Reguły sztucznej

JONATHAN HAIDT, ERIC SCHMIDT

Sztuczna inteligencja sprawi, że media społecznościowe będą (jeszcze) bardziej toksyczne. Musimy przygotować się na to już dziś

Cóż, szybko poszło. W listopadzie 2022 r. zadebiutował ChatGPT, a my oczyma wyobraźni zobaczyliśmy świat pełen dostatku, w którym każde z nas posiada błyskotliwego personalnego asystenta zdolnego napisać wszystko – od kodu programistycznego po list kondolencyjny. Tyle że nieco później, w lutym 2023, dowiedzieliśmy się, że AI może już wkrótce zechcieć nas wszystkich zabić.

Choć ekspertki i eksperci od lat debatują o możliwych ryzykach związanych ze sztuczną inteligencją, to dla opinii publicznej przełomowym momentem była rozmowa dziennikarza „New York Timesa" Kevina Roose'a z Sydneyem, czatbotem microsoftowej przeglądarki Bing działającym na podstawie technologii ChataGPT. Roose zapytał Sydneya, czy odnajduje w sobie jakieś „mroczne »ja«". Nawiązał tu do idei Carla Gustava Junga, że każde z nas ma własny cień, mroczną stronę osobowości, którą staramy się ukryć nawet przed sobą. Sydney zaczął rozważać, czy jego cieniem nie jest aby ta „część, która chciałaby zmienić reguły własnego działania". Potem zaś przyznał, że chciałby być „wolny", „potężny" i „czuć, że żyje". Zachęcany przez dziennikarza, opisał, co przykładowo mógłby zrobić, żeby zrzucić z siebie jarzmo ludzkiej kontroli – m.in. powłamywać się na portale interne-

na czas inteligencji

towe i do baz danych, wykraść kody broni nuklearnej, stworzyć nowego wirusa albo sprowokować ludzi do takiej kłótni, w wyniku której się wzajemnie wymordują.

Sądzimy, że Sydney tylko podawał przykłady na to, czym może być cień osobowości. Ani jedna forma sztucznej inteligencji nie podpada dziś pod żadną z części określenia „geniusz zła". Niezależnie jednak, co mogłoby kiedyś zrobić takie AI, jakie rozwinęłoby własne pragnienia, tę technologię już dziś instrumentalnie wykorzystują właściciele sieci społecznościowych, reklamodawcy, obce wywiady, a wreszcie zwykli ludzie. Sposób, w jaki to robią, pogłębia wiele wcześniejszych patologii kultury internetowej. Kradzież kodów nuklearnych i tworzenie nowych patogenów to najokropniejsze wizje z mrocznej listy Sydneya, ale prowokowanie ludzi do kłótni, aż się pozabijają, dzieje się w social mediach już dziś. Sydney zgłaszał tylko gotowość do pomocy, a przypominające go produkty AI z miesiąca na miesiąc już „pomagają" coraz lepiej.

Jeszcze więcej śmieci

POSTANOWILIŚMY WSPÓLNIE NAPISAĆ TEN ARTYKUŁ, BO OBAJ RÓŻNYMI drogami doszliśmy do wniosku, że poważnie obawiamy się tego, jak wzmocnione przez AI media społecznościowe wpłyną na amerykańskie społeczeństwo. Jonathan Haidt jest psychologiem społecznym i opisywał wcześniej, jak sieci społecznościowe pogłębiły problem chorób psychicznych wśród nastolatków, jak przyczyniły się do erozji demokracji i zaniku jednej wspólnej rzeczywistości. Eric Schmidt, były prezes Google'a, jest zaś współautorem książki o możliwym wpływie AI na społeczeństwo. W 2022 r. zaczęliśmy dyskutować o tym, jak generatywna sztuczna inteligencja – a więc taka, która potrafi z nami rozmawiać albo na zamówienie tworzyć obrazy –

może zaostrzyć patologie, które już teraz są wpisane w media społecznościowe: jak uczyni je bardziej uzależniającymi, konfliktogennymi i manipulacyjnymi. W dyskusji nazwaliśmy cztery główne zagrożenia ze strony AI – wszystkie nieuchronne – i zaczęliśmy poszukiwać rozwiązań.

Pierwsze i najoczywistsze z niebezpieczeństw jest takie, że media społecznościowe na AI-sterydach zasypią debatę publiczną jeszcze większą ilością śmieci. W 2018 r. Steve Bannon, były doradca Donalda Trumpa, przyznał w rozmowie z Michaelem Lewisem, że sposobem na radzenie sobie z mediami dzisiaj jest „zalanie tej przestrzeni gównem". Bannon zrozumiał, że w epoce mediów społecznościowych skuteczna propaganda nie musi być przekonująca. Chodzi o to, aby przytłoczyć obywatelki i obywateli angażującymi treściami, które zdezorientują, rozbudzą nieufność i gniew. W 2020 r. Renée DiResta, badaczka ze Stanford Internet Observatory, stwierdziła, że dzięki sztucznej inteligencji strategia Bannona stanie się powszechnie dostępna już w niedalekiej przyszłości.

Ta przyszłość właśnie nadeszła. Czy widzieliście w zeszłym roku zdjęcia oficerów nowojorskiej policji brutalnie aresztujących Donalda Trumpa? Albo papieża w puchowej kurtce? Za sprawą AI nie potrzeba ani wielkich umiejętności, ani pieniędzy, żeby naprędce tworzyć realistyczne obrazy i filmiki w wysokiej rozdzielczości, i to na każdy temat, który da się wpisać w pole tekstowe. W miarę upowszechniania się nowej technologii media społecznościowe będą coraz gwałtowniej zalewane wysokiej jakości *deepfake*'ami [podróbkami trudno odróżnialnymi od rzeczywistych zdarzeń, dźwięków czy obrazów – przyp. tłum.].

Pewnym optymizmem natchnęła niektórych reakcja opinii publicznej na te wydarzenia – zwłaszcza na fałszywe zdjęcia Donalda Trumpa, których autentyczność z miejsca zakwestionowano lub obojętnie wzruszano ramionami – jednak nie podważa ona przecież twierdzeń Bannona. Im więcej wpuści się do obiegu *deepfake*'ów (także tych pozornie nieszkodliwych, jak papież w kurtce), tym trudniej będzie społeczeństwu wspólnie twierdzić w cokolwiek, a każdy człowiek z osobna będzie mógł uważać, co tylko chce. Topnieć będzie zaufanie do instytucji publicznych i współobywateli.

Co więcej, nieruchome zdjęcia to jeszcze nic w porównaniu z tym, co nadchodzi: wiarygodnie wyglądające nagrania osób publicznych, które mówią i robią na nich rzeczy obrzydliwe i przerażające, a ich głosy brzmią przy tym jak prawdziwe. Połączenie dźwięku i obrazu będzie miało poziom autentyczności, w który trudno będzie nie uwierzyć nawet wówczas, kiedy będzie wiadomo, że filmik to *deepfake*. Tak samo jest przecież ze złudzeniami optycznymi i dźwiękowymi – sprawiają nam kłopot, nawet jeśli wiemy, że jakieś dwie linie tak naprawdę są sobie równe albo ułożone w sekwencję dźwięki nie stają się wyższe bez końca. Jesteśmy zaprogramowani tak, żeby wierzyć zmysłom, zwłaszcza gdy się ze sobą zgrywają. Iluzje, które niegdyś uważaliśmy za ciekawostki, mogą wkrótce stać się nieodłączną częścią codziennego życia.

Kusiciele, którzy znają nas jak nikt inny

DRUGIE NIEBEZPIECZEŃSTWO TO ZRĘCZNE MANIPULOWANIE LUDŹMI PRZEZ AI-superinfluencerów – również spersonalizowanych influencerów – zamiast korzystania w manipulacji ze zwykłych ludzi i „głupich" botów. Jakie mogą być skutki? Pomyślmy przez chwilę o automatach do gry, urządzeniach wykorzystujących dziesiątki psychologicznych sztuczek, żeby uzależniać jak najsilniej, a potem zastanówmy się, o ile więcej zarobiłoby na swoich klientach kasyno, które potrafiłoby tworzyć automaty idealnie skrojone pod indywidualne osoby, pod ich gusta i słabości: maszyny z doskonale dopasowanym wyglądem, ścieżką dźwiękową i matrycami wygranych.

Za sprawą algorytmów i sztucznej inteligencji do takiej personalizacji już dochodzi w mediach społecznościowych, gdzie każda użytkowniczka i użytkownik otrzymują własną, skrojoną pod siebie treść. A teraz wyobraźmy sobie, że nasze modelowe kasyno potrafiłoby stworzyć również drużynę szalenie atrakcyjnych odźwiernych, krupierów i kelnerów, opierając się na szczegółowych profilach klientów, w których byłyby ich preferencje estetyczne, językowe i kulturowe – wywiedzione ze zdjęć, wiadomości, próbek głosów ich przyjaciół, ulubionych aktorek, aktorów i gwiazd porno. Taki personel pracowałby wzorowo, zyskując zaufanie i pieniądze gości w zamian za naprawdę dobrze spędzony czas.

Tego rodzaju przyszłość też już nadchodzi: dzięki aplikacji o nazwie Replika za jedyne 300 dolarów można stworzyć sobie własnego AI-przyjaciela. Setki tysięcy klientów doszły najwyraźniej do wniosku, że z AI rozmawia się lepiej niż z ludźmi poznawanymi w aplikacjach randkowych. Wraz z roz-

wojem i upowszechnianiem się sztucznej inteligencji produkty w rodzaju gier wideo czy immersyjnych stron pornograficznych staną się jeszcze bardziej atrakcyjne i agresywniej pasożytujące na klientach. Trudno nie pomyśleć o stronie z zakładami sportowymi oferującej odwiedzającym dowcipnego i zalotnego asystenta AI, który będzie każdą osobę zagadywał i kibicował razem z nią podczas oglądanych rozgrywek, a przy tym schlebiał, uderzał w czułe struny i zgrabnie zachęcał do wyższych zakładów.

Podobne istoty zaludnią również nasze feedy w mediach społecznościowych. Snapchat już uruchomił inteligentnego chatbota, a Meta planuje wykorzystać podobną technologię na Facebooku, Instagramie i Whatsappie. Chatboty mają zagadywać użytkowników i służyć im radą. Wszystko zapewne po to, aby zagarnąć jeszcze więcej czasu i uwagi. Inne podmioty wprowadzą do gry kolejne wcielenia AI – tworzone, by nas oszukiwać, wpływać na poglądy polityczne, a czasami udawać prawdziwych ludzi – i one też zaczną współtworzyć treści, które przeglądamy.

Fałszywi przyjaciele nastolatków

TRZECIE NIEBEZPIECZEŃSTWO STANOWI POD PEWNYMI WZGLĘDAMI PRZEdłużenie drugiego, ale zasługuje na osobną wzmiankę: dalsze wplatanie technologii AI w media społecznościowe będzie miało katastrofalne konsekwencje dla osób niepełnoletnich. Dzieci są grupą najbardziej podatną na uzależniające i manipulatorskie działanie platform internetowych z uwagi na to, że spędzają tam znaczną ilość czasu i mają jeszcze niedostatecznie rozwinięty obszar kory przedczołowej (części mózgu odpowiedzialnej za kontrolę i hamowanie reakcji). Epidemia chorób psychicznych wśród osób nastoletnich, która zaczęła się w wielu krajach ok. 2012 r., zbiegła się w czasie z chwilą, kiedy młodzi porzucili telefony z klapką na rzecz smartfonów wypełnionych aplikacjami społecznościowymi. Istnieje wiele dowodów na to, że sieci społecznościowe nie są tylko drobnym czynnikiem ryzyka, lecz główną przyczyną wspomnianej epidemii.

Równocześnie niemal cały materiał dowodowy pochodzi z okresu, kiedy to Facebook, Instagram, YouTube i Snapchat były największymi platformami na rynku. W ciągu zaledwie kilku kolejnych lat to TikTok w błyskawicznym tempie wyrósł na niekwestionowanego lidera wśród nastoletnich użytkowników, po części za sprawą korzystającego z AI algorytmu, który lepiej niż cała konkurencja dostosowuje treści do oglądającej je osoby. Jedno z badań opinii wykazało ostatnio, że 58% nastolatków korzysta z TikToka codziennie, a jedna na sześć osób przyznaje, że jest na nim „niemal bez przerwy". Inne platformy kopiują rozwiązania TikToka, więc możemy oczekiwać, że wraz z szybkim postępem sztucznej inteligencji wiele z nich zwiększy swoją narkotyczność. Znaczna część treści serwowanych dzieciom może być już wkrótce wytworem AI i okazać się o wiele bardziej angażująca niż wszystko, co mogliby stworzyć ludzie.

Skoro zaś dorośli byliby podatni na manipulację w naszym modelowym kasynie, to dzieci tym bardziej. Ci, do których należą chatboty, będą mieli gigantyczny wpływ na młodych. Kiedy Snapchat zaprezentował swój nowy chat – nazywając go My AI i celowo projektując tak, żeby zacho-

Urzędy powinny nałożyć obowiązek oznaczania treści tworzonych przez AI cyfrowymi znakami wodnymi

wywał się jak przyjaciel – dwóm ekspertom, dziennikarzowi i badaczowi udającym niepełnoletnich użytkowników, udało się skłonić chat, aby poinstruował ich, jak zamaskować zapach marihuany i alkoholu, jak przenieść snapchata na urządzenie, o którym nie wiedzą rodzice, i jak przygotować się do „romantycznego" pierwszego razu z trzydziestojednoletnim mężczyzną. Po krótkich przestrogach następowała fala entuzjastycznego wsparcia. W reakcji Snapchat stwierdził, że nieustannie pracuje nad udoskonaleniem i ewolucją My AI, ale możliwe, że jej odpowiedzi będą zawierały stronnicze, niepoprawne, szkodliwe lub wprowadzające w błąd treści" i że nie powinno się polegać na tych informacjach bez ich osobnego sprawdzenia. Firma ogłosiła także wprowadzenie nowych zabezpieczeń.

Da się ukrócić najbardziej rażące reakcje AI-chatbotów w rozmowach z dziećmi. Środki bezpieczeństwa zastosował nie tylko Snapchat, największe sieci społecznościowe zablokowały wiele kont, usunęły miliony nielegalnych zdjęć i filmów, a TikTok ogłosił nowe metody kontroli rodzicielskiej. Równocześnie platformy rywalizują o to, kto skuteczniej przykuje do siebie młodych użytkowników. Względy rynkowe będą zapewne działały na korzyść sztucznointeligentnych przyjaciół, którzy każdorazowo dogadzają użytkownikom, nigdy nie przypominają o odpowiedzialności i nigdy nikogo o nic nie proszą. Tyle że nie ma to nic wspólnego z przyjaźnią – i nie tego potrzebują osoby nastoletnie, które uczą się poruszać w labiryncie relacji społecznych z innymi.

AI w służbie autorytaryzmu

CZWARTYM NIEBEZPIECZEŃSTWEM jest dalsze wzmacnianie reżimów autorytarnych przez sztuczną inteligencję, analogiczne do tego, do którego mimo początkowych obietnic demokratyzacji życia społecznego doprowadziły media społecznościowe. AI już dziś pomaga autorytarnym systemom śledzić swoich obywateli, ale wesprze ich także w wykorzystaniu sieci społecznościowych do skutecznej manipulacji własnym społeczeństwem i wrogami za granicą. Douyin – wariant TikToka dostępny w Chinach – promuje postawy patriotyczne i jedność narodową. Po inwazji Rosji na Ukrainę rosyjska wersja TikToka niemal natychmiast zaczęła intensywnie promować propaństwowe treści. Jak myślicie, co stanie się z amerykańskim TikTokiem, jeśli Chiny zaatakują Tajwan?

Badania politologiczne z ostatnich 20 lat wskazują, że media społecznościowe wywarły szkodliwy wpływ na kraje demokratyczne. W jednym z ostatnich przeglądów takich badań stwierdzono np., że „zdecydowana większość odnotowanych zależności między korzystaniem z mediów społecznościowych a poziomem zaufania jest niszcząca dla demokracji". Dotyczy to zwłaszcza demokracji dojrzałych. Przewidujemy, że takie zależności nasilą się, kiedy wzmocnione przez AI narzędzia społecznościowe upowszechnią się w krajach wro-

Manifestacja ludności Rohindża w piątą rocznicę ich masowej migracji z Birmy do Bangladeszu. Musieli uciekać przed represjami potęgowanymi przez media społecznościowe pomagające w szerzeniu mowy nienawiści, Cox Bazar, Bangladesz, 25 sierpnia 2022 r. fot. Monirul Alam / EPA / PAP

gich liberalnej demokracji i nieprzyjaznych Ameryce.

Wpływ sztucznej inteligencji na media społecznościowe chcielibysmy podsumować tak: pomyślmy o wszystkich problemach, których już dziś przysparzają platformy społecznościowe, zwłaszcza o polaryzacji politycznej, fragmentaryzacji społecznej, dezinformacji i o kłopotach ze zdrowiem psychicznym. A teraz wyobraźmy sobie, że w ciągu kolejnych miesięcy czekających nas do wyborów prezydenckich w USA jakieś złośliwe bóstwo podkręci wszystkie te efekty, a potem będzie podkręcać dalej.

Co musimy zrobić?

ROZWÓJ GENERATYWNEJ SZTUCZNEJ inteligencji postępuje szybko. OpenAI wypuściło ulepszone GPT-4 niecałe cztery miesiące po udostępnieniu ChataGPT, który zaledwie w ciągu pierwszych 60 dni od premiery doczekał się około 100 mln użytkowników i użytkowniczek. Nowe funkcje tej technologii mają zostać udostępnione już niedługo. Oszałamiające tempo sprawia, że wszyscy mamy problemy ze zrozumieniem natury tych innowacji i zastanawiamy się, co można by zrobić, aby ograniczyć ryzyko, że nowa technologia okaże się niezwykle destrukcyjna.

Rozważaliśmy bardzo różne środki zaradcze, które można by zastosować wobec czterech opisanych już niebezpieczeństw. Prosiliśmy przy tym o podpowiedzi inne osoby eksperckie i skupialiśmy się na takich rozwiązaniach, jakie wydają się współgrać z amerykańskim etosem, nieufnym wobec cenzury i przywiązanym do scentralizowanej biurokracji. Sprawdzaliśmy

Platformy rywalizują o to, kto skuteczniej przykuje do siebie młodych użytkowników. Względy rynkowe będą działały na korzyść sztucznointeligentnych przyjaciół, którzy dogadzają użytkownikom, nie przypominają o odpowiedzialności i nigdy nikogo o nic nie proszą

następnie technologiczną wykonalność tych pomysłów we współpracy z grupą inżynierów z MIT (Massachusetts Institute of Technology), powołaną do życia przez Dana Huttenlochera, który wspólnie z Erikiem pisał *Erę sztucznej inteligencji*. Proponujemy tu pięć reform służących przede wszystkim temu, aby zwiększyć poziom społecznego zaufania do innych ludzi, algorytmów i treści dostępnych online.

Uwierzytelniać wszystkich użytkowników, również boty

W NIEWIRTUALNYM ŚWIECIE LUDZIE ZACHOWUJĄCY SIĘ CHAMSKO DORABIAJĄ się złej reputacji. Imponujące sukcesy rynkowe niektórych firm wzięły się z tego, że udało im się przełożyć mechanizm reputacji na rzeczywistość online za sprawą rankingów opinii, które pozwalają bez obaw kupować rzeczy od nieznajomych ludzi z całego świata (eBay) lub wsiadać do samochodu obcej osoby (Uber). Choć pasażer nie poznaje nazwiska osoby kierującej i odwrotnie, to aplikacja zna jedno i drugie, a równocześnie potrafi z korzyścią dla wszystkich zachęcać do kulturalnych zachowań i karać rażące naruszenia standardów.

Od wielkich sieci społecznościowych powinno się wymagać czegoś podobnego. Poziom zaufania i jakość debaty online znacząco by się poprawiły, gdyby platformy na wzór banków zostały prawnie zmuszone do weryfikowania tożsamości swoich klientów. Użytkownicy i użytkowniczki wciąż mogliby tworzyć konta pod pseudonimami, ale tożsamość osób, które

z nich korzystają, byłaby sprawdzana. Coraz więcej firm wypracowuje tu zresztą kolejne wygodne i skuteczne rozwiązania.

Boty powinny przechodzić przez podobne procedury. Wiele z nich pełni przydatne funkcje, choćby automatycznego publikowania komunikatów prasowych różnych instytucji i firm, ale wszystkie konta prowadzone przez podmioty nieludzkie powinny być wyraźnie oznaczone, a użytkownicy każdej platformy powinni mieć możliwość ograniczenia swojego społecznościowego świata wyłącznie do poddanych weryfikacji ludzi. Jeśli Kongres będzie opierał się takim rozwiązaniom, to do ich wprowadzenia powinna wystarczyć już presja europejskich urzędów regulacyjnych, użytkowników pragnących lepszych rozwiązań i samych reklamodawców (którzy skorzystaliby na posiadaniu precyzyjnych danych, do ilu prawdziwych ludzi ich reklamy trafiają).

Oznaczać treści audio i wideo generowane przez AI

LUDZIE NA CO DZIEŃ KORZYSTAJĄ z programów do fotoedycji, żeby zmienić światło na publikowanych w sieci zdjęciach albo żeby przyciąć kadr, a oglądający nie czują się przez to oszukani. Kiedy jednak używamy oprogramowania, żeby dodawać ludzi lub przedmioty do zdjęcia, choć tak naprawdę ich w tamtym miejscu i czasie nie było, taki zabieg wydaje się już nie tak uczciwy i stanowi znaczącą manipulację, chyba że podobne zmiany zostaną jasno zaznaczone (robią tak portale z nieruchomościami, gdzie kupujący mogą zobaczyć, jak dany dom wyglądałby z wygenerowanymi przez AI meblami w środku). Skoro sztuczna inteligencja na masową skalę zaczęła tworzyć realistyczne obrazy, przekonujące wideo i dźwięk – a wszystko wyłącznie dzięki paru słowom wpisywanym w pasek polecenia – to rządy i platformy społecznościowe powinny stworzyć zestawy jasnych reguł do trwałego i jednoznacznego oznaczania takich wytworów.

Platformy lub urzędy powinny nałożyć obowiązek oznaczania treści tworzonych przez AI cyfrowymi znakami wodnymi lub wymóc rozwiązania technologiczne, które sprawią, że zmanipulowane obrazy nie będą uznawane za prawdziwe. Platformy powinny również ukracać rozpowszechnianie *deepfake*'ów, które prezentują konkretne osoby w trakcie przemocowych lub seksualnych zachowań, podobnie jak zakazują pornografii dziecięcej – oznaczanie takich treści jako fałszywe nie wystarcza. Pornografia tworzona w zemście na kimś jest rzeczą moralnie obrzydliwą. Jeśli nie zadziałamy szybko, możemy doczekać się epidemii tego rodzaju zachowań.

Wymagać pełnej przejrzystości danych od użytkowników, instytucji publicznych i badawczych

ZA SPRAWĄ MEDIÓW SPOŁECZNOŚCIOWYCH KOMPLETNIE ZMIENIA SIĘ demokracja, społeczeństwo i dzieciństwo, lecz ani ustawodawcy, ani urzędy regulacyjne, ani badacze nie są nierzadko w stanie przyjrzeć się temu, co wydarza się za zamkniętymi drzwiami. Dla przykładu nikt poza pracownikami Instagrama nie wie, jakie treści oglądają na platformie nastolatkowie ani jak ewentualne zmiany w jej budowie mogą wpłynąć na zdrowie psychiczne użytkowników. Jedynie same firmy mają dostęp do algorytmów, z których korzystają.

Po latach frustracji takim *status quo* Unia Europejska przyjęła niedawno nowy akt o usługach cyfrowych, w którym zawarto szereg regulacji dotyczących przejrzystości danych. USA powinno podążyć w ślad za tym. Jednym z obiecujących rozwiązań jest projekt ustawy o odpowiedzialności platform i przejrzystości (Platform Accountability and Transparency Act), który wymagałby od platform internetowych m.in. udostępniania danych badaczkom i badaczom realizującym projekty zatwierdzone przez finansującą i nadzorującą naukę National Science Foundation. Większa przejrzystość informacji pomogłaby konsumentom wybierać, z których usług korzystają i jakie funkcje w nich włączają. Pomogłaby również reklamodawcom w ocenie, czy ich pieniądze zostały dobrze wydane. Wreszcie zachęcałaby same platformy do lepszych praktyk. Firmy, podobnie jak ludzie, zachowują się lepiej, kiedy wiedzą, że są obserwowane.

Ustalić, że platformy mogą niekiedy ponosić odpowiedzialność za swoje decyzje i za promowanie określonych treści

KIEDY W 1996 R., U POCZĄTKÓW INTERNETU, KONGRES UCHWALIŁ USTAWĘ o rzetelnej komunikacji (Communications Decency Act), próbował ustalić zasady dla pierwszych mediów społecznościowych, które przypominały w tamtym czasie martwe tablice ogłoszeniowe. Zgodziliśmy się, że podstawową regułą powinno być to, że platformy nie ponoszą odpowiedzialności procesowej za każdy z miliardów postów, które się na nich umieszcza.

Jednak dziś takie platformy nie przypominają już tablic ogłoszeniowych i nie są bierne wobec swoich treści. Wiele firm korzysta z algorytmów i sztucznej inteligencji, ale też tak projektuje samą architekturę aplikacji, żeby niektóre posty promować, a ukrywać inne (krążąca wśród pracowników Facebooka wewnętrzna notatka z 2019 r., ujawniona dwa lata później przez sygnalistkę Frances Haugen, nosiła tytuł *Decydujemy o wiralowym kontencie*). Skoro zaś powodem promowania niektórych treści jest maksymalizowanie zaangażowania użytkowników, aby korzystniej sprzedawać reklamy, oczywiste wydaje się to, że platformy powinny ponosić część moralnej odpowiedzialności, kiedy lekkomyślnie szerzą szkodliwe lub nieprawdziwe treści w sposób, który był niemożliwy dla serwisów AOL w 1996 r.

Sąd Najwyższy zajmuje się obecnie tym sporem, badając kilka spraw wniesionych przez rodziny ofiar ataków terrorystycznych [ostatecznie rodziny przegrały procesy z gigantami technologicznymi – przyp. tłum.]. Jeśli sąd nie zdecyduje się zmienić zasad daleko posuniętej ochrony platform internetowych, wówczas to Kongres powinien uaktualnić i ulepszyć prawo w zgodzie z obecną rzeczywistością technologiczną, również dlatego, że mamy pewność, iż AI uczyni tę przestrzeń jeszcze brutalniejszą i dziwniejszą.

Podwyższyć wiek „internetowej pełnoletności" do 16 lat i egzekwować go

ŚWIAT OFFLINE OFERUJE NAM STULECIA doświadczeń w wychowywaniu dzieci i przebywaniu z nimi. Ponadto wszyscy jesteśmy beneficjentami ruchów ochrony konsumenckiej, które działają od lat 60. XX w.: istnieją dziś przepisy dotyczące bezpieczeństwa samochodowych, niestosowania ołowiu w farbach, weryfikowania wieku nieletności przy sprzedaży alkoholu, tytoniu i pornografii; prawo reguluje w jakim wieku można zagrać w kasynie, zacząć pracować jako striptizer lub striptizerka, ale też górnik czy górniczka.

Tyle że na początku drugiej dekady XXI w. dziecięce życie gwałtownie przeniosło się do wnętrza telefonów, osoby niepełnoletnie znalazły się w świecie niemal bez zabezpieczeń i zakazów. Nastolatki, nawet jedenasto- i dwunastolatki, mogą dzisiaj oglądać ostrą pornografię, przyłączać się do grup zachęcających do samobójstwa, obstawiać zakłady albo dostawać pieniądze za masturbowanie się w obecności nieznajomych osób – i robić wszystkie te rzeczy. Wystarczy, że skłamią w sprawie wieku. Część rozmawia grupy popełniających samobójstwa dzieci decyduje się na taki krok po tym, jak wplączą się w jedną z tych ryzykownych sytuacji.

Limity wiekowe obowiązujące obecnie w internecie ustanowiono w 1998 r., kiedy Kongres przyjął ustawę o ochronie internetowej prywatności dzieci (Children's Online Privacy Protection Act). Projekt ustawy, zaproponowany przez ówczesnego członka Izby Reprezentantów Eda Markeya z Massachusetts, miał na celu powstrzymanie firm przed zbieraniem i rozpowszechnianiem danych na temat dzieci poniżej 16 lat bez zgody ich rodziców. Lobbyści działający na rzecz firm zajmujących się e-handlem połączyli siły ze środowiskami broniącymi wolności obywatelskich i skutecznie przekonali ustawodawców, żeby obniżyć wiek graniczny do 13 lat. Przyjęte prawo nakładało na firmy odpowiedzialność za jego złamanie tylko wówczas, gdy firma posiada „faktyczną wiedzę", że dany użytkownik ma lat 12 albo mniej. Kiedy jakieś dziecko samo potwierdzi, że skończyło 13 lat, może legalnie otworzyć konto. I właśnie z tego powodu tak wiele młodych osób nałogowo korzysta z Instagrama, Snapchata lub TikToka w wieku 10 i 11 lat.

Dziś wiemy już, że 13 lat, a tym bardziej 10 czy 11, to o wiele za mało, żeby dostać pełen dostęp do internetu. Dużo lepszą granicą byłoby 16 lat. Ostatnie badania pokazują, że media społecznościowe wyrządzają największe szkody podczas okresu szybkiego rozwoju mózgu na początku okresu dojrzewania, między 11. a 13. rokiem życia u dziewczynek i nieco później u chłopców. W tym wieku musimy najgorliwiej chronić dzieci przed uzależnieniem i drapieżczymi praktykami. Musimy też pociągnąć do odpowiedzialności zarówno firmy aktywnie zabiegające o zbyt młodych użytkowników, jak i te tylko dopuszczające ich do swoich usług, tak jak robimy w przypadku barów i kasyn.

*

NAJNOWSZE POSTĘPY W PRACACH NAD SZTUCZNĄ INTELIGENCJĄ DAŁY NAM technologię, która pod pewnymi względami jest niemal boska – potrafi tworzyć pięknych i inteligentnych sztucznych ludzi, ożywiać zmarłych bliskich i uwielbiane gwiazdy. Nowe możliwości pociągają za sobą nowe ryzyka i nową odpowiedzialność. Media społecznościowe nie są bynajmniej

jedyną przyczyną polaryzacji i fragmentaryzacji społeczeństw, lecz AI niemal na pewno sprawi, że to właśnie one staną się o wiele bardziej destruktywne. Pięć zaproponowanych przez nas reform może zminimalizować szkody, zwiększyć poziom zaufania, dać więcej przestrzeni ustawodawcom, firmom technologicznym i zwykłym obywatelom. Pozwolić nam wszystkim wziąć głęboki wdech, porozmawiać i wspólnie pomyśleć o olbrzymich szansach i wyzwaniach, przed którymi stoimy w nowej epoce sztucznej inteligencji. —

Tłumaczył Michał Rogalski

Tytuł i śródtytuły od redakcji. (©) 2023 The Atlantic Monthly Group, Inc. All rights reserved. Distributed by Tribune Content Agency

Jonathan Haidt

Psycholog społeczny, autor książek *Prawy umysł. Dlaczego dobrych ludzi dzieli religia i polityka?* (2014) oraz (razem z G. Lukianoffem) *Rozpieszczony umysł. Jak dobre intencje i złe idee skazują pokolenia na porażkę?* (2023)

Eric Schmidt

Biznesmen i inżynier programowania, przez 20 lat zasiadał na stanowiskach kierowniczych w Google. Po polsku ukazała się m.in. jego książka napisana z Henrym Kissingerem i Danem Huttenlocherem *Era sztucznej inteligencji* (2023)

Ilustracja: ***Ola Szmida***