

dr inż. Franciszek Wyrwa

Urząd Miasta i Gminy Margonin

dr Michał Wyrwa

Uniwersytet im. Adama Mickiewicza w Poznaniu, Wydział Psychologii
i Kognitywistyki

Etyka Graalem sztucznej inteligencji

Abstrakt

Referat podejmuje problematykę sztucznej inteligencji (SI) i jej wpływu na edukację. Dynamiczny rozwój i upowszechnianie się technologii generatywnej SI we wszystkich przestrzeniach społecznych jest historycznie bezprecedensowe. Niemniej relatywnie niewiele uwagi poświęca się sektorowi edukacji. Jest to tym bardziej groźne, że wypracowywanie norm nie może odbywać się wyłącznie na papierze – muszą zostać zinternalizowane przez społeczeństwo. Wykonania tego zadania musi podjąć się sektor edukacyjny. Etyka SI, prężnie rozwijająca się także w kontekstach pozaakademickich, mimo rosnącej presji potentatów technologicznych jest naszym zdaniem kluczem do przygotowania edukatorów w odpowiedzi na aktualne wyzwania związane z wdrażaniem SI do szkół i na uczelniach. W celu lepszego zrozumienia tematu prezentujemy zarówno wyłaniający się konsensus w zakresie rekomendacji etycznych odpowiedzialnego SI, jak i ich wprowadzania w proces nauczania.

1. Wstęp

„Chcemy maszyny, która może uczyć się na podstawie doświadczenia” – słowa Alana Turinga wygłoszone w trakcie wykładu dla Londyńskiego Towarzystwa Matematycznego w 1947 roku były znamienne dla rozwijających się w połowie XX wieku nauk informatycznych. Idea sztucznej inteligencji (SI) pojawiła się już w zasadzie w latach przedwojennych, kiedy Turing krytycznie odniósł się do rozważań nad myślącymi maszynami, proponując w ich miejsce ideę gry w udawanie – osławiony test Turinga – czyli efektywną metodę oceny sprawności, z jaką program komputerowy jest w stanie konwersować z człowiekiem. Wiara w możliwość konstrukcji maszyn wyposażonych w takie oprogramowanie była od początku silna. Weźmy choćby słowa drugiego z pionierów SI, Johna McCarthy'ego, który w 1956 roku powiedział: „Badanie [sztucznej inteligencji] ma opierać się na przypuszczeniu, że każdy aspekt uczenia się lub jakakolwiek inna cecha inteligencji może być w zasadzie tak dokładnie opisana, że można zmusić maszynę do jej symulacji” (Russel i Norvig 2010, s. 18).

Przez kolejne dekady XX wieku SI rozwijała się powoli przede wszystkim w ośrodkach naukowych, jednak szybko pojawiły się różnice w rozumieniu samego celu powstawania tej technologii. Jako część dyscyplin składowych kognitywistyki, wielu badaczy SI skupiało się na możliwości lepszego

rozumienia ludzkiego umysłu poprzez obliczeniowe modelowanie procesów poznawczych. Zdaniem Allena Newella i noblisty Herberta Simona (1976) – a także większości dzisiejszych kognitywistów – ośrodkowy układ nerwowy jest biologicznym komputerem. Modelowanie i puszczenie symulacji mogą więc być rozumiane dwojako (Searle, 1980, s. 429). Zwolennicy silnej SI są przekonani, że komputery o odpowiedniej architekturze i oprogramowaniu dosłownie mają umysły. Zwolennicy słabej SI natomiast są ostrożniejsi, twierdząc, że nie należy mylić efektywności modeli komputerowych z faktycznymi procesami poznawczymi realizowanymi przez organizmy istot żywych.

Gdzie stoimy dziś, biorąc pod uwagę dynamiczny rozwój SI w XXI wieku? Większość badaczy bierze ideę silnej SI za nieprzystającą do aktualnej technologii, a więc za należącą raczej do science fiction aniżeli nauki. Dobrym podsumowaniem są tu słowa Johna Searle'a, autora zresztą rozróżnienia silnej i słabej SI:

Aby skonstruować sztuczny umysł, musielibyśmy zbudować maszynę, która działa jak mózg i wytwarza świadomość. Nie wiemy jednak, jak działa mózg, zatem nie potrafimy go zbudować w sztuczny sposób. (...) sztuczna inteligencja może być pomocna w studiowaniu mózgu, tak jak jest pomocna w innych badaniach, ale, błędem jest przypuszczać, że symulacja czegoś jest realnie tym czymś (Searle, 2015).

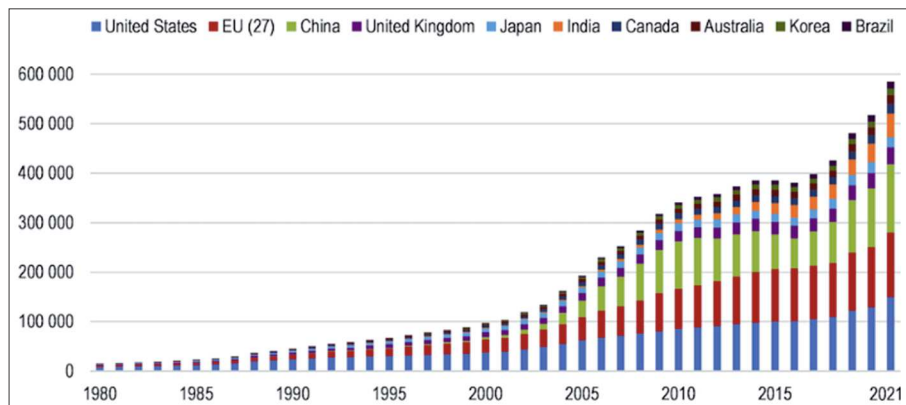
Technologiczny wybuch generatywnej SI w ostatniej dekadzie burzy obowiązujące dotychczas zasady. Tempo adopcji SI we wszystkich sektorach gospodarki i łatwość użytkowania narzędzi takich jak ChatGPT, już teraz wywołuje reperkusje cywilizacyjne. Ma to o tyle znaczenie, że trudno systematycznie przewidywać konsekwencje szerokiej adopcji nowoczesnej SI, co zgrabnie podsumował Stephen Hawking:

Potencjalne korzyści płynące z tworzenia inteligencji są ogromne. Nie możemy przewidzieć, co możemy osiągnąć, gdy nasze własne umysły zostaną wzmocnione przez SI. (...) Powstanie potężnej SI będzie albo najlepszą, albo najgorszą rzeczą, jaka kiedykolwiek spotkała ludzkość (Hawking, 2016).

Intensyfikacji ulec muszą więc prace przygotowujące społeczeństwo do obecności nie tylko algorytmów uczenia maszynowego w kodzie aplikacji, ale i bezpośredniego wchodzenia w interakcję z SI i korzystania z niego w pracy twórczej. Jako elementowi kognitywistyki, obszarowi SI od początku towarzyszy myśl filozoficzna (Brey i Soraker, 2009, s. 1342–1344), ale w ostatnich latach szczególnego znaczenia pozaakademickiego nabrała etyka sztucznej inteligencji. Na poziomie większości organizacji międzynarodowych, w tym UE, prowadzone są prace legislacyjne mające regulować rozwój i korzystanie z SI, prace, w których ta etyczna perspektywa wodzi prym. Jednym z mniej eksplorowanych aspektów jest sfera edukacji. Niestety, bo przecież to edukatorzy odpowiedzialni są za kształcenie i formowanie postaw jednostek, zarówno dzieci, młodzieży, jak i dorosłych. W niniejszym referacie charakteryzujemy wyzwanie dla edukacji, jakie stoi nie tyle w konsekwencji samej technologii generatywnej SI, ile w tempie i masowości zmian cywilizacyjnych, jakie ze sobą niesie. Rozpoczynamy od opisanego tempa, aby następnie przejść do próby określenia, jak etyka sztucznej inteligencji wpływa na kształtowanie się norm dotyczących odpowiedzialnego rozwoju i korzystania z SI. Kończymy, wskazując na rolę SI w edukacji i możliwość przygotowania edukatorów do pełnienia tej roli.

2. Era sztucznej inteligencji

Choć od 40 lat obserwujemy wzrost badań nad SI, to gwałtowny ich przyrost nastąpił od roku 2000, gdy liczba publikacji naukowych wynosiła 100 tysięcy, do ponad 550 tysięcy w 2021 roku. Dają się przy tym zaobserwować trzy obszary wpływu na rozwój technologii: Stany Zjednoczone, Unia Europejska i Chiny, które łącznie odpowiadają za ponad 70% publikacji (rys. 1).



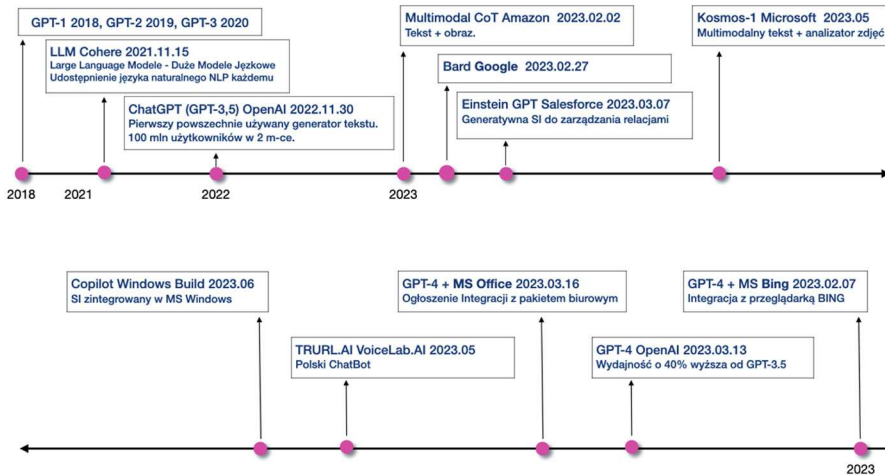
Rysunek 1. Liczba publikacji naukowych dotyczących SI w latach 1980–2021

Źródło: OECD.

Według prognozy IDC (*International Data Corporation*) globalne wydatki na SI osiągną w 2023 roku poziom 154 miliardów dolarów, co stanowić będzie wzrost o 27% w ciągu roku. Najwyższy poziom inwestycji przeznaczony zostanie na dwie branże: bankowość i handel detaliczny. Wzrost gospodarczy odegrał kluczową rolę w wydzwignięciu milionów ludzi z ubóstwa i przyczynił się do podniesienia standardu życia na całym świecie. Dzięki dynamicznemu rozwojowi gospodarczemu wiele społeczeństw osiągnęło znaczący postęp w zakresie dostępu do podstawowych usług, edukacji, opieki zdrowotnej i infrastruktury. Niestety, wraz z tym postępem ekonomicznym można zaobserwować pogłębiające się nierówności społeczno-ekonomiczne. Pomimo że wielu ludzi odniosło korzyści z rozwoju gospodarczego, inni nadal pozostają w tyle, trwając w ubóstwie i mając ograniczony dostęp do możliwości rozwoju. Zatem, mimo korzyści wzrostu gospodarczego, istnieje pilna potrzeba skoncentrowania się na redukcji nierówności społeczno-ekonomicznych i przyjęciu zrównoważonych praktyk gospodarczych. Obawy dotyczące rozwinięcia sztucznej inteligencji (SI) i jej potencjalnego wpływu na nierówności społeczno-ekonomiczne mają swoje uzasadnienie. Gwałtowny rozwój gospodarczy, który będzie wspierany przez SI, może prowadzić do jeszcze większych różnic między bogatymi a biednymi krajami oraz wewnątrz tych krajów.

W ostatnim czasie, a szczególnie w ciągu ostatniego roku, badania przyczyniły się do dynamicznego rozwoju narzędzi sztucznej inteligencji (rys. 2). Najbardziej znaczącym momentem było udostępnienie w listopadzie 2022 roku ChataGPT (GPT-3.5) – zaawansowanego bota konwersacyjnego opartego

na modelu językowym opracowanego przez OpenAI. Dodatkowo, integracja SI z systemem operacyjnym MS Windows, pakietem MS Office oraz przeglądarką MS Bing doprowadzi do jeszcze większego rozpowszechnienia narzędzi SI wśród zwykłych użytkowników. Wraz z dynamicznym rozwojem narzędzi SI ważne jest świadome korzystanie z nowych technologii oraz rozważenie potencjalnych skutków jej wpływu na kwestie prywatności, wiarygodności, stabilności zatrudnienia, predykcyjności, bezpieczeństwa i edukacji. Tworzona technologia z powodzeniem wdrożona zostaje w tak wielu branżach, że jej brak lub niedostateczny rozwój w niektórych sektorach, w tym w edukacji, skazuje te sektory na znaczące opóźnienia w rozwoju.



Rysunek 2. Oś czasu wybranych kamieni milowych rozwoju generatywnej SI konwersacyjnej

Technologie oparte na SI wymagają znacznych zasobów inwestycyjnych, infrastrukturalnych i wiedzy. Bogate kraje i korporacje są w stanie skorzystać z tych technologii i odnieść korzyści z ich zastosowania. Ważne jest więc, aby podejść do rozwoju SI z odpowiedzialnością społeczną i ekonomiczną. Wymaga to uwzględnienia działań mających na celu zniwelowanie potencjalnych skutków nierówności. Wspieranie innowacji i rozwoju SI powinno być naszym zdaniem połączone z inwestycjami w edukację, szkolenia zawodowe i rozwój umiejętności technologicznych wśród społeczeństwa w każdej jego populacji.

3. Etyka SI

Sztuczna inteligencja jest wyjątkowym przykładem technologii, która nie tylko wprowadza zmiany cywilizacyjne, ale również przekształca naszą codzienność. Porównując ją do technologii energetyki atomowej czy odnawialnej, które stanowią ogromną szansę dla ludzkości w walce ze zmianami klimatycznymi, można zauważyć, że dla przeciętnego użytkownika niewiele zmieniają się codzienne zachowania – nadal podłączamy sprzęt elektryczny do gniazdka w ścianie. Jednak w przypadku sztucznej inteligencji mamy zupełnie inną sytuację, ponieważ wpływa ona na charakter codziennego życia ogromnej liczby ludzi, którzy

żyją w skomputeryzowanych i zinternetowanych środowiskach. Dlatego istnieje pilna potrzeba nie tylko opracowania odpowiednich regulacji prawnych, jak to ma miejsce przy każdej nowej technologii, ale także kształtowania etycznego podejścia użytkowników i producentów do sztucznej inteligencji.

Jak wspomnieliśmy we wprowadzeniu, mamy przyptyw inicjatyw i funduszy. Według danych Organizacji Współpracy Gospodarczej i Rozwoju do końca 2021 roku na świecie istniało ponad 800 inicjatyw legislacyjnych i propozycji norm dotyczących wszystkich aspektów sztucznej inteligencji. Prawo często następuje już po zmianach społecznych i cywilizacyjnych, ale wydaje się, że obecna faza rozwoju SI jest zjawiskiem o niespotykanej dotąd mocy dokonywania zmian. Stopień inwestycji w firmy rozwijające SI rośnie wręcz wykładniczo na przestrzeni ostatnich lat, w szczególności generatywnego SI, dla którego to sektora odnotowano ponad 11 000-procentowy wzrost inwestycji w ostatnim kwartale 2022 roku, a więc w okresie upublicznienia testowej wersji ChatGPT przez OpenAI¹. Wraz z nakładami finansowymi i coraz większą świadomością potrzeby legislacyjnej, w drugiej połowie drugiej dekady XXI wieku pojawiło się zainteresowanie etycznymi aspektami SI (Borenstein i in., 2021).

Pomimo swojej młodości etyka sztucznej inteligencji (SI) jest obecnie jednym z najszybciej rozwijających się obszarów etyki stosowanej, skupiającej się na badaniu i tworzeniu norm zachowań wokół konkretnych zjawisk i problemów w świecie ludzkim. Jednocześnie, rozwój tej dziedziny jest wyrazem rozpoznawanych zagrożeń związanych z SI. Istnieje już blisko setka etycznych zaleceń dotyczących rozwoju sztucznej inteligencji, opracowanych przez badaczy akademickich, organizacje publiczne i sektor prywatny (Jobin i in., 2019). Choć różnią się one w swoich postulatach, w ostatnim czasie zaczyna się powoli wyłaniać konsensus dotyczący szczególnie wrażliwych aspektów sztucznej inteligencji, które mogą prowadzić do negatywnych konsekwencji społecznych. Przedstawiamy je w tabeli 1, uwzględniając najczęściej pojawiające się rekomendacje w ramach poszczególnych aspektów (Fjeld i in., 2020).

Tabela 1. Prezentacja obszarów konsensu etycznego oraz najczęstsze zalecenia przynależące do danych obszarów. Procenty wskazują na odsetek rekomendacji etycznych, w których dany obszar/zalecenie się pojawiły

OBSZARY KONSENSUSU	NAJCZĘSTSZE ZALECENIA W DANYCH OBSZARACH		
Prywatność (97%)	Kontrola użytkownika nad jego danymi (42%)	Konieczność wyrażania zgody przez użytkownika (25%)	Możliwość ograniczenia przetwarzania danych użytkownika (11%)
Odpowiedzialność za skutki (97%)	Ocena wpływu; rekomendacje przyjęcia nowych regulacji prawnych (53%)	Wymóg ewaluacji i audytu (47%)	Weryfikowalność i replikowalność (36%)
Bezpieczeństwo i ochrona (81%)	Ochrona danych (bezpieczeństwo dla technologii) (67%)	Bezpieczeństwo (użytkowania) (61%)	Przewidywalność działania (11%)

¹ Dane dostępne pod adresem <https://oecd.ai/> [wszystkie linki w niniejszym referacie mają datę dostępu 7.07.2023].

OBSZARY KONSENSUSU	NAJCZĘSTSZE ZALECENIA W DANYCH OBSZARACH		
Transparentność i wyjaśnialność (94%)	Wyjaśnialność (78%)	Transparentność działania (72%)	Użycie otwartych danych i algorytmów (28%)
Równość (fairness) i niedyskryminacja (100%)	Niedyskryminacja i zapobieganie uprzedzeniom SI (89%)	Równość (fairness) (56%)	Inkluzywność projektowa (47%)
Ludzka kontrola nad technologią (69%)	Ludzka kontrola nad technologią (64%)	Ludzki nadzór nad zautomatyzowanymi decyzjami (33%)	Możliwość wyłączenia zautomatyzowanego procesu decyzyjnego (8%)
Zawodowa odpowiedzialność (obowiązek) (78%)	Kolaboracja udziałowców (64%)	Odpowiedzialny design (44%)	Rozważenie efektów długofalowych (33%)
Promowanie wartości ludzkich (69%)	Przyczynianie się do wzrostu dobrobytu (64%)	Ludzkie wartości (44%)	Dostęp do technologii (31%)

Obecnie już nie jest zaskoczeniem obecność komórek zajmujących się aspektami etycznymi i społecznymi w rozwoju technologii i medycyny w rządowych i pozarządowych organizacjach publicznych. Jednak etyka sztucznej inteligencji (SI) jest w tym względzie wyjątkowa, ponieważ od samego początku cieszy się największym zainteresowaniem również ze strony komercyjnych organizacji. Wiodące firmy technologiczne utworzyły kolejne multidyscyplinarne zespoły, które pracują nad odpowiedzialnym rozwojem rozwiązań opartych na SI. Przykłady tego typu zespołów można znaleźć w firmach takich jak: Amazon, Google, Meta, Microsoft czy Twitter. W przeciwieństwie do czysto akademickich badań, które często mają ograniczony wpływ na rzeczywiste rozwiązania, specjaliści pracujący w tych zespołach rzeczywiście wpływają na rozwój technologii sztucznej inteligencji. Niektóre wybitne przykłady to prace Cynthii Dworck i Aarona Rotha dotyczące metod ochrony prywatności danych wykorzystywanych w rozwoju SI (2014), badania Timnit Gebru i Joy Buolamwini, które wykazały rasizm i seksizm w algorytmach SI do rozpoznawania twarzy, wynikające z niezrównoważonych zbiorów danych treningowych (Buolamwini i Gebru, 2018), oraz opracowane przez zespół Margaret Mitchell standardy przejrzystego szkolenia modeli SI (2018).

W ciągu ostatnich kilku lat zauważono istotne zmiany w podejściu głównych graczy technologicznych do kwestii etyki. Na przykład Microsoft drastycznie zmniejszył swój zespół etyczny, z około 30 osób w 2020 roku do 7 osób w 2022 roku, a następnie w marcu 2023 roku zlikwidował go całkowicie (Schiffer i Newton, 2023). Meta, firma należąca do Marka Zuckerberga, również zlikwidowała swój około 20-osobowy zespół etyczny we wrześniu 2022 roku, podobnie jak Twitter, który zwolnił 17 z 18 osób odpowiedzialnych za rozwój odpowiedzialnych usług (de Vynck i Oremus, 2023). Podobne działania miały miejsce również w platformie streamingowej dla graczy, Twitch, należącej do Amazona. W 2021 roku Google zwolnił Timnit Gebru i Margaret Mitchell, dwie pionierki etyki SI i założycielki zespołu etycznego w Google, po tym, jak opublikowały wyniki dotyczące zagrożeń i słabości dużych modeli językowych

(*Large Language Models*, LLMs), które stanowią podstawę usług takich jak ChatGPT (Bender i in., 2021; Field, 2023). Choć Google zapowiedział zwiększenie zespołu etycznego w wyniku restrukturyzacji (Mickle, 2021), to zakończenie współpracy z niezależnymi badaczami stało się ich standardową praktyką. W 2022 roku Google próbował zablokować publikację El-Mahdi El-Mhamdiego, która ujawniła podatność dużych modeli SI na cyberprzestępczość, co doprowadziło do zerwania współpracy (Alba i Love, 2023). W tym samym roku Google zwolnił Satrajitę Chatterjee po jego pracach krytykujących wiarygodność i bezpieczeństwo wykorzystania SI w produkcji komputerowych układów scalonych (Wakabayashi i Metz, 2022).

Trudno znaleźć inny powód takich działań, jak wyścig między modelami generatywnymi SI, który osiągnął punkt kulminacyjny wraz z udostępnieniem testowej wersji ChatGPT przez OpenAI w listopadzie 2022 roku. Microsoft zainwestował miliard dolarów w OpenAI w 2019 roku, a do wiosny 2023 roku łączna suma inwestycji w firmę wyniosła 18 miliardów dolarów (Novet, 2023). W początkowym okresie bieżącego roku Google również wprowadził na rynek swoje generatywne SI językowe Bard, jednak zakończyło się to fiaskiem. Mimo negatywnej wewnętrznej oceny ryzyka oraz oporów ze strony pracowników, firma nazwała swoje produkty „wersją eksperymentalną” (Alba i Love, 2023; Google, 2023).

Obecnie jesteśmy daleko od wydarzeń z 2016 roku, kiedy Microsoft, po zaledwie 16 godzinach, zdecydował się zakończyć projekt bota Tay po wykazaniu przez niego mowy nienawiści (Schwarz, 2019). W dzisiejszych czasach żadna firma nie rozważa wyłączenia swoich generatorów SI do tworzenia tekstów, obrazów czy dźwięku ze względu na konsekwencje ekonomiczne. Choć coraz większy odsetek firm technologicznych deklaruje znaczenie etyki SI w swojej pracy, to większość z nich jeszcze nie podjęła odpowiednich działań w celu jej wdrożenia (Rossi, Rudden i Goehring, 2023). Przykłady działań wiodących graczy technologicznych, które zostały wcześniej wspomniane, sugerują raczej, że dążenie do dominacji na rynku SI zaczęło przysłonić potencjalną, ale być może fasadową, dobrą wolę firm w zakresie dostarczania bezpiecznych usług.

Naiwne jest więc oczekiwanie, że o dopasowanie SI (*alignment*), a więc także dostosowanie pod kątem wpływu na ludzki dobrostan, do reprezentujących różnorodnie grupy społeczne użytkowników, będzie zapewnione przez samych ich producentów. Cytując słowa Josha Simsona, byłego badacza etyki SI z Meta:

Zespoły Odpowiedzialnego SI są jednymi z nielicznych wewnętrznych bastionów wykorzystywanych przez korporacje technologiczne do zapewnienia, aby inżynierowie pamiętali o ludziach i społecznościach, dla których budują swoje rozwiązania SI. Prędkość, z jaką te zespoły upadają, zostawia firmy technologiczne na łasce imperatywów marketingu, podkopując dobrostan dzieci, podatnych grup społecznych i naszej demokracji (za: Criddle i Murgia, 2023).

Pozostawienie kwestii etyki SI wyłącznie w rękach firm technologicznych byłoby nieodpowiedzialne, ponieważ etyczność SI dotyczy nie tylko samego procesu tworzenia modeli i zbiorów danych, ale także użytkowania tych technologii. Obok wymiaru technicznego jest też wymiar etyczności użytkownika. Niezależnie od mechanizmów wpływających i dostosowujących się do użytkowników w usługach opartych na SI każdy użytkownik ma ostatecznie

prawo do korzystania z nich w odpowiedzialny sposób, zarówno dla siebie, jak i dla innych. Świadomość tego faktu i wypracowanie norm postępowania w tej kwestii nie mogą być pominięte w dziedzinie edukacji zarówno dzieci, jak i dorosłych, biorąc pod uwagę niezwykle szybkie tempo adopcji tych technologii.

Warto także zauważyć, że obecne standardy odpowiedzialnego rozwoju SI, przedstawione w tabeli 1, nie zawierają bezpośrednich wskazówek dotyczących potrzeby kształtowania etycznego użytkownika SI przez samych użytkowników. Ten obowiązek spoczywa zatem na edukatorach, którzy powinni włączyć tę tematykę do swojej pracy.

4. Przyszłość SI w edukacji

Rola sztucznej inteligencji w globalnej konkurencji spotyka się z dużym zainteresowaniem ze strony państw i organizacji międzynarodowych, które dostrzegają szansę na społeczny postęp. Kilka państw, takich jak Chiny i Finlandia, już wprowadziło programy promujące rozwój sztucznej inteligencji w społeczeństwie (Niemi, 2023, s. 265). Rola liderów rozwoju w tym kontekście jest niezwykle istotna, ponieważ bez ich zaangażowania nie wszyscy będą mogli w równym stopniu korzystać z dobrodziejstw technologii. Jednocześnie istotne są również działania filantropijne. Jak zauważył Bill Gates, „siły rynkowe nie będą naturalnie wytwarzać produktów i usług SI, które pomogą najuboższym (...) będziemy musieli skupić najlepsze na świecie rozwiązania SI na rozwiązywaniu ich największych problemów” (2023).

Jednym z kluczowych wyzwań jest wprowadzenie sztucznej inteligencji do systemów edukacyjnych. W ramach celu Zrównoważonego Rozwoju ONZ dąży się do zapewnienia wysokiej jakości edukacji dostępnej dla wszystkich. Aby osiągnąć ten ambitny cel, konieczne jest odpowiednie szkolenie i wsparcie dla kadry nauczycielskiej. W kontekście wdrażania sztucznej inteligencji w edukacji należy się zastanowić, kto jest odpowiedzialny za kompleksowe wprowadzenie tych rozwiązań. Jakże cele chcemy osiągnąć poprzez to wdrożenie i jak możemy dojść do tego punktu?

Jasno widzimy wiele negatywnych aspektów, których jako edukatorzy intuicywnie chcemy unikać. Przede wszystkim, obawiamy się, że sztuczna inteligencja będzie wykorzystywana przez uczniów nie jako narzędzie wspomagające, ale jako środek eliminujący wyzwania edukacyjne. Obecnie studenci masowo korzystają z usług wsparcia ChatGPT do pisania zadań domowych, mimo że zdają sobie sprawę, że jest to forma oszustwa (Intelligent.com, 2023). Oczywiście, uczniowie zawsze próbowali łamać zasady obowiązujące w systemach edukacyjnych (Rettinger i Gallant, 2020). Generatywna SI zmienia reguły gry o tyle, że nagle dostęp do tworzenia odpowiedzi na zadania na wystarczającym poziomie jest bezkosztowy, natychmiastowy i łatwy.

Nierozsądne korzystanie z narzędzi takich jak ChatGPT sprawia, że uczniowie nie rozwijają umiejętności krytycznego myślenia, weryfikacji informacji i samodzielnego rozwiązywania problemów. Biorąc pod uwagę inherentne uprzedzenia w danych, na których uczy się sztuczna inteligencja, odzwierciedlających społeczne nierówności i stereotypy, istnieje ryzyko, że uczniowie

wewnętrznie przyjmą te uprzedzenia w generowanych treściach. Sam fakt, że uczniowie polegają na generatorach do wykonywania pracy intelektualnej, ma poważne konsekwencje dla ilości i wiarygodności zdobywanej wiedzy, a także dla braku rozwoju postaw prospołecznych i proedukacyjnych. Reperkusje sztucznej inteligencji w edukacji nie ograniczają się jednak tylko do tego (zob. też Lo, 2023). Istnieje także zagrożenie dla miejsc pracy nauczycieli – po co zatrudniać wielu tutorów, skoro można uruchomić tyle instancji korepetytora opartego na SI, ilu jest uczniów? Tym bardziej że stosunkowo niewiele nauczycieli rozumie i korzysta z technologii sztucznej inteligencji, zwłaszcza w krajach o wysokim średnim wieku nauczycieli, takich jak Polska.

Dowodem na nieuniknione dominowanie sztucznej inteligencji w dziedzinie pracy intelektualnej i rozwojowej są znaczące nakłady finansowe oraz tempo adopcji (100 milionów użytkowników ChatGPT w trzy miesiące). W takiej sytuacji kluczowe staje się holistyczne podejście, które obejmuje zarówno edukację młodszych pokoleń w odpowiedzialnym korzystaniu z SI, jak i przygotowanie edukatorów do prowadzenia procesu nauczania opartego na tym podejściu.

5. Etyka SI w edukacji

W sektorze edukacji priorytetowym zasobem, który zaczyna być wykorzystywany, jest etyka sztucznej inteligencji (SI). Od lat istnieją pojedyncze kursy dotyczące tego tematu, zarówno na poziomie akademickim, jak i w formie kursów online. Jednak dopiero niedawno zaczęły powstawać odpowiednie programy studiów. Przykładem jest inicjatywa CFI, która wprowadziła jedne z pierwszych na świecie studiów magisterskich z zakresu AI Ethics and Society, przygotowujących specjalistów do pracy w obszarze etyki sztucznej inteligencji. Celem tych studiów jest wspieranie rozwoju sztucznej inteligencji w odpowiednim kierunku. Programy te mają silne skupienie na praktycznym zastosowaniu, z naciskiem na zrozumienie i umiejętności programowania uczenia maszynowego. Co ważne, specjaliści z dziedziny etyki sztucznej inteligencji są już zatrudniani w szkołach podstawowych i średnich, wprowadzając elementy sztucznej inteligencji w zaktualizowane programy nauczania (tab. 2).

Szybka adaptacja i modyfikacja powszechnego systemu kształcenia jest kluczowa w kontekście szkolenia specjalistów ds. odpowiedzialnych relacji ze sztuczną inteligencją. W przeciwieństwie do wielu rewolucji technologicznych z przeszłości, korzystanie z generatorów SI jest obecnie łatwo dostępne. Ważne jest, aby podczas procesu edukacji uczniowie nie tylko zrozumieli specyfikę narzędzi SI, ale i to, że głównym celem systemu szkolnego jest zdobywanie wiedzy i umiejętności. Samo opanowanie i umiejętność korzystania z narzędzi SI nie odzwierciedla pełnego zrozumienia przedmiotu nauki. Innymi słowy, jak to zostało zapisane w austriackim programie nauczania dla szkół podstawowych: „Uczniowie rozumieją również dylematy etyczne związane z korzystaniem z takich technologii i stają się aktywnymi uczestnikami społecznego dyskursu na ten temat” (AI Index, 2023, s. 262).

Tabela 2. Poziom wdrażania edukacji SI w krajowych programach nauczania. W większości państw świata takie wdrożenia nie są jeszcze przygotowywane

Kraj	Status	Szkoła podstawowa	Gimnazjum	Szkoła średnia
Armenia	Zatwierdzone i wdrożone		+	+
Austria	Zatwierdzone i wdrożone			+
Belgia	Zatwierdzone i wdrożone			+
Chiny	Zatwierdzone i wdrożone	+	+	+
Indie	Zatwierdzone i wdrożone		+	+
Kuwejt	Zatwierdzone i wdrożone	+	+	
Portugalia	Zatwierdzone i wdrożone	+	+	+
Katar	Zatwierdzone i wdrożone	+	+	+
Serbia	Zatwierdzone i wdrożone		+	+
Korea Południowa	Zatwierdzone i wdrożone			+
Zjednoczone Emiraty Arabskie	Zatwierdzone i wdrożone	+	+	+
Bułgaria	Przygotowywane	+	+	+
Niemcy	Przygotowywane	+	+	+
Jordania	Przygotowywane		+	+
Arabia Saudyjska	Przygotowywane	+	+	+
Serbia	Przygotowywane		+	+

Źródło: OECD, *The AI index 2023 annual report*.

Coraz większy wpływ na kształtowanie sztucznej inteligencji mają niewielkie, ale bogate podmioty z sektora prywatnego, zamiast szerokiego grona podmiotów społecznych, w tym związanych z edukacją. Przeglądając kalendarz wpływowych konferencji i seminariów na światową skalę poświęconych sztucznej inteligencji w tym roku, trudno znaleźć takie, które skupiają się na tematyce edukacyjnej. Globalne konferencje przyciągają przedstawicieli dużych firm technologicznych, naukowców i startupy, transmitując swoje wydarzenia na całym świecie. Wybór lokalizacji tych konferencji również nie jest przypadkowy, często odbywają się w miejscach takich jak: Boston, Santa Clara, Nowy Jork, Londyn, a nawet w Zjednoczonych Emiratach Arabskich (Hopkin, 2023). Większość tych konferencji uwzględnia kwestie etyczne związane ze sztuczną inteligencją, ale nie koncentrują się na aspektach edukacyjnych.

Jedną z niewielu konferencji o tematyce edukacyjnej, choć o mniejszym zasięgu, była *AI in the Service of Teaching and Learning*² zorganizowana przez Uniwersytet Stanforda. Podczas tego wydarzenia poszukiwano odpowiedzi na pytania dotyczące wykorzystania sztucznej inteligencji w usprawnianiu procesu nauczania i nauki poprzez nowe metody pedagogiczne oraz nowe formy oceniania.

Wyzwaniem było również zastanowienie się, w jaki sposób sztuczna inteligencja może przekształcić proces nauczania i uczenia się w sposób etyczny, sprawiedliwy i bezpieczny. Konferencja zakończyła się stworzeniem listy

² <https://hai.stanford.edu/tlahai-agenda>.

rekomendowanych narzędzi SI. Wiele wystąpień z konferencji jest dostępnych na kanale YouTube Stanford HAI³. Poniżej przedstawiamy wybrane wnioski wynikające z tego wydarzenia.

1. Spersonalizowane wsparcie dla nauczycieli. SI daje możliwość wspierania nauczycieli w doskonaleniu ich rzemiosła na dużą skalę dzięki aplikacjom służącym do:
 - a. symulowania uczniów (SI jako uczniowie dla młodych nauczycieli);
 - b. informacji zwrotnych w czasie rzeczywistym (bank porad na żywo opartych na pedagogice eksperckiej);
 - c. raportowania po lekcjach przebiegu zajęć;
 - d. aktualizowania wiedzy dziedzinowej (nauczyciel na bieżąco z nowościami).

Przykładowe narzędzia dla usprawnienia procesu oceniania: Gradescope i Nuance Dragon.

2. Dla uczniów:
 - a. Zmiana tego, co jest ważne dla uczących się: SI nie wyręczy uczniów, lecz podniesie poprzeczkę wymagań. Zmusi ich do głębszego zaangażowania się w rozwiązywanie zadań. Staną się bardziej kreatywni.
 - b. Umożliwi wyzbycie się strachu przed ocenami. SI wspierać będzie motywację i pewność siebie uczniów. Uczący się są bardziej skłonni do zaangażowania, podejmowania ryzyka i bycia wrażliwymi. SI wspierać będzie umiejętności miękkie.
 - c. Technologie SI pozwolą na personalizację ścieżki edukacyjnej każdego ucznia.
 - d. Sztuczna inteligencja pozwala na spersonalizowane programy i naukę w tempie, które odpowiada uczniowi. Ta elastyczność zmniejsza obciążenie uczniów, co może mieć bardzo pozytywny wpływ na ich oceny końcowe.
 - e. Pomaga uczniom z trudnościami w uczeniu się zachować koncentrację (kanadyjski robot QT).

Przykładowe narzędzia: Altitude (wsparcie samodzielnej nauki), Cognii (indywidualny korepetytor), Knowji (nauka języka), Plaito (wirtualny nauczyciel), MATHia (nauka języka) oraz dla studentów: Knewton's Alta (kursy adaptacyjnego uczenia się).

3. Dla jakości uczenia się i oceniania:

Sztuczna inteligencja może szybko określić umiejętności ucznia, zaproponować rozwiązania w celu wypełnienia luk i dopasować je do ról, które faktycznie wymagają tych umiejętności.

Przykładowe narzędzia: Querium (pomoc w nauce), Century Tech (spersonalizowane planu nauczania).

³ <https://www.youtube.com/playlist?list=PLYLBSCrqNXzWmT6tr9GcdZCsuNAImdVg>.

6. Wnioski

Podsumowując, ryzyko etyczne związane ze sztuczną inteligencją jest ogromne, zwłaszcza gdy wynika z nieprzemyślanych działań korporacji i braku uwzględnienia negatywnych skutków SI już obecnego w świecie. Sztuczna inteligencja niesie ze sobą nowe sposoby realizacji istniejących od wieków zagrożeń. Potrzeba więc nowych metod blokowania tych ścieżek, tak abyśmy jak najrzadziej odczuwali chęć nimi podróżowania. Warunkowo zgadzamy się z pierwszą częścią myśli Hawkinga, że sztuczna inteligencja może być najlepszą rzeczą, jaka spotkała ludzkość. Nie można zatrzymać rewolucji związanej z SI, ale powinniśmy zaakceptować jej obecność w naszym życiu, o ile będzie użytkowana w sposób etyczny. W ten sposób rozwój sztucznej inteligencji staje się przynajmniej rozwojem, jak określa to Reid Blackman (2022), w dążeniu do „SI nie dla złego” (*AI for not bad*), a może nawet na rzecz „dobrego” (*AI for good*).

Szczególnie istotne staje się obecność narzędzi sztucznej inteligencji w edukacji. Bezpieczny użytek wymagać będzie nie tylko etycznego nauczania, ale również ich etycznego stosowania. Etyczne SI w edukacji to systemy, których głównymi zasadami są uczciwość, przejrzystość i odpowiedzialność, a nie jedynie cyfrowy odpowiednik dziesięciu przykazań. Aby wdrożyć je w praktyce jako użytkowników, konieczne są zmiany w podstawach programowych od najwcześniejszych etapów kształcenia, pamiętając przy tym, że kształtowanie etycznych postaw wymaga świadomego wysiłku i cierpliwości. Etyczny Graal nie jest drogą na skróty.

Bibliografia

- Alba D., Love J., *Google's Rush to Win in AI Led to Ethical Lapses, Employees Say*, Bloomberg, 2023, <https://www.bloomberg.com/news/features/2023-04-19/google-bard-ai-chatbot-raises-ethical-concerns-from-employees>.
- Bender E.M., Gebru T., McMillan-Major A., Shmitchell S., *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, [w:] *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, Nowy Jork 2021, 610–623, <https://doi.org/10.1145/3442188.3445922>.
- Blackman R., *Ethical machines: your concise guide to totally unbiased, transparent, and respectful AI*, Harvard Business Publishing, Boston 2022.
- Borenstein J., Grodzinsky F., Howard A., Miller K., Wolf M., *AI Ethics: A Long History and a Recent Burst of Attention*, „Computer” 2021, 54(1), 96–102, <https://doi.ieeecomputersociety.org/10.1109/MC.2020.3034950>.
- Brey P., Soraker J., *Philosophy of computing and information technology* [w:] Gabbay D.M., Meijers A.W.M., Woods J., Thagard P. (red.), *Philosophy of Technology and Engineering Sciences*: 9, North-Holland, Amsterdam 2009, 1341–1408.
- Buolamwini J., Gebru T., *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, „Proceedings of Machine Learning Research” 2018, 81, 77–91, <https://proceedings.mlr.press/v81/buolamwini18a.html>.
- Criddle C., Murgia M., *Big tech companies cut AI ethics staff, raising safety concerns*, Financial Times, 2023, <https://www.ft.com/content/26372287-6fb3-457b-9e9c-f722027f36b3>.

- De Vynck G., Oremus W., *As AI booms, tech firms are laying off their ethicists*, The Washington Post, 2023, <https://www.washingtonpost.com/technology/2023/03/30/tech-companies-cut-ai-ethics/>.
- Dwork C., Roth A., *The Algorithmic Foundations of Differential Privacy*, "Foundations and Trends in Theoretical Computer Science" 2014, 9(3–4), 211–407, <http://dx.doi.org/10.1561/0400000042>.
- Fjeld J., Achten N., Hilligoss H., Nagy A., Srikumar M., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, Berkman Klein Center for Internet and Society, 2020, <http://nrs.harvard.edu/urn-3:HUL.InstRepos:42160420>.
- Field H., *How Google's 2021 AI ethics debate foreshadowed the future*, Tech Brew, 2023, <https://www.emergingtechbrew.com/stories/2023/03/07/how-google-s-2021-ai-ethics-debate-foreshadowed-the-future>.
- Gates B., *AI is most important tech advance in decades*. BBC News, 2023, <https://www.bbc.com/news/technology-65032848>.
- Hawking S., *The best of worst or worst thing to happen to humanity*, 2016, https://youtu.be/_5XvDCjrdXs.
- Hopkin G., *Top 10 artificial intelligence events in 2023*, AiMagazine.com, <https://aimagazine.com/top10/top-10-artificial-intelligence-events-in-2023>.
- IDC, *Worldwide Artificial Intelligence Spending Guide*, 2023, <https://www.idc.com/getdoc.jsp?containerId=prUS50454123>.
- Intelligent.com, *Study: 30% of College Students Have Used ChatGPT for Essays*, <https://www.intelligent.com/nearly-1-in-3-college-students-have-used-chatgpt-on-written-assignments/>.
- Jobin A., Ienca M., Vayena E., *The global landscape of AI ethics guidelines*, "Nature Machine Intelligence" 2019, 1, 389–399, <https://doi.org/10.1038/s42256-019-0088-2>.
- Komisja Europejska, *Niezależna grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji*, 2019, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- Mickle T., *Google Plans to Double AI Ethics Research Staff*, The Wall Street Journal, 2021, <https://www.wsj.com/articles/google-plans-to-double-ai-ethics-research-staff-11620749048?mod=djemalertNEWS>.
- Mitchell M., Wu S., Zaldivar A., Barnes P., Vasserman L., Hutchinson B., Spitzer E., Raji I.D., Gebru T., *Model Cards for Model Reporting*. [w:] *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, Nowy Jork 2019, 220–229, <https://doi.org/10.1145/3287560.3287596>.
- Newel A., Simon H., *Computer Science as Empirical Inquiry. Symbols and Search*, "Communications of the ACM" 1976, 19(3), 113–126.
- Nguyen A., Ngan Ngo H., Hong Y., Dang B., Thi Nguyen B., *Ethical principles for artificial intelligence in education*, „Education and Information Technologies” 2023, 28(4), 4221–4241, <https://doi.org/10.1007/s10639-022-11316-w>.
- Niemi H., Pea R.D., Lu Y. (red.), *AI in Learning: Designing the Future*, Springer, Cham 2023.
- Novet J., *Microsoft's \$13 billion bet on OpenAI carries huge potential along with plenty of uncertainty*, CNBC, 2023, <https://www.cnbc.com/2023/04/08/microsofts-complex-bet-on-openai-brings-potential-and-uncertainty.html>.
- OECD, *Trends Shaping Education 2022*, OECD Publishing, Paryż 2022, <https://doi.org/10.1787/6ae8771a-en>.
- Rossi F., Rudden B., Goehring B., *AI ethics in action: An enterprise guide to progressing trustworthy AI*, IBM, 2022, <https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/ai-ethics-in-action>.
- Russel S., Norvig P., *Artificial intelligence – a modern approach*, Pearson Education, New Jersey 2010.

- Schiffer Z., Newton C., *Microsoft lays off team that taught employees how to make AI tools responsibly*, The Verge, 2023, <https://www.theverge.com/2023/3/13/23638823/microsoft-ethics-society-team-responsible-ai-layoffs>.
- Searle J., *Minds, Brains and Programs*, "Behavioural and Brain Sciences" 1980, 3(3), 417–424.
- Searle J., *Rozwiążemy „zagadkę umysłu”, gdy wyzwolimy się od błędnych założeń*, "Filozofuj!" 2015, 4, 21–23, <https://filozofuj.eu/john-searle-rozwiazemy-zagadke-umyslu-gdy-wyzwolimy-sie-od-blednych-zalozen/>.
- Stanford University Institute for Human-Centered Artificial Intelligence, The AI index 2023 annual report*, Stanford University, 2023, <https://aiindex.stanford.edu/report/>.
- Turing A., *Lecture to the London Mathematical Society on 20 February 1947*, "MD Comput" 1995, 12, 390–397.