

## **Sebastian Szymański: Czy należy obawiać się sztucznej inteligencji?**

Wiele doświadczeń z wykorzystywanymi systemami AI wskazuje, że jej upowszechnienie może prowadzić do osłabienia naszej pozycji jako członków demokratycznego społeczeństwa – pisze Sebastian Szymański w „Teologii Politycznej Co Tydzień”: „Sztuczna inteligencja. Przedzione przyszłości?”

### **Spółeczny eksperyment z nową technologią**

Obawy dotyczące potencjalnych zagrożeń związanych z rozwojem i upowszechnianiem sztucznej inteligencji od początku towarzyszyły badaniom nad jej tworzeniem. W połowie lat 60. XX wieku Irving John Good, brytyjski matematyk, który podczas II wojny światowej pracował jako główny statystyk w Bletchley Park z Alanem Turingiem, zwrócił uwagę na następujący problem:

Zdefiniujmy maszynę ultrainteligentną jako maszynę, która dalece przewyższa na polu wszelkich aktywności intelektualnych dowolnego, choćby najinteligentniejszego człowieka. Ponieważ projektowanie maszyn jest jedną z tych aktywności intelektualnych, maszyna ultrainteligentna potrafi projektować jeszcze lepsze maszyny; dojdzie zatem bez wątpienia do „eksplozji inteligencji”, w wyniku której inteligencja człowieka pozostanie daleko w tyle. A zatem pierwsza maszyna ultrainteligentna będzie

ostatnim wynalazkiem, którego człowiek kiedykolwiek dokona – zakładając, że maszyna ta okaże się wystarczająco potulna, by powiedzieć nam, jak ją utrzymać pod kontrolą[1].

Współcześnie przedstawiciele wielu środowisk – w tym gwiazdy przemysłu informatycznego z Billem Gatesem i Elonem Muskiem na czele oraz przedstawiciele świata nauki tacy jak nieżyjący już Stephen Hawking – wyrażają podobne do Gooda obawy. Echa tych lęków pobrzmiwają również w kulturze masowej i publicystyce. Często słyszymy i czytamy o robotach, które odbiorą ludziom pracę, o autonomicznych systemach, które przejmą kontrolę nad instytucjami wojskowymi i politycznymi czy sztucznej inteligencji, która zniewoli lub zetrze ludzkość z powierzchni ziemi. Tego rodzaju apokaliptyczne przewidywania podsycane są przez komentatorów, których wyobrażenia rodem z *science fiction* mają niewielki związek z rzeczywistością i aktualnym stanem rozwoju techniki.

*Upowszechnianie sztucznej inteligencji (artificial intelligence, AI), którego jesteśmy świadkami, to dalekosiężny eksperyment społeczny, którego skutki nie są łatwe do przewidzenia*

Upowszechnienie każdej nowej technologii, od ognia i koła po komputery, ma dwie, bardzo istotne z etycznego punktu widzenia cechy: zabieg taki można przeprowadzić tylko

raz i nie ma on ścisłych analogii w procesach przeprowadzonych wcześniej w odniesieniu do innej technologii. Upowszechnianie sztucznej inteligencji (*artificial intelligence, AI*), którego jesteśmy

świadkami, to dalekosiężny eksperyment społeczny, którego skutki nie są łatwe do przewidzenia. Trudno przewidywać konsekwencje upowszechniania AI na podstawie zmian, jakie wywołało oparcie przemysłu na maszynie parowej czy komputeryzacja. Co więcej, nie jest to eksperyment etycznie obojętny – już dziś eksperci wskazują na szereg zagrożeń związanych z procesem rozpowszechniania rozwiązań opartych na sztucznej inteligencji, na przykład osłabianie pozycji użytkowników przez zmianę rozkładu odpowiedzialności. Z uwagi na brak solidnej wiedzy dotyczącej następstw zmian społecznych wywoływanych przez sztuczną inteligencję, podstawową zasadą etyczną w tej sferze powinna być zasada ostrożności. Głosi ona, że istnieje obowiązek ochrony ogółu przed szkodą albo krzywdą, kiedy wiadomo, że z badaniami naukowymi albo innowacjami technicznymi związane jest wiarygodne ryzyko. A o tym, że takie ryzyko istnieje i jest poważne, trudno dziś wątpić.

### **AI wąska i AI ogólna w potocznej wyobraźni.**

Szeroko rozpowszechniona, towarzysząca nam regularnie w codziennym życiu, sztuczna inteligencja to tak zwana wąska AI (*artificial narrow intelligence*, ANI) stworzona z myślą o realizacji jednego, wybranego przez projektanta celu. ANI wykonuje jedynie konkretne zadania, do których została stworzona, takie jak gra w szachy albo rozpoznawanie obrazów. Wielu użytkowników nie uświadamia sobie, że z wąskiej AI korzystają od dawna, na przykład korzystając z wyszukiwarki internetowej lub z asystentów głosowych takich jak Siri, Cortana, czy Alexa, optymalizując trasę przejazdu za pomocą map Google albo filtrując spam w swojej skrzynce mailowej. Obok innych nowych technologii wąska AI stała się już nieodłączną częścią naszego życia i wywiera na nie bardzo silny wpływ, zwykle

niedostrzegalny dla użytkowników. Jednak potoczna wyobraźnia utożsamia AI z tak zwaną ogólną AI (*artificial general intelligence*, AGI), czyli urządzeniem albo oprogramowaniem zdolnym do zastępowania wszystkich zdolności ludzkiego umysłu. Tymczasem ogólna AI nie istnieje, a sama możliwość jej stworzenia budzi szerokie kontrowersje wśród naukowców i inżynierów. Lęki dotyczące AI nakierowane są na tę jej odmianę, której – przynajmniej na razie – obawiać się nie ma powodu. Natomiast wąska AI, którą powszechnie wykorzystujemy już teraz, rodzi poważne problemy etyczne.

### **Czego się obawiać? Wyścig z odwróconym handicapem.**

*Gracze, którzy już na starcie są silni, stają się w nim coraz silniejsi. (...) Google, które już praktycznie dysponuje monopolem na wyszukiwanie w globalnej sieci, należy do firm przodujących w pracach nad stworzeniem komputera kwantowego*

W dziedzinie AI wiodącą rolę odgrywają globalne firmy, które górują nad indywidualnymi użytkownikami zasobem wiedzy i siłą rynkową. Co ważniejsze indywidualny użytkownik jest od nich podwójnie uzależniony: jako

źródło danych i jako konsument przetworzonych danych.

Konsekwencją tego jest gwałtowny wzrost nierówności sił w grze rynkowej. Prowadzi to do sytuacji, w której bierzemy udział w **wyścigu z odwróconym handicapem** – gracze, którzy już na starcie są silni, stają się w nim coraz silniejsi. Wystarczy przypomnieć, że Google, które

już praktycznie dysponuje monopolem na wyszukiwanie w globalnej sieci, należy do firm przodujących w pracach nad stworzeniem komputera kwantowego[2]. Połączenie tych dwóch technologii oznaczać będzie uzyskanie siły politycznej i ekonomicznej trudnej obecnie do oszacowania. Jedną z fundamentalnych wartości demokratycznych jest wzmacnianie pozycji jednostek. Mechanizmy kontrolne, prawa człowieka i innego rodzaju gwarancje instytucjonalne to rozwiązania, które mają sprawiać, że jednostki nie będą bezsilne wobec procesów politycznych i gospodarczych, wykraczających poza ich możliwości oddziaływania. Rozwój i upowszechnienie AI może w znaczący sposób wpłynąć na sytuację jednostek w demokracji. Wśród potencjalnych wyzwań, które rodzą istotne problemy etyczne, można wskazać trzy najważniejsze obszary.

## 1. Wpływ AI na pozycję użytkowników

Najbardziej dotkliwą konsekwencją upowszechnienia AI może się okazać **osłabienie pozycji użytkowników tej technologii w różnych sferach życia spowodowane nierównościami w dostępie do wiedzy i władzy oraz nieproporcjonalnym rozkładem ryzyka**. Doskonale widać to na przykładzie ryzyka związanego z danymi. Dla jednostki ogół jej danych stanowi integralny element tożsamości, a ich niekontrolowane ujawnienie może wywołać poważne, a często nieodwracalne szkody. Na przykład ujawnienie danych dotyczących zdrowia może diametralnie zmienić pozycję jednostki wobec ubezpieczyciela lub pracodawcy. Tymczasem dla firmy przetwarzającej duże ilości danych, pełny zestaw informacji o jednostce jest mało znaczący z punktu widzenia całego zbioru danych. Ten zbieg nierówności prowadzi do podwójnej niesprawiedliwości – raz poszkodowani tracą w dwójnasób, a wygrywający zyskują podwójnie.

Natomiast nierówności w dostępie do wiedzy prowadzą do rozpowszechnionego przekonania, że AI to *black box*, „czarna skrzynka”, której działania nie rozumieją i nie potrafią kontrolować nawet jej twórcy. Jednak tylko działanie dość wąskiej klasy algorytmów wykorzystujących losowość może być nieprzewidywalne nawet dla specjalistów. AI jest traktowana jako *black box* przede wszystkim dlatego, że zwykli użytkownicy, których interakcje z „czarną skrzynką” ograniczają się do dostarczania i użytkowania danych nie rozumieją jej i nie potrafią kontrolować. To brak edukacji w zakresie funkcjonowania AI, a nie sam sposób jej działania jest głównym źródłem ryzyka dla użytkowników

## 2. Wpływ AI na pozycję pracowników

*Oczywistym ryzykiem jest pogłębienie już istniejących i powstanie nowych nierówności. Praca z wykorzystaniem AI, czy może lepiej: współpraca z AI, wymaga szczególnych kompetencji i umiejętności*

Upowszechnienie AI może również przyczynić się do **osłabienia naszej pozycji jako pracowników.** Istnieją już dziesiątki raportów przewidujących poważne zmiany w strukturze zatrudnienia

związane z algorytmizacją pewnych dziedzin gospodarki. Przewidywania dotyczące liczby pracowników, których dotkną te zmiany, są bardzo rozbieżne i nie zawsze wiarygodne. Jednak nie ma wątpliwości, że będzie to proces o rozległych i głębokich

konsekwencjach, również etycznych. Najbardziej oczywistym ryzykiem jest  **pogłębienie już istniejących i powstanie nowych nierówności**. Praca z wykorzystaniem AI, czy może lepiej: współpraca z AI, wymaga szczególnych kompetencji i umiejętności. Chodzi tu nie tyle o kwalifikacje wymagane od programistów lecz o te potrzebne zwykłym użytkownikom tej technologii. Kolejnym źródłem niespotykanych dotąd problemów etycznych jest rozwój algorytmów uczących się i samouczących, które wykorzystywane są na przykład w pojazdach autonomicznych. W tym przypadku użytkowanie ich oznacza w praktyce ich uczenie. Z etycznego punktu widzenia oznacza to **zmianę rozkładu odpowiedzialności**, który dotychczas uznawano za oczywisty: producent odpowiada za funkcjonalności produktu, a użytkownik za to, jak je wykorzysta. W przypadku technologii opartych na algorytmach (samo)uczących się użytkownicy mają wpływ na ich funkcjonalności i zapewne będą ponosić część odpowiedzialności za ich działanie. Świadczą o tym między innymi prace nad regulacjami prawnymi w zakresie inteligentnej robotyki i AI prowadzone obecnie w Komisji Europejskiej i Parlamencie Europejskim, w których zakłada się, że integralną częścią regulacji prawnych będą kodeksy etyczne, również dla użytkowników.

### **3. Wpływ AI na pozycję obywateli i procesy polityczne**

Wiele doświadczeń z wykorzystywanymi systemami AI wskazuje, że jej upowszechnienie może prowadzić do **osłabienia naszej pozycji jako członków demokratycznego społeczeństwa**. Dbłość o etyczne wykorzystywanie AI wynikać powinna między innymi z faktu, że technologia ta **pozwala wywierać ogromny wpływ na procesy polityczne**. Najbardziej znanym tego przykładem jest głośny skandal związany z firmą Cambridge Analytica, która wykorzystywała dane

użytkowników Facebooka bez ich wiedzy i zgody do mikrotargetowania reklam między innymi podczas wyborów prezydenckich w Stanach Zjednoczonych. Choć co do faktycznego wpływu tych konkretnych zabiegów panują rozbieżne opinie, nie ma wątpliwości, że możliwość gromadzenia i przetwarzania ogromnej ilości danych o użytkownikach pozwala na manipulowanie opinią publiczną na niespotykaną dotychczas skalę. Nie bez znaczenie jest tu również fakt, że najwięksi i najważniejsi gracze na polu AI, „Big Five” – Google, Facebook, Amazon, Apple i Microsoft – to wielonarodowe korporacje mające siedziby w Stanach Zjednoczonych. Oznacza to, że za pomocą narzędzi tworzonych i wykorzystywanych przez te podmioty wpływa się na obywateli krajów, które mają zazwyczaj niewielkie możliwości oddziaływania na sposób funkcjonowania tych korporacji.

### **Godna zaufania AI – etyka na ratunek**

*Zaufanie do technologii AI  
buduje się na respektowaniu  
kluczowych wymagań:  
poszanowaniu podmiotowości  
człowieka i ludzkiemu  
nadzorowi nad technologią,  
technicznej solidności i  
bezpieczeństwa,  
poszanowaniu prywatności*

Najbardziej zauważalną konsekwencją upowszechnienia AI jest zatem osłabienie pozycji nas, „użytkowników końcowych” tej technologii na różnych płaszczyznach. Tym groźnym trendom przeciwdziała się za

pomocą regulacji etycznych, które nakładają wymóg respektowania

określonych wartości w projektowaniu, tworzeniu i wykorzystywaniu systemów sztucznej inteligencji. Najważniejszym przykładem takich rozwiązań są „Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji” przygotowane przez Grupę Ekspertów Wysokiego Szczebla ds. Sztucznej Inteligencji (AI HLEG) przy Komisji Europejskiej[3]. Celem tej regulacji jest zapewnienie, że w Europie będzie powstawać bezpieczna i godna zaufania AI. Zaufanie do technologii AI buduje się na respektowaniu kluczowych wymagań: poszanowaniu podmiotowości człowieka i ludzkiemu nadzorowi nad technologią, technicznej solidności i bezpieczeństwa, poszanowaniu prywatności w zarządzaniu danymi, promowaniu przejrzystości, różnorodności, niedyskryminacji i sprawiedliwości, dobrobytu społecznego i środowiskowego oraz odpowiedzialności za działanie AI. W wytycznych zawarto również metodę oceny oraz wskazówki dotyczące praktycznego wdrożenia każdego z wymogów. Metoda oceniania zgodności systemów AI z wartościami zawartymi w „Wytycznych” będzie testowana w procesie pilotażowym, w którym mogą wziąć udział wszystkie zainteresowane strony. Ponadto powołano do życia Europejski Sojusz Sztucznej Inteligencji (AI Alliance) jako forum wymiany najlepszych praktyk dotyczących wdrażania wiarygodnej AI oraz obserwatorium sztucznej inteligencji (AI Watch), które mają pomóc we wdrażaniu skoordynowanego planu działania w zakresie godnej zaufania sztucznej inteligencji made in EU już w 2020 roku. Analogiczne rozwiązania zawarte są w projekcie „Polityki Rozwoju Sztucznej Inteligencji w Polsce na lata 2019–2027”[4].

Nie mamy powodu, żeby obawiać się AI jako użytecznego narzędzia, którego wykorzystanie przyniosło już i przynosi nadal niezliczone korzyści na wielu polach, od ochrony zdrowia po badania naukowe. Naprawdę powinniśmy się obawiać tego, że dysponentami najdoskonalszych rozwiązań z zakresu AI są nieliczne firmy

posiadające ogromną siłą polityczną i ekonomiczną. Epoka sztucznej inteligencji może się stać epoką nowych, bezprecedensowych nierówności, jeśli tworzenie i wykorzystywanie sztucznej inteligencji nie zostanie wpisane w solidne ramy etyczne.

[1] I. J. Good, *Speculations concerning the first ultraintelligent machine*, w: F. Alt, M. Ruminoff (red.), *Advances in Computers*, t. 6, New York 1965, Academic Press, s 33;  
<https://exhibits.stanford.edu/feigenbaum/catalog/gz727rg3869>

[2] <https://www.technologyreview.com/f/614416/google-researchers-have-reportedly-achieved-quantum-supremacy/>

[3] <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

[4] <https://www.gov.pl/web/cyfryzacja/konsultacje-spoeczne-projektu-polityki-rozwoju-sztucznej-inteligencji-w-polsce-na-lata-2019--2028>  
Dokument opiera się na „Założeniach do strategii AI w Polsce”, w przygotowaniu których brałem udział, przewodnicząc grupie roboczej zajmującej się kwestiami etyki i prawa:  
[https://www.gov.pl/documents/31305/436699/Założenia\\_do\\_strategii\\_AI\\_w\\_Polsce\\_-\\_raport.pdf/a03eb166-0ce5-e53c-52a4-3bfb903edf0a](https://www.gov.pl/documents/31305/436699/Za%C5%82o%C5%BCzenia_do_strategii_AI_w_Polsce_-_raport.pdf/a03eb166-0ce5-e53c-52a4-3bfb903edf0a)



Sfinansowano przez Narodowy Instytut  
Wolności - Centrum Rozwoju  
Społeczeństwa Obywatelskiego  
ze środków Programu Rozwoju  
Organizacji Obywatelskich  
na lata 2018 – 2030



Ministerstwo  
**Kultury**  
i Dziedzictwa  
Narodowego.

Dofinansowano  
ze środków Ministra  
Kultury i Dziedzictwa  
Narodowego