osoby podatne na sugestię i manipulację na przykład ze względu na wiek lub stan emocjonalny.**

Chodzi o system, który będzie dostarczał obowiązkowych kryteriów oceny etycznej podejmowanych przedsięwzięć i umożliwił taką ocenę nie tylko na początku procesu, ale również w trakcie jego realizacji. Pozwoli to zidentyfikować pojawiające się nowe potencjalne czynniki oddziaływania na człowieka i społeczeństwo, a zwłaszcza na sposobność realizacji praw podstawowych. Po zdobyciu takiej wiedzy będzie można zaś podjąć działania prowadzące do minimalizacji negatywnego wpływu – czy to samemu, czy we współpracy z podmiotami z trzeciego sektora lub administracji państwowej. Firmy mogłyby powołać specjalny interdyscyplinarny zespół składający się z ekspertów wewnętrznych i zewnętrznych z różnych dziedzin: inżynierów, socjologów, etyków, psychologów. Jego zadaniem byłaby identyfikacja ryzyka etycznego prowadzonych działań badawczo-rozwojowych.

Wymagałoby to odwagi ze strony zarządzających firmami, którą Fromm nazywa hartem ducha; odwagi życia pozwalającej trwać w nadziei i uniknąć powierzchownego optymizmu lub irracjonalnej wiary. Ta odwaga, ten hart ducha to, jak powiada myśliciel, umiejętność powiedzenia „nie”, gdy świat chce usłyszeć „tak”.

Nie jestem w swoich opiniach dotyczących niebezpieczeństwa, jakie płynie ze strony rozwoju technologii i sztucznej inteligencji, tak radykalny jak Fromm. Pisał on 50 lat temu, że jesteśmy w samym środku kryzysu nowoczesnego człowieka i nie zostało nam wiele czasu. Że jeśli nie zaczniemy wprowadzać zmian od razu, to prawdopodobnie wkrótce będzie za późno. Niektóre jego obawy się nie sprawdziły. Część opisywanych zjawisk, jak choćby zmiana klimatu, przyjęła natomiast nieznane wtedy rozmiary. Wciąż warto jednak powtarzać za Frommem, że jest jeszcze nadzieja, ponieważ istnieje realna możliwość, że człowiek określi się na nowo i uczyni ludzkim to, co technologiczne. Choć nie od nas zależy wykonanie zadania, nie mamy żadnego prawa, aby się przed nim uchylać.