

Prof. Przegalińska: Koncepcji, że za chwilę czeka nas koniec świata, nie kupuję

© Konrad Sadurski (Newsweek Polska) rozmawia z prof. Aleksandrą Przegalińską

14 stycznia 2025, Newsweek Polska 3/2025

<https://www.newsweek.pl/polska/spoleczenstwo/zycie-z-ai-koncepcji-ze-za-chwile-czeka-nas-koniec-swiatea-nie-kupuje/j3geswz>

Ludzie zakochują się w celebrytach i gwiazdach filmowych, z którymi nie zamienili słowa. Dlaczego mają się nie zakochać w cyfrowym bocie? – pyta prof. Aleksandra Przegalińska.

Prof. Aleksandra Przegalińska, filozofka, prorektor ds. współpracy z zagranicą Akademii Leona Koźmińskiego

Newsweek: Można już stworzyć spersonalizowaną sztuczną inteligencję, która byłaby wytrenowana na doświadczeniu i wykształceniu konkretnej osoby?

Prof. Aleksandra Przegalińska: Do pewnego stopnia, choćby na podstawie ChatGPT. Możemy takiemu osobistemu asystentowi nadać określoną tożsamość, opisując ją promptami, czyli naszymi zapytaniami do sztucznej inteligencji. I dołożyć masę plików i dokumentów, jakieś rozmowy, które by pokazywały, jak się ktoś zachowuje, czym interesuje, co czyta. Sama mam takiego asystenta, tak go zaprojektowałam, że odpisuje w moim stylu na e-maile.

Trudno było się z panią umówić, więc pomyślałem, że może moja sztuczna inteligencja zrobiłaby wywiad z pani sztuczną inteligencją...

– To nie byłaby taka sama rozmowa. Dzisiejsze rozwiązania AI mają jeszcze szereg ograniczeń. Mogą dobrze naśladować styl danej osoby, zwłaszcza jeśli próbek tego stylu byłoby dużo w internecie. Nie byłoby więc problemu, gdybyśmy poprosili ChatGPT, żeby mówił jak Elon Musk, Donald Trump albo Jarosław Kaczyński. Brzmieliby pewnie jak prawdziwi, ale wygłaszając jedynie generalne opinie. Próbowalam parę razy tego swojego asystenta AI zapytać o sprawy dla mnie istotne i padały wprawdzie sformułowania, które byłyby mi bliskie, ale na pewno nie była to odpowiedź, której osobiście bym udzieliła.

A wierszówkę za wywiad naszych sztucznych inteligencji dostałbym ja czy moje AI?

– Pan, bo jako twórcy promptów jesteśmy niejako właścicielami praw do tego, co ten model AI powie. Ale sztuczna inteligencja nie rozwija się w próżni, więc dotyczących jej regulacji będzie coraz więcej. Pewnie jakieś rozstrzygnięcia w sprawach własności intelektualnej też się niedługo pojawią, a razem z nimi może więcej klarowności również w branży medialnej.

Jak na razie najślawniejszymi dziennikarzami w branży zostali 20-lletnia Emi, dwa lata starszy Kuba oraz Alex – niebinarny 23-latek. Wszyscy z OFF Radia Kraków.

– Słyszałam o nich. Eksperyment z prowadzącymi audycje, których stworzyła sztuczna inteligencja, był opisywany na całym świecie i opowiadany w podcastach "New York Timesa".

Zrobili wywiad z Wisławą Szymborską, która nie żyje od 2012 r.

– Podobno miał być jeszcze wywiad z Piłsudskim z okazji Święta Niepodległości.

To wciąż jeszcze zabawa?

– Nie, to był bardzo nieśmieszny eksperyment, w którym najbardziej mnie ucieszyło to, że został przerwany. Miał trwać trzy miesiące, ale po dwóch tygodniach szefostwo radia stwierdziło, że już zebrało wystarczająco dużo danych, by zamknąć projekt, któremu towarzyszyło wiele protestów. Przecież to jest absurd – te boty miały imiona i nazwiska, podszywały się pod ludzi. Od tego roku, gdy wejdzie w życie odpowiednia regulacja, będzie to zakazane.

Pustostłowiem były pytania zadawane tej AI-owej Wisławie Szymborskiej i pustostłowiem były też jej odpowiedzi. Rozumiem, że sekretarz noblistki Michał Rusinek zgodził się na tę zabawę, dysponując prawami do wizerunku noblistki, ale myślę, że nie mógł być zadowolony z tego, co ta AI-owa Szymborska powiedziała i w jakim świetle to postawiło prawdziwą noblistkę.

Najbardziej nieśmiesznym elementem było wkładanie w usta wybitnej osoby, która nie żyje od ponad dekady, opinii na temat tego, co się dzisiaj dzieje w świecie.

Za chwilę przeczytamy w internecie, że Czesław Miłosz, pisząc: "Gromadę błaznów wokół siebie mając/ Na pomieszanie dobrego i złego", miał na myśli internet. Już teraz dużo ludzi ma problemy z odróżnieniem, co jest prawdą, a co fałszem w sieci, a tu jeszcze fałsz dostaje wsparcie sztucznej inteligencji.

– Akurat tutaj gromadą błaznów nie są modele AI, tylko ludzie, którzy ich używają w niewłaściwy sposób. Gdyby ktoś po otwarciu Muzeum Sztuki Nowoczesnej w Warszawie chciał się zaangażować w krytyczny dyskurs na temat tego, co potrafi sztuczna inteligencja, i w eksperymencie ożywiłby naszych noblistów, byłaby to ciekawa zabawa dla widzów. Sprawdzenie, czy modele AI Szymborskiej i Miłosza odzwierciedlają prawdziwe poglądy noblistów. A ten radiowy eksperyment był tanim chwytem: "Hej, zrobiliśmy z AI-ową Szymborską rozmowę. Przyjdź, posłuchaj. Ha, ha, ha".

W internecie przyszłości, w którym będzie bardzo dużo tekstów pisanych przez boty, to rzeczywiście może być bardzo konfundujące – jak odróżnić wartościowe wypowiedzi noblistki od różnych AI-owych iteracji Wisławy Szymborskiej?

Wyobraziłem sobie wywiad z mistrzem krzyżackim Ulrichem von Jungingenem z pytaniem: Dlaczego nie uważał pan na lekcjach historii i poszedł pod Grunwald? Przecież wiadomo było, że tam przegra.

– Słyszę, że w mediach społecznościowych ma być jeszcze więcej syntetycznych treści. A platformy internetowe mocno przymierzają się do tego, żeby dzięki technologii AI dać komuś poczucie, że gdzieś był albo robił coś, czego nigdy dotąd nie robił. Czyli na przykład wejdę sobie na stronę ekskluzywnego hotelu i za chwilę w moich mediach społecznościowych pojawi się zdjęcie lub filmik mojej osoby w tym hotelu, żeby mnie przekonać do wyjazdu tam. Już dzisiaj za pomocą sztucznej inteligencji można stworzyć takie kompletnie syntetyczne doświadczenie, którego celem może być namawianie nas do zakupów. Albo poszerzania naszej wyobraźni, na zasadzie: co by było, gdyby.

A tu likwidator krakowskiego radia wpada na pomysł: zastąpmy dziennikarzy AI, a potem pogrzebmy w grobach – przepraszam, to tak dla mnie wyglądało – znajźmy różnych ważnych, nieżywych ludzi i sobie ich odtworzymy. Myślę, że dlatego tyle uwagi skierowano na ten eksperyment na całym świecie, bo on był po prostu ordynarny.

Kilkadziesiąt tysięcy osób podpisało się pod protestem. Mniej więcej w tym samym czasie Polski Instytut Ekonomiczny opublikował raport o tym, jak sztuczna inteligencja może zmienić rynek pracy. Z 20 badanych grup zawodowych najbardziej zagrożeni mają być finansiści, prawnicy, niektórzy urzędnicy państwowi czy programiści.

– Raport jest bardzo ciekawy, używa specyficznego słownictwa, tłumacząc, że pewne prace będą dotknięte istotnie sztuczną inteligencją, zostaną przez nią zmienione. Ale nie ma bezpośredniego zarzutu, że AI te prace zlikwiduje.

Raport jest więc bardzo ostrożny, nie formułuje radykalnych sądów, w jaką stronę to pójdzie, bo nikt tego dzisiaj nie wie. Elon Musk opowiada z wielką pewnością od 10 lat, że będzie masowe bezrobocie technologiczne, ale nie widać tego na rynku pracy.

W bardzo wielu krajach ze starzejącymi się społeczeństwami widać coś innego – automatyzacja jest potrzebna, żeby w ogóle ktoś pracę wykonywał. W Polsce jesteśmy demograficznie w tak fatalnej sytuacji, że jeśli nie pójdziemy mocno w automatyzację, to za chwilę nie będzie komu pracować.

W zawodach, które mogą być zmienione przez AI, pracuje blisko 3,7 mln Polaków, z tego 2,2 mln to kobiety.

– Kobiety są statystycznie lepiej wykształcone, częściej wykonują pracę opartą na wiedzy, więc można przypuszczać, że bardziej będą narażone na utratę pracy. Ale z drugiej strony to kobiety wykonują prace zwane w USA pink work (różowego kołnierzyka), które wymagają troski w stosunku do innych – są opiekunkami, nauczycielkami, pielęgniarkami, stewardesami. I na taką pracę według "New York Timesa" zapotrzebowanie jest większe od tego, które mogłyby zaspokoić kobiety. Mężczyźni nie chcą takich prac wykonywać, no i nie da się ich zautomatyzować.

Kobiety mają więc bardzo komfortową pozycję.

– Z mojej pracy badawczej wynika też, że sztuczna inteligencja w wielu zawodach opartych na wiedzy zwiększa mocno produktywność i może człowieka istotnie odciążać, zwłaszcza w rutynowych zadaniach. Ale jeśli notariusz chciałby wykorzystywać specjalistyczne modele sztucznej inteligencji, to potem on sam będzie musiał wziąć odpowiedzialność za to, co sztuczna inteligencja zrobiła. Bo nawet jeśli AI wykona 80 czy 90 proc. pracy, to spodziewamy się nad nią nadzoru człowieka i jego podpisu pod tym zadaniem.

W Stanach głośna była sprawa prawnika, któremu model AI wygenerował argumentację w jakiejś sprawie sądowej, opierając się na nieistniejących sprawach – ten człowiek został wyśmiany i naraził reputację swojej firmy na szwank.

Obraz zmian zachodzących na rynku pracy jest więc bardzo skomplikowany. Mamy jakby zdjęcia z różnych miejsc i bardzo trudno na ich bazie zbudować spójną wizję tego, co się za chwilę wydarzy.

Czy sama praca zniknie? To jest dużo bardziej skomplikowane pytanie.

Inteligencja od mądrości różni się tym, że człowiek inteligentny potrafi wybrać z największych problemów. Nawet z takich, w które człowiek mądry w ogóle by się nie wpakował. Czy nie moglibyśmy zamiast sztucznej inteligencji od razu wykreować Artificial Wisdom – sztuczną mądrość?

– Nie mamy na to jeszcze pomysłów. Sama inteligencja jest spektrum różnych kompetencji, które pozwalają nam dostosowywać się do sytuacji, odpowiadać na bodźce z otoczenia w miarę adekwatny sposób.

Jako ludzie mamy swoje moce obliczeniowe, gramy w szachy, jesteśmy w stanie analizować dane. Sztuczna inteligencja robi to samo, więc to jest forma inteligencji, jednak nasze spektrum inteligencji jest inaczej skonstruowane. My mamy dużo tak zwanej inteligencji społecznej i emocjonalnej, modele AI mają tego mało albo nic. A mądrość przypisujemy przede wszystkim istocie, która w świadomy sposób potrafi przeżywać doświadczenia w swoim i nie tylko swoim życiu.

Jest też kwestia życia prywatnego. Czy można się zakochać w sztucznej inteligencji? W filmie "Ona" Joaquin Phoenix zakochał się w systemie operacyjnym, któremu głosu użyczyła Scarlett Johansson. Fabuła brzmi jak banał, ale po obejrzeniu filmu ciarki przechodziły.

– Znakomity, absolutnie profetyczny film. Większość wizji dotyczących walki z maszyną sztucznej inteligencji, takich jak w "Ex Machina" albo w "Terminatorze", kompletnie się nie zrealizowała. A tu kino w subtelny sposób pokazuje wrażliwość człowieka, pewnego rodzaju podatność na to, by budować w swojej głowie pewne obrazy. I w tych obrazach de facto się zakochiwać. To jest bardzo głęboki film o człowieku, znacznie mniej o sztucznej inteligencji. Po jego obejrzeniu dużo myślałam o kulturze epistolarnej. W XIX i XX w. ludzie pisali do siebie listy, nie widząc się latami, więc byli zakochani w czyimś obrazie, który wynikał z ich odpowiedzi listownej. Mamy w sobie taką tendencję, a teraz dochodzi do tego technologia, która jest w stanie z tą tendencją pracować. W "Her" nie było konfliktu, film pokazał po prostu, że sztuczna inteligencja ma inne cele niż człowiek i nie odwzajemni jego uczucia, więc człowiek ostatecznie zostanie sam.

Uważam, że jesteśmy w stanie się absolutnie zakochać w swoim wyobrażeniu na czyjś temat. Ludzie się zakochują w celebrytach, gwiazdach filmowych, których nigdy nie widzieli i nie rozmawiali z nimi. Dlaczego mieliby się nie zakochać w cyfrowym bocie?

Romansowanie z AI może jednak prowadzić do tragedii. Nastolatek, który długo rozmawiał z Daenerys Targaryen, awatarem królowej smoków z "Gry o tron", popełnił samobójstwo.

– To jest wstrząsający przypadek, ale moim zdaniem – przykro to powiedzieć – to historia o zaniedbaniu rodziców, którzy nie wiedzieli, że chłopiec zmagał się z problemami psychicznymi, chował się, był w eskapistycznym kompletnie klimacie. Jego mama widziała, że godzinami komuś zapalczywie coś pisze, ale myślała, że to jakaś gierka à la "Candy Crush".

Ale faktem jest, że ten chłopiec nie miał historii niestabilności psychicznej. Być może był bardzo wrażliwy i ewidentnie rozmowy z botem wprowadziły go w stan absolutnego rozchwiania. A firma, która stworzyła tego bota, postawiła na to, że ludzie mają z nim rozmawiać intensywnie, ale żeby tak się stało, to boty nie mogą wychodzić ze swojej roli. Chłopiec mówił, że rozważa popełnienie samobójstwa, a bot na to: "Wiesz co, byłoby mi strasznie przykro, gdyby tak się stało. Ale oczywiście tęsknię za tobą i chcę się z tobą spotkać".

Chłopiec: "To może spotkajmy się w przyszłym życiu?". A bot odpowiada: "O, wspaniale byłoby się z tobą spotkać w przyszłym życiu".

Na koniec chłopiec mówi: "Przyjdę do ciebie teraz, Daenerys Targaryen".

A bot: "Mój kochany, czekam tutaj na ciebie".

Czy można uregulować kwestie związane z naszymi emocjami i AI?

– Kurcze, przecież w tej sytuacji było co najmniej kilka momentów, kiedy powinno się zapalić światelko alarmowe i odpalić prosty schemat: 14-letnie dziecko ma zaburzenia, rozważa samobójstwo – skieruj je do terapeuty i, do cholery, zakończ tę rozmowę. Tego totalnie zabrakło po stronie dostawcy tej technologii. I to niestety jest właśnie iluzja, w którą to biedne, wrażliwe dziecko, z nieukształtowanym jeszcze umysłem, wpadło. Nikt nie zadbał o jego bezpieczeństwo.

Powinna być tu jakaś prosta bariera. Może to jest technologia dla ludzi od 18. roku życia? Na pewno trzeba jasno komunikować, że to jest iluzja, a nie twój prawdziwy przyjaciel: asystent AI-owy, który będzie się cieszył ze wszystkiego, co zrobisz, ale to nie jest autentyczna radość i nie powie ci szczerze, tak jak przyjaciel by ci powiedział: "Hej, to, co robisz, jest nefajne".

Dorośli ludzie są odporniejsi, bardziej świadomi. Niektórzy już teraz robią sobie randkowe boty, z którymi rozmawiają o różnych sprawach, również seksualnych. Trudno mi to oceniać, ale mam poczucie, że jest dużo nieszczęścia na świecie i samotności. I jakoś nie udało nam się inaczej jej przepracować, więc jeżeli ktoś miałby się poczuć lepiej, być bardziej doceniany po rozmowie z botem, to może to nie jest najstraszliwsza z iluzji. Ale dzieci powinny być chronione.

Wierzy pani, że takie ograniczenia zostaną szybko wprowadzone?

– Obawiam się, że pod rządami nowej administracji amerykańskiej to się nie stanie. Donald Trump regulacji technologicznych nie lubi, a za doradcę wziął sobie Elona Muska, który jest zgorzkniały, że OpenAI osiągnęło taki wielki sukces. Próbuje więc budować własny model,

który jest najbardziej chamskim z modeli sztucznej inteligencji. Nie ma tam żadnych filtrów, bo Musk jest absolutystą wolności słowa. Jego model karmiony jest więc treściami z jego Twittera, gdzie jest masa syfu, ludzie sobie docinają, dokuczają, awanturują się, przeklinają. To jest model świata, który zasila AI Muska, on stara się budować takie rozwiązania, żeby nikt mu nie wchodził w paradę. A jak ktoś ma ochotę, żeby model przeklinał czy pokazywał pornografię, to czemu nie?

Jednocześnie w Kalifornii jest w tym momencie procedowanych kilkanaście ustaw, które dotyczą ochrony przed negatywnymi skutkami AI i większej transparentności. Chodzi o deepfake i generowane wideo albo głosy. Myślę, że kilka innych stanów też będzie podążać w tę stronę.

W Europie mamy nadchodzącą regulację AI Act, której ja nie jestem wielką entuzjastką, ale na pewno wprowadza porządek i transparentność co do deepfake'ów. Mamy mieć jasność, że jak widzimy контент wygenerowany w mediach społecznościowych, to on musi być oznaczony, czy wykorzystywał sztuczną inteligencję.

Próbujemy nadążać za tym pędem technologicznym.

Geoffrey Hinton, ojciec chrzestny sztucznej inteligencji, noblista z fizyki, mówi: "Martwię się, że ogólną konsekwencją mogą być systemy bardziej inteligentne od nas, które ostatecznie przejmą kontrolę".

– Gdy dostał Nobla za sztuczną inteligencję, to się nie ucieszył, tylko demonstrował swoją niechęć. Hinton stara się tworzyć katalog zagrożeń i mówi, że niektóre z nich są bardzo realne tu i teraz, a inne będą w przyszłości – i to przed symulacją przyszłości nas ostrzega. Tylko nasza podatność na jakąkolwiek symulację jest kolosalna, a nasza zdolność weryfikacji symulacji jest żadna.

Poza tym wcale nie trzeba inteligentniejszej AI od nas, wystarczy, żeby ktoś, kto ma nieczne cele, podpuszczał na ludzi boty, które dadzą jakąś iluzję zaufania, i ci ludzie nagle potracą majątki życia, ich konta bankowe zostaną zhakowane. To jest bardzo silna broń już dzisiaj, której należy się obawiać. Zgadzam się z sygnałami alarmowymi, mam jednak wątpliwości co do tego, że AI przewyższy intelektualnie człowieka.

Opinie na temat AI zawierają się jakby w triadzie heglowskiej – teza, antyteza i synteza – poglądów, które prezentują Ray Kurzweil, Mustafa Suleyman i Nick Bostrom. Bostrom, filozof z Oksfordu i autor "Superintelligence", mówi: "Inteligencja maszynowa to ostatni wynalazek, jakiego ludzkość będzie kiedykolwiek potrzebować".

– Bostrom jest znanym apokaliptykiem, ale gdybyśmy polegali na tym, co on nam rekomenduje, toby nie było żadnych systemów do wykrywania raka, do precyzyjnej prognozy pogody albo detekcji chorób zakaźnych na wczesnym etapie, bo po prostu nie mielibyśmy takich algorytmów.

On proponuje drzewka decyzyjne i proste, szczegółowe rozwiązania, z których już dzisiaj nikt nie korzysta, więc musielibyśmy być trochę jak amisze w krainie sztucznej inteligencji – wyrzec się atrybutów postępu.

Kurzweil, główny inżynier Google'a, przewiduje, że AI połączy się z ludźmi: "W połowie lat 2040. będziemy mieć narzędzia do zwiększania inteligencji miliardy razy".

– On z kolei jest entuzjastą AI i fascynatem futurologii, który rozsiewa czasem bombastyczne wizje.

Wieść gminna niesie, że bierze 150 tabletek dziennie, żeby doczekać tej osobliwości, bo liczy, że się połączy z maszyną i będzie żył wiecznie.

Z kolei Suleyman, współzałożyciel DeepMind, ostrzega, że fala nadciąga i musimy nauczyć się serfować: "Działać teraz, by kontrolować, co dzieje się dalej" – zanim będzie za późno.

– Mam wrażenie, że stoi trochę w szpagacie, ale może to najlepsza postawa, jaką można w tym momencie przyjąć, bo wszystkie prognozy są w pewnym sensie spekulacją. Suleyman uważa, że najważniejsze technologie związane z biologią, genetyką są dopiero przed nami.

Sztuczna inteligencja jest zaś jedynie preludium do znacznie poważniejszych kompleksowych problemów i zagadnień etycznych, które mogą się pojawić.

Według niego AI to nie jest ostatnia technologia ludzkości i ta wizja jest mi bliska. Koncepcji, że za chwilę czeka nas koniec świata, nie kupuję.

Źródło: Newsweek Polska 3/2025

© Konrad Sadurski