

Lidia Krawczyk

Uniwersytet Komisji Edukacji Narodowej w Krakowie

ORCID 0000-0003-3253-5520

Maszynowe projektowanie statystycznych renderingów, czyli kultura wizualna w służbie sztucznej inteligencji

Od ponad dwudziestu lat uczestniczymy w zakrojonej na olbrzymią skalę rewolucji generatywnej, która nie tylko przeorganizowała doświadczenia, relacje i sposoby komunikowania się między ludźmi, ale także sukcesywnie zawężyła obszar sprawczości. Nasza podmiotowość, sprowadzona do technologicznej i algorytmicznej klasyfikacji, zawarta w kodzie, coraz intensywniej zastępuje możliwość samostanowienia i posiadania intymności tylko na wyłączność.

Kultura wizualna jest polem, na którym zarysowują się szczególne metody cyfrowych technologów. Są one stosowane zarówno w perswazyjnych produktach angażujących użytkowników za pośrednictwem m.in. wzrokowych i słuchowych impulsów, poprzez tworzenie aplikacji wymagających udostępniania danych behawioralnych i biometrycznych, aż po wykorzystywanie zdobytych informacji do trenowania algorytmów sztucznej inteligencji by sprzedawać nam kolejne produkty, nadzorować oraz kreować statystycznie uproszczoną geometrię rzeczywistości. Nasza obecność w tym cyklu jest niezbędna. Stajemy się jej siłą napędową, dostarczycielami kontentu, klientami, użytkownikami, uczestnikami i tymi, nad którymi sprawuje się kontrolę. Wizualność ma w tym swoją kluczową rolę. Mam tu na myśli funkcjonalność projektową (intencjonalne projektowanie aplikacji przy użyciu perswazyjnego designu), użytkową (usługi oparte na udostępnianiu materiałów wizualnych na takich platformach jak YouTube, Instagram, Flickr, Snapchat, Facebook, Tik Tok, które są naszego autorstwa), ale także epistemologiczną (udostępniane materiały wizualne są wykorzystywane do trenowania modeli widzenia maszynowego).

Technologia, o której tutaj piszę, jest wynikiem statystycznego opracowania danych. Jej podstawą są wytrenowane na danych algorytmy, wnioskowanie indukcyjne i dedukcyjne, prognoza, matematyczna pewność skutkująca wygenerowaniem końcowego wyniku. Shoshana Zuboff, autorka książki *Wiek kapitalizmu inwigilacji. Walka o przyszłość ludzkości na nowej granicy władzy*¹ całą odpowiedzialność za pojawienie się datafikacji sieci obarcza przede wszystkim firmę Google, która jako pierwsza w 2003 r. zgłosiła patent pod nazwą „Generowanie informacji

1 S. Zuboff, *Wiek kapitalizmu inwigilacji. Walka o przyszłość ludzkości na nowej granicy władzy*, tłum. A. Unterschuetz, Warszawa 2019.

o użytkownikach do wykorzystania w ukierunkowanej reklamie". To dzięki temu dokumentowi umożliwiony został dostęp do danych behawioralnych pomocnych w zdobyciu wiedzy o tym „co konkretna osoba w określonym czasie i miejscu myśli, czuje i robi”². Firma posiada matematyczną pewność dzięki systemowi określanemu jako „UPI” (*user profile information*). Zuboff, w pełnym złośliwości fragmencie, specjalistów od inżynierii danych nazywa „wyspecjalizowaną rasą fizyków od kliknięć” w formie „świątynnej googlowskiej gwardii chroniącej urząd pierwszego kapłana w rodzącym się kościele behawioralnej predykcji”³, którzy zapewnili „oszałamiający, 3590-procentowy wzrost w ciągu niespełna czterech lat”⁴.

Taki uporządkowany proces ekstrakcji, przetwarzania i produkcji potrzebuje konkretnych, nie znosi abstrakcji i rozmytych kategorii, nawet kosztem ułomnych i szkodliwych definicji. Wymaga pewności, etykiet, poprawnie zapisanego kodu. W konsekwencji na podstawie średniej wyciągane są wnioski dotyczące ogółu. Technologie generatywne, chociaż efektywne w oferowaniu nam danych wyjściowych (obrazu, dźwięku, wideo, tekstu), w swoim rdzeniu opierają się na uporządkowanej taksonomii słów, na zamkniętym zbiorze etykiet przyporządkowywanych do obrazów. Cykl podporządkowywania nas matematycznym regułem to proces na wskroś polityczny, oparty na klasyfikacji, eliminacji i dominacji. By móc to zobaczyć „musimy wykroczyć poza sieci neuronowe, statystyczne rozpoznawanie wzorów i zapytać o to, **co i dla kogo** się optymalizuje, a także **kto** podejmuje decyzje”⁵.

Moim celem zawartym w tym opracowaniu jest opisanie ząbających się porządków opartych na matematycznych i statystycznych regułach porządkowania, gdzie procedura eliminacji przebiega w ramach procesu wizualnego segregowania rzeczywistości. Uwypuklę te zjawiska, które mają wpływ na definiowanie za pomocą obrazów kategorii człowieka czy kanonu piękna. Przy użyciu narzędzi krytycznych przyjrzę się zasadom działania sztucznych sieci neuronowych oraz efektom ich pracy w postaci generowanych obrazów. Zajrzę do wnętrza kilku modeli fundamentalnych służących szkoleniu sztucznej inteligencji, aby przy pomocy przytaczanych badań Kate Crawford, Trevora Paglena, Emily Bender i innych, uwidocznić perswazyjną funkcję powstałą w bazach danych wizji świata.

Wątpliwe fakty, pewne mity

Zahipnotyzowani nawykowymi aplikacjami, udostępniając niezliczone ilości danych, w tym *selfie*, będące nowym gatunkiem „performatywnej tożsamości przeplatających się osobowości offline i online”⁶, przeoczamy zagrożenie jakim jest nadzór sprawowany przez firmy technologiczne, te same, które zachęcają i prowokują

2 Tamże, s. 122.

3 Tamże, s. 123.

4 Tamże, s. 124.

5 K. Crawford, *Atlas sztucznej inteligencji. Władza, pieniądze i środowisko naturalne*, tłum. T. Chawziuk, Kraków 2021, s. 19.

6 J. Deeb-Swihart, C. Polack, E. Gilbert, I. Essa, *Selfie-presentation in everyday life: A large-scale characterization of selfie contexts on Instagram*, „Eleventh International AAAI Conference on Web and Social Media” 2017, vol. 11, no. 1, s. 42–51, <https://doi.org/10.1609/icwsm.v11i1.14896>.

do sukcesywnego przekazywania prywatnych danych. Ufając, że korzyści, które z tego mamy przewyższają straty, ulegamy urzekającym zapewnieniom, że celem postępu technologicznego jest wzmocnienie społecznego dobrostanu, ułatwienie nam pracy, pomoc w administrowaniu codziennością, dostarczanie rozrywki i ułatwianie kontaktu z innymi. Tymczasem warto przyjrzeć się technologiom uczenia maszynowego w kontekście inwigilacji i kontroli, które są masowo wykorzystywane przez instytucje, korporacje i rządy stając się zagrożeniem dla wolności obywatelskiej i praw własności biometrycznej.

W kontekście powszechnie głoszonemu spektakularnemu progresowi i optymizmowi wynikającemu z tworzenia coraz sprawniej działających sztucznych sieci neuronowych, narzuca nam się definicje opracowane przez inżynierów i informatyków z Doliny Krzemowej⁷. Chimeryczne i naukowo skomplikowane wydają się przewyższać nasze zdolności poznawcze. Onieśmieleni technologiczną nomenklaturą nie do końca wiemy z czym mamy właściwie do czynienia. Podejmijmy zatem próbę zrozumienia czym jest oraz czym nie jest sztuczna inteligencja z perspektywy myślenia krytycznego, które pozwoli nam obalić kilka wybranych, ale i kluczowych hipotez.

Zacznijmy od naszej nieporadności w kwestii zdecydowania się z czym właściwie mamy do czynienia, gdy myślimy o sztucznej inteligencji. Poglądy ocierające się o intuicję i podsłuchane ciekawostki medialne mogą m.in. przybrać taką postać:

AI jest sexy, AI jest cool. AI utrwała nierówności, wywraca rynek pracy do góry nogami i niszczy edukację. AI to przejażdżka w parku rozrywki, AI to sztuczka magiczna. AI to nasz ostatni wynalazek, AI to moralny obowiązek. AI to hasło dekady, AI to marketingowy żargon z 1955 roku. AI jest podobna do człowieka, AI jest obca. AI jest superinteligentna i głupia jak but. Boom na AI wzmocni gospodarkę, bańka AI zaraz pęknie. AI zwiększy obfitość i umożliwi ludzkości maksymalny rozwój we wszechświecie. AI zabije nas wszystkich⁸.

Brzmi znajomo? Nie do końca wiemy komu przyznać rację. Pesywiście przewidującym maksymalizację zysków kosztem bezpieczeństwa użytkowników czy może optywiście widzącym wyłącznie świetlaną przyszłość sztucznej inteligencji, która wybawi nas od żmudnej pracy, przyspieszy produkcję i efektywność? Jak wynika z raportu *Global Views on A.I.*⁹ z 2023 r., prawie siedemdziesiąt procent Polaków i Polek deklaruje dobrą znajomość zasad działania sztucznej inteligencji. W dużym stopniu ufają korporacjom technologicznym, a szczególnie w ich zapewnienia, że przekazywane dane osobowe są dobrze chronione (55%) oraz że sztuczna inteligencja nie generuje wyników uprzedzonych czy dyskryminujących (59%). Co ciekawe, inne badanie przeprowadzone cztery lata wcześniej *Sztuczna*

⁷ W.D. Heaven, *What is AI?*, „MIT Technology Review”, 10.07.2024, [on-line] <https://www.technologyreview.com/2024/07/10/1094475/what-is-artificial-intelligence-ai-definitive-guide> – 20.08.2024.

⁸ Tamże.

⁹ *Global Views on A.I. 2023*, Ipsos, July 2023, [on-line] https://www.ipsos.com/sites/default/files/ct/news/documents/2023-07/Ipsos%20Global%20AI%202023%20Report-WEB_1.pdf – 15.09.2024.

*inteligencja w społeczeństwie i gospodarce*¹⁰, pozwoli nam spojrzeć na źródła owej wiedzy i przekonań dotyczących wiarygodności technologii tak intensywnie ingerującej w nasze życie. Sześćdziesiąt jeden procent badanych swoje kompetencje buduje na programach telewizyjnych oraz filmach science fiction. Uważa, że najczęstsze zastosowanie ma w kierowaniu pojazdami i w rozrywce, a także że działa bez udziału człowieka (55%). Zakładając, że powszechny stan wiedzy o metodach administrowania naszymi danymi oraz zachowaniami, pochodzi głównie ze szczątkowych medialnych doniesień i popkulturowej fantazji, tym bardziej warto szczegółowo analizować procesy i mechanizmy pracy „fizyków od kliknięć”. Szczególnie, że osoby ją tworzące, jak na przykład Demis Hassabis, współzałożyciel (razem z Mustafą Sulejmanem) start-upu DeepMind, firmy będącej obecnie częścią Google, otrzymał w październiku 2024 r. Nagrodę Nobla¹¹. Tą samą nagrodą został uhonorowany Geoffrey Hinton, wynalazca sztucznych sieci neuronowych. To wyróżnienie wskazuje na to, że jest najwyższy czas by zerwać z powszechnymi stereotypami i zrewidować zasłyszane pogłoski, bowiem perspektywa zrozumienia metod działania uczenia maszynowego jest pilna i nieunikniona.

Statystyczne prawdopodobieństwo

Sztuczna inteligencja (SI) jest terminem spajającym różne metodologie i odmiany działań technologii opartych na obliczeniach matematycznych, których zadaniem jest symulowanie działania ludzkiego umysłu¹². Podstawą takich systemów jest algorytm – „ciąg instrukcji logicznych, które mówią od początku do końca, jak wykonać określone zadanie. [...] Korzystając z równań, arytmetyki, algebry, analizy, logiki i rachunku prawdopodobieństwa, przetwarzają ciąg operacji matematycznych na kod komputerowy”¹³. Posiadając konkretny, skończony zakres instrukcji zadaniem systemu opartego na algorytmicznych obliczeniach jest rozwiązanie przedstawionego problemu w postaci danych wyjściowych.

W ramach szeroko pojmowanej sztucznej inteligencji znajdują się podkategorie takie jak uczenie maszynowe (*machine learning*, ML), a następnie, bardziej rozwinięty i złożony system – głębokie uczenie (*deep learning*, DL) oparte na architekturze sztucznych sieci neuronowych najbardziej zbliżonych do metod działania ludzkiego umysłu. Do tej kategorii są zaliczane m.in. generatywna sztuczna inteligencja (*Generative Artificial Intelligence*, GAI) oraz ogólna sztuczna inteligencja

10 *Sztuczna inteligencja w społeczeństwie i gospodarce. Raport z badań społecznych. Analiza wyników ogólnopolskiego badania opinii polskich internautów*, NASK Państwowy Instytut Badawczy, Warszawa 2019.

11 W 2004 r. Nagroda Nobla w dziedzinie chemii została podzielona: połowę otrzymał amerykański biochemik David Baker, a pozostałą część otrzymali wspólnie – Amerykanin John M. Jumper i Brytyjczyk Demis Hassabis.

12 W celu zrozumienia tego rozróżnienia można go porównać do znaczenia inkluzji w geometrii, gdzie każdy kwadrat jest czworokątem, lecz nie każdy czworokąt jest kwadratem – każde uczenie głębokie jest sztuczną inteligencją, lecz nie każda sztuczna inteligencja jest uczeniem głębokim.

13 H. Fry, *Hello world. Jak być człowiekiem w epoce maszyn*, tłum. S. Musielak, Kraków 2018, s. 18–19.

(*General Artificial Intelligence*, AGI) czy generatywne sieci przeciwstawne (*Generative Adversarial Network*, GAN).

Uczenie maszynowe przy użyciu metod statystycznych umożliwia algorytmom klasyfikowanie, prognozowanie i odnajdywanie wzorów. Służy do dostarczania wyników (danych wyjściowych), będących efektem porównywania, katalogowania, porządkowania i optymalizowania surowych danych zarówno w oparciu o model nadzorowany (posiadający etykiety, czyli nazwy przyporządkowane wcześniej przez człowieka) oraz nienadzorowany (samodzielnie precyzując elementy wspólne bądź różnicujące)¹⁴. ML używane jest do wytrenowania modelu, który będzie stosowany w identyfikowaniu wzorców i regularności w danych oraz ich ciągłej optymalizacji¹⁵. Do tego celu wymagane jest oznaczanie zbiorów, nadanie etykiet danym podstawowym by sztuczna inteligencja nauczyła się rozpoznawać obiekty. Ten często wykonywany przez człowieka proces stanowi duże utrudnienie oraz zagrożenie z powodu ogromnej ilości danych do skatalogowania i opisanie.

Na uczeniu maszynowym oparte są silniki wyszukiwania (Google, Baidu), rekomendacji (Spotify, Netflix, YouTube), targetowania (Facebook, Instagram) czy rozpoznawania obrazów, w tym twarzy (Clearview AI, PimEyes). Posiadając olbrzymie ilości danych (nasze preferencje w postaci śladów, które zostawiamy za każdym razem, gdy klikamy w link, reagujemy na post czy udostępniamy własne materiały) platformy korzystają z uczenia maszynowego by znaleźć wzór, którego produktem wyjściowym będzie np. predykcja, czyli przewidywanie tego, co w następnej kolejności może nam się do tego stopnia spodobać, że zdecydujemy się nadal używać aplikacji bądź serwisu¹⁶. To z kolei spowoduje, że niezwykle skutecznie nasza uwaga zostanie przytrzymana a korzystanie z produktu stanie się nawykowe. Dowodzi to efektywności opracowanego w Laboratorium Techniki Perswazji na Uniwersytecie Stanforda modelu behawioralnego stosowanego w marketingu, projektowaniu aplikacji i usług cyfrowych. Cel – gwarancja wywołania ściśle określonych reakcji i utrwalanie konkretnych zachowań¹⁷.

Głębokie uczenie jest poddziedziną uczenia maszynowego używającą do obliczeń sztucznych sieci neuronowych. To ML na „sterydach”, dzięki któremu systemy

14 Korzystając z metafory przepisu kulinarnego możemy owe działania porównać do gotowania idealnej potrawy. Posiadając składniki ML wykonuje określone czynności, aby efektem końcowym była np. lasagna. Powtarza kroki zgodnie z podanym przepisem (model nadzorowany) do momentu, aż danie będzie miało jak najlepszy smak zgodny z gustem osoby spożywającej (optymalizacja). Model nienadzorowany nie posiada przepisu, zatem maszyna metodą prób i błędów porównuje oczekiwania i gust konsumenta odszukując wzory i podobieństwa, łącząc składniki a następnie samodzielnie tworząc zoptymalizowane danie.

15 Z uwagi na to, że celem tego tekstu nie jest dokładna analiza i opis pojęcia jakim jest sztuczna inteligencja, lecz zbadanie jej wpływu na kulturę wizualną, skupię się przede wszystkim na tych modelach, które wywierają szczególny wpływ na wzajemną zależność, mając równocześnie świadomość istnienia wielu innych metod i podkategorii, np. uczenie transferowe, uczenie przez wzmocnienie i in.

16 K. Haoarchive, *What is machine learning?*, „MIT Technology Review”, 17.11.2018, [on-line] <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart> – 26.06.2024.

17 Zob. np. B.J. Fogg, *Persuasive technology. Using computers to change what we think and do*, San Francisco 2012; N. Eyal, R. Hoover, *Skuszeni. Jak tworzyć produkty, które kształtują nawyki konsumenckie*, tłum. A. Kucharczyk-Barycza, Gliwice 2024.

posiadają „zdolność do znajdowania i wzmacniania nawet najmniejszych wzorców”¹⁸. Takim rodzajem jest generatywna sztuczna inteligencja, w tym duże modele językowe (*large language models*, LLM), które nie wyszukują¹⁹ „informacji w bazach danych ani w Internecie, zamiast tego generują nowe sekwencje słów na podstawie przewidywania kolejnych wyrazów w zdaniu”²⁰. Wewnątrz takiego modelu znajduje się ogrom słów, które zostały opracowane według statystycznego prawdopodobieństwa występowania ich koło siebie. Tym samym poprzez taką formę generowania treści nastąpiła widoczna zmiana polegająca na „odrzućeniu kontekstu na rzecz danych, znaczenia – na rzecz rozpoznawania statystycznych regularności”²¹, gdzie „duże modele językowe to matematyczne automaty do gier. Przekręć rączkę a pojawi się słowo”²². W sytuacji generowania obrazów proces przebiega w ten sam sposób – dane wyjściowe są generowane na podstawie wyuczonego metodami statystycznymi prawdopodobieństwa w oparciu o posiadane dane znajdujące się w modelu.

W kontekście kultury wizualnej na szczególną uwagę zasługuje uczenie głębokie wykorzystujące generatywne sieci przeciwnawstane (GAN) czy transformery umożliwiające tworzenie obrazów na podstawie tekstowych opisów (*text-to-image*). Oba stanowią fundament wielu nowoczesnych systemów generatywnych, w tym GPT od OpenAI, DALL-E, Midjourney, Firefly 3 czy Stable Diffusion.

Chcąc wytłumaczyć zasadę działania architektury GAN przywołam projekt opracowany przez jednego z inżynierów Ubera – Philipa Wanga. W 2018 r. Nvidia opracowała i udostępniła kod funkcjonujący pod nazwą StyleGAN. Ten nowy model sztucznych sieci neuronowych oparty na *Generative Adversarial Network*, polega na współpracy ze sobą dwóch sieci, które są przeciwnawstane wobec siebie. Wang zastosował je do utworzenia szczególnego generatora twarzy, gdzie jedna sieć produkuje obrazy, a druga sprawdza, czy są one wyprodukowane przez SI, czy przedstawiają wizerunek prawdziwego człowieka, którego portret znajduje się w olbrzymiej bazie danych nakarmionej zdjęciami pobranymi z sieci. Ten trening kończy się w momencie, gdy pierwsza sieć zaczyna regularnie oszukiwać drugą. Co oznacza, że żaden z wygenerowanych obrazów nie funkcjonuje w bazie, jest w pełni stworzony przez sieci neuronowe. Wang opublikował swój model jako stronę internetową zatytułowaną „Random Face Generator – This Person Does Not Exist”²³. Dzięki niej możemy zobaczyć twarze ludzi będące wyłącznie

18 K. Haoarchive, dz. cyt.

19 O tym, jak bardzo trudno jest napisać tekst o sztucznej inteligencji, który byłby aktualny, świadczy chociażby to, że pod koniec października 2024 r. OpenAI wypuściło swój nowy produkt SearchGPT, co powoduje, że informacja przekazana przeze mnie na początku października tego samego roku, gdy ukończyłam pisanie tego artykułu, stała się niepełna. Zob. np. M. Heikkilä, M. Honan, *OpenAI brings a new web search tool to ChatGPT*, „MIT Technology Review”, 31.10.2024, [on-line] <https://www.technologyreview.com/2024/10/31/1106472/chatgpt-now-lets-you-search-the-internet> – 5.11.2024.

20 W.D. Heaven, *Why does AI hallucinate?*, „MIT Technology Review”, 18.06.2024, [on-line] <https://www.technologyreview.com/2024/06/18/1093440/what-causes-ai-hallucinate-chatbots> – 24.08.2024.

21 K. Crawford, *Atlas*, dz. cyt., s. 107.

22 W.D. Heaven, *Why does AI hallucinate?*, dz. cyt.

23 [on-line] <https://this-person-does-not-exist.com> – 22.11.2024.

złożoną figurą matematyczną. W ramach tego systemu udostępniona jest również możliwość modyfikacji poprzez zmienianie nawet niewielkich wartości ustawień, co wpływa na zmianę wyglądu. Każdy taki parametr dostarcza nam pewności, że nigdzie na świecie nie ma portretu żadnej z tych osób.

Maszynowa studnia bez dna

Ważne do pełnego zrozumienia mechanizmów działania systemów sztucznej inteligencji jest przyjrzenie się temu co znajduje się wewnątrz owego modelu generującego końcowy wynik obliczeń. A zatem na jakim zbiorze tekstów oraz obrazów było trenowane uczenie maszynowe. W końcu „widzenie komputerowe to widzenie przez pryzmat danych treningowych”²⁴, jak tłumaczy Adam Harvey, artysta i badacz algorytmów rozpoznawania twarzy, podkreślając, że logika nowej technologii opiera się nie na lepszych algorytmach, lecz na większej ilości danych²⁵. W dodatku danych wartościowych. Aby algorytmy zaczęły „widzieć” nasz świat należy je wytrenować na obrazach wraz z przypisanymi do nich etykietami, uporządkowanymi według kategorii i klas. Ich funkcją jest nazwanie tego, co jest przedmiotem cyfrowego przedstawienia. Jeśli chcemy, by sztuczna inteligencja odróżniła na przykład drzewo od lampy ulicznej, należy dostarczyć jej tysiące wizerunków przedstawiających w różnych kontekstach i figuracjach skategoryzowane i oznakowane fotografie obiektów. Takie zbiory danych wizualnych wraz z etykietami są gromadzone w modelach nazywanych prawdą fundamentalną. To one zawierają całą wiedzę, którą przekazują w dalszej kolejności trenowanej na nich sztucznej inteligencji. Jeśli w danych treningowych zabraknie na przykład jakiegoś gatunku drzewa, dla sztucznej inteligencji ono nie istnieje, nie wygeneruje żadnego wyniku, który uwzględniłby go w swoich danych wyjściowych. Zakres informacji znajdujących się w zbiorach jest kluczowy do generowania wyników końcowych w postaci obrazu, tekstu bądź wideo. Decyduje o jakości i wartości wyprodukowanych treści²⁶. Taki model fundamentalny powinien być od pięciu do dziesięciu razy większy od trenowanej na niej sztucznej inteligencji, dlatego do szkolenia systemów rozpoznawania twarzy czy emocji używa się od kilku do kilkunastu milionów fotografii, aby efekt końcowy był jak najbardziej precyzyjny.

W przypadku procedur identyfikacji człowieka potrzebne są cyfrowe reprezentacje twarzy „w różnorodnych pozach, wieku, płci, odcieniu skóry, mimice i jakości obrazu, przy czym większa liczba obrazów zazwyczaj prowadzi do wyższej wydajności”²⁷. Skąd firmy pozyskują materiały do trenowania sztucznej inteligencji? Właściwie skąd się da, bowiem obszar danych dostępnych w Internecie jest powierzchownie i pozornie uregulowany. Mustafa Sulejman, szef działu AI w Microsoft będącego właścicielem OpenAI – mimo licznych pozwów w sprawie kradzieży chronionych prawem autorskim materiałów – uważa, że „w momencie, gdy

24 A. Harvey, *On computer vision*, UMBAU, 1/2021, [on-line] <https://umbau.hfg-karlsruhe.de/posts/on-computer-vision> – 26.06.2024.

25 Tamże.

26 C. Buschek, J. Thorp, *Models all the way down*, [on-line] <https://knowingmachines.org/models-all-the-way#section4> – 5.07.2024.

27 A. Harvey, *Today's selfie is tomorrow's biometric profile*, [w:] *House of Mirrors. Artificial Intelligence As Phantasm*, 2022.

cokolwiek opublikujesz w otwartej sieci, staje się to freeware, które każdy może swobodnie kopiować i używać²⁸. Takie powszechne podejście technologów do praw własności tłumaczy pochodzenie danych wizualnych pobieranych z miliarda stron internetowych, m.in. z wyszukiwarki Google, Flickr (IBM posiada prawie sto milionów zdjęć znajdujących się w bazie YFCC100M), YouTube'a (z filmów zamieszczanych na tej platformie bezprawnie korzystają takie firmy jak Runway, OpenAI, Anthropic, Apple, Nvidia, Salesforce²⁹), Facebooka, Venmo, miliony *selfie* pochodzą z Instagrama (skąd model UCF *Selfie* został zasilony ponad czterdziestoma sześcioma tysiącami *selfie*), fotografie celebrytów z IMDB.com, a także z Pinteresta (skąd LAION-5B pobrał sto piętnaście milionów fotografii wraz z podpisami, które uzupełniły jego kolekcję pięciu miliardów zaetykietowanych obrazów), Shopify, SlidePlayer (serwis z prezentacjami w PowerPoint) i wielu innych. Dla porównania skali weźmy duże modele językowe, jak na przykład uruchomiony w marcu 2024 r. open source'owy DBRX, który został przeszkolony na ośmiu bilionach słów³⁰, co pokazuje z jaką objętością danych mamy do czynienia. W sytuacji modeli opartych na obrazach wizualnych ilości dostępnych materiałów ćwiczeniowych jest stanowczo mniejsza, co z jednej strony powoduje gorszą jakość uzyskiwanych wyników, ale także prowokuje korporacje cyfrowe do łamania przepisów poprzez ignorowanie praw własności do tych wizerunków, które udało im się pobrać. Dane biometryczne każdego użytkownika Internetu mogą być bez jego wiedzy i świadomości dowolnie używane do trenowania modeli generatywnej sztucznej inteligencji i systemów widzenia komputerowego. Pozyskiwane dane podnoszą wydajność algorytmów, optymalizują efekty wyjściowe i równocześnie doskonalą sposoby instytucjonalnego nadzoru poprzez większy bank danych, w tym wizerunków twarzy.

Jednym z podmiotów odpowiedzialnych za nadzór wizyjny, który posiada w swoich zbiorach zdobyte w sposób nielegalny dane, jest Iarpa Janus Benchmark-C (IJB-C), system amerykańskiej agencji rządowej przetrzymujący łącznie ponad

28 S. Hollister, *Microsoft's AI boss thinks it's perfectly okay to steal content if it's on the open web*, „The Verge”, 28.06.2024, [on-line] <https://www.theverge.com/2024/6/28/24188391/microsoft-ai-suleyman-social-contract-freeware> – 22.11.2024.

29 E. Roth, *Runway's AI video generator trained on thousands of scraped YouTube videos*, „The Verge”, 25.07.2024, [on-line] <https://www.theverge.com/2024/7/25/24206120/runway-ai-video-generator-scraped-youtube-videos-report> – 20.08.2024. Do treningu wykorzystano sto tysięcy siedemset filmów z czterdziestu ośmiu tysięcy kanałów. Zob. np. M. Sato, *Apple, Anthropic, and other companies used YouTube videos to train AI*, „The Verge”, 16.07.2024, [on-line] <https://www.theverge.com/2024/7/16/24199636/apple-anthropic-nvidia-salesforce-youtube-videos-training-data-copyright> – 20.08.2024. „404 Media”, które przeprowadziło śledztwo w sprawie Runway, udostępniło arkusz danych szkoleniowych, gdzie można sprawdzić używane w treningu materiały filmowe: [on-line] https://docs.google.com/spreadsheets/d/1eO5cwguMHeu63F0vsKRXs_dLLcuy_P4F/edit?ref=404media.co&gid=532885050#gid=532885050 – 20.08.2024. S. Cole, *AI Video Generator Runway trained on thousands of YouTube videos without permission*, „404 Media”, 25.07.2024, [on-line] <https://www.404media.co/runway-ai-image-generator-training-data-youtube> – 23.08.2024.

30 Mosaic AI Research, *Introducing DBRX: A new state-of-the-art open LLM*, „Databricks”, 27.06.2024, [on-line] <https://www.databricks.com/blog/introducing-dbrx-new-state-art-open-llm> – 30.06.2024.

dwadzieścia tysięcy portretów, zdjęć i filmów przedstawiających ponad trzy i pół tysiąca osób, średnio sześć fotografii i trzy filmy na osobę. Około jedenaście tysięcy wideo zostało pobranych z platformy YouTube, mimo że Google kilkakrotnie informowało, iż jest to wyraźne naruszenie ich Warunków Użytkowania³¹. Autorzy bazy skorzystali z materiałów udostępnianych przez użytkowników platformy, którzy „przesyłają dobrze oznakowane filmy z takich wydarzeń jak Światowe Forum Ekonomiczne czy Międzynarodowa Federacja Sportu Uniwersyteckiego”³². Materiały nie tylko posiadają dobrze opisane miejsca i przyporządkowane imiona i nazwiska występujących osób, ale także są precyzyjnie filmowane, przy użyciu statycznej kamery, z dobrze oświetlonym prelegentem, ustawionym często centralnie do kamery. Najczęściej na filmach znajdują się politycy, naukowcy, dziennikarze, teoretycy, artyści, aktywiści w trakcie prowadzenia wykładów, prezentacji, biorących udział w konferencji. Znajdują się na nich widoczne pod różnymi kątami i w różnym oświetleniu twarze m.in. takich osób jak Ai Weiwei, Slavoj Žižek, Paola Antonelli czy Tracey Emin³³.

Następną firmą posiadającą olbrzymią bazę danych i efektywny algorytm rozpoznawania twarzy jest Clearview AI. Pierwotnie miała w planie udostępnić swój system prywatnym firmom będącym zainteresowanymi na przykład identyfikacją gości w ekskluzywnych salonach bądź hotelach, gdzie chciano ich witać po imieniu, a dzięki automatycznemu wglądowi w ich konta na platformach społecznościowych, mogliby traktować ich adekwatnie do statusu społecznego³⁴. Ostatecznie w wyniku wyroku sądowego z 2022 r., w którym zakazano jej udostępniania dostępu do niej osobom i firmom prywatnym³⁵, firma współpracuje obecnie wyłącznie z organami ścigania Stanów Zjednoczonych (np. Departamentem Bezpieczeństwa Krajowego i FBI) na zasadzie miesięcznej subskrypcji. W swojej bazie danych posiada pobrane z Internetu trzy miliardy zdjęć z mediów społecznościowych i Internetu, które pozwalają na identyfikację osób podejrzanych o różnego rodzaju przestępstwa czy nadzór nad resztą, która mogłaby ewentualnie stwarzać zagrożenie. Metoda działania algorytmu jest według jej autorów skuteczna w siedemdziesięciu pięciu procentach. Wystarczy załadować fotograficzny portret poszukiwanego do systemu, aby ukazał się szereg wyników ze zdjęciami oraz linkami odsyłającymi do stron, na których znajduje się wzmianka o danej osobie. Jak podkreśla Kashmir Hill, dziennikarka śledcza „The New York Times”, która poświęciła wiele miesięcy na zbadanie sposobów działania Clearview AI, warto mieć na uwadze to, że aby taka kontrola i rozpoznanie działało bez zarzutu, nie wystarczą wyłącznie dane o osobach

31 Updates to YouTube’s Terms of Service, [on-line] <https://support.google.com/youtube/thread/83733719?hl=en> – 22.11.2024.

32 A. Harvey, J. LaPlace, *Exposing.ai, IARPA Janus Benchmark C*, 2021, [on-line] [https://exposing.ai/ijb_c/#\[^Whitelam2017IARPAJB\]](https://exposing.ai/ijb_c/#[^Whitelam2017IARPAJB]) – 30.06.2024.

33 M. Murgia, *Who’s using your face? The ugly truth about facial recognition*, „Financial Times” 18.09.2019, [on-line] <https://www.ft.com/content/cf19b956-60a2-11e9-b285-3acd-5d43599e> – 24.07.2024.

34 K. Hill, *Your face belongs to us. A secretive startup’s quest to end privacy as we know it*, New York 2023, s. 148.

35 *Clearview AI settles ACLU Illinois lawsuit confirming continuity of business supporting public safety*, [on-line] <https://www.clearview.ai/press-room/clearview-ai-settles-aclu-illinois-lawsuit-confirming-continuity-of-business-supporting-public-safety> – 15.08.2024.

podejrzewanych o przestępstwa. Żeby poradziła sobie z precyzyjnym „rozpoznanie wzorców w ludzkim zachowaniu, musi mieć tych danych dużo, dużo więcej. Najlepiej o wszystkich”³⁶. Clearview AI posiada jeden z najlepszych systemów rozpoznawania twarzy, który działa w warunkach słabego oświetlenia. Potrafi zidentyfikować osobę nawet wtedy, gdy nosi kapelusz, założyła okulary i zapuściła zarost. Może zidentyfikować kogoś w tłumie lub gdy znajduje się wyłącznie na drugim, bądź jeszcze dalszym planie zdjęcia, bądź gdy jest odwrócona w przeciwną stronę. System potrafi również rozróżniać od siebie bliźniaków³⁷.

Kontrowersje dotyczące jej działalności są dosyć często powodem wytaczania przeciwko niej pozwów sądowych. W czerwcu 2024 r. firma podpisała w sądzie w Chicago ugodę w związku ze skargą milionów Amerykanów, których portrety znalazły się w bazie danych³⁸. Na podstawie tego dokumentu i z powodu olbrzymich kosztów odszkodowania firma zaproponowała pokrzywdzonym udziały w postaci procentu wartości, jaką Clearview mógłby osiągnąć w przyszłości. Natomiast jak podaje „Forbes”, we wrześniu 2024 r. firma została skazana przez holenderski organ ochrony danych osobowych karą w wysokości ponad trzydziestu milionów euro za „automatyczne pozyskanie miliardów zdjęć osób z Internetu, które następnie przekształcała w unikalny kod biometryczny dla każdej twarzy”³⁹. Organ podkreśla, że baza danych jest nielegalna i nigdy nie powinna powstać.

Firm oferujących modele treningowe uczenia maszynowego dla systemów rozpoznawania twarzy jest wiele. Jedną z nich był Microsoft, który w 2016 r. uruchomił MS-Celeb-1M. Zgromadził w nim prawie dziesięć milionów fotografii ponad stu tysięcy osób. Wśród nich byli przede wszystkim amerykańscy i europejscy aktorzy. Tymczasem znaleźli się tam także dziennikarze, artyści, muzycy, aktywiści i badacze, w tym Trevor Paglen, Laura Poitras, Shoshana Zuboff czy Hito Steyerl⁴⁰. MS-Celeb był używany na całym świecie przez organizacje komercyjne, wojskowe i akademickie, w tym przez korporację SenseTime, „firmę zajmującą się sztuczną inteligencją, która [...] dostarczała chińskim władzom oprogramowanie do nadzoru wykorzystywane do monitorowania i śledzenia Ujgurów”⁴¹. Eksperti technologiczni stwierdzili, że Microsoft mógłby naruszyć Ogólne Rozporządzenie o Ochronie Danych Osobowych (RODO) UE, gdyby po wejściu w życie przepisów z 2019 r. kontynuował dystrybucję swojego produktu. Dlatego usługa została wstrzymana,

36 A. Iwańska, *Czy europosłowie i europosłanki zagłosują przeciwko biometrycznej inwigilacji?*, Fundacja Panoptykon, 12.06.2023, [on-line] <https://panoptykon.org/glosowanie-ai-act-biometria> – 3.07.2024.

37 Por. K. Hill, *Your face*, dz. cyt., s. 148.

38 K. Hill, *Clearview AI used your face. Now you may get a stake in the company*, „The New York Times”, 13.06.2024, [on-line] <https://www.nytimes.com/2024/06/13/business/clearview-ai-facial-recognition-settlement.html> – 30.06.2024.

39 R. Hart, *Clearview AI – controversial facial recognition firm – fined \$33 million for ‘illegal database’*, „Forbes” 3.09.2024, [on-line] <https://www.forbes.com/sites/roberthart/2024/09/03/clearview-ai-controversial-facial-recognition-firm-fined-33-million-for-illegal-database> – 10.09.2024.

40 A. Harvey, J. LaPlace, *Exposing.ai. MS-Celeb*, 2021, [on-line] <https://exposing.ai/ms-celeb> – 24.07.2024.

41 H. Steyerl, *Mean images*, „New Left Review”, iss. 140/141, Mar/June 2023.

a cała baza usunięta jeszcze w tym samym roku⁴². Co prawda Microsoft nie udziela już do niego dostępu, jednak nadal można go znaleźć krążącego w kilku miejscach w Internecie, w tym na GitHubie⁴³.

Jeszcze inna grupa badawcza, tym razem z Yahoo Labs udostępniła w 2014 r. zestaw danych Yahoo Flickr Creative Commons 100 Million Dataset tzw. YFCC100M, olbrzymią, publicznie dostępną kolekcję multimedialną przeznaczoną do uczenia maszynowego. Zawiera ona prawie sto milionów zdjęć i osiemset tysięcy filmów wraz z metadanymi stworzonymi przez użytkowników serwisu Flickr⁴⁴. PimEyes⁴⁵, stworzony w Polsce przez dwóch absolwentów Uniwersytetu Wrocławskiego⁴⁶ oferuje usługi rozpoznawania twarzy na podobnej zasadzie jak Clearview AI z tą różnicą, że ma o wiele mniejszą bazę danych (około jednego miliarda fotografii pochodzących ze stu pięćdziesięciu milionów stron internetowych⁴⁷) oraz jest publicznie dostępna. Każdy może z niej skorzystać i sprawdzić, na jakich stronach znajdują się portrety wybranej osoby oraz dowiedzieć się kim ona jest, gdzie mieszka, jaki jest jej status⁴⁸. Działa podobnie jak wyszukiwarka Google – algorytmy odnajdują wszystkie strony, na których znajduje się osoba, której referencyjny portret załadowaliśmy do serwisu. Na koniec tej listy warto jeszcze wspomnieć o udostępnionym w pierwszej połowie 2023 r. zbiorze CommonPool posiadającym ponad dwanaście miliardów zdjęć wraz z opisami⁴⁹. Jest to na chwilę obecną największy zbiór materiałów wizualnych używanych do trenowania widzenia komputerowego i rozpoznawania twarzy.

Wymienione systemy rozpoznawania twarzy oraz bazy danych treningowych to tylko wycinek pokazujący niektóre z największych modeli fundamentalnych wykorzystywanych do szkolenia sztucznej inteligencji oraz systemów widzenia

42 M. Murgiam, *Microsoft quietly deletes largest public face recognition data set*, „Financial Times”, 6.06.2019, [on-line] <https://www.ft.com/content/7d3e0d6a-87a0-11e9-a028-86cea8523dc2> – 1.07.2024.

43 J. Pearson, *Microsoft deleted a massive facial recognition database, but it's not dead*, „Vice”, 6.06.2019, [on-line] <https://www.vice.com/en/article/microsoft-deleted-a-facial-recognition-database-but-its-not-dead> – 1.07.2024.

44 A. Harvey, J. LaPlace, *Exposing.ai. MegaFace*, 2021, [on-line] <https://exposing.ai/megaface> – 24.07.2024.

45 [on-line] <https://en.wikipedia.org/wiki/PimEyes> – 30.06.2024.

46 Od 2020 r. firma przeszła w ręce Face Recognition Solutions Ltd., firmy „zarejestrowanej w jednym pokoju na Seszelach, wyspiarskim kraju na Oceanie Indyjskim, który stał się popularnym rajem offshore dla firm chcących ukryć swój majątek i szczegóły korporacyjne”. D. Harwell, *This facial recognition website can turn anyone into a cop – or a stalker*, „The Washington Post”, 14.05.2021, [on-line] <https://www.washingtonpost.com/technology/2021/05/14/pimeyes-facial-recognition-search-secrecy> – 30.06.2024.

47 D. Harwell, dz. cyt.

48 K. Hill, *A face search engine anyone can use is alarmingly accurate*, „The New York Times”, 26.05.2022, [on-line] <https://www.nytimes.com/2022/05/26/technology/pimeyes-facial-recognition-search.html> – 30.06.2024.

49 K. Crawford, *9 ways to see a dataset: What's at stake in examining datasets?*, „Knowing Machines”, [on-line] <https://knowingmachines.org/publications/9-ways-to-see/essays/9-ways-to-see-a-dataset> – 5.07.2024; Gabriel Ilharco, *Announcing Datacomp: In Search Of The Next Generation Of Multimodal Datasets*, 27.04.2023, [on-line] <https://laion.ai/blog/data-comp> – 5.07.2024.

komputerowego, ale jest ich oczywiście o wiele więcej⁵⁰. Firmy technologiczne tworzące produkty dla instytucji państwowych i prywatnych firm sprawnie rozwijają metody ekstrakcji danych umiejętnie obchodząc regulacje prawne. Ograniczając naszą prywatność otrzymują darmowe materiały w postaci udostępnianych przez nas na szeroką skalę wizerunki twarzy⁵¹. Nasze *selfie* mogą zostać użyte do „trenowania i testowania algorytmów służących różnym celom związanym z biometrią – rozpoznawaniu twarzy przy kontroli paszportowej, nadzorowi nad tłumem, automatycznemu prowadzeniu pojazdów, robotyce, a nawet analizie emocji na potrzeby reklamy”⁵². Są także wykorzystywane do zbierania danych na temat osób, które nie są tego świadome, na przykład przez stalkerów, którzy z łatwością są w stanie wyśledzić swoją ofiarę.

Maszynowa wizja świata

Właściwie jesteśmy na początku rewolucji generatywnej, a już osiągnęliśmy „punkt, w którym całe terytorium Internetu przemieniło się w mapę sztucznej inteligencji”,⁵³. Dlatego tak ważne jest uświadomienie sobie, że „zakres, w jakim algorytmy widzenia komputerowego interpretują świat, jest ograniczony, sterowany i programowany przez zbiory treningowe”⁵⁴. Zbadanie tego, co się w nich znajduje jest pomocne w zrozumieniu sposobów kreowania szczególnej, maszynowej wizji naszej rzeczywistości oraz tego, w jaki sposób udostępnianie przez nas *selfie* zataczając koło wracają do nas w zupełnie innej formie mając odmienne zastosowanie.

Widzenie komputerowe w systemie nadzorowanym wymaga szkolenia na modelu posiadającym zestawy obrazów w połączeniu z opisem. Tego typu bazą treningową jest ImageNet⁵⁵. Autorką projektu jest profesor Fei-Fei Li z Artificial Intelligence Laboratory na Uniwersytecie Stanforda. Model został pomyślany jako widzenie komputerowe mogące być porównywalne do sposobów ludzkiej percepcji⁵⁶. Celem założycielki było stworzenie takiego produktu posiadającego uporządkowane według kategorii i opisów obrazy, dzięki któremu będzie można trenować uczenie maszynowe do rozpoznawania każdego obiektu. Do 2020 r. zbiór ImageNet posiadał ponad czternaście milionów fotografii podzielonych na około dwadzieścia tysięcy kategorii. Badanie Alexandry Sashy Luccioni oraz Kate Crawford wykazało,

50 Bardziej szczegółowa lista wraz z ich analizą krytyczną znajduje się na stronie: <https://exposing.ai/datasets/> i jest częścią projektu Adama Harveya *Exposing.ai*.

51 C. Metz, K. Hill, *Here's a way to learn if facial recognition systems used your photos*, „The New York Times”, 31.01.2021, [on-line] <https://www.nytimes.com/2021/01/31/technology/facial-recognition-photo-tool.html> – 5.07.2024.

52 M. Murgia, *Who's using your face?*, dz. cyt.

53 K. Crawford, *9 ways*, dz. cyt.

54 A. Harvey, *On computer vision*, dz. cyt.

55 [on-line] <https://www.image-net.org> – 22.11.2024.

56 F.-F. Li, *How we're teaching computers to understand pictures*, TED2015, March 2015, [on-line] https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures?subtitle=en – 5.07.2024; F.-F. Li, K. Yang, K. Qinami, O. Russakovsky, J. Deng, *Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the ImageNet hierarchy*, [w:] *Conference on Fairness, Accountability, and Transparency (FAT* '20)*, January 27–30, 2020, Barcelona, <https://doi.org/10.1145/3351095.3375709>.

że znajduje się w nim „562 626 twarzy na 1 431 093 obrazach, czasami z wieloma twarzami na jednym obrazie”⁵⁷.

Każda fotografia, gdy znajdzie się w takiej bazie, zostaje przypisywana do konkretnej klasy przedmiotów. Nadaje się jej stosowną etykietę w postaci rzeczownika zaczerpniętego z listy funkcjonującej w ramach bazy WordNet. Z powodu ogromu materiałów wizualnych, które należało w ten sposób opisać, Uniwersytet zatrudnił pracowników tymczasowych za pośrednictwem Amazon Mechanical Turk. Co roku werbowano od dwudziestu do trzydziestu tysięcy tzw. turkerów, którzy nadawali kategorii około pięciu zdjęciom na minutę otrzymując niewielką zapłatę za każdą z nich⁵⁸. Łącznie pracę posegregowania i opisanie bazy danych wykonało prawie pięćdziesiąt tysięcy ludzi ze stu sześćdziesięciu siedmiu krajów⁵⁹. W efekcie udostępniane na platformach internetowych *selfie* oraz fotografie przedstawiające ludzi w różnych prywatnych i publicznych sytuacjach w momencie, gdy znalazły się w bazie ImageNet, zostały użyte do „wyczerpującego zilustrowania wszystkich kategorii za pomocą obrazów”⁶⁰.

Tę obszerną bazę danych zanalizowała Kate Crawford we współpracy z Trevorem Paglenem w projekcie badawczo-artystycznym *Excavating AI* w 2019 r. Dotarli oni do listy etykiet, które nadawano wizerunkom człowieka. W efekcie ich badania ukazało niezwykle obraz językowej konstrukcji sfery wizualnej świata. Kategoria „osoba” zawierała niecałe trzy tysiące synset, wśród których były takie określenia jak: zła osoba, dziewczyna na telefon, narkoman, gej w ukryciu, skazaniec, szaleniec, nieudacznik, frustrat, drań, hipokryta, kleptomanka, przegrany, melancholik, primadonna, schizofrenik, epileptyk, stara panna, ulicznica, alkoholik, dziwka, bękart, samiec alfa, palant, osoba rozwiązała, tchórz, dłużnik, znajomy, daltonista, łamaga, patałach, profesor, przyjaciel, obcy, kochanka, neuronaukowiec kognitywny⁶¹. W bazie treningowej pojawiały się sceny nagości oraz przemocy, kategorie „prezydent” oraz „były prezydent” reprezentowały wyłącznie przywódców Stanów Zjednoczonych, najczęściej George’a W. Busha, „raper” to tylko zdjęcia czarnoskórych mężczyzn, a fotografie z etykietą „modelka” przedstawiały „prawie wyłącznie stereotypowe obrazy uprzedmiotowionych kobiet, z których wiele ma charakter pornograficzny”⁶². „Przedstawiciel prawni” to mężczyzna w garniturze, a kategorię „poród” reprezentowały drastyczne obrazy noworodków pokrytych krwią oraz fragmenty kobiecego ciała⁶³.

Crawford i Paglen demonstrują jak wiele z obrazów wykorzystywanych do trenowania jest obraźliwych, przemocowych i seksistowskich oraz – z powodu dużej stopnia ich abstrakcyjności – niemożliwych do zwizualizowania. Ta staranność

57 A.S. Luccioni, K. Crawford, *The nine lives of ImageNet: A sociotechnical retrospective of a foundation dataset and the limits of automated essentialism*, „Journal of Data-centric Machine Learning Research”, 29.02.2024, s. 5, [on-line] <https://openreview.net/forum?id=jh-0ck1bPGF> – 30.06.2024.

58 K. Crawford, T. Paglen, *Excavating AI: The politics of training sets for machine learning*, September 19, 2019, [on-line] <https://excavating.ai> – 20.07.2024.

59 F.-F. Li, dz. cyt.

60 K. Crawford, T. Paglen, dz. cyt.

61 Tamże; A.S. Luccioni, K. Crawford, dz. cyt.

62 Tamże, s. 9–10.

63 Tamże, s. 9.

w znalezieniu każdego wizualnego odpowiednika wszystkich synsety znajdujących się w leksykalnej bazie danych języka angielskiego WordNet ukazuje jak bardzo uczenie maszynowe wymaga jasnych wskazówek i uporządkowania kategorii bez względu na wszystko, kosztem banalizacji i trywializowania. Doprowadza to do szczególnej algorytmicznej esencjonalizacji człowieka nie mającej nic wspólnego z tym, kto właściwie jest uchwycony na fotografii, w jakiej sytuacji zrobiono zdjęcie, jaka jest historia tej osoby, skąd się wywodzi, jakie ma doświadczenia. Sprowadzono nas do statystycznych reprezentantów kulturowych pojęć nie mających z nami nic wspólnego. Sztuczna inteligencja przy użyciu ludzkich agentów nadających etykiety dostała spory pakiet konkretnych demonstrujących to, „jak wygląda »przegraną«, »dziwka« i »przestępca«”⁶⁴, a wyboru dokonały nieznanne nam osoby, z różnych stron świata, kierując się swoją opinią, uprzedzeniami, intuicją i moralnością. Z niezwykle złożonej tkanki społecznej i performatywnej tożsamości złożono katalog skończonych przedmiotów, które tylko wydają się tym, na co wyglądają. Pojęcia, które są wynikiem relacji i społecznego kontekstu (np. sprzedawca detaliczny, fryzjer, emeryt, zwierzchnik) zostały z dynamicznych układów sprowadzone do poziomu nieziemnego i wiecznego okazu w gablocie. Kate Crawford porównuje to do scentralizowanej produkcji tożsamości, nadzorowanej przez techniki uczenia maszynowego, które są dostępne w danym czasie⁶⁵. Jesteś tym, na kogo wyglądasz dla osoby nadającej etykiety bądź tym, za kogo ma cię algorytm widzenia komputerowego, bo tak go nauczono, takimi danymi dysponował. Wizerunki osób zostały użyte do opisanego świata wizualnego w systemie geometrycznej regularności i subiektywnej oceny, a horyzont rzeczywistości wyznaczony został przez to, na co może sobie pozwolić narzędzie oraz zakres słów w słowniku⁶⁶.

ImageNet to dobry przykład pokazujący, że tego typu klasyfikacje sprowadzają ludzi do roli przedmiotu⁶⁷ funkcjonującego w ramach tej samej porządkującej statystycznej formuły opartej na znajdowaniu regularności, wzorów i podobieństw. Crawford i Paglen kierowali się potrzebą zmapowania „logiki leżącej u podstaw sposobów wykorzystywania obrazów do trenowania systemów rozpoznawania twarzy i tego, w jaki sposób sztuczna inteligencja »widzi« świat”⁶⁸. Ostatecznie udowodnili, że polityka klasyfikacji ludzi stosowana w ten sam sposób co rozpoznawanie przedmiotów jest rasistowska, dyskryminująca, krzywdząca i dehumanizująca⁶⁹. Crawford w innym tekście dodaje także, że „nawet największe zbiory danych nie zapewniają uniknięcia zasadniczych błędów, które muszą występować, gdy nieskończenie złożony świat jest upraszczany i dzielony na kategorie”⁷⁰. Bez względu zatem na ilość informacji znajdujących się w modelu, każda próba opisanego świata przy użyciu matematycznych kategorii jedynie ograniczy nasze perspektywy jego

64 C. Metz, 'Nerd,' 'Nonsmoker,' 'Wrongdoer': How might A.I. label you?, „The New York Times”, 20.09.2019, [on-line] <https://www.nytimes.com/2019/09/20/arts/design/imagenet-trevor-paglen-ai-facial-recognition.html> – 1.07.2024.

65 K. Crawford, *Atlas*, dz. cyt., s. 138.

66 Tamże.

67 K. Crawford, T. Paglen, dz. cyt.

68 Tamże.

69 K. Crawford, *9 Ways*, dz. cyt.

70 Tejże, *Atlas*, dz. cyt., s. 105.

widzenia i rozumienia. Kultura wizualna, a szczególnie fotografia społecznościowa pazernie wykorzystywana w procedurach głębokiego uczenia reprezentuje pole, na którym testowane są algorytmiczne regularności i próby nowego zdefiniowania rzeczywistości tak intensywnie wpływającego na to jakie są platformy i dyskursy kształtujące naszą tożsamość.

Sztuczna sztuczna inteligencja⁷¹

Algorytmy uczenia maszynowego stają się tym samym narzędziami konstruującymi rasę, płeć, pochodzenie, charakter definiując następstwa dla osób tak sklasyfikowanych⁷². Te systemy są z nami codziennie, na każdym kroku. Model ImageNet jest najczęstszym systemem wykorzystywanym w produktach i usługach firm technologicznych oferujących np. możliwość rozpoznawania twarzy. Błędne przypisanie danego wizerunku do wybranej kategorii może mieć kolosalne znaczenie dla zarządzania i nadzoru nad milionami osób. Aktualnie technolodzy, którzy stworzyli ImageNet w reakcji na zarzuty o stronniczość i dyskryminację zaktualizowali swój model fundamentalny przyznając, że ponad połowa kategorii „osoba” posiadała obraźliwe etykiety, a z pozostałej grupy wyłącznie sto pięćdziesiąt osiem nie było obarczonych uprzedzeniami. Model został zmieniony, a wiele setek tysięcy fotografii usuniętych⁷³. Nie zmienia to jednak faktu, że nadal w ramach tej statystycznej kultury zostaliśmy sprowadzeni do wizualizacji jednej z wybranych odgórnie kategorii, którą algorytmy metodą predykcji dystrybuują dalej trenując kolejne systemy sztucznej inteligencji. Hito Steyerl nazywa je „uśrednionymi/podłymi obrazami” (*mean images*), ponieważ są „przewidywalnymi produktami populizmu danych” reprezentującymi to, co społeczne i polityczne, łącząc moralność ze statystyką⁷⁴. Są średnią z cyfrowych reprezentacji nas samych będąc przy tym podłą próbą dyscyplinowania nas poprzez odgórnie i apriorycznie nadane definicje. Nie dziwi zatem, że „świat według Stable Diffusion jest zarządzany przez białych mężczyzn będących prezesami. Kobiety rzadko są lekarzami, prawniczkami czy sędzinami. Mężczyźni o ciemnej skórze popełniają przestępstwa, podczas gdy kobiety o ciemnej skórze smażą burgery”⁷⁵. To nie jest wyłącznie wizja subiektywnej rzeczywistości osób, które kuratorują zbiory, ale także cyfrowo-statystyczna architektura uprzedzeń i stereotypów.

Angażowanie ludzi do tego, aby opracowywali zbiory treningowe w ramach darmowej bądź niskopłatnej pracy nie jest rzadkością. Ten system stosuje większość korporacji cyfrowych oferujących usługi generatywnej sztucznej inteligencji.

71 Jest to określenie użyte przez Jeffa Bezosa do opisu metod działania Amazon Mechanical Turk, który posiłkując się pracą ludzi wykonujących zadania zbyt trudne dla algorytmów, symuluje działanie sztucznej inteligencji. Zob. np. J. Pontin, *Artificial intelligence with help from the humans*, „The New York Times”, 25.03.2007, [on-line] <https://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html> – 18.08.2024.

72 K. Crawford, *Atlas*, dz. cyt., s. 150.

73 A.S. Luccioni, K. Crawford, dz. cyt.

74 H. Steyerl, dz. cyt.

75 L. Nicoletti, D. Bass, *Humans are biased. Generative AI is even worse*, „Bloomberg”, 9.06.2023, [on-line] <https://www.bloomberg.com/graphics/2023-generative-ai-bias> – 23.08.2024.

Akordowi pracownicy z Kairu⁷⁶, Kolumbii⁷⁷ czy Chin⁷⁸ nie tylko stają się odpowiedzialni za to, jak korzystający z aplikacji i serwisów zostaną sportretowani, z jakimi emocjami utożsamieni, ale także, oszczędzając nam widoku drastycznych i traumatycznych scen w procesie „czyszczenia Internetu”, biorą na siebie ryzyko doświadczenia dużej dawki stresu prowadzącego do zaburzeń emocjonalnych i psychicznych⁷⁹.

LAION-Aesthetics⁸⁰, podzbiór mający za zadanie skatalogowanie obrazów charakteryzujących się wysoką jakością artystyczną, jest ciekawym przykładem odgórnego kreowania kategorii piękna. Christo Buschek oraz Jer Thorp w ramach projektu *Knowing Machines*⁸¹ śledząc procedury opracowywania tej bazy danych treningowych zaobserwowali, że ten konkretny zbiór jest szczególnie nie tylko dlatego, że to na nim szkolona była generatywna sztuczna inteligencja Stable Diffusion oraz Midjourney, ale także z powodu wyboru osób odpowiedzialnych za ten cyfrowy kanon. Byli to „użytkownicy społeczności platformy Discord, [...] którzy są określane jako WEIRD (*Western, Educated, Industrialized, Rich i Democratic*)”⁸², ale także wybrani twórcy sztuki AI, przez niektórych nazywani „ekscentrycznymi nerdami”. Do wytrenowania generatywnej sztucznej inteligencji zostały użyte takie dzieła, których wyboru dokonały osoby w większości pochodzące z USA, Kanady i Europy. Ci „głównie średniozaawansowani entuzjaści fotografii z małych amerykańskich miast”⁸³ stoją za wygenerowanymi obrazami. Ich gust oraz opinia na temat tego, czym jest piękno, co ma wysoką wartość estetyczną jest zapisany w kodzie. Tym samym stworzyli sztuczny kanon odpowiadający za styl i prawidłowość definicji dzieła sztuki.

Systemy błędnego kategoryzowania

Wygenerowane obrazy nigdy nie są uniwersalne czy obiektywne. To efekt pracy ludzkich agentów kierujących się swoimi poglądami, odczuciami, pochodzących

76 B. Perrigo, *OpenAI used Kenyan workers on less than \$2 per hour to make ChatGPT less toxic*, „Time”, 18.01.2023, [on-line] <https://time.com/6247678/openai-chatgpt-kenya-workers> – 18.08.2024.

77 B. Perrigo, *TikTok's subcontractor in Colombia under investigation for traumatic work*, „Time”, 10.11.2022, [on-line] <https://time.com/6231625/tiktok-teleperformance-colombia-investigation> – 15.08.2024.

78 Y. He, H. She, *Beyond the great power competition narrative: Exploring labor politics and resistance behind AI innovation in China*, „Georgetown Journal of Asian Affairs” 2021, vol. 7.

79 R. Spence, A. Bifulco, P. Bradbury, E. Martellozzo, J. DeMarco, *Content moderator mental health, secondary trauma, and well-being: A cross-sectional study*, „Cyberpsychology, Behavior, and Social Networking” 2024, vol. 27, no. 2, <https://doi.org/10.1089/cyber.2023.02>.

80 Modelem nadrzędnym był LAION-5B (Large-scale AI Open Network), za którego odpowiedzialna była firma Nublis.

81 [on-line] <https://knowingmachines.org> – 22.11.2024.

82 A. Harvey, *Today's selfie*, dz. cyt.

83 Tamże. Więcej informacji o błędach towarzyszących modelowi treningowemu LAION można znaleźć w tekście J. Schultz, *9 ways to see a dataset: What can LAION teach us about copyright law?*, „Knowing Machines”, [on-line] <https://knowingmachines.org/9-ways-to-see/LAION-copyright> – 5.07.2024.

z różnych rejonów świata, mających zatem bardzo konkretne doświadczenia będące pochodną kulturowych i społecznych uwarunkowań. Same modele stają się tym samym „opiniami opisanymi językiem matematyki”⁸⁴, które nie podlegają dyskusji ani odwołaniu⁸⁵. Są obliczone na ograniczanie kosztów produkcji przy równoczesnym banalizowaniu skutków jakie mogą wywoływać. W sytuacji, gdy produkty oparte na sztucznej inteligencji coraz intensywniej anektują zakres swojego oddziaływania na nasze życie, świadomość metod stojących za procesem ich powstawania jest niezwykle ważna. Istota sposobów tworzących ten system ma bowiem przełożenie także na mechanikę tożsamościową, klasową, sposoby organizujące nasze codzienne życie. Porządkując fotografie na podstawie uogólnień, subiektywnych ocen, kategorii i klas zamykamy nieskończoność opcji, abstrakcyjność emocji i wieloznaczność interpretacji obiektów kultury wizualnej w konkretnej, politycznej architekturze sztucznych sieci neuronowych. Przecież „nawet dobrze zaprogramowany model, jeśli zostanie nakarmiony niereprezentatywnymi danymi, wygeneruje fałszywe wyniki”⁸⁶ – jak przekonują nas założycielki Fundacji Panoptykon stojącej na straży wykorzystywania sztucznej inteligencji zgodnie z prawem i zasadami etycznymi.

Dane początkowe złej jakości dadzą niewłaściwy, mylący wynik końcowy (*garbage in, garbage out*⁸⁷). To mogą być materiały, które są błędne, niepełne bądź uprzedzone, co skutkuje na przykład źle rozpoznany podejrzany, który, mimo że nie popełnił przestępstwa, może zostać aresztowany⁸⁸. Tego rodzaju zdarzenia są szczególnie niebezpieczne w sytuacji systemów rozpoznawania twarzy bądź emocji, które mają duże przełożenie na nasze zatrudnienie czy kontrolę graniczną. Miejmy na względzie to, że trenowanie tego typu widzenia komputerowego przebiega w procesie nadzorowanego uczenia maszynowego, któremu dostarczane są dane w postaci portretów wraz z etykietami emocji, które są również przypisywane przez ludzi. Osoby, którym zlecono zadanie uporządkowania danych, korzystają z listy sformułowanej przez Paula Ekmana, koncepcji sześciu podstawowych, uniwersalnych emocji tj. radość, smutek, złość, strach, zaskoczenie, wstręt mimowolnie wyrażanych za pomocą mimiki twarzy. Badacz opracował „The Facial Action Coding System” (FACS) wykorzystywany do identyfikacji widocznych ruchów mięśni twarzy. Stworzył prototyp mapujący widoczne napięcia będące bezwiedną reakcją na afekty, np. zmarszczone brwi, szeroko otwarte oczy, zaciśnięte usta w jego wykazie oznacza odczuwanie złości. Ekspresja twarzy według badacza jest tożsama

84 C. O’Neil, *Broń matematycznej zagłady. Jak algorytmy zwiększają nierówności i zagrażają demokracji*, tłum. M.Z. Zieliński, Warszawa 2017, s. 48.

85 Tamże, s. 27.

86 A. Obem, K. Szymielewicz, *Sztuczna inteligencja non-fiction*, Fundacja Panoptykon, 2020, [on-line] <https://panoptykon.org/sztuczna-inteligencja-non-fiction> – 2.04.2024.

87 M. Rouse, *Garbage in, garbage out*, „Techopedia”, 4.01.2017, [on-line] <https://www.techopedia.com/definition/3801/garbage-in-garbage-out-gigo> – 5.07.2024. W języku polskim zjawisko to nazywane jest „śmieci na wejściu, śmieci na wyjściu”. Zob. np. K. Kopańko, *Śmieci na wejściu, śmieci na wyjściu. Tak programiści kwitują sztuczną inteligencję*, „Spider’sWeb / bizblog”, 29.07.2024, [on-line] <https://bizblog.spidersweb.pl/smieci-na-wejsciu-smieci-na-wyjsciu-tak-programisci-kwituja-sztuczna-inteligencje> – 2.08.2024.

88 K. Hill, *Facial recognition led to wrongful arrests. So Detroit is making changes*, „The New York Times”, 29.06.2024, [on-line] <https://www.nytimes.com/2024/06/29/technology/detroit-facial-recognition-false-arrests.html> – 3.08.2024.

z emocjami, które przeżywa dana osoba. Dane używane w modelach treningowych zostały pozyskane poprzez fotografowanie ludzi, których poproszono o wykonanie grymasu zgodnego z opisem. Ten emocjonalno-wizualny alfabet oparty na upozowanych zdjęciach został przełożony na dane matematyczne, których zadaniem jest automatyczne przyporządkowywanie konkretnych napięć i ruchów mięśni rozpoznawanych na twarzach do jednego z sześciu afektów. Systemy tego typu są wykorzystywane w identyfikowaniu poszukiwanych osób, kontroli dostępu do budynków, odblokowywania telefonów komórkowych, wykrywania podejrzanych zachowań, personalizacji reklam, badaniu reakcji na produkty marketingowe, w opiece zdrowotnej, monitorowaniu stanu pacjenta itd.

Istnieje szereg przesłanek ku temu, aby podważyć wiarygodność zarówno efektów badań Ekmana, jak i zakwestionować nieomyślność systemów z powodu braku reprezentatywności całego złożonego spektrum kulturalno-społecznego, psychologicznego czy rasowego tych modeli. Istnieje szereg artykułów i opracowań naukowych, w których zostały uwzględnione podstawowe problemy wynikające z takiego typu technologii kontroli. W tekście *The unbearable (technical) unreliability of automated facial emotion recognition* autorzy przedstawiają wyniki badań, w których udowodnili, że systemy rozpoznawania twarzy trenowane na etykietowanych danych są obciążone dużą dozą subiektywności i nierzadko z powodu pozbawienia kontekstu społecznego i sytuacyjnego są mylnie oznaczane. Naukowcy przyglądali się przede wszystkim „wiarygodności tzw. »ground truth«, sprawdzając stopień, w jakim różni ludzie zgadzają się co do emocji, które rozpoznają na twarzach badanych osób”⁸⁹, czyli porównywaniu efektywności działań algorytmów, ich poprawności i zgodności. Według tego opracowania raterzy odpowiedzialni za oznaczanie zdjęć nazwami emocji z uwagi na to, że „nie otrzymują żadnych informacji na temat przedstawionych osób, kontekstu społecznego ani warunków, w jakich funkcjonuje dana osoba oraz w jakiej sytuacji zrobiono zdjęcie”⁹⁰, są pozbawieni możliwości przyporządkowania adekwatnej etykiety, co prawdopodobnie podniosłoby skuteczność oznaczania. Można bowiem płakać nie tylko z powodu doświadczanego smutku czy żałoby, ale także w trakcie odbierania ważnej nagrody czy w kontakcie ze swoim idolem. W eksperymencie potwierdzono także, że istnieje niska zgodność między raterami w rozpoznawaniu emocji. Niewiele fotografii zostało ocenionych w ten sam sposób. Świadczy to o naszych trudnościach w bezbłędnej ocenie stanu emocjonalnego drugiej osoby wyłącznie na podstawie skadrowanej, statycznej fotografii. Jest możliwe, że różnice w nazywaniu emocji są skutkiem tzw. efektu Kuleszowa, polegającego na takiej metodzie montowania ze sobą różnych scen i obrazów, których sąsiedztwo wpływa na interpretację. Takie zestawienie tworzy „historię”, opowieść, która może ewoluować w czasie co wywołuje w nas zmienne reakcje, ale i mylne interpretacje⁹¹.

Do podobnych wniosków doszła Lisa Feldman Barrett, profesorka psychologii na Uniwersytecie Northeastern, wraz z zespołem, udowadniając, że ani odegrane,

89 F. Cabitza, A. Campagner, M. Mattioli, *The unbearable (technical) unreliability of automated facial emotion recognition*, „Big Data & Society”, July–December, 2022, s. 3, <https://doi.org/10.1177/20539517221129549>.

90 Tamże, s. 4.

91 Tamże, s. 12.

upozowane ani spontaniczne wyrazy twarzy, czyli specyficzne konfiguracje mięśni, nie są tym samym co odciski palców, dlatego też „nie można z pewnością wnioskować o radości na podstawie uśmiechu, złości na podstawie grymasu czy smutku na podstawie zmarszczenia brwi”⁹². Uśmiech w zależności od kontekstu i osobowości może wyrażać radość, ale też pogardę, współczucie, dumę, zakłopotanie, kpinę czy strach pojawiający się w sytuacji próby załagodzenia sytuacji. Uczenie maszynowe nie dostaje całego spektrum złożoności i kontekstualności, lecz uporządkowane i jasne kategorie. Systemy sztucznej inteligencji potrafią co prawda wykryć grymas, „ale to nie to samo, co wykrycie złości”⁹³. Wyciągając średnią statystyczną nie zastanawia się nad adekwatnością sytuacyjną danej ekspresji, nie polemizuje i nie podaje w wątpliwość. Systemy automatycznego rozpoznawania emocji nie są oparte na naukowych dowodach, lecz powszechnych poglądach wynikających z błędnych przekonań, że emocje można bezbłędnie wyczytać z twarzy. Algorytmy rozpoznawania twarzy udostępniane przez firmy takie jak Microsoft, IBM czy Amazon mogą na podstawie mimiki uchwyconego w obiektywie kamery bądź aparatu człowieka wnioskować o towarzyszących mu emocjach. Korporacje, które dysponują taką formą nadzoru mogą rozpoznać zagrożenie w postaci wysłedzenia człowieka, który zmarszczył brwi i zacisnął usta.

Caty O’Neill dla takich szkodliwych odmian modeli wymyśliła osobliwą nazwę: „Broń Matematycznej Zagłady, w skrócie Beemzet”⁹⁴. Te, które „definiują własną rzeczywistość, a następnie wykorzystują ją do uzasadnienia generowanych wyników”⁹⁵ są najbardziej niebezpieczne i destrukcyjne dla sprawiedliwości społecznej. Służby bezpieczeństwa, departamenty policji i instytucje rządowe szczególnie w Stanach Zjednoczonych oraz w Chinach na masową skalę wykorzystują systemy rozpoznawania twarzy do kontroli i identyfikacji podejrzanych czy osób przekraczających granice. Jednak jak pokazuje badanie amerykańskiego Narodowego Instytutu Standardów i Technologii (The National Institute of Standards and Technology, NIST) opublikowane w 2019 r., mimo posiadania olbrzymich zbiorów systemy niepoprawnie identyfikowały twarze rdzennych Amerykanów, Afroamerykanów i Azjatów. Wiele fałszywie negatywnych wyników pojawiało się także w stosunku do kobiet, osób starszych oraz dzieci⁹⁶. Podobny efekt przedstawiają rezultaty badań Joy Buolamwini i Timnit Gebru, które udowodniły, że „kobiety o ciemniejszym kolorze skóry są najczęściej źle klasyfikowaną grupą (z wskaźnikami błędów sięgającymi do 34,7%)”, podczas gdy „[m]aksymalny wskaźnik błędów dla mężczyzn

92 L. Feldman Barrett, R. Adolphs, S. Marsella, A.M. Martinez, S.D. Pollak, *Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements*, „Psychological Science in the Public Interest” 2019, vol. 20, iss. 1, s. 46, <https://doi.org/10.1177/1529100619832930>.

93 J. Vincent, *Discover the stupidity of AI emotion recognition with this little browser game*, „The Verge”, 6.04.2021, [on-line] <https://www.theverge.com/2021/4/6/22369698/ai-emotion-recognition-unscientific-emoji-web-browser-game> – 14.09.2024.

94 C. O’Neil, dz. cyt., s. 27.

95 Tamże, s. 31.

96 P. Grother, M. Ngan, K. Hanaoka, *Face Recognition Vendor Test (FRVT). Part 3: Demographic effects*, National Institute of Standards and Technology, 2019, <https://doi.org/10.6028/NIST.IR.8280>.

o jaśniejszej karnacji wynosi 0,8%⁹⁷. Powoduje to, że ryzyko pomyłki osób o ciemnym kolorze skóry w sytuacjach nadzoru i kontroli jest o wiele większe. Przy tym wiąże się z krzywdą i cierpieniem tej grupy osób, które zostały niesprawiedliwie potraktowane, co pokazuje jak bardzo systemy są ułomne, niesprawiedliwe przez swoją nieobiektywność.

Przykładów występowania uprzedzonych algorytmów w obszarze sztucznej inteligencji i uczenia maszynowego jest znacznie więcej, dlatego warto zastanowić się czy przypadkiem nie jest tak, że „[p]rawda okazuje się [...] nie tyle odzwierciedleniem faktów lub rzeczywistością intersubiektywną, ile stosem obrazów pobranych z rozmaitych dostępnych internetowych źródeł”⁹⁸. Reprezentuje zatem nie naszą rzeczywistość, ale jej algorytmiczną wersję stworzoną przez konkretne osoby, w konkretnym miejscu i czasie, przy użyciu specyficznych metod, na wybranym zestawie danych, których celem nie jest ukazanie ludzkiej performatywnej, kulturowo-biologicznej i kontekstualnej złożoności, lecz utrwalenie schematyczności i wzorów zwizualizowanych na fotografiach „przedmiotów”, w tym matematycznego uporządkowania i statystycznego zdefiniowania tego, kim jest człowiek na podstawie tylko i wyłącznie jego wyglądu. Skutki w postaci wygenerowanych obrazów ukazujących uproszczony i uprzedzony obraz świata są efektem decyzji twórców systemów, a nie samej sztucznej inteligencji.

Maszyny do autouzupelniania

Z tej perspektywy ważne jest przyjrzenie się leksykalnej stronie technologii. Kate Crawford pisze, że sztuczna inteligencja nie jest sztuczna, wymaga bowiem olbrzymiej infrastruktury, „powstaje z naturalnych zasobów, paliwa, ludzkiej pracy, [...] logistyki, historii i klasyfikacji”⁹⁹, ale także nie jest inteligentna, algorytmy bowiem nie są w stanie podejmować samodzielnych, racjonalnych decyzji bez „kosztownego obliczeniowo trenowania na wielkich zbiorach danych”¹⁰⁰. Nie istnieje bez początkowych ludzkich decyzji ani danych przez nich wytworzonych. W książce *Atlas sztucznej inteligencji* oraz w innych tekstach i projektach artystycznych¹⁰¹ badaczka udowadnia, że uczenie maszynowe jest rozbudowanym systemem wydobywczym opartym na eksploatacji zasobów naszej planety, taniej siły roboczej¹⁰², kontroli ciał

97 J. Buolamwini, T. Gebru, *Gender shades: Intersectional accuracy disparities in commercial gender classification*, „Proceedings of Machine Learning Research” 2018, vol. 81, s. 1; zob. także projekt Gender Shades dostępny na stronie: <http://gendershades.org/overview.html> – 3.08.2024.

98 K. Crawford, *Atlas*, dz. cyt., s. 103.

99 Tamże, s. 17.

100 Tamże.

101 K. Crawford, V. Joler, *Anatomy of an AI System. An anatomical case study of the Amazon echo as an artificial intelligence system made of human labor*, 2018, [on-line] <https://anatomyof.ai> – 22.11.2024; K. Crawford, T. Paglen, *Excavating AI. The politics of images in machine learning training Sets*, 2019, [on-line] <https://excavating.ai> – 22.11.2024.

102 Zarówno tych, którzy udostępniają dane poprzez aplikacje i platformy, jak i cyfrowych pracowników akordowych, którzy „czyszczą” zawartość mediów społecznościowych z treści niebezpiecznych i kontrowersyjnych oraz tych odpowiedzialnych za etykietowanie i klasyfikowanie danych używanych do tworzenia modeli fundamentalnych w celu trenowania uczenia maszynowego.

przez nielegalne przejmowanie gigantycznych zbiorów danych, w tym intymnych i biometrycznych, służących trenowaniu sztucznych sieci neuronowych czy systemów rozpoznawania twarzy i emocji. SI jest karmiona naszą prywatnością, której zawłaszczanie jest dokonywane w imię nauki i postępu technologicznego mających służyć usprawnieniu działania instytucji i zwiększeniu poczucia bezpieczeństwa. A w gruncie rzeczy jest przede wszystkim reżimem rozumowania normatywnego oraz:

istotnym czynnikiem kształtującym wiedzę, komunikację i stosunki władzy. Rekonfiguracje dotyczące poziomu epistemologicznego, zasad sprawiedliwości, organizacji społecznej, politycznej ekspresji, kultury, rozumienia ludzkiego ciała, podmiotowości i tożsamości – kwestii tego, czym jesteśmy i czym możemy się stać¹⁰³.

Podobne stanowisko ma Will Douglas Heaven, który twierdzi, że uczenie maszynowe to „pojęcie, wizja, pewien rodzaj spełnienia życzeń. Te idee kształtują się pod wpływem innych idei jak moralność, quasi-religijne przekonania, światopogląd, polityka i intuicja”¹⁰⁴. Jego dominujące rozpoznania reprezentują zatem szczególną formę dyskursu racjonalizującego destrukcyjne rozwiązania.

Zwróćmy zatem uwagę na wybrane opracowania badawcze, które wskażą parę tropów, za którymi moglibyśmy podążać chcąc zrozumieć ową hegemoniczną strukturę dyskursu technologicznego. Emily M. Bender i Alexander Koller sugerują ostrożność w definiowaniu sztucznej inteligencji jako porównywalnej w swoim działaniu z ludzkim umysłem. Optymistyczne hasła „prowadzą do przesadnych oczekiwań, w których modele [duże modele językowe] są opisywane jako »rozumiające« język lub jako wychwytyjące »znaczenie«”¹⁰⁵. Tymczasem są one uczone „wyłącznie na samej formie języka nie mając dostępu do znaczeń będących wynikiem aktów komunikacji”¹⁰⁶. W kolejnym badaniu Emily M. Bender prowadzonym wraz z Timnit Gebru, Angeliną McMillan-Major, Shmargaret Shmitchell nazwały LLMy „papugą stochastyczną”¹⁰⁷, które w sposób przypadkowy prognozują następowanie po sobie, których nie można przewidzieć ze stuprocentową pewnością. Wygenerowane wyniki stają się tym samym losowymi elementami opartymi na statystycznie sprofilowanym prawdopodobieństwie. Te syntetyczne, pozornie spójne teksty wytrenowane na hegemonicznych przekazach funkcjonują poza regułami rzetelnych i sprawdzonych opracowań. Chatboty będąc coraz lepszymi w tym co robią, stając się coraz dokładniejszymi dzięki zwiększanej ilości materiałów tekstowych do trenowania oraz naszemu używaniu, powodują, że tracimy czujność i błędy możemy brać za coś pewnego, potwierdzonego. Generatywna sztuczna inteligencja jest

103 K. Crawford, *Atlas*, dz. cyt., s. 28.

104 W.D. Heaven, *What is AI?*, dz. cyt.

105 E.M. Bender, A. Koller, *Climbing towards NLU: On meaning, form, and understanding in the age of data*, [w:] *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2020, s. 5185.

106 Tamże, s. 5193.

107 E.M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, *On the dangers of stochastic parrots: Can language models be too big?*, [w:] *Conference on Fairness, Accountability, and Transparency (FAccT '21)*, March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, <https://doi.org/10.1145/3442188.3445922>.

pozbawiona odpowiedzialności za informacje, nie mamy do kogo się zwrócić w celu sprostowania fałszywych treści¹⁰⁸. Duże modele językowe, takie jak GPT-4¹⁰⁹, są niezwykle sprawne w rozwiązywaniu zadań maturalnych, testów IQ, zdawaniu egzaminów uniwersyteckich czy adwokackich. Świetnie radzą sobie z testami programistycznymi organizowanymi przez Codeforces. Jednak jest tam szczególnie haczyk. Zdają tylko te egzaminy, których wyniki zostały opublikowane w Internecie przed 2021 rokiem. Wyniki są gorsze, gdy dotyczą materiałów po tej dacie¹¹⁰.

Wracając do nazewnictwa technologii, z którą mamy do czynienia – to nie inteligencja, lecz umiejętność zapamiętywania i kopiowania, a nie wnioskowania. Podobne problemy z „inteligencją” sztucznej inteligencji badacze zauważyli w zadaniach na poziomie edukacji przedszkolnej, co może być skutkiem zbyt małej ilości odpowiednich danych treningowych oraz ograniczonych możliwości rozumowania logicznego czy znajomości zasad interakcji społecznych¹¹¹, której sztucznej inteligencji po prostu brak.

Heaven wypowiada się o nich, jako „superwydajnych maszynach do autouzupełniania”, które przestrzeń informacyjną przemieniły w statystyczną reprezentację następujących po sobie słów o „miliardach wartości numerycznych”¹¹². To losowy wymiar algorytmicznych reguł matematycznych, a nie wiedza jako taka. Zbliżone zdanie ma Gary Marcus, który w tekście *Deep Learning Is Hitting a Wall* pisze, że „uczenie głębokie będące w istocie techniką rozpoznawania wzorców sprawdza się najlepiej w momentach, gdy potrzebujemy jedynie wstępnych, przybliżonych wyników”¹¹³. Czyli wszędzie tam, gdzie pomyłka nie wywoła sporych kryzysów, a brak perfekcji nie będzie żadnym utrudnieniem. Ryzykowne jest jej używanie w sytuacjach mających bezpośredni wpływ na ludzkie życie. Marcus zaznacza, że „niewiele dziedzin jest bardziej przepełnionych szumem niż sztuczna inteligencja”¹¹⁴. Szumem, do którego możemy zaliczyć także powszechne cyfrowe halucynacje, błędy wydające się pozornie wiarygodnymi informacjami. Galactica, model firmy Meta, zaliczył symptomatyczną serię wpadek. Wytrenowany na czterdziestu ośmiu milionach różnego rodzaju artykułów naukowych, podręcznikach, notatkach z wykładów akademickich i encyklopedii został stworzony do celów badawczych, jako pomoc dla studentów i naukowców. Jednak bardzo szybko okazało się, że podstawowym problemem Galactiki była nieumiejętność odróżnienia prawdy od kłamstwa. Odkryto, że „tworzyła ona fałszywe artykuły (czasem przypisując je prawdziwym autorom), generowała teksty na temat historii niedźwiedzi w kosmosie z taką samą wiarygodnością, jak w przypadku opisu działania związków chemicznych czy informacji na

108 Tamże.

109 Od momentu ukończenia pracy nad pisaniem tego tekstu OpenAI zdążył uruchomić jego dwie nowe wersje tj. GPT-4o oraz GPT-4o1, które mają zaktualizowaną bazę danych tekstowych do 2023 r.

110 W.D. Heaven, *AI hype is built on high test scores. Those tests are flawed*, „MIT Technology Review”, 30.08.2023, [on-line] <https://www.technologyreview.com/2023/08/30/1078670/large-language-models-arent-people-lets-stop-testing-them-like-they-were> – 25.08.2024.

111 Tamże.

112 W.D. Heaven, *What is AI?*, dz. cyt.

113 G. Marcus, *Deep learning is hitting a wall*, „Nautilus”, 10.03.2022, [on-line] <https://nautilus.us/deep-learning-is-hitting-a-wall-238440> – 24.08.2024.

114 E.M. Bender, T. Gebru, A. McMillan-Major, S. Shmitchell, dz. cyt.

temat prędkości światła¹¹⁵. Niebezpieczeństwo pojawia się wtedy, gdy mamy małą wiedzę na dany temat. Może to skutkować rozsiewaniem zmyślonych przez sztuczną inteligencję prawd, którą bierzemy za pewnik, pomyłkami, które traktujemy jako fakty czy błędne wskazówki, którymi bez uważności możemy się kierować.

Niektórzy nazywają uczenie maszynowe „inteligencją reklamową”¹¹⁶ opartą na błyskotliwych hasłach rozwoju technologii przeznaczonej wyłącznie dla bogatych bądź „fałszomatyzacją”, ponieważ „system jedynie koordynuje ludzką pracę odbywającą się w ukryciu. [...] Wiele systemów stanowi połączenie pracy źle opłacanych akordowych pracowników cyfrowych i konsumentów wykonujących nieodpłatnie zadania”¹¹⁷, w tym udostępnianie fotografii, filmów, tekstów, dzieł cyfrowych czy artykułów. Produkty oparte na SI są budowane dla uśrednionego użytkownika, a dane treningowe bazują na mówiących po angielsku osobach z zachodniego świata posiadających wyraźnie normatywne cechy¹¹⁸.

Zespół badawczy z Uniwersytetu w Oxfordzie, Uniwersytetu w Cambridge oraz Uniwersytetu w Toronto zauważył, że generatywna sztuczna inteligencja produkując coraz więcej treści będących wynikiem trenowania na danych wyjściowych innych modeli prowadzi do efektu, który nazwali „zapaścią modelu” (*model collapse*). Zachodzi on w sytuacji namnożenia syntetycznych tekstów przez co model zaczyna generować niespójne, bezużyteczne treści tracąc dostęp do danych pierwotnych¹¹⁹, w tym wypadku różnorodnych tekstów specjalistycznych, badawczych, literackich, reportaży i innych artykułów opartych na wiarygodnych i sprawdzonych informacjach napisanych przez człowieka. Heaven dodaje, że złożoność modeli jest tak duża, że sami badacze nie są do końca pewni z czym mają do czynienia, a wiele z ich obserwacji „stoi w sprzeczności z klasyczną statystyką, która wcześniej dostarczała najlepszego zestawu wyjaśnień dotyczących zachowania modeli predykcyjnych”¹²⁰. Podobnego zdania jest również Kasia Chmielinski nazywając rzeczywistość cyfrową „Dzikim Zachodem danych”¹²¹, w którym nie funkcjonują żadne standardy oceny jakości zarówno tego, co znajduje się wewnątrz modeli fundamentalnych, na których trenowane są systemy sztucznej inteligencji, jak i nieprzejrzystość algorytmów

115 W.D. Heaven, *Why Meta's latest large language model survived only three days online*, „MIT Technology Review”. 18.11.2022, [on-line] <https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science> – 15.05.2024.

116 D. Sumpter, *Osaczeni przez liczby. O algorytmach, które kontrolują nasze życie. Od Facebooka i Google'a po fakty newsy i bańki filtrujące*, tłum. Radosław Kosarzycki, Copernicus Center Press, 2019, s. 301

117 K. Crawford, *Atlas*, dz. cyt., s. 75.

118 K. Chmielinski, *Why AI needs a „nutrition label”*, TED Salon: The Rockefeller Foundation, May 2024, [on-line] https://www.ted.com/talks/kasia_chmielinski_why_ai_needs_a_nutrition_label/transcript?subtitle=en – 21.08.2024.

119 I. Shumailov, Z. Shumaylov, Y. Zhao, N. Papernot, R. Anderson, Y. Gal, *AI models collapse when trained on recursively generated data*, „Nature”, vol. 631, 24/07.2024, s. 757, <https://doi.org/10.1038/s41586-024-07566-y>.

120 W.D. Heaven, *Large language models can do jaw-dropping things. But nobody knows exactly why*, „MIT Technology Review”, 4.03.2024, [on-line] <https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why> – 2008.2024.

121 K. Chmielinski, dz. cyt.

chronionych przez korporacje jako ich własność intelektualna. Czy chciałbyś jeść w restauracji, która serwuje smaczne dania, lecz nie masz pojęcia co się w nich znajduje, z czego są przyrządzone? – zadaje kluczowe pytanie Chmielinski.

Z jednej strony obserwujemy nieograniczony prawem swobodny dostęp do efektów ludzkiej pracy wykonywanej w ramach cyfrowej technologii, niekontrolowalność procesów zachodzących wewnątrz systemów sztucznych sieci neuronowych, koszty środowiskowe oraz społeczne. Z drugiej – otrzymaliśmy zaszumioną przestrzeń dystrybucyjną i proces weryfikacji, gdzie widzenie komputerowe, a w szczególności systemy rozpoznawania twarzy, stosuje się do sprawowania społecznej kontroli za pomocą nadzoru wizyjnego, ale także projektowanie dyskursu obiektywnej przydatności. W takich warunkach na podstawie naszych portretów są opracowywane prognozy dotyczące skłonności przestępczej, rekrutacji do pracy (Amazon przy użyciu sztucznej inteligencji porównywał twarze kandydatów do pracy z tymi, którzy są najlepiej ocenianymi pracownikami, sprawdza się w ten sposób na przykład uczciwość i zaangażowanie), tymi metodami posługują się służby na lotniskach, w systemach wykrywających choroby psychiczne, czy w krajowych systemach bezpieczeństwa. Tak pobrane i opracowane dane biometryczne mają swoje zastosowanie w kabinach samochodowych kurierów monitorując ich koncentrację, podobnie pracowników biurowych, których gałki oczne są śledzone pod kątem uwagi jaką poświęcają na swoją pracę. Integracja technologii rozpoznawania twarzy wraz z systemami biometrycznymi są opisywane jako mające zapewnić bezpieczeństwo czy mierzyć produktywność, podczas gdy w rzeczywistości stają się narzędziami kontroli i dyscyplinowania, do których tylko nieliczni mają wgląd¹²².

Poruszamy się w architekturze cyfrowej będącej wynikiem precyzyjnych matematycznych obliczeń, gdzie tzw. korelacja statystyczna wykrywa głównie wzory, a nie wiarygodne reprezentacje tak przecież złożonego świata. Jest coś kuszącego i fascynującego w narzędziach sztucznej inteligencji pomagających w pracy, wyszukiwaniu informacji, opracowywaniu nowych kodów, rozpoznawaniu chorób czy wzbogacających naszą rozrywkę. Nie możemy jednak zapominać, że każdy z tych systemów jest wypracowany w konkretnych warunkach społeczno-ekonomicznych, przez ludzi, których zadaniem jest zaspokojenie głodu finansowego akcjonariuszy. Wiąże się to z szeregiem niebezpieczeństw dlatego nie powinniśmy na tym etapie ufać jej w takim stopniu, jak pokazuje to wspomniany na początku raport z 2023 r. *Global Views on A.I.*¹²³. Nauczmy się z uwagą przyglądać najnowszym technologiom i z większą dozą krytycznego dystansu nieustannie poddawać ją testom i w sposób bardziej stanowczy wymagać prawnych regulacji.

Bibliografia

Bender E.M., Koller A., *Climbing towards NLU: On meaning, form, and understanding in the age of data*, [w:] *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2020, s. 5185–5198.

122 *Biometric surveillance is quietly expanding: Bright-line rules are key*, AI Now Institute, 11.04.2023, [on-line] <https://ainowinstitute.org/publication/biometric-surveillance-is-quietly-expanding> – 20.07.2024.

123 *Global views on A.I. 2023*, dz. cyt.

- Bender E.M., Gebru T., McMillan-Major A., Shmitchell S., *On the dangers of stochastic parrots: Can language models be too big?*, [w:] *Conference on Fairness, Accountability, and Transparency (FAccT '21), March 3–10, 2021*, Virtual Event, Canada. ACM, New York, NY, USA, s. 610–623, <https://doi.org/10.1145/3442188.3445922>.
- Biometric surveillance is quietly expanding: Bright-line rules are key*, AI Now Institute, 11.04.2023, [on-line] <https://ainowinstitute.org/publication/biometric-surveillance-is-quietly-expanding> – 20.07.2024.
- Buolamwini J., Gebru T., *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, „Proceedings of Machine Learning Research” 2018, vol. 81.
- Buschek Ch., Thorp J., *Models all the way down*, „Knowing machines”, [on-line] <https://knowingmachines.org/models-all-the-way#section4> – 5.07.2024.
- Cabitz F., Campagner A., Mattioli M., *The unbearable (technical) unreliability of automated facial emotion recognition*, „Big Data & Society”, July–December, 2022, <https://doi.org/10.1177/20539517221129549>.
- Chmielinski K., *Why AI needs a „nutrition label”*, TED Salon: The Rockefeller Foundation, May 2024, [on-line] https://www.ted.com/talks/kasia_chmielinski_why_ai_needs_a_nutrition_label/transcript?subtitle=en – 21.08.2024.
- Cole S., *AI Video Generator Runway trained on thousands of YouTube videos without permission*, 404 Media, 25.07.2024, [on-line] <https://www.404media.co/runway-ai-image-generator-training-data-youtube> – 23.08.2024.
- Crawford K., Joler V., *Anatomy of an AI system. An anatomical case study of the Amazon echo as an artificial intelligence system made of human labor*, „Anatomy of AI”, 2018, [on-line] <https://anatomyof.ai> – 22.11.2024.
- Crawford K., *Atlas sztucznej inteligencji. Władza, pieniądze i środowisko naturalne*, tłum. T. Chawziuk, Kraków 2021.
- Crawford K., *9 ways to see a dataset: What's at stake in examining datasets?*, „Knowing Machines”, [on-line] <https://knowingmachines.org/publications/9-ways-to-see/essays/9-ways-to-see-a-dataset> – 22.11.2024.
- Crawford K., Paglen T., *Excavating AI: The politics of training sets for machine learning*, „Excavating AI”, September 19, 2019, [on-line] <https://excavating.ai> – 22.11.2024.
- Deeb-Swihart J., Polack Ch., Gilbert E., Essa I., *Selfie-presentation in everyday life: A large-scale characterization of selfie contexts on Instagram*, „Eleventh International AAAI Conference on Web and Social Media” 2017, vol. 11, no. 1, s. 42–51, <https://doi.org/10.1609/icwsm.v11i1.14896>.
- Eyal N., Hoover R., *Skuszeni. Jak tworzyć produkty, które kształtują nawyki konsumenckie*, tłum. A. Kucharczyk-Barycza, Gliwice 2024.
- Feldman Barrett L., Adolphs R., Marsella S., Martinez A.M., Pollak S.D., *Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements*, „Psychological Science in the Public Interest” 2019, vol. 20, iss. 1, s. 1–68, <https://doi.org/10.1177/1529100619832930>.
- Fogg B.J., *Persuasive technology. Using computers to change what we think and do*, San Francisco 2012.
- Fry H., *Hello world. Jak być człowiekiem w epoce maszyn*, tłum. S. Musielak, Kraków 2018.
- Global views on A.I. 2023*, Ipsos, July 2023, [on-line] https://www.ipsos.com/sites/default/files/ct/news/documents/2023-07/Ipsos%20Global%20AI%202023%20Report-WEB_1.pdf – 15.09.2024.

- Grother P., Ngan M., Hanaoka K., *Face Recognition Vendor Test (FRVT). Part 3: Demographic effects*, National Institute of Standards and Technology, 2019, <https://doi.org/10.6028/NIST.IR.8280>.
- Haoarchive K., *What is machine learning?*, „MIT Technology Review”, 17.11.2018, [on-line] <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart> – 26.06.2024.
- Hart R., *Clearview AI – controversial facial recognition firm – fined \$33 million for ‘illegal database’*, „Forbes”, 3.08.2024, [on-line] <https://www.forbes.com/sites/roberthart/2024/09/03/clearview-ai-controversial-facial-recognition-firm-fined-33-million-for-illegal-database> – 10.09.2024.
- Harwell D., *This facial recognition website can turn anyone into a cop — or a stalker*, „The Washington Post”, 14.05.2021, [on-line] <https://www.washingtonpost.com/technology/2021/05/14/pimeyes-facial-recognition-search-secrecy> – 30.06.2024.
- Harvey A., *On computer vision*, UMBAU, 1/2021, [on-line] <https://umbau.hfg-karlsruhe.de/posts/on-computer-vision> – 26.06.2024.
- Harvey A., *Today’s selfie is tomorrow’s biometric profile*, [w:] *House of Mirrors. Artificial Intelligence As Phantasm*, 2022.
- Harvey A., LaPlace J., *Exposing.ai*, 2021, [on-line] <https://exposing.ai> – 30.06.2024.
- He Y., She H., *Beyond the great power competition narrative: Exploring labor politics and resistance behind AI innovation in China*, „Georgetown Journal of Asian Affairs” 2021, vol. 7,.
- Heaven W.D., *Why does AI hallucinate?*, „MIT Technology Review”, 18.06.2024, [on-line] <https://www.technologyreview.com/2024/06/18/1093440/what-causes-ai-hallucinate-chatbots> – 24.08.2024.
- Heaven W.D., *What is AI?*, „MIT Technology Review”, 10.07.2024, [on-line] <https://www.technologyreview.com/2024/07/10/1094475/what-is-artificial-intelligence-a-i-definitive-guide> – 20.08.2024.
- Heaven W.D., *AI hype is built on high test scores. Those tests are flawed*, „MIT Technology Review”, 30.08.2023, [on-line] <https://www.technologyreview.com/2023/08/30/1078670/large-language-models-arent-people-lets-stop-testing-them-like-they-were> – 25.08.2024.
- Heaven W.D., *Why Meta’s latest large language model survived only three days online*, „MIT Technology Review”, 18.11.2022, [on-line] <https://www.technologyreview.com/2022/11/18/1063487/meta-large-language-model-ai-only-survived-three-days-gpt-3-science> – 15.05.2024.
- Heaven W.D., *Large language models can do jaw-dropping things. But nobody knows exactly why*, „MIT Technology Review”, 4.03.2024, [on-line] <https://www.technologyreview.com/2024/03/04/1089403/large-language-models-amazing-but-nobody-knows-why> – 20.08.2024.
- Heikkilä M., Honan M., *OpenAI brings a new web search tool to ChatGPT*, „MIT Technology Review”, 31.10.2024, [on-line] <https://www.technologyreview.com/2024/10/31/1106472/chatgpt-now-lets-you-search-the-internet> – 5.11.2024.
- Hill K., *A face search engine anyone can use is alarmingly accurate*, „The New York Times”, 26.05.2022, [on-line] <https://www.nytimes.com/2022/05/26/technology/pimeyes-facial-recognition-search.html> – 30.06.2024.

- Hill K., *Facial recognition led to wrongful arrests. So Detroit is making changes*, „The New York Times”, 29.06.2024, [on-line] <https://www.nytimes.com/2024/06/29/technology/detroit-facial-recognition-false-arrests.html> – 3.08.2024.
- Hill K., *Your face belongs to us. A secretive startup’s quest to end privacy as we know it*, New York 2023.
- Hill K., *Clearview AI used your face. Now you may get a stake in the company*, „The New York Times”, 13.06.2024, [on-line] <https://www.nytimes.com/2024/06/13/business/clearview-ai-facial-recognition-settlement.html> – 30.06.2024.
- Hollister S., *Microsoft’s AI boss thinks it’s perfectly okay to steal content if it’s on the open web*, „TheVerge”, 28.06.2024, [on-line] <https://www.theverge.com/2024/6/28/24188391/microsoft-ai-suleyman-social-contract-freeware> – 22.11.2024.
- Iwańska A., *Czy europoście i europościanki zgłaszają przeciwko biometrycznej inwigilacji?*, Fundacja Panoptykon, 12.06.2023, [on-line] <https://panoptykon.org/glosowanie-ai-act-biometria> – 3.07.2024.
- Kopańko K., *Śmieci na wejściu, śmieci na wyjściu. Tak programiści kwitują sztuczną inteligencję*, „Spider’sWeb / bizblog”, 29.07.2024, [on-line] <https://bizblog.spidersweb.pl/smieci-na-wejsciu-smieci-na-wyjsciu-tak-programisci-kwituja-sztuczna-inteligencje> – 2.08.2024.
- Li F.F., *How we’re teaching computers to understand pictures*, TED2015, March 2015, [on-line] https://www.ted.com/talks/fei_fei_li_how_we_re_teaching_computers_to_understand_pictures?subtitle=en 5.07.2024.
- Li F.F., Yang K., Qinami K., Russakovsky O., Deng J., *Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the ImageNet hierarchy*, [w:] *Conference on Fairness, Accountability, and Transparency (FAT* ’20)*, January 27–30, 2020, Barcelona, <https://doi.org/10.1145/3351095.3375709>.
- Luccioni A.S., Crawford K., *The Nine Lives of ImageNet: A Sociotechnical Retrospective of a Foundation Dataset and the Limits of Automated Essentialism*, „Journal of Data-centric Machine Learning Research”, 29.02.2024, [on-line] <https://openreview.net/forum?id=jh0ck1bPGF> – 30.06.2024.
- Marcus G., *Deep learning is hitting a wall*, „Nautilus”, 10.03.2022, [on-line] <https://nautilus.deep-learning-is-hitting-a-wall-238440> – 24.08.2024.
- Metz C., Hill K., *Here’s a way to learn if facial recognition systems used your photos*, „The New York Times”, 31.01.2021, [on-line] <https://www.nytimes.com/2021/01/31/technology/facial-recognition-photo-tool.html> – 5.07.2024.
- Metz C., *‘Nerd,’ ‘Nonsmoker,’ ‘Wrongdoer’: How might A.I. label you?*, „The New York Times”, 20.09.2019, [on-line] <https://www.nytimes.com/2019/09/20/arts/design/imagenet-trevor-paglen-ai-facial-recognition.html> – 1.07.2024.
- Murgia M., *Who’s using your face? The ugly truth about facial recognition*, „Financial Times”, 18.09.2019, [on-line] <https://www.ft.com/content/cf19b956-60a2-11e9-b285-3acd5d43599e> – 24.07.2024.
- Murgia M., *Microsoft quietly deletes largest public face recognition data set*, „Financial Times”, 6.06.2019, [on-line] <https://www.ft.com/content/7d3e0d6a-87a0-11e9-a028-86cea8523dc2> – 1.07.2024.
- Nicoletti L., Bass D., *Humans are biased. Generative AI is even worse*, „Bloomberg”, 9.06.2023, [on-line] <https://www.bloomberg.com/graphics/2023-generative-ai-bias> – 23.08.2024.

- Obem A., Szymielewicz K., *Sztuczna inteligencja non-fiction*, Fundacja Panoptykon, 2020, [on-line] <https://panoptykon.org/sztuczna-inteligencja-non-fiction> – 2.04.2024.
- O'Neil C., *Broń matematycznej zagłady. Jak algorytmy zwiększają nierówności i zagrażają demokracji*, tłum. M.Z. Zieliński, Warszawa 2017.
- Pearson J., *Microsoft deleted a massive facial recognition database, but it's not dead*, „Vice”, 6.06.2019, [on-line] <https://www.vice.com/en/article/microsoft-deleted-a-facial-recognition-database-but-its-not-dead> – 1.07.2024.
- Perrigo B., *OpenAI used Kenyan workers on less than \$2 per hour to make ChatGPT less toxic*, „Time”, 18.01.2023, [on-line] <https://time.com/6247678/openai-chatgpt-kenya-workers> 18.08.2024.
- Perrigo B., *TikTok's subcontractor in Colombia under investigation for traumatic work*, „Time”, 10.11.2022, [on-line] <https://time.com/6231625/tiktok-teleperformance-colombia-investigation> 15.08.2024.
- Pontin J., *Artificial intelligence with help from the humans*, „The New York Times”, 25.03.2007, [on-line] <https://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html> – 18.08.2024.
- Roth E., *Runway's AI video generator trained on thousands of scraped YouTube videos*, „The Verge”, 25.07.2024, [on-line] <https://www.theverge.com/2024/7/25/24206120/runway-ai-video-generator-scraped-youtube-videos-report> – 20.08.2024.
- Rouse M., *Garbage in, garbage out*, „Techopedia”, 4.01.2017, [on-line] <https://www.techopedia.com/definition/3801/garbage-in-garbage-out-gigo> – 5.07.2024.
- Sato M., *Apple, Anthropic, and other companies used YouTube videos to train AI*, The Verge, 16.07.2024, [on-line] <https://www.theverge.com/2024/7/16/24199636/apple-anthropic-nvidia-salesforce-youtube-videos-training-data-copyright> – 20.08.2024.
- Schultz J., *9 ways to see a dataset: What can LAION teach us about copyright law?*, „Knowing Machines”, [on-line] <https://knowingmachines.org/9-ways-to-see/LAION-copyright> – 5.07.2024.
- Shumailov I., Shumaylov Z., Zhao Y., Papernot N., Anderson R., Gal Y., *AI models collapse when trained on recursively generated data*, „Nature” vol. 631, 24.07.2024, <https://doi.org/10.1038/s41586-024-07566-y>.
- Spence R., Bifulco A., Bradbury P., Martellozzo E., DeMarco J., *Content moderator mental health, secondary trauma, and well-being: A cross-sectional study*, „Cyberpsychology, Behavior, and Social Networking” 2024, vol. 27, no. 2, <https://doi.org/10.1089/cyber.2023.02>.
- Steyerl H., *Mean images*, „New Left Review”, iss. 140/141, Mar/June 2023.
- Vincent J., *Discover the stupidity of AI emotion recognition with this little browser game*, „The Verge”, 6.04.2021, [on-line] <https://www.theverge.com/2021/4/6/22369698/ai-emotion-recognition-unscientific-emojify-web-browser-game> – 14.09.2024.
- Zuboff S., *Wiek kapitalizmu inwigilacji. Walka o przyszłość ludzkości na nowej granicy władzy*, tłum. A. Unterschuetz, Warszawa 2019.

Streszczenie

Funkcjonujemy w rzeczywistości, która jest utkana z cyfrowego zapośredniczenia naszych aktywności. Korzystamy z aplikacji i serwisów Internetowych. Nasze teksty, maile, wiadomości, dzieła, fotografie krążą w wirtualnych chmurach tworząc bazy danych korporacji technologicznych. Posługujemy się sztuczną inteligencją nierzadko nawet o tym nie wiedząc. Tymczasem coraz więcej śladów naszej prywatności i intymności w postaci danych behawioralnych i biometrycznych zasila wielkie modele fundamentalne tworząc tzw. *ground truth*. Na tej podstawie w dalszej kolejności metodą statystycznej predykcji i optymalizowania wzorów, generowane są nowe wizje rzeczywistości sklecone ze skrawków naszej wirtualnej obecności na platformach. Generatywna sztuczna inteligencja, w tym duże modele językowe oraz uczenie maszynowe wykorzystywane do systemów rozpoznawania twarzy, posiłkują się informacjami, które nawykowo, powszechnie i dobrowolnie udostępniamy w przestrzeni Internetu. Artykuł porusza problem obrotu danymi, tego, co się z nimi dzieje, gdy już zostaną zassane przez systemy technologiczne.

Słowa kluczowe: sztuczna inteligencja, algorytmy, kultura wizualna, bazy danych, widzenie komputerowe, rozpoznawanie twarzy

Machine-designed statistical renderings, or visual culture in the service of artificial intelligence**Abstract**

We exist in a reality shaped by the digital mediation of our activities. We use software applications and online services. Our texts, emails, messages, works, and photographs circulate in virtual clouds, forming databases for technological corporations. We often use artificial intelligence without even being aware of it. Meanwhile, increasing traces of our privacy and intimacy—in the form of behavioral and biometric data—feed large foundational models, creating the so-called „ground truth.” Based on this, subsequent methods of statistical prediction and pattern optimization generate new visions of reality, pieced together from fragments of our virtual presence on platforms. Generative artificial intelligence—including large language models and machine learning algorithms used in facial recognition systems—depends on the information we habitually, commonly, and voluntarily share in the online space. The article addresses the issue of data circulation and what happens to it once it has been absorbed by technological systems.

Keywords: artificial intelligence, algorithms, visual culture, databases, computer vision, facial recognition