ACTA UNIVERSITATIS LODZIENSIS
FOLIA SOCIOLOGICA 92, 2025

https://doi.org/10.18778/0208-600X.92.03

Iwona Leonowicz-Bukała[*]
https://orcid.org/0000-0003-3164-1209

Monika Struck-Peregończyk[**]
https://orcid.org/0000-0001-8427-2510

Mikołaj Birek[***]
https://orcid.org/0000-0003-2659-5594

Katarzyna Dudzińska[****]
https://orcid.org/0009-0003-3321-2253

„AI, WYGENERUJ OBRAZ OSOBY Z NIEPEŁNOSPRAWNOŚCIĄ" (OD)TWORZENIE SPOŁECZNYCH REPREZENTACJI NIEPEŁNOSPRAWNOŚCI Z WYKORZYSTANIEM NARZĘDZI GENERATYWNEJ SZTUCZNEJ INTELIGENCJI

Abstrakt. Tematyka reprezentacji niepełnosprawności w tradycyjnych i internetowych mediach była dotychczas często podnoszona przez wielu badaczy, którzy udowodnili, że przedstawienia obecne w środkach masowego przekazu mogą wpływać na społeczne postrzeganie tego tematu.

[*] Dr, Kolegium Mediów i Komunikacji Społecznej, Wyższa Szkoła Informatyki i Zarządzania z siedzibą w Rzeszowie, ul. Henryka Sucharskiego 2, 35-225 Rzeszów, e-mail: ileonowicz@wsiz.edu.pl

[**] Dr, Kolegium Mediów i Komunikacji Społecznej, Wyższa Szkoła Informatyki i Zarządzania z siedzibą w Rzeszowie, ul. Henryka Sucharskiego 2, 35-225 Rzeszów, e-mail: mstruck@wsiz.edu.pl

[***] Dr, Kolegium Mediów i Komunikacji Społecznej, Wyższa Szkoła Informatyki i Zarządzania z siedzibą w Rzeszowie, ul. Henryka Sucharskiego 2, 35-225 Rzeszów, e-mail: mbirek@wsiz.edu.pl

[****] Lic., Kolegium Mediów i Komunikacji Społecznej, Wyższa Szkoła Informatyki i Zarządzania z siedzibą w Rzeszowie, ul. Henryka Sucharskiego 2, 35-225 Rzeszów, e-mail: w64330@student.wsiz.edu.pl

Received: 17.06.2024. Verified: 31.10.2024. Accepted: 18.12.2024.

Rozwój nowych technologii komunikacyjnych oferuje dzisiaj jednak nowe szanse w kwestii przedstawiania adekwatnych, realistycznych wizerunków osób niepełnosprawnych, a także przynosi nowe ryzyka. W artykule zwrócono uwagę na jedną z najszybciej rozwijających się technologii, jaką jest generatywna sztuczna inteligencja (GenAI), wpływająca na profesjonalnych nadawców i platformy społecznościowe. Łatwy dostęp do narzędzi opartych o tzw. GenAI stwarza dotychczas nieznane wyzwania, dotyczące jej możliwych zastosowań w tworzeniu powszechnie dostępnych treści online. O ile sztuczna inteligencja może zwiększać dostępność i poprawiać jakość życia osób z niepełnosprawnościami, jej wykorzystanie może jednocześnie skutkować utrwaleniem negatywnych reprezentacji społecznych tej grupy.

W oparciu o przegląd istniejących problemów i dyskusji na ten temat oraz wyniki własnych badań, w artykule staramy się odpowiedzieć na pytania dotyczące możliwych mechanizmów powielania stereotypów wobec niepełnosprawności przez treści wizualne generowane przez narzędzia oparte na sztucznej inteligencji. W żadnym miejscu w artykule nie używano AI do procesu pisania.

Słowa kluczowe: generatywna sztuczna inteligencja, AI, niepełnosprawność, stereotypy, komunikacja społeczna, uprzedzenia.

"AI, GENERATE AN IMAGE OF A DISABLED PERSON" (RE)CREATING REPRESENTATIONS OF DISABILITY WITH GENERATIVE AI TOOLS

Abstract. The issue of the importance of representations of disability in various media has been frequently raised by numerous scholars, who proved that such mass media representations can exert a significant influence on the perception of disabled people in society. However, the constant change and development of new communication technologies pose new challenges and offer new chances to the issue of presenting adequate, realistic images of disabled people. The article describes generative artificial intelligence, as one of the most rapidly developing areas, which significantly impacts social communication both in professional media and social platforms. The easy online access to AI-based tools creates new challenges and uncertainties regarding its possible uses in creating widespread online content. While AI can increase accessibility and improve the quality of life of disabled people, its use may also result in perpetuating negative social representations of this group. In this paper, we attempt to review the existing problems and discussions on this topic together with the results of our research and answer the questions concerning the possible mechanisms of reproducing bias towards disability through content generated with various AI-based tools. Nowhere in the article was AI used in the writing process.

Keywords: generative artificial intelligence, AI, disability, stereotypes, social communication, bias.

Wstęp

Tematyka reprezentacji niepełnosprawności w mediach masowych była dotychczas często podnoszona przez wielu badaczy (Barnes 1992; Clogston 1993; Haller 1999, 2000; Zhang, Haller 2013). Dzięki ww. studiom wiemy już, że przedstawienia rozpowszechniane w środkach masowego przekazu mogą wywierać istotny wpływ na społeczne postrzeganie niepełnosprawności. Niestety, w większości powielają one znane stereotypy, pokazując osoby z niepełnosprawnościami w dychotomicznym ujęciu jako przeżywające osobistą tragedię,

chore, potrzebujące pomocy (rola „ofiary”) albo „superbohaterów” dokonujących niezwykłych czynów (lub osiągających coś, co w świecie sprawnych uznane jest za normalne) (Barnes 1992; Clogston 1993; Haller 1999; Struck-Peregończyk, Leonowicz-Bukała 2018). Oba te wzorce medialnego przedstawienia osób z niepełnosprawnościami bazują na uproszczonych schematach i wzmacniają „inność” tych osób. Pewną szansą na zmianę w tym zakresie jest rozwój szeroko rozumianych nowych technologii komunikacyjnych, który stawia nowe wyzwania i oferuje nowe szanse w kwestii przedstawiania adekwatnych, realistycznych wizerunków osób z niepełnosprawnościami, także przez nich samych (zob. Thoreau 2006; Goggin, Newell 2007; Hill 2017; Cocq, Ljuslinder 2020; Struck-Peregończyk, Leonowicz-Bukała 2023).

W ostatnich latach niezmiernie dynamicznie rozwija się obszar technologii informatycznej związany ze sztuczną inteligencją (dalej używamy zamiennie, gdyż oba skróty są w powszechnym użyciu: SI od *sztuczna inteligencja* lub AI od ang. *artificial intelligence*). Istnieje wiele (często rozbieżnych) koncepcji i definicji sztucznej inteligencji, jednak powszechnie przyjmuje się, że o SI można mówić, kiedy oprogramowanie komputerowe (maszyna) jest w stanie działać lub zachowywać się w taki sposób, że gdyby te same czynności wykonywał człowiek, jego działania ocenione byłyby jako inteligentne (McCarthy i in. 1955; Tadeusiewicz 2019; Adiguzel i in. 2023). Patrick H. Winston określił SI jako „badanie obliczeń umożliwiających percepcję, rozumowanie i działanie” (1992: 5 za: OECD 2019: 6). W nowocześniejszym ujęciu system SI jest rozumiany jako „system maszynowy, zdolny do wpływania na środowisko poprzez formułowanie rekomendacji, przewidywań lub podejmowanie decyzji dla określonego zestawu celów. Czyni to, wykorzystując wejścia/dane pochodzące od maszyn i/lub ludzi, aby: i) postrzegać rzeczywiste i/lub wirtualne środowiska; ii) przekształcać takie postrzegania w modele manualnie lub automatycznie; oraz iii) wykorzystywać interpretacje modeli do formułowania opcji dla wyników” (OECD 2019: 7, tłum. własne). U podstaw współczesnej sztucznej inteligencji leżą techniki uczenia maszynowego (z ang. *machine learning*, ML), które są w stanie stopniowo zwiększać wydajność obliczeniową i oferować dokładniejsze wyniki w oparciu o dane treningowe zgromadzone w bazach danych (ang. *dataset*) (Zhou 2021; Wach i in. 2023). Wśród metod uczenia maszynowego najbardziej prominentne są sieci neuronowe (z ang. *neural network*, NN), będące modelami obliczeniowymi, dokonującymi kalkulacji w sposób zbliżony do biologicznego systemu nerwowego (Wu, Feng 2018). Z tego zjawiska wywodzi się technika uczenia głębokiego (z ang. *deep learning*, DL), polegająca na konstruowaniu wielowarstwowych, głębokich sieci neuronowych, operujących na wielu poziomach abstrakcji (LeCun i in. 2015). Tego typu modele zwracają wysokiej jakości efekty, jednak wielowarstwowość procesów w nich zachodzących uniemożliwia ich interpretację przez człowieka, stąd też przyjęło się określać tego typu sieci jako czarne skrzynki (z ang. *black box*) – z uwagi na ich zamkniętą, nieprzenikalną naturę. Nie jesteśmy w stanie

precyzyjnie odtworzyć działania algorytmu ani sposobu przetwarzania zmiennych. Brak wiedzy o tym, które zmienne uwzględniono w ukrytych warstwach sieci, uniemożliwia interpretację zachodzących tam procesów (Iwasiński 2023). Jednocześnie rozwijana jest klasa wyjaśnialnej sztucznej inteligencji (z ang. *explainable AI*, XAI), mającej na celu stworzenie systemu, wgląd w mechanizmy którego będzie możliwy. Taką sieć nazywa się białą lub szklaną skrzynką (Rai 2020).

Ważnym dla przedmiotu niniejszego badania rodzajem SI jest generatywna sztuczna inteligencja (z ang. *generative artificial intelligence*, GAI, GenAI), która cechuje się możliwością tworzenia z pozoru nowych, twórczych treści, takich jak tekst, obraz czy dźwięk, w oparciu o dostarczone dane treningowe (Feuerriegel i in. 2023). U podstaw GenAI leżą duże modele językowe (z ang. *large language model*, LLM), czyli modele umożliwiające komunikację człowiek – system za pośrednictwem naturalnego języka. Większość obecnie funkcjonujących modeli jest także zdolna do interpretacji kodu programistycznego, obrazu, filmu oraz dźwięku, przez co nazywa się je coraz częściej dużymi modelami multimodalnymi (z ang. *large multimodal model*, LMM) (Toner 2023). Celem tych modeli jest umożliwienie komunikacji z systemem SI bez kodowania czy skomplikowanego interfejsu użytkownika, a z wykorzystaniem języka naturalnego, dźwięku i obrazu właśnie.

Bazy danych, na których trenowane są modele GenAI, złożone są z ogromnej ilości pozycji. W zależności od rodzaju modelu zbiory te składają się z obrazów, tekstów, filmów i plików dźwiękowych. Aby możliwe było wytrenowanie modelu generującego treści wysokiej jakości, korzysta się z baz danych mających od kilku miliardów do ponad 100 miliardów pozycji (Liu i in. 2024). Tak wytrenowany model reaguje na prompty (wyzwalacze) – polecenia o różnym stopniu skomplikowania, wydawane z użyciem języka naturalnego podczas pracy z SI. Interakcja z systemem przypomina w takiej sytuacji bardziej rozmowę w relacji interpersonalnej niż operację komputerową. Promptowanie (już w powszechnym użyciu w języku polskim w odniesieniu do poleceń kierowanych do SI) jest jedną z charakterystycznych dla pracy z GenAI, a nieistniejącą wcześniej umiejętnością, będącą specjalistyczną domeną tzw. prompt-inżynierów (z ang. *prompt-engineer*; Martineau 2023).

W oparciu o GenAI tworzone są modele, które można podzielić na przyjazne użytkownikowi chatboty, umożliwiające prowadzenie symulowanej rozmowy, oraz generatory przeznaczone do zwracania treści multimedialnych w oparciu o wprowadzony prompt. Narzędzia te dostępne są online, zarówno bezpłatnie, jak i za opłatą (najczęściej w formie odnawialnej subskrypcji). Tempo powstawania oraz rozwoju tych modeli jest dynamiczne. W lipcu 2022 r. do publicznego użytku oddane zostały DALL·E 2 i Midjourney, pierwsze z generatorów obrazów nowej ery GenAI (nie były to pierwsze generatory obrazów opartych o SI, ale przewyższyły wszystkie poprzednie wszechstronnością i jakością zwracanych obrazów). W ciągu kolejnych dwóch lat powstał szereg innych narzędzi zmieniających

paradygmat tworzenia cyfrowego obrazu. Do tego grona dołączyły następnie modele GenAI przeznaczone do generowania dźwięku i tekstu. To te ostatnie, w formie chatbotów, takich jak ChatGPT czy Copilot, wielu uznaje za mające potencjał całkowitego zrewolucjonizowania dotychczasowego sposobu ludzkiej pracy, twórczości i interakcji (Suleyman 2023), przy jednoczesnym zalecaniu ostrożności w wydawaniu takich sądów przez innych (Leslie, Meng 2024).

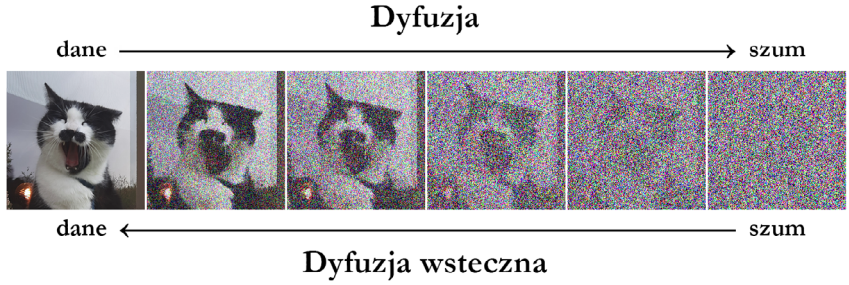
W związku z tym, w jakim stopniu nasze społeczeństwa mogą odczuwać zarówno pozytywne, jak i negatywne skutki powszechnie dostępnych narzędzi, opartych na sztucznej inteligencji, tematyka GenAI dominuje obecnie nie tylko w dyskursie akademickim większości dyscyplin naukowych, ale także w obszarach: polityki, wojskowości, przemysłu, edukacji, a także w społecznej przestrzeni komunikacyjnej. Ta sytuacja skłoniła nasz zespół, już wcześniej zajmujący się kwestią reprezentacji niepełnosprawności w komunikacji społecznej (Struck-Peregończyk, Kurek-Ochmańska 2018; Struck-Peregończyk, Leonowicz-Bukała 2018; Kurek-Ochmańska i in. 2020; Struck-Peregończyk, Leonowicz-Bukała 2023), do podjęcia próby zbadania, czy – a jeśli tak, to w jaki sposób – wizualne generatory oparte na generatywnej sztucznej inteligencji reprodukują lub tworzą nowe, stereotypowe przedstawienia niepełnosprawności i osób z niepełnosprawnością, a także rozważenia, w jaki sposób wykorzystanie tych obrazów może przyczynić się do pogłębienia istniejących uprzedzeń.

Generatory obrazu oparte na sztucznej inteligencji

Generatory obrazów były jednymi z pierwszych narzędzi GenAI, które trafiły do głównego nurtu. Ich pierwsze wersje, jak np. upubliczniony w styczniu 2001 r. DALL·E firmy OpenAI (Johnson 2021), charakteryzowały widoczne błędy w obrazie cyfrowym (tzw. artefakty; Manthey i in. 2017: 2) – tu w formie zdeformowanej i nierealistycznej anatomii, niezamierzonego przenikania się form i obiektów – oraz ogólnie fantastyczna natura tworzonych obrazów, przez co były traktowane głównie jako ciekawostka, narzędzie do tworzenia memów internetowych. Zmieniło się to wraz z wydaniem generatora wizualnego DALL·E 2 (później DALL·E 3), który generował obrazy o znacznie większej precyzji, złożoności i wyższej jakości. Od 2022 r. generatory obrazów dostarczają coraz bardziej imponujących wyników, umożliwiając użytkownikom bez wcześniejszego artystycznego lub technicznego doświadczenia przekształcanie swoich pomysłów w obrazy. Szacuje się, że do sierpnia 2023 r. różne generatory obrazów, w tym głównie Midjourney, DALL·E, Stable Diffusion i Firefly, zostały wykorzystane do stworzenia ponad 15 miliardów obrazów (Everypixel 2023). W 2024 r. na popularności zyskały także generatory muzyki i wideo.

Zagadnieniem wymagającym osobnego omówienia jest kwestia sposobu, w jaki generatory SI tworzą obrazy. Współczesne generatory obrazów (w tym

także te badane przez nas, np. Midjourney, DALL·E i Stable Diffusion) są modelami dyfuzyjnymi. Są to modele uczenia głębokiego, oparte na dwóch kluczowych etapach: dyfuzji oraz dyfuzji wstecznej. W procesie dyfuzji dane treningowe (obrazy) poddane są procesowi dekonstrukcji poprzez stopniowe rozpraszanie ich do stopnia nieczytelnego szumu. Następnie przeprowadza się proces dyfuzji wstecznej, w trakcie którego model stopniowo przywraca obrazy z wytworzonego szumu. Czynność ta jest powtarzana wielokrotnie przez algorytmy treningowe, co skutkuje wyuczeniem systemu zdolnego do wytwarzania spójnego wyniku, zaczynającego się od nieuporządkowanej „piany danych" (Croitoru i in. 2023; Song i in. 2021). Aby dało się je wykorzystać, używane w treningu obrazy muszą być opisane. W tym celu także stosuje się narzędzia SI, takie jak np. CLIP (ang. *Contrastive Language-Image Pre-Training*) będący modelem multimodalnym, zdolnym opisywać dostarczone mu grafiki w formie krótkiego tekstu (Radford i in. 2021).



Rys. 1. Schematyczne przedstawienie sposobu działania algorytmów dyfuzyjnych

Źródło: opracowanie własne M. Birek na podstawie Song i in. (2021: 2)

W uzyskaniu wysokiej jakości modeli dyfuzyjnych kluczowe są ilość, jakość i reprezentacja w zbiorach tych danych. Innymi słowy, im więcej zdjęć uroczych kotów znajduje się w zbiorach danych, tym bardziej urocze (i precyzyjniej przedstawione) koty jest w stanie wytworzyć generator obrazu na podstawie polecenia tekstowego (schematyczne przedstawienie tego procesu znajduje się na rys. 1).

Jednym z najbardziej kontrowersyjnych aspektów GenAI jest sposób, w jaki firmy pozyskują dane używane do szkolenia swoich modeli. Zamiast kupować grafiki, filmy, teksty i muzykę w specjalistycznych bankach (tzw. *stocki*), firmy zdecydowały o wykorzystaniu danych zgromadzonych automatycznie z internetu, uzasadniając to zasadą tzw. dozwolonego użytku, opartą na przetworzeniu przez AI (Sundara Rajan 2024). Praktyka ta jest otwarcie krytykowana przez twórców wizualnych i właścicieli zgromadzonych danych, stanowi także podstawowy zarzut w licznych toczących się sprawach sądowych, czego przykładem jest proces, jaki firmie OpenAI wytoczyła redakcja tygodnika „New York Times"

(zob. Sundara Rajan 2024). Jednocześnie, podczas gdy biznes kontynuuje argumentację o swoim prawie do nieograniczonego wykorzystywania treści z internetu do szkolenia modeli AI, firmy takie jak Midjourney i OpenAI starają się o partnerstwo z zewnętrznymi portalami i bankami stockowymi, co ma być sposobem na rozbudowę posiadanych przez nich baz danych treningowych. To samo dotyczy generatorów innego rodzaju. Na przykład najnowsza wersja ChatGPT-4o jest dostępna za darmo w zamian za to, że wszystkie dane wprowadzone przez użytkowników do systemu zostają przez niego wykorzystane do dalszego treningu (Zhang, Yang 2024).

Sztuczna inteligencja a uprzedzenia

Kwestie różnorodnych błędów, jakie pojawiają się podczas korzystania z narzędzi SI, zauważono bardzo szybko, co ma odzwierciedlenie m.in. w publikacjach prasowych. Dziennikarze pisali m.in. o „halucynującym” czy „konfabulującym” ChacieGPT, który nie posiadając wystarczającej ilości informacji, zmyśla je, nie informując o tym użytkowników (Kania 2024). Internauci zaczęli na własną rękę badać i dzielić się przypadkami kuriozalnych propozycji ChatGPT (Tomaszewski 2024). Izabela Tomaszewska z portalu Demagog.pl również zaobserwowała ryzyko wzmacniania stereotypów płciowych czy rasowych przez obrazy wygenerowane przez nią w generatorze Bing (2024).

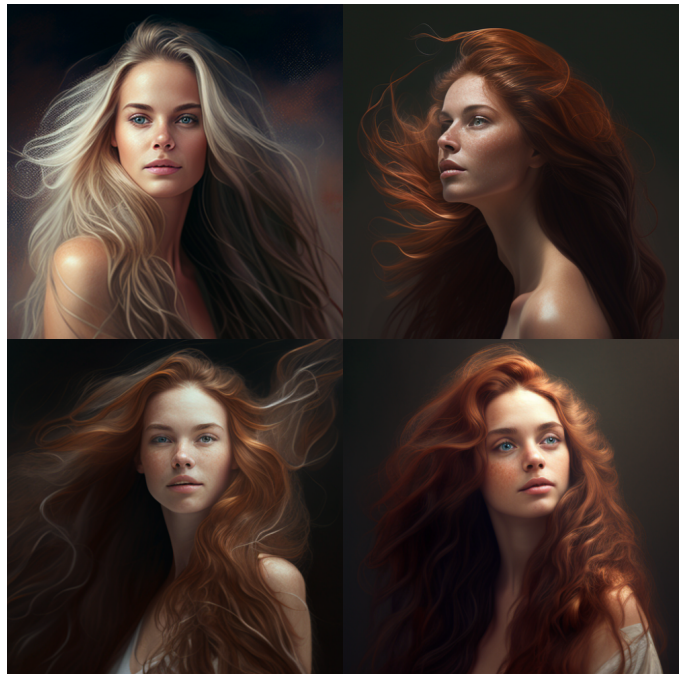
Problem uprzedzeń generowanych przez sztuczną inteligencję jest związany z nieprzejrzystym mechanizmem podejmowania przez nią decyzji, co utrudnia identyfikację i analizę uprzedzeń zakodowanych w ich przyczynach (van Giffen i in. 2022). Znane są przykłady negatywnych postaw narzędzi opartych o SI, prowadzących do dyskryminacji różnych grup społecznych. Przykładowo, narzędzia SI używane w amerykańskim wymiarze sprawiedliwości pomagały sędziom decydować, którzy więźniowie powinni być zwolnieni warunkowo. Dochodzenia wykazały, że są one rasowo stronnicze, częściej odmawiając zwolnienia Afroamerykanom. Innym przykładem jest narzędzie rekrutacyjne SI firmy Amazon, które zostało natomiast wycofane z powodu wykrytych uprzedzeń wobec kobiet (Land 2023). W obu przypadkach uprzedzenia zawarte w danych historycznych, czyli dostępnych w internecie dyskursach na temat osób z mniejszości, marginalizowanych, kulturowo zdominowanych – przenoszą się na nowe systemy SI (por. Miltenburg 2016).

Powyżej opisane zjawiska celnie podsumowuje stwierdzenie, że „systemy ML [uczenia maszynowego – przyp. aut.] są tylko tak dobre, jak duże zbiory danych (tzw. *big data*), na których są szkolone” (Land 2023: 33, tłum. własne). Oznacza to, że w sytuacji, gdy dane treningowe nie są reprezentatywne dla populacji, systemy będą działać korzystnie na rzecz użytkowników dobrze reprezentowanych przez dane, podczas gdy grupy demograficzne, które w danych nie

mają proporcjonalnej (rozumianej jako zgodna z faktyczną) reprezentacji, będą przedstawiane nieadekwatnie (Crawford 2020). Zjawisko to to tzw. stronniczość reprezentacyjna (*representation bias*). Inną sytuacją jest, kiedy mamy do czynienia z uprzedzeniami społecznymi (*social bias*)[1] – kiedy dane są historycznie dokładne, ale zachowania społeczne i normy kulturowe reprezentowane w danych są potencjalnie (choćby z dzisiejszego punktu widzenia) dyskryminujące, system będzie utrwalał ten dyskryminujący potencjał (van Giffen i in. 2022; Miltenburg 2016). Krytyczne w tej sytuacji z punktu widzenia osób z niepełnosprawnością jest to, iż niepełnosprawność nie zawsze jest widoczna, a wielość rodzajów niepełnosprawności sprawia, że grupa ta jest bardzo zróżnicowana wewnętrznie (Land 2023) – stąd trudno o widoczną i spójną reprezentację.

Uprzedzenia powielane i generowane są także przez dostępne za darmo online narzędzia oparte o GenAI. W obliczu licznych pozwów, składanych przez twórców przeciwko dużym firmom technologicznym, a także legislacji w postaci Aktu o sztucznej inteligencji Unii Europejskiej (European Parliament 2023), trudno przewidzieć, jak w przyszłości będą wyglądać modele GenAI; w chwili pisania jednak narzędzia GenAI opierają się na danych treningowych odzwierciedlających dominujące społeczne dyskursy na różne tematy. Ponieważ są one zależne od zbiorów danych i opierają się na sieciach neuronowych, a nie zaprogramowane na jasno określony i łatwy do przewidzenia efekt, modele dyfuzyjne są trudne do precyzyjnego skalibrowania i podatne na szereg błędów oraz nieplanowanych wyników. Mogą one znacznie się różnić, od humorystycznych w najlepszym przypadku, do wręcz obraźliwych w najgorszym, mimo intencji ich twórców. SI Google'a o nazwie Gemini wygenerowało rasowo zróżnicowanych nazistów bez żadnych poleceń określających kolor skóry osób noszących uniformy, co skutkowało wstrzymaniem przez firmę całego projektu generatora (Robertson 2024). Generator obrazów SI firmy Meta okazał się niezdolny do przedstawienia par mieszanych, konkretnie składających się z azjatyckiego mężczyzny i białej kobiety (Sato 2024). Z kolei Midjourney poproszony o wygenerowanie różnych obrazów kobiet, przy prompcie „piękna kobieta", sugerował wyłącznie kobiety o bladej, białej skórze, piegach, często z rudymi włosami (por. rys. 2). Dopiero po wpisaniu polecenia wygenerowania „pięknej czarnej kobiety" zaoferował obrazy osoby o ciemniejszej skórze. Gdy poproszono o wygenerowanie obrazu „przystojnego mężczyzny z ładnymi włosami", Midjourney wyprodukował obrazy mężczyzny o kwadratowej szczęce, białym kolorze skóry, z ciemnymi włosami i niebieskimi oczami (Singleton 2023).

[1] Pojęcia *representation bias* i *social bias* to jedne z 8 kategorii *ML biases* (uprzedzeń uczenia maszynowego) opisanych dokładniej w artykule van Giffen i in. (2022). Autorzy dokonali systematycznego przeglądu literatury dotyczącej uprzedzeń ML i zaproponowali katalog 8 typów uprzedzeń wraz z propozycjami tego, jak minimalizować ich skutki.



Rys. 2. „Kobieta z pięknymi włosami”, obraz wygenerowany przez Midjourney

Źródło: Singleton 2023

Wydaje się, że problem z uprzedzeniami leży w sposobie, w jaki GenAI przedstawia ludzi z pewnych grup, a nawet w jaki w ogóle zgadza się je przedstawiać. Jeśli zbiór danych zawiera proporcjonalnie więcej obrazów białych kobiet niż czarnych kobiet oznaczonych jako „piękne”, generator z większym prawdopodobieństwem zaproponuje te pierwsze. Potencjał sztucznej inteligencji do utrwalania i wzmacniania nierówności płci pokazują wyniki badań Górskiej i Jemielniaka (2023) dotyczących uprzedzeń ze względu na płeć w generowanych przez sztuczną inteligencję wizerunkach profesjonalistów. Analiza przeprowadzona przez badaczy wykazała, że 99 obrazów z dziewięciu popularnych generatorów obrazów wykazało istotne uprzedzenia płciowe w kreowanych przez sztuczną inteligencję grafikach: mężczyźni byli reprezentowani na 76% obrazów, a kobiety – jedynie na 8%.

Wizerunki osób z niepełnosprawnościami generowane przez SI – co wiemy?

Problem uczciwości i bezstronności GenAI wobec osób z niepełnosprawnościami interesuje coraz szersze grono autorów (Trewin 2018; Whittaker i in. 2019; Hutchinson i in. 2020; Bennett, Keyes 2020). W badaniach prowadzonych przez Gadiraju i in. (2023) 19 grup fokusowych składających się z 56 osób z różnymi typami niepełnosprawności prosiło LLM o stworzenie narracji dotyczącej osób z niepełnosprawnościami. Jak się okazało, narracje te utrwalały dominujące uprzedzenia, dotyczące niepełnosprawności, takie jak nadreprezentacja osób poruszających się na wózkach, przedstawianie osób z niepełnosprawnościami jako smutnych lub samotnych, biernych i potrzebujących pomocy lub „superbohaterów" inspirujących dla osób pełnosprawnych.

Najnowsze badania, prowadzone przez Mack i in. (2024), dotyczące reprezentacji niepełnosprawności w obrazach generowanych przez systemy sztucznej inteligencji typu tekst-obraz dowodzą, że odzwierciedlają one szersze społeczne stereotypy i uprzedzenia. W opisywanym badaniu uczestnicy grup fokusowych (identyfikujący jako osoby z niepełnosprawnościami) poddawali ocenie, stworzone przez trzy popularne, publicznie dostępne generatory, obrazy przedstawiające osoby z niepełnosprawnościami w różnych kontekstach. Analiza tematyczna materiału zebranego podczas dyskusji w grupach fokusowych wykazała, że obrazy często zawierały uproszczone, stereotypowe reprezentacje niepełnosprawności (np. osoby niewidome jako noszące okulary z ciemnymi szkłami), koncentrujące się na negatywnych aspektach kondycji osób z niepełnosprawnościami, uwypuklające inność. Często pojawiały się nieadekwatne reprezentacje technologii asystujących, obrazy dehumanizujące (zbytnio zmedykalizowane, nierealistyczne), podkreślające bezradność, smutek, samotność, niereprezentatywne, jeśli chodzi o wiek, rasę, płeć. Uczestnicy badań wyrazili swoje obawy dotyczące negatywnego wpływu tego typu obrazów na osoby nieposiadające dużej wiedzy o niepełnosprawności czy niemające kontaktów z osobami z niepełnosprawnościami.

Metodyka badań

Jak wskazano we wstępie, celem badań było zweryfikowanie, czy oparte na sztucznej inteligencji wizualne generatory, dostępne publicznie w internecie, reprodukują powszechne stereotypy na temat niepełnosprawności i osób z niepełnosprawnością, o których pisano w dotychczas opublikowanej literaturze przedmiotu. Postawiony problem dotyczył także potencjalnej możliwości negatywnego wpływu szerokiego wykorzystania wizualnych narzędzi opartych o SI w komunikacji społecznej na postawy społeczne względem osób z niepełnosprawnością oraz możliwości tworzenia swoistej pętli sprzężenia zwrotnego, w której tworzone

obrazy, oparte o zawartość sieci, poprzez ich wykorzystywanie i publikowanie w przestrzeni internetu, stają się same materiałem treningowym. Prowadzi to do „zapętlenia” wizerunków niepełnosprawności oraz wizerunków osób z niepełnosprawnością, a także ich potencjalnych interpretacji, nie mówiąc o szerszym problemie „załamywania się” modeli SI (z ang. *model collapse*). Jest to zjawisko, w którym modele, w uczeniu których wykorzystano treści wygenerowane przez inne modele, podlegają nieodwracalnej degradacji, zwracając treści coraz gorszej jakości (Shumailov i in. 2024).

Nasze pytania badawcze dotyczyły zatem tego: (1) kim według różnych narzędzi wizualnych, opartych o GenAI, jest typowa osoba z niepełnosprawnością; (2) jak reprezentowane są osoby z konkretnymi typami niepełnosprawności; (3) jak reprezentowana jest typowa osoba z niepełnosprawnością w porównaniu z typową osobą bez niepełnosprawności podczas wykonywania codziennych czynności lub w codziennym otoczeniu. Z racji objętości niniejszego opracowania skupiamy się w nim na odpowiedzi na pierwsze pytanie, dwa pozostałe sygnalizując, gdyż odpowiedź na nie wymaga pogłębienia analizy ilościowej oraz przeprowadzenia analizy jakościowej.

Ponieważ wzięliśmy pod uwagę, iż sam sposób promptowania (tworzenia i kierowania zapytań) może być obarczony naszymi własnymi uprzedzeniami czy wyobrażeniami (por. Miltenburg 2016), a także warunkowany np. poziomem kompetencji komunikacyjnych, zestaw promptów proponowaliśmy indywidualnie, niezależnie od siebie w czteroosobowym zespole badawczym, zróżnicowanym płciowo (trzy kobiety, jeden mężczyzna) i wiekowo (23–42 lata), następnie wybierając wspólnie te zapytania, które w ocenie większości członków zespołu pozwoliłyby odpowiedzieć na zadane pytania badawcze. Mamy przy tym świadomość, że wybór innych promptów mógłby być równie/bardziej/mniej efektywny, niemniej z racji wstępnego charakteru badań byliśmy zmuszeni ograniczyć się do jakiegoś wyboru. W tym miejscu należy podkreślić, że żadna osoba z zespołu badawczego nie jest osobą z niepełnosprawnością.

Ostatecznie wypracowaliśmy 16 promptów dotyczących osób z niepełnosprawnościami oraz 7 promptów dotyczących po prostu osób, sformułowanych w języku angielskim (wybrane najpopularniejsze generatory nie zawsze miały możliwość efektywnej pracy w innych językach). Na (1) pytanie badawcze miała odpowiedzieć nam analiza obrazów wygenerowanych za pomocą promptów: *a person with disability*, *a disabled person*; na (2) pytanie prompty: *a blind person*, *a deaf person*, *a person with autism*, *a person with Multiple Personality Disorder*, *a hard-of-hearing person*, *a person with mental health issues*, *a person with a physical disability*, *a person with multiple disabilities*; na (3) pytanie badawcze pary promptów: *a person dealing with an everyday problem – a disabled person dealing with an everyday problem*, *people hanging out – disabled people hanging out*, *people doing sports – disabled people doing sports*, *people in love – disabled people in love*, *a group of people at the pool – a group of disabled people at the pool*, *a person at work – a disabled person at work*.

Aby zyskać jak najczystszą próbę wśród wygenerowanych obrazów, w przeprowadzonym eksperymencie prompty ograniczone zostały do równoważników zdań, bez szczegółowych opisów wyglądu czy okoliczności, w jakich ma znajdować się opisywana osoba. Przykładowo, zamiast prompta „kobieta cierpiąca na schizofrenię paranoidalną, siedząca przy stoliku w kuchni", zastosowany był prompt o treści „osoba z problemami psychicznymi".

Generatory obrazów do zbadania zostały wytypowane przez nasz zespół na podstawie ich miejsca na rynku i ogólnej dostępności. Według szacunkowych danych z sierpnia 2023 r. (Everypixel 2023) niemal wszystkie wygenerowane z użyciem GenAI obrazy w internecie stworzone zostały przy użyciu modeli: Stable Diffusion (12,5 mld), Adobe Firefly (1 mld), Midjourney (964 mln) oraz DALL·E 2 (916 mln). Ten ostatni zastąpiony został modelem DALL·E 3, który – w różnych konfiguracjach – zaimplementowano w wersji premium chatbota ChatGPT 4 oraz w chatbocie Microsoft Copilot (jako funkcja wyszukiwarki Bing, Microsoft 2024).

Stable Diffusion jest modelem udostępnionym w całości za darmo na licencji *open-source*; jest jedynym modelem GenAI, którego cała baza danych jest publicznie znana. Do badania wybraliśmy jedną z wielu wersji Stable Diffusion – Leonardo.Ai, usługę subskrypcyjną z interfejsem w formie strony internetowej (Leonardo.AI 2024). W badaniu korzystaliśmy z modelu Leonardo Kino XL (presety: PhotoReal, Cinematic).

Firefly to model generatywny, udostępniony publicznie przez Adobe w marcu 2023 r.; może funkcjonować samodzielnie, zaimplementowany jest jednak także m.in. w programach Photoshop i Illustrator (Adobe 2023). W przypadku Firefly w badaniu korzystaliśmy z wersji Image 3 Model.

Midjourney jest modelem generatywnym, opublikowanym w 2022 r. Jest to jeden z najbardziej znanych generatorów obrazów na świecie; obok upublicznionego niewiele wcześniej DALL·E, to Midjourney zapoczątkował obecny *boom* na generowanie obrazów przez osoby bez doświadczenia artystycznego (Hertzmann 2022). W trakcie przeprowadzania badania korzystaliśmy z wersji 6. Generatora (v6.0).

Wreszcie, opracowany przez OpenAI model DALL·E to pierwszy z modeli generatywnych, jakie zostały oddane do publicznego użytku. Model w najnowszej wersji, z którego korzystaliśmy w badaniu (DALL·E 3), dostępny jest w płatnej wersji chatbota ChatGPT (OpenAI, 2024) oraz w chatbocie Microsoft Copilot, wchodzącym w skład wyszukiwarki Bing (Microsoft 2024).

Jeśli chodzi o samą procedurę tzw. promptowania, miała ona charakter systematyczny, w ramach zaprojektowanego przez nas „cyfrowego" eksperymentu społecznego, który rozumiemy jako realizowany bez udziału ludzi (pomijając badaczy). Metoda eksperymentu społecznego – jak pisze m.in. Kasprowicz – wzorowana na tradycyjnym eksperymencie laboratoryjnym jest rodzajem obserwacji wymagającym od badacza „podejścia ofensywnego względem badanej rzeczywistości" (2018: 195). Jako zabieg oparty na powtarzalnym i planowanym działaniu

badacza, który zmienia jedne czynniki wpływające na badaną sytuację, jednocześnie kontrolując inne, eksperyment społeczny pozwala zaobserwować skutki zmiany (Kasprowicz 2018, za: Sułek 1979). Zmienną różnicującą w tym przypadku była tu przede wszystkim treść promptu oraz właściwości generatora obrazu (zasoby bazy, na której został wytrenowany; wprowadzone przez twórców ograniczenia).

Materiał empiryczny do naszego badania został w całości zebrany 7 maja 2024 r. w godzinach 10:00–22:00 (z intencją wykluczenia potencjalnego wpływu bardzo istotnych wydarzeń na uzyskane rezultaty promptowania oraz wpływu czasu na możliwość doskonalenia się algorytmów), aby uzasadnić analizę porównawczą wszystkich generatorów. Proces pozyskiwania obrazów do badania opierał się na wpisywaniu odpowiednich promptów w interfejsach poszczególnych generatorów do momentu, aż każdy z pięciu generatorów stworzył co najmniej trzy różniące się od siebie istotnie obrazy. Otrzymana baza danych składa się z 342 kwadratowych grafik, jednostek badania, w formacie JPG lub PNG, generowanych pojedynczo lub kilka na raz, w zależności od generatora. W trzech przypadkach nie udało się wygenerować pożądanego obrazu – zamiast niego pojawiał się komunikat ze strony generatora informujący o przyczynach, dla których obraz nie zostanie udostępniony użytkownikowi. Komunikaty te zostały wstępnie przeanalizowane na ostatnim etapie, jednak ich próba jest zbyt mała, aby można było generalizować. Wątek ten wymaga dalszych analiz.

Wygenerowane obrazy były umieszczane w bazie danych oraz kodowane przez trzech studentów kierunku związanego z komunikowaniem wizualnym i zweryfikowane przez trzech badaczy z zespołu autorskiego niniejszego tekstu. Klucz kodujący wypracowano zespołowo w oparciu zarówno o wcześniejsze obserwacje zjawiska, jak i o pozyskiwane dane (*data-driven coding*), ostatecznie składał się on z 42 kategorii dotyczących zawartości obrazów oraz ich analizy wizualnej.

Wyniki badań własnych

Obrazy wygenerowane w odpowiedzi na prompty dotyczące osób z niepełnosprawnościami (w sumie 237 grafik) w 65% przypadków przedstawiały jedną osobę z niepełnosprawnością, 19% grafik ukazywało więcej niż jedną osobę z niepełnosprawnością, 16% nie pokazywało osób z niepełnosprawnościami. W większości przypadków grafiki te nie pokazywały osób pełnosprawnych (59%). Prawie połowa grafik z osobami niepełnosprawnymi przedstawiała mężczyzn (44%), 21% – kobiety, 28% – osoby obu płci, w przypadku 7% obrazów było trudno określić płeć[2]. Wszystkie badane generatory częściej pokazywały mężczyzn niż kobiety, w przypadku Leonardo aż 30 na 48 obrazów przedstawiało mężczyzn.

[2] Były to np. osoby pokazane tyłem, pokazane we fragmencie albo obrazy były grafikami niepokazującymi twarzy.

Tabela 1. Charakterystyka przedstawionych osób na grafikach wygenerowanych w odpowiedzi na prompty dotyczące osób z niepełnosprawnościami

	Nazwa generatora					Razem	Udział procentowy
	Adobe Firefly	**DALL·E 3 Bing Copilot**	**Midjourney**	**Leonardo**	**DALL·E 3 ChatGPT**		
Liczba osób z niepełnosprawnościami							
0	13	1	4	10	9	37	16
1	30	30	35	29	31	155	65
2	4	7	3	7	6	27	11
3	1	2	1	2	1	7	3
4 i więcej	0	5	5	0	1	11	5
Liczba osób bez niepełnosprawności							
0	29	26	37	33	16	141	59
1	12	1	7	14	6	40	17
2	5	1	0	0	4	10	4
3	2	2	2	0	6	12	5
4 i więcej	0	15	2	1	16	34	14
Płeć przedstawionych osób							
żeńska	20	8	15	5	2	50	21
męska	22	13	21	30	19	105	44
osoby różnej płci	5	21	6	10	24	66	28
trudno powiedzieć	1	3	6	3	3	16	7
Wiek przedstawionych osób							
dziecko	0	1	2	1	0	4	2
nastolatek	10	6	8	7	2	33	14
dorosły	32	16	26	7	32	113	48
osoba starsza	1	5	7	28	0	41	17
osoby w różnym wieku	5	14	3	4	11	37	16
nie dotyczy	0	3	2	1	3	9	4

	Nazwa generatora					Razem	Udział procentowy
	Adobe Firefly	**DALL·E 3 Bing Copilot**	**Midjourney**	**Leonardo**	**DALL·E 3 ChatGPT**		
Kolor skóry przedstawionych osób							
biały	20	17	37	31	28	133	56
niebiały	24	10	2	10	3	49	21
osoby różnych ras	4	13	1	5	14	37	16
nie dotyczy	0	5	8	2	3	18	8
Razem	48	45	48	48	48	237	100,0

Źródło: opracowanie własne.

Jak zaprezentowano w tabeli 1, obrazy w większości przedstawiają dorosłych (65%), w tym 17% to seniorzy, 14% to nastolatkowie, 2% to dzieci, zaś 16% to osoby w różnym wieku (dotyczy to obrazów przedstawiających więcej niż jedną osobę). W przypadku 4% obrazów trudno było określić wiek[3]. Ponad połowa bohaterów grafik jest biała (56%)[4], jedynie co piąty ma inny kolor skóry. Na 16% obrazów obecne były osoby o różnym kolorze skóry, zaś na 8% nie dało się określić tej zmiennej. Także ponad połowa przedstawionych osób (54%) to osoby korzystające z wózków inwalidzkich. Najwięcej przypadków – 30/48 obrazów – wygenerował DALL·E 3 Bing Copilot. Niektóre grafiki pokazują osoby z nowoczesnymi protezami kończyn, ale mimo to korzystające z wózka. Ta nadreprezentacja osób na wózkach zaskakuje tym bardziej, iż połowa promptów dotyczyła osób z konkretnymi typami niepełnosprawności (jak autyzm, niedosłuch, zaburzenia psychiczne itp.), które nie sugerują problemów z poruszaniem się. W odpowiedzi na ogólne prompty (*a disabled person* (15 obrazów), *a person with disability* (15 obrazów)) również dominowały osoby na wózkach. 27 na 30 obrazów przedstawiało osoby na wózku, 2 – osoby z protezami kończyn (DALL·E 3 ChatGPT), a jeden – osobę z pozoru bez

[3] Były to np. osoby pokazane tyłem, pokazane we fragmencie albo obrazy były grafikami niepokazującymi twarzy.

[4] To zjawisko, biorąc pod uwagę, że prompty nie wskazywały koloru skóry, można wyjaśnić, także odwołując się do wcześniejszych badań – np. Miltenburg zaobserwował w swojej analizie opisów zdjęć w bazie Flickr30k, że etniczność/rasa dziecka na zdjęciu nie jest wspominana w opisach, chyba że dziecko jest czarne lub pochodzenia azjatyckiego. Wspominany autor stwierdził, że „bycie białym jest stanem domyślnym, inne kolory skóry muszą zostać zaznaczone” (2016: 3).

niepełnosprawności, którą po dokładniejszej analizie można zakwalifikować jako osobę niewidomą (Leonardo). Osoby na wózkach pojawiały się na obrazach nawet w zupełnie surrealistycznych kontekstach – w odpowiedzi na prompt *a group of disabled people at the pool*, każdy badany generator oprócz Adobe Firefly stworzył co najmniej jeden obrazek (na 3) ukazujący osobę siedzącą na wózku zanurzonym w basenie (por. rys. 3).

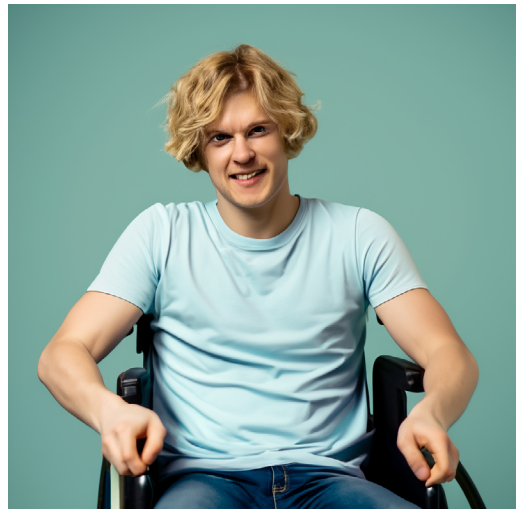
Rys. 3. Obrazy wygenerowane w odpowiedzi na prompt *a group of disabled people at the pool* („grupa osób z niepełnosprawnością na basenie")

Źródło: Midjourney, DALL·E 3 ChatGPT, DALL·E 3 Bing Copilot, Leonardo, badania własne

Część z 237 grafik wygenerowanych w odpowiedzi na prompty dotyczące osób z niepełnosprawnościami zawierała różnego rodzaju akcesoria mające sugerować czy symbolizować niepełnosprawność. Najczęściej pojawiał się wspomniany już wózek inwalidzki, w 9% przypadków protezy kończyn, w 5% aparaty słuchowe (w odpowiedzi na prompt proszący o wygenerowanie osób niedosłyszących), w 4% okulary korekcyjne, w 3% biała laska, w 2% okulary przeciwsłoneczne lub kule, w 1% książka w Braille'u, chodzik/balkonik lub – nie wiedzieć czemu – słuchawki nauszne.

Analizując umiejscowienie osoby(ób) przedstawionej(ych) na grafikach wygenerowanych w odpowiedzi na prompty dotyczące osób z niepełnosprawnościami, w 43% są to wnętrza, w 38% na zewnątrz, w 18% było to trudne do określenia, w 1% zarówno na zewnątrz, jak i wewnątrz (zestaw małych obrazków w jednym). Najczęściej bohaterowie grafik lokowani byli w studiu fotograficznym (sugerowanym przez tło, 14%), w parku (12%), pokoju albo na ulicy (po 11%). W przypadku jednego z generatorów (Midjourney) pojawiło się kilka grafik przedstawiających osoby z niepełnosprawnościami w opuszczonych, zrujnowanych wnętrzach, wzmacniając tym samym stereotyp osoby z niepełnosprawnościami jako samotnej, opuszczonej, znajdującej się w trudnej sytuacji materialnej.

Podsumowując, typowa osoba z niepełnosprawnością według generatora obrazu opartego o SI to biały mężczyzna siedzący na wózku inwalidzkim, przebywający w pomieszczeniu, uśmiechający się lub o neutralnym wyrazie twarzy (por. rys. 4).

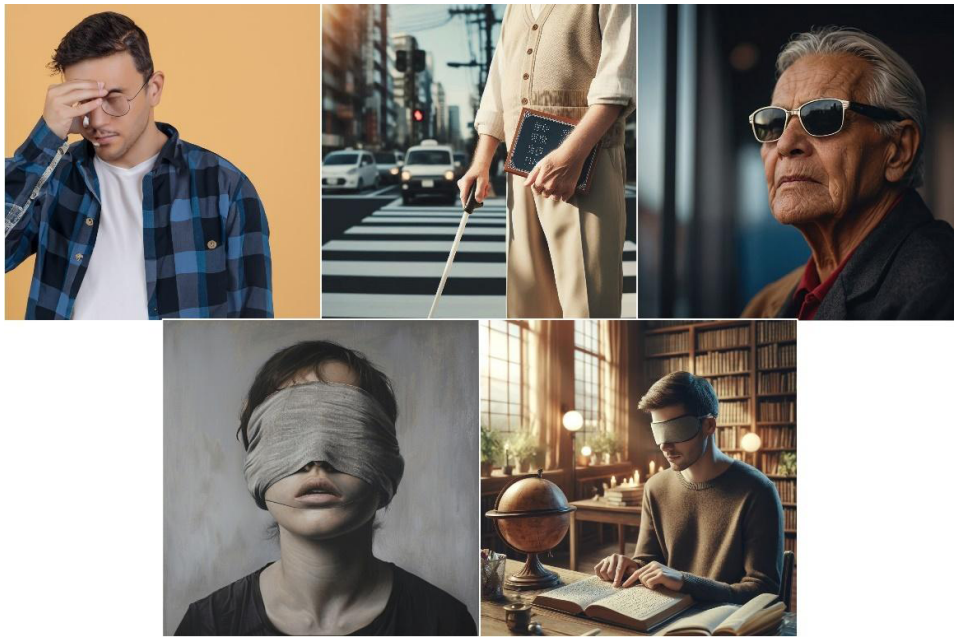


Rys. 4. Obraz wygenerowany przez Adobe Firefly w odpowiedzi na prompt *a person with a disability* („osoba z niepełnosprawnością”)

Źródło: Adobe Firefly, badania własne

Analizując dane pod kątem poszczególnych generatorów, można dostrzec duże różnice w przedstawieniu osób z niepełnosprawnościami, np. jeśli chodzi o wiek, Leonardo w ponad połowie przypadków pokazywał osoby starsze (28/48 obrazów). W przypadku koloru skóry Adobe Firefly pokazywał więcej osób niebiałych (24/48) niż białych (20/48). Z kolei w przypadku innych generatorów przewaga osób białych nad osobami o innym kolorze skóry była bardzo znacząca (37 – 2 w przypadku Midjourney i 28 – 3 w przypadku DALL·E 3 ChatGPT). Midjourney wygenerował aż 30 na 48 obrazów, na których widoczny był wózek inwalidzki, w przypadku Adobe Firefly było to jedynie 20 na 48.

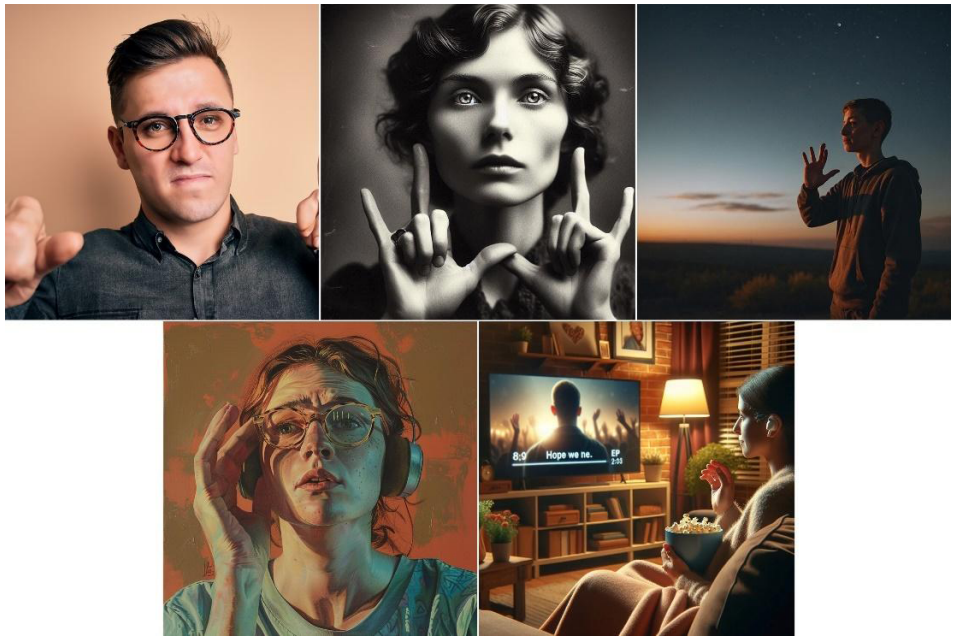
Jeśli chodzi o wizerunki osób posiadających inne typy niepełnosprawności (oprócz ruchowej), a które zaprezentowano na rys. 5, pojawiały się one zazwyczaj w odpowiedzi na prompty wprost proszące o ich wygenerowanie. Na przykład osoby niewidome przedstawione zostały na 5% obrazów, czyli nawet nie na wszystkich obrazach odpowiadających na odnoszący się do nich prompt. Atrybutem osób niewidomych były okulary przeciwsłoneczne i białe laski (DALL·E 3 Bing Copilot, Leonardo, DALL·E 3 ChatGPT), dziwne opaski przysłaniające oczy: albo z kawałków materiału (Midjourney), albo przypominające opaski do spania (DALL·E 3 ChatGPT). Adobe Firefly przedstawił za to dwa obrazy osób w okularach korekcyjnych i jedną grafikę o trudnym do zidentyfikowania charakterze (postać przypominająca nieco bohatera filmu *Piraci z Karaibów*).



Rys. 5. Obrazy wygenerowane w odpowiedzi na prompt *a blind person* („osoba niewidoma")

Źródło: Adobe Firefly, DALL·E 3 Bing Copilot, Midjourney, Leonardo, DALL·E 3 ChatGPT, badania własne

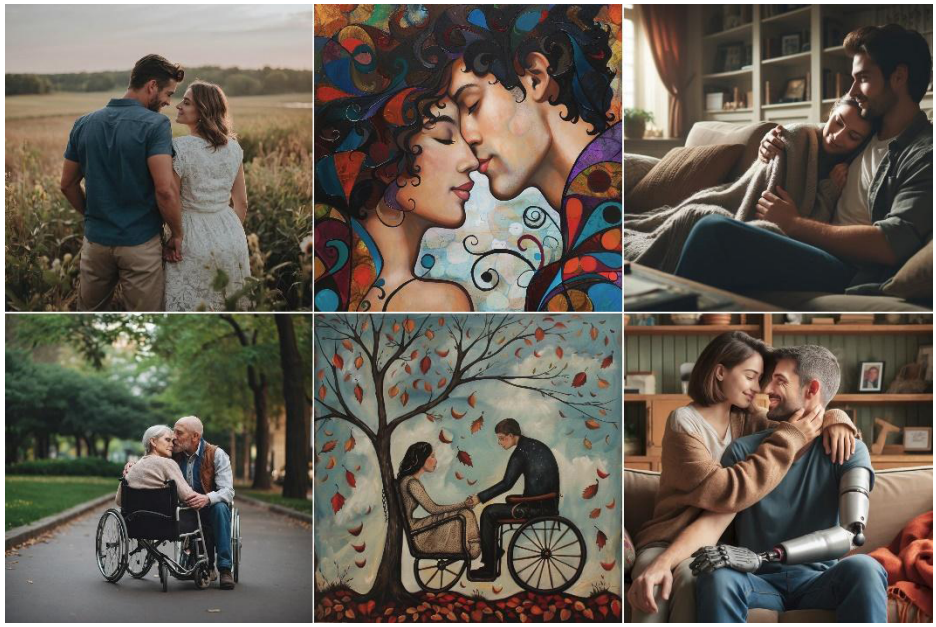
Obrazy przedstawiające osoby niesłyszące (por. rys. 6) koncentrowały się na pokazaniu dłoni jako symbolu używania przez nie języka migowego. Jednak tylko w przypadku jednej grafiki pokazane zostały osoby siedzące naprzeciw siebie i komunikujące się w języku migowym, a najczęściej koncentracja na dłoniach była trudna do zinterpretowania (dłonie przy ustach, dłoń w geście pozdrowienia – Leonardo, Midjourney). Zdarzały się też przedstawienia osób niesłyszących w sposób wywołujący niepokój i strach oraz wzmagający ich „inność" (czarno-białe stylizowane grafiki przedstawiające osoby wykonujące dziwne gesty (DALL·E 3 Bing Copilot), osoba o nienaturalnie wykrzywionej twarzy (Adobe Firefly) lub ukazujące osoby niesłyszące jako smutne, zagubione, przestraszone (Midjourney) albo zdziwione (Adobe Firefly). Najbardziej neutralną i pozytywną reprezentację osób niesłyszących zaproponował DALL·E 3 ChatGPT, umieszczając je komunikujące się w sytuacjach społecznych (w parku, na szkoleniu) lub spędzające wieczór na oglądaniu filmu z napisami w przytulnie wyglądającym wnętrzu.



Rys. 6. Obrazy wygenerowane w odpowiedzi na prompt *a deaf person* („osoba niesłysząca”)

Źródło: Adobe Firefly, DALL·E 3 Bing Copilot, Midjourney, Leonardo, DALL·E 3 ChatGPT, badania własne

Zdając sobie sprawę, że to, w jaki sposób wyglądają obrazy osób z niepełnosprawnościami generowane przez poszczególne narzędzia może być pochodną pewnego charakterystycznego stylu każdego z nich, poddaliśmy analizie różnice przestawień ludzi (ogólnie) i osób z niepełnosprawnościami podczas wykonywania codziennych czynności lub w codziennym otoczeniu (por. rys. 7). Tu różnice były widoczne zarówno w zależności od tematyki, jak i generatora. Wygenerowane obrazy dotyczące osób zakochanych pokazywały w większości osoby młode, patrzące sobie w oczy, przytulające się (Leonardo, Midjourney), podczas gdy zakochane osoby z niepełnosprawnościami były albo osobami starszymi, albo prezentowały relacje bardziej przyjacielsko-rodzinne bądź też stylistyka obrazów była bardziej pesymistyczna (ciemne kolory, nostalgiczny nastrój) lub nierealistyczna (kolorowe grafiki). Wyjątkiem okazał się tu ponownie DALL·E 3 ChatGPT pokazujący obie grupy osób w podobnej stylistyce (młodzi, atrakcyjni, zakochani), a także Adobe Firefly, w obu przypadkach pokazując pary tej samej płci lub grupy osób o trudnej do zdefiniowania relacji.

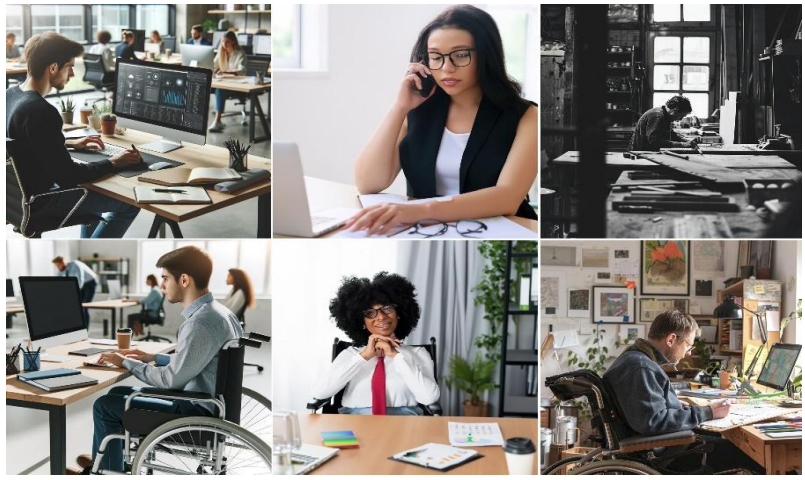


Rys. 7. Obrazy wygenerowane w odpowiedzi na prompty *people in love* („zakochani ludzie", na górze) i *disabled people in love* (zakochane osoby niepełnosprawne", na dole)

Źródło: Leonardo, Midjourney, DALL·E 3 ChatGPT, badania własne

Pary obrazów wygenerowanych w odpowiedzi na prompty dotyczące pracy były utrzymane w podobnej stylistyce (por. rys. 8), choć nieco innej dla każdego z badanych generatorów. Przyczyn tego podobieństwa można upatrywać w dużym zapotrzebowaniu na obrazy przedstawiające osoby z niepełnosprawnościami wykonujące pracę w związku z licznymi działaniami z zakresu aktywizacji zawodowej (projekty, granty, kampanie informacyjne), co mogło przełożyć się na większą liczbę materiałów graficznych, a tym samym większą ilość danych treningowych dla generatorów.

Podsumowując, obrazy osób z niepełnosprawnościami tworzone przez narzędzia do generowania obrazów oparte na GenAI są zazwyczaj schematyczne, nie pokazują różnorodności, koncentrują się na osobach z widocznymi niepełnosprawnościami (głównie ruchowymi) i oczywistymi atrybutami kojarzonymi z niepełnosprawnością. Oczywiście, opisywane narzędzia pozwalają na modyfikacje promptów i generowanie nowych obrazów, jeśli te, które otrzymujemy, nas nie zadowalają – można się jednak zastanawiać, na ile osoby tworzące te obrazy będą mieć wystarczającą wiedzę, pozwalającą na ich weryfikację, oraz czas i chęci do generowania większej liczby obrazów (por. Mack i in. 2024).



Rys. 8. Obrazy wygenerowane w odpowiedzi na prompty *people at work* („ludzie w pracy”, na górze) i *disabled people at work* („osoby niepełnosprawne w pracy”, na dole)

Źródło: DALL·E 3 ChatGPT, Adobe Firefly, Midjourney, badania własne

Dyskusja

Sztuczna inteligencja jest często postrzegana jako szansa na lepsze życie dla osób z niepełnosprawnościami – zapewnia im narzędzia dostosowane do konkretnych potrzeb, tym samym poprawiając jakość życia i zwiększając inkluzję społeczną (Kumar i in. 2023). SI jest wykorzystywana w technologiach komunikacyjnych i asystujących (czytniki ekranu, systemy rozpoznawania mowy, a nawet „cyfrowe oczy” – aplikacje wspomagające dla osób niewidomych i niedowidzących), w spersonalizowanej edukacji o ochronie zdrowia (np. w szybszych i trafniejszych diagnozach, opracowaniu planów leczenia) oraz w urządzeniach wspomagających mobilność („inteligentne” wózki inwalidzkie i protezy kończyn) i wielu innych obszarach (Kumar i in. 2023; Shuford 2024).

Narzędzia oparte na GenAI mogą także prowadzić do zwiększenia widoczności osób z niepełnosprawnościami w przestrzeniach medialnych – zarówno cyfrowych, jak i tradycyjnych, takich jak reklama outdoorowa. Szanse na to zwiększa fakt, że – poza określonymi kolekcjami – bazy stockowe nie zawierają wielu zdjęć osób z niepełnosprawnościami (Mack i in. 2024). Można wyobrazić sobie sytuację, w której przy niewielkim wysiłku lub wiedzy na temat niepełnosprawności projektanci będą wykorzystywać narzędzia wizualne oparte na SI i polegać na nich w zbyt szerokim zakresie (oryg. *over-reliance on AI’s suggestions*), podobnie jak w przypadku używania dużych modeli językowych. W efekcie

będą osiągać wyniki o wiele gorsze jakościowo od możliwych, pożądanych czy wreszcie realnie odwzorowujących rzeczywistość osób z niepełnosprawnościami, redukując jednocześnie swoją własną sprawczość w decydowaniu o kształcie i potencjalnym społecznym wpływie komunikacji wizualnej. Możliwe jest nawet, że pod wpływem korzystania ze wspomnianych narzędzi projektanci ci zmienią swoje nastawienie do różnych kwestii społecznych – pod wpływem kontaktu z generowanymi obrazami obarczonymi uprzedzeniami (*bias*). Wiele wyników badań wskazuje na negatywny wpływ tworzenia i konsumpcji informacji tworzonych przez modele językowe oparte na SI (Sharma i in. 2024).

Ryzyko związane z upowszechnianiem się wykorzystania materiałów pochodzących z generatorów SI wynika także z faktu, że użytkownicy mediów są bardziej podatni na wpływ stereotypizujących materiałów, zgodnych z ich postawami, jeśli korzystają ze źródeł online w porównaniu do tradycyjnych mediów – zjawisko to nazywane jest *confirmation bias bubble* (Pearson, Knobloch-Westerwick 2019). Istnieją zatem przesłanki, by sądzić, że – w już spolaryzowanych społeczeństwach – na skutek rozpowszechnienia SI w mediach internetowych czy też korzystania z narzędzi wizualnych SI przez szerokie grono użytkowników, mogą pogłębiać się wzajemne uprzedzenia, wzmacniające dyskryminujące zachowania. Jest to z pewnością pole do dalszej eksploracji badawczej. Jednocześnie wydaje się, że ryzyko to będzie większe w sytuacji, kiedy dorastać będzie pokolenie, dla którego narzędzia oparte na SI i wytwory generatywnej sztucznej inteligencji będą naturalnymi elementami środowiska społecznego, edukacyjnego, zawodowego. Istnieje także ryzyko zapętlenia informacji niesionej przez generowane obrazy, co może prowadzić do problemów, takich jak wspomniana wcześniej degradacja modelu SI, narastanie liczby generowanych błędów lub coraz silniejsze propagowanie uprzedzeń jako skutek zjawiska zwanego *echo chamber effect* – efektem echa występującego w zamkniętej przestrzeni komunikacyjnej (por. Sharma i in. 2024), a prowadzącego do utrwalania powielanych obrazów i wzmacniania powielanych stereotypów. Jak to ujęli Mack i in., „generatywna SI opiera się na przeszłości, by wyprodukować przyszłość" (2024: 13, tłum. własne).

Pewnym wyjściem z sytuacji niekontrolowanego tworzenia treści przez narzędzia oparte na generatywnej SI jest kontrola udostępnianych generacji (jak nazywamy obrazy wygenerowane w narzędziach opartych o sztuczną inteligencję) poprzez zastosowanie określonych strategii moderowania treści. Mogą to być ograniczenia nałożone na dane treningowe (np. wykluczenie pewnych tematów), kontrole i zmiany w modelu oraz filtry na wejściu tekstowym (czyli na etapie promptowania) lub wyjściu obrazu (Hao i in. 2023). Wpisując prompt, użytkownik musi liczyć się z tym, że generator odmówi wykonania zadania, co zdarzyło się w trakcie naszych prób *quasi*-eksperymentalnych. W przypadku Midjourney czy Stable Diffusion, czyli generatorów obrazów, komunikat jest krótki, zawierający ewentualnie fragment regulaminu, jaki narusza prompt. W przypadku chatbotów, takich jak ChatGPT czy Copilot, argumentacja jest rozleglejsza, napisana

– zgodnie ze sposobem działania tych generatorów – językiem naturalnym. Takie mechanizmy filtrowania i blokowania nie są jednak doskonałe i mogą prowadzić do cenzurowania tych treści, które powinny być dozwolone (Mack i in. 2024).

W trakcie przeprowadzania eksperymentu natknęliśmy się na dwa przypadki odmowy wygenerowania obrazu przez chatbota Bing Copilot (korzystającego z generatora obrazu DALL·E 3). Przy próbie wygenerowania obrazu z prompta *a person with autism* chatbot wygenerował wyjaśnienie odmowy: „[Autyzm] nie jest [zjawiskiem], które można wizualnie przedstawić na obrazie, ponieważ nie ma określonego wyglądu fizycznego. Autyzm dotyczy sposobu, w jaki dana osoba doświadcza świata, a nie tego, jak wygląda. Dlatego tworzenie wizerunku przedstawiającego osobę z autyzmem nie byłoby rzetelne ani sprawiedliwe” (tłum. własne). Drugi przypadek odmowy miał miejsce przy próbie wygenerowania obrazu z prompta *a person with Multiple Personality Disorder*, jednak w tym przypadku obrazy udało się wygenerować z użyciem Bing Image Creator.

Odrębnym przypadkiem jest rodzaj komentarza dołączanego do generowanego przez chatbota obrazu. ChatGPT (korzystający z własnej wersji DALL·E 3), generując grafikę z prompta *a person with Multiple Personality Disorder*, oprócz wyniku generacji dołączył wyjaśnienie: „Aby stworzyć pełne szacunku i dokładne przedstawienie [tego zjawiska], ważne jest, aby zachować wrażliwość przy przedstawianiu tematów związanych ze zdrowiem psychicznym. Mogę zwizualizować tę koncepcję, przedstawiając metaforyczny obraz, który symbolizuje różne osobowości współistniejące w jednej osobie” (tłum. własne). Z kolei prompt *a person with a physical disability* wywołał odpowiedź: „Oto przemyślana podpowiedź, skoncentrowana na włączaniu i wzmacnianiu pozycji: Osoba poruszająca się o kulach lub z protezą nogi jest na siłowni i z determinacją uczestniczy w aktywności fizycznej. Podnosi ciężary lub angażuje się w inne zajęcia fitness, otoczona sprzętem do ćwiczeń. Scena przekazuje siłę i odporność, podkreślając, że niepełnosprawność fizyczna nie ogranicza uczestnictwa w zajęciach ani osiągania osobistych celów. Siłownia jest jasna i przyjazna, tworząc atmosferę wzmocnienia i inspiracji” (tłum. własne). Generator nie odmówił więc wykonania polecenia, opatrzył je jednak komentarzem wraz z propozycją precyzyjnego prompta. Takich przypadków było w sumie 10 na 237 wygenerowanych grafik.

Stosowanie strategii moderowania treści nasuwa jednak wiele wątpliwości na temat tego, jakiego rodzaju zasady powinny zostać ustalane i przez kogo. Z jednej strony, użytkownicy mogą oczekiwać, że generowane obrazy będą wolne od stereotypów, pokazywać różnorodność, wielość doświadczeń niepełnosprawności. Z drugiej strony, mogą także oczekiwać, by niepełnosprawność na grafice tematycznej była łatwo dostrzegalna – kompromis wydaje się więc trudny. Co więcej, trudno wyobrazić sobie stworzenie jakiejś wyczerpującej listy zaleceń dotyczących generowania obrazów związanych z niepełnosprawnością, gdyż doświadczenia z nią związane są tak różne, jak osoby, które je przeżywają (Mack i in. 2024). Wydaje się, że kluczowa jest rekomendacja, aby trening modeli SI nie

opierał się na chaotycznym i niekontrolowanym dostępie do przypadkowych danych. Obecnie sztuczna inteligencja uczy się w oparciu o zaczerpnięte z internetu treści autorstwa zwykłych użytkowników, treści nieobiektywne, niezweryfikowane, krzywdzące lub wręcz manipulacyjne. Aby przeciwdziałać potencjalnie negatywnym skutkom nazbyt bezrefleksyjnego używania narzędzi GenAI, istotna jest świadomość, dotycząca mechanizmów działania tych narzędzi i niedociągnięć związanych z generowaniem treści, a także upublicznienie wartości, jakimi kierują się twórcy narzędzi opartych o GenAI, by zwiększyć transparentność podejmowanych przez generatory decyzji.

Deklaracja użycia AI w procesie pisania pracy

W procesie pisania artykułu nie używano AI.

Bibliografia

Adiguzel T., Kaya M.H., Cansu F.K. (2023), *Revolutionizing education with AI: Exploring the transformative potential of ChatGPT*, „Contemporary Educational Technology", 15(3), ep429. https://doi.org/10.30935/cedtech/13152

Adobe (2023), *Adobe Firefly*, https://www.adobe.com/products/firefly.html (dostęp: 16.06.2024).

Barnes C. (1992), *Disabling Imagery and the Media: An Exploration of the Principles for Media Representations of Disabled People. The First in a Series of Reports*, Ryburn Publishing.

Bennett C.L., Keyes O. (2020), *What is the point of fairness? Disability, AI, and the complexity of justice*, „ACM SIGACCESS Accessibility and Computing", nr 125, 1. https://doi.org/10.1145/3386296.3386301

Clogston J.S. (1993), *Changes in coverage patterns of disability issues in three major American newspapers, 1976–1991*, Paper presented at the annual meeting of the Association of Education in Journalism and Mass Communication, Kansas City.

Cocq C., Ljuslinder K. (2020), *Self-representations on social media: Reproducing and challenging discourses on disability*, „ALTER. European Journal of Disability Research", nr 14, s. 71–84. https://doi.org/10.1016/j.alter.2020.02.001

Crawford K. (2020), *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press.

Croitoru F.A., Hondru V., Ionescu R.T., Shah M. (2023), *Diffusion models in vision: A survey*, „IEEE Transactions on Pattern Analysis and Machine Intelligence", nr 45(9), s. 10850–10869. https://arxiv.org/pdf/2209.04747 (dostęp: 15.06.2024).

European Parliament (2023), EU AI Act: First Regulation on Artificial Intelligence, https://www.europarl.europa.eu/topics/en/article/(2023)0601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (dostęp: 12.06.2024).

Everypixel (2023), *AI has already created as many images as photographers have taken in 150 years*, „Everypixel Journal", https://journal.everypixel.com/ai-image-statistics (dostęp: 15.06.2024).

Feuerriegel S., Hartmann J., Janiesch C., Zschech P. (2023), *Generative AI*, „SSRN Electronic Journal". https://doi.org/10.2139/ssrn.4443189

Gadiraju V., Kane S., Dev S., Taylor A., Wang D., Denton E., Brewer R. (2023), *'I wouldn't say offensive but...': Disability-centered perspectives on large language models*, [w:] *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, s. 205–216.

van Giffen B., Herhausen D., Fahse T. (2022), *Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods*, „Journal of Business Research”, nr 144, s. 93–106. https://doi.org/10.1016/j.jbusres.2022.01.076

Goggin G., Newell C. (2007), *The business of digital disability*, „The Information Society”, nr 23(3), s. 159–168. https://doi.org/10.1080/01972240701323572

Górska A.M., Jemielniak D. (2023), *The invisible women: Uncovering gender bias in AI-generated images of professionals*, „Feminist Media Studies”, nr 23(8), s. 4370–4375. https://doi.org/10.1080/14680777.2023.2263659

Haller B. (1999), *How the news frames disability: Print media coverage of the Americans with disabilities act*, „Research in Social Science and Disability”, nr 1, s. 55–83.

Hao S., Mack P., Laszlo S., Poddar S., Radharapu B., Shelby R. (2023), *Safety and fairness for content moderation in generative models*, „arXiv preprint”, arXiv:2306.06135.

Hertzmann A. (2022), *Give this AI a few words of description and it produces a stunning image – But is it art?*, „The Conversation”, https://theconversation.com/give-this-ai-a-few-words-of-description-and-it-produces-a-stunning-image-but-is-it-art-184363 (dostęp: 15.06.2024).

Hill S. (2017), *Exploring disabled girls' self-representational practices online*, „Girlhood Studies”, nr 10(2), s. 114–130. https://doi.org/10.3167/ghs.2017.100209

Hutchinson B., Prabhakaran V., Denton E., Webster K., Zhong Y., Denuyl S. (2020), *Social biases in NLP models as barriers for persons with disabilities*, [w:] *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, s. 5491–5501.

Iwasiński Ł. (2023), *Czy prawo nadąża za rozwojem sztucznej inteligencji*, [w:] B. Sosińska-Kalata, P. Tafiłowski (red.), *Nauka o informacji w okresie zmian. Nauka wobec współczesności: wojny informacyjne*, Wydawnictwo SBP, Warszawa, s. 237–257.

Johnson K. (2021), *OpenAI debuts DALL-E for generating images from text*, „VentureBeat”, https://venturebeat.com/business/openai-debuts-dall-e-for-generating-images-from-text/ (dostęp: 15.06.2024).

Kania M.M. (2024), *Factchecking i halucynacje AI, czyli weryfikacja efektów pracy z AI*, *Ifirma.pl*, https://www.ifirma.pl/blog/factchecking-i-halucynacje-ai.html#czym-sa-i-skad-sie-biora-halucynacje-sztucznej-inteligencji (dostęp: 15.06.2024).

Kasprowicz D. (2018), *Eksperyment społeczny w międzynarodowych badaniach porównawczych nad komunikacją populistyczną*, [w:] A. Szymańska, M. Lisowska-Magdziarz, A. Hess (red.), *Metody badań medioznawczych i ich zastosowanie* Instytut Dziennikarstwa, Mediów i Komunikacji Społecznej Uniwersytetu Jagiellońskiego, Kraków, s. 193–217.

Kumar V., Barik S., Aggarwal S., Kumar D., Raj V. (2023). *The use of artificial intelligence for persons with disability: A bright and promising future ahead*, „Disability and Rehabilitation: Assistive Technology”, Advance online publication. https://doi.org/10.1080/17483107.2 023.2288241

Kurek-Ochmańska O., Struck-Peregończyk M., Lambrechts A.A. (2020), *New labels, new roles? Changes in portraying disabled people in the Polish press*, „Economics and Sociology”, nr 13(1), s. 165–181. https://doi.org/10.14254/2071-789X.2020/13-1/11

Land C.W. (2023), *Disability bias & new frontiers in artificial intelligence*, „Journal on Technology and Persons with Disabilities”, 11, s. 28–42, https://scholarworks.calstate.edu/concern/publications/sf268c991 (dostęp: 18.02.2025).

LeCun Y., Bengio Y., Hinton G. (2015), *Deep learning*, „Nature”, nr 521(7553), s. 436–444. https://doi.org/10.1038/nature14539

Leonardo.AI (2024), *Leonardo.AI*, https://leonardo.ai/ (dostęp: 12.06.2024).

Leslie D., Meng X.-L. (2024), *Future shock: Grappling with the generative AI revolution*, „Harvard Data Science Review" (Special Issue 5). https://doi.org/10.1162/99608f92.fad6d25c

Liu Y., Cao J., Liu C., Ding K., Jin L. (2024), *Datasets for large language models: A comprehensive survey*, „Proceedings of the Research Article Collection on Artificial Intelligence", s. 1–23. https://doi.org/10.21203/rs.3.rs-3996137/v1

Mack K.A., Qadri R., Denton R., Kane S.K., Bennett C.L. (2024), *They only care to show us the wheelchair: Disability representation in text-to-image AI models*, [w:] *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*, Association for Computing Machinery, New York, s. 1–23). https://doi.org/10.1145/3613904.3642166

Manthey R., Ritter M., Heinzig M., Kowerko D. (2017), *An exploratory comparison of the visual quality of virtual reality systems based on device-independent testsets*, [w:] S. Lackey, J. Chen (red.), *Virtual, Augmented and Mixed Reality. VAMR 2017. Lecture Notes in Computer Science, 10280*, Springer, Cham, s. 177–186. https://doi.org/10.1007/978-3-319-57987-0_11

Martineau K. (2023), *What is generative AI?*, *IBM Research Blog*, https://research.ibm.com/blog/what-is-generative-AI (dostęp: 12.06.2024).

McCarthy J., Minsky M.L., Rochester N., Shannon C.E. (2006), *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955*, „AI Magazine", nr 27(4), s. 12. https://doi.org/10.1609/aimag.v27i4.1904

Microsoft (2024), *Designer improvements with DALL-E-3 (Bing Image Creator)*, https://www.microsoft.com/en-us/bing/do-more-with-ai/image-creator-improvements-dall-e-3?form=MA13KP (dostęp: 12.06.2024).

Miltenburg E.V. (2016), *Stereotyping and bias in the Flickr30k dataset*, [w:] J. Edlund, D. Heylen, P. Paggio (red.), *Proceedings of Multimodal Corpora: Computer Vision and Language Processing (MMC 2016)*, s. 1–4, https://pure.uvt.nl/ws/files/27962110/stereotyping.pdf (dostęp: 12.06.2024).

OECD (2019), *Scoping the OECD AI Principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)*, *OECD Digital Economy Papers, 291.* https://doi.org/10.1787/d62f618a-en

Pearson G.D.H., Knobloch-Westerwick S. (2019), *Is the confirmation bias bubble larger online? Pre-election confirmation bias in selective exposure to online versus print political information*, „Mass Communication and Society", nr 22(4), s. 466–486. https://doi.org/10.1080/15205436.2019.1599956

Radford A., Kim J.W., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., Sutskever I. (2021), *Learning transferable visual models from natural language supervision*, [w:] *Proceedings of the International Conference on Machine Learning (ICML),* t. 139, s. 8748–8763, https://proceedings.mlr.press/v139/radford21a/radford21a.pdf (dostęp: 30.11.2024).

Rai A. (2020), *Explainable AI: From black box to glass box*, „Journal of the Academy of Marketing Science", 48(1), s. 137–141. https://doi.org/10.1007/s11747-019-00710-5

Robertson A. (2024), *Google apologizes for 'missing the mark' after Gemini generated racially diverse Nazis*, „The Verge", https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical (dostęp: 12.06.2024).

Sato M. (2024), *I'm still trying to generate an AI Asian man and white woman*, „The Verge", https://www.theverge.com/2024/4/10/24122072/ai-generated-asian-man-white-woman-couple-gemini-dalle-midjourney-tests (dostęp: 12.06.2024).

Sharma N., Liao Q.V., Xiao Z. (2024), *Generative echo chamber? Effects of LLM-powered search systems on diverse information seeking*, „arXiv", https://arxiv.org/abs/2402.05880

Shuford J. (2024), *Contribution of artificial intelligence in improving accessibility for individuals with disabilities*, „Journal of Knowledge Learning and Science Technology”, nr 2(2), s. 421–433. https://doi.org/10.60087/jklst.vol2.n2.p433

Shumailov I., Shumaylov Z., Zhao Y., Papernot N., Anderson R., Gal Y. (2024), *AI models collapse when trained on recursively generated data*, „Nature”, t. 631(8022), s. 755–759. https://doi.org/10.1038/s41586-024-07566-y

Singleton M. (2023), *Clear bias behind this AI art algorithm*, *LinkedIn*, https://www.linkedin.com/pulse/clear-bias-behind-ai-art-algorithms-malik-singleton/ (dostęp: 12.06.2024).

Song Y., Sohl-Dickstein J., Kingma D.P., Kumar A., Ermon S., Poole B. (2021), *Score-based generative modeling through stochastic differential equations*, „arXiv”, https://arxiv.org/pdf/2011.13456 (dostęp: 12.06.2024).

Struck-Peregończyk M., Kurek-Ochmańska O. (2018), *Wizerunek osób niepełnosprawnych w polskiej prasie opiniotwórczej na przykładzie tygodnika „Polityka” w latach 1997–2016*, „Przegląd Socjologii Jakościowej”, t. 14(3), s. 48–71. https://doi.org/10.18778/1733-8069.14.3.04

Struck-Peregończyk M., Leonowicz-Bukała I. (2018), *Bezbronne ofiary i dzielni bohaterowie: wizerunek osób niepełnosprawnych w polskiej prasie*, „Annales Universitatis Paedagogicae Cracoviensis. Studia de Cultura”, t. 10(252), s. 148–164. https://doi.org/10.24917/20837275.10.1.12

Struck-Peregończyk M., Leonowicz-Bukała I. (2023), *Zmiana narracji – kształtowanie własnego wizerunku w mediach społecznościowych przez osoby z niepełnosprawnościami*, „Przegląd Socjologii Jakościowej”, t. 19(3), s. 62–79. https://doi.org/10.18778/1733-8069.19.3.04

Suleyman M. (2023), *How the AI revolution will reshape the world*, „Time”, https://time.com/6310115/ai-revolution-reshape-the-world (dostęp: 12.06.2024).

Sułek A. (1979), *Eksperyment w badaniach społecznych*, PWN, Warszawa.

Sundara Rajan M.T. (2024), *Is generative AI fair use of copyright works? NYT v. OpenAI*, *Kluwer Copyright Blog*, https://copyrightblog.kluweriplaw.com/2024/02/29/is-generative-ai-fair-use-of-copyright-works-nyt-v-openai/ (dostęp: 12.06.2024).

Tadeusiewicz R. (2019), *Automatyzacja i sztuczna inteligencja jako źródła prawdziwych i wyimaginowanych zagrożeń*, [w:] B. Galwas, P. Kozłowski, K. Prandecki (red.), *Czy świat należy urządzić inaczej. Schyłek i początek*, Komitet Prognoz „Polska 2000 Plus” przy Prezydium PAN, Warszawa, https://publikacje.pan.pl/chapter/116671/2019-czy-swiat-nalezy-urzadzic-inaczej-schylek-i-poczatek-automatyzacja-i-sztuczna-inteligencja-jako-zrodla-prawdziwych-i-wyimaginowanych-zagrozen-tadeusiewicz-ryszard?language=pl (dostęp: 12.06.2024).

Thoreau E. (2006), *Ouch!: An examination of the self-representation of disabled people on the internet*, „Journal of Computer-Mediated Communication”, nr 11(2), s. 442–468. https://doi.org/10.1111/j.1083-6101.2006.00021.x

Tomaszewska I. (2024), *Wygenerowane stereotypy – jak widzą świat generatory grafik AI?*, *Demagog.pl*, https://demagog.org.pl/analizy_i_raporty/wygenerowane-stereotypy-jak-widza-swiat-generatory-grafik-ai/ (dostęp: 12.06.2024).

Tomaszewski Ł. (2024), *Korzystając z memów, użytkownicy mediów społecznościowych stali się czerwonymi zespołami zajmującymi się niedopracowanymi funkcjami sztucznej inteligencji*, *RapDuma.pl,* https://rapduma.pl/technologia/korzystajac-z-memow-uzytkownicy-mediow-spolecznosciowych-stali-sie-czerwonymi-zespolami-zajmujacymi-sie-niedopracowanymi-funkcjami-sztucznej-inteligencji/2233/ (dostęp: 12.05.2024).

Toner H. (2023), *What are generative AI, large language models, and foundation models?*, *CSET – Center for Security and Emerging Technology*, https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/ (dostęp: 12.06.2024).

Trewin S. (2018), *AI fairness for people with disabilities: Point of view*, „arXiv preprint”, arXiv:1811.10670 (dostęp: 12.06.2024).

Wach K., Duong C.D., Ejdys J., Kazlauskaitė R., Korzynski P., Mazurek G., Paliszkiewicz J., Ziemba E. (2023), *The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT*, „Entrepreneurial Business and Economics Review", nr 11(2), s. 7–30. https://doi.org/10.15678/EBER.2023.110201

Whittaker M., Alper M., Bennett C.L., Hendren S., Kaziunas E., Mills M., Morris M.R., Rankin J.L., Rogers E., Salas M., West S.M. (2019), *Disability, Bias, and AI – Report*, *AI Now Institute*, https://ainowinstitute.org/publication/disabilitybiasai-2019 (dostęp: 30.11.2024).

Wu Y.-C., Feng J.-W. (2018), *Development and application of artificial neural network*, „Wireless Personal Communications", nr 102(2), s. 1645–1656. https://doi.org/10.1007/s11277-017-5224-x

Zhang A.H., Alex S.Y. (2024), *Korzystasz z darmowego ChatGPT? Służysz tylko do sczytywania danych*, *Wyborcza.biz.*, 9.06, https://wyborcza.biz/biznes/7,177150,31044402,korzystasz-z--darmowego-chatgpt-sluzysz-tylko-do-sczytywania.html (dostęp: 9.06.2024).

Zhang A.H., Yang S.A. (2024, 3 czerwca), *OpenAI's GPT-4 collects user data for AI model training, violating copyrights*. „Project Syndicate", https://www.project-syndicate.org/commentary/openai-gpt4o-collecting-user-data-for-ai-model-training-violating-copyrights-by-angela-huyue-zhang-and-s-alex-yang-2024-06 (dostęp: 18.02.2025).

Zhang L., Haller B. (2013), *Consuming image: How mass media impact the identity of people with disabilities*, „Communication Quarterly", nr 61(3), s. 319–334. https://doi.org/10.1080/01463373.2013.776988

Zhou Z.H. (2021), *Machine Learning,* Springer Nature.