Annales Universitatis Paedagogicae Cracoviensis 410

Studia ad Bibliothecarum Scientiam Pertinentia 23 ▪ 2025
ISSN 2081-1861
DOI 10.24917/20811861.23.23

Veslava Osińska
Uniwersytet Mikołaja Kopernika w Toruniu
ORCID: 0000-0002-1306-7832

Weronika Kortas
Uniwersytet Mikołaja Kopernika w Toruniu
ORCID: 0000-0002-4276-7651

Adam Szalach
Akademia Kultury Społecznej i Medialnej w Toruniu
ORCID: 0000-0001-8040-001X

Percepcja obrazów generowanych przez sztuczną inteligencję: w stronę zrozumienia odbiorcy

Wstęp

Zdaniem ekspertów w dziedzinie sztucznej inteligencji, jednym z najbardziej obiecujących i spektakularnych osiągnięć ostatniej dekady są sieci GAN (Generative Adversarial Networks), czyli generatywne sieci współzawodniczące. Mechanizm GAN wykorzystuje dwie niezależne sieci: generator i dyskryminator, które ze sobą współzawodniczą: pierwszy próbuje oszukać drugi, fałszując rzeczywiste obrazy, a drugi weryfikuje je na podstawie prawdziwych danych. Obie sieci ze sobą rywalizują, jednak pojedynek wygrywa jedna z nich, generując obrazy maksymalnie zbliżone do rzeczywistych. Koncepcja ta została wymyślona przez Iana Goodfellowa i jego

współpracowników[1]. Od tego momentu problematyka sieci neuronowych w zakresie rozpoznawania i klasyfikacji obrazów poszerzyła się o zagadnienie tworzenia realistycznej grafiki.

Trenowany systematycznie model GAN uzyskał spektakularne wyniki. Dzięki zasilaniu dużymi zbiorami danych (takimi jak obrazy) dostępnymi w Internecie oraz zastosowaniu mechanizmu rywalizacji, generowane przez nie efekty są niezwykle realistyczne. Sieci GAN znajdują zastosowanie w wielu dziedzinach, wykorzystując zdolność do realistycznego generowania danych. Przykładem jest aplikacja FaceApp, umożliwiająca modyfikację zdjęć, m.in. poprzez odmładzanie, postarzanie twarzy czy zmianę cech wyglądu. Technologia ta wspiera również projektowanie wnętrz, pozwalając na testowanie różnych aranżacji w czasie rzeczywistym przy użyciu smartfona. GAN-y wykorzystywane są także jako szybkie i ekonomiczne narzędzie w organizacji sesji zdjęciowych, zastępując potrzebę angażowania modeli. W obszarze naukowym służą m.in. do rozpoznawania tekstu oraz tworzenia modeli 3D maszyn i urządzeń, a także znajdują zastosowanie w obrazowaniu medycznym. Obiecującym kierunkiem ich rozwoju pozostaje generowanie utworów muzycznych i materiałów filmowych.

W ostatnich latach obserwujemy niezwykły rozwój technologii GAN widoczny w poprawie fotorealizmu generowanych obrazów[2]. Algorytm jest w stanie wytworzyć wysokiej rozdzielczości obrazy twarzy i ciała człowieka, koty i psy, samochody i inne kategorie obiektów, których niewprawne oko nie jest w stanie odróżnić od prawdziwych zdjęć. Na stronie ThisPersonDoesNotExist.com, po każdym odświeżeniu strony, odwiedzający zobaczą wizerunek ludzkiej twarzy ostrością nie ustępujący zdjęciom wykonywanym nowoczesnym aparatem cyfrowym. Te twarze wyglądają niezwykle realistycznie: rysy twarzy, faktura skóry, włosy, a nawet defekty skóry (np. pory są wyraźne). Oczy i wyraz twarzy, wyglądające bardzo realistycznie, utrudniają ich odróżnienie od prawdziwych postaci.

W mediach społecznościowych powstaje sporo zamieszania wokół hiperrealistycznych obrazów ukazujących znane postacie i dotyczących bieżących wydarzeń. Takie aplikacje jak Midjourney, DALL-E, Stable

[1] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, *Generative Adversarial Networks*, [w:] *Advances in Neural Information Processing Systems*, 2014, vol. 2, https://doi.org/10.48550/arXiv.1406.2661.

[2] T. Karrasi, T. Aila, S. Laine, J. Lehtinen, *Progressive growing of GANs for improved quality, stability, and variation*, [w:] *International Conference on Learning Representations*, 2018, [on-line:] https://www.researchgate.net/publication/320707565 – 18.11.2025; J. Xiang, *On generated artistic styles: Image generation experiments with GAN algorithms*, „Visual Informatics" 2023, vol. 7, issue 4, s. 36–40, https://doi.org/10.1016/j.visinf.2023.10.005.

Diffusion, Craiyon pozwalają na wygenerowanie takich zdjęć przy użyciu słów kluczowych, tzw. promptów. W sieci pojawia się coraz więcej poradników zawierających wskazówki, jak rozpoznać obrazy generowane przez sztuczną inteligencję i odróżnić je od fotografii rzeczywistych (określanych dalej jako artefakty). W odpowiedzi na to wyzwanie opracowywane są również specjalistyczne detektory, dostępne online, które automatycznie analizują i identyfikują generowane obrazy. Przykładem takiego narzędzia jest „AI or NOT"[3].

W badaniach dotyczących relacji między człowiekiem a sztuczną inteligencją, często analizuje się percepcję i zdolności do rozróżniania obrazów generowanych przez SI (sztuczną inteligencję) od zdjęć rzeczywistych. Materiał badawczy jest katygoryzowany, m.in. uczestnicy mierzą się z rozpoznawaniem autorstwa (człowiek lub SI) zdjęć samochodów, krajobrazów, budynków oraz pomieszczeń, np. sypialni[4].

Podejmując próbę identyfikacji i nazwania typów artefaktów występujących na obrazach generowanych przez SI, autorzy skorzystali z aktualnej literatury przedmiotu[5], a także z materiałów internetowych i poradników przygotowanych przez i dla fact-checkerów oraz grafików[6].

W celu usystematyzowania typów artefaktów występujących w obrazach generowanych przez sztuczną inteligencję posłużyliśmy się innym obiecującym modelem AI – GPT-4, zaimplementowanym w narzędziu ChatGPT. Model ten kompiluje aktualnie dostępną wiedzę internetową w danym zakresie tematycznym. Na pytanie: „Po czym rozpoznajemy obrazy wygenerowane za pomocą SI?" odpowiedział: „Rozpoznawanie obrazów wygenerowanych za pomocą algorytmów sztucznej inteligencji może być trudne, ponieważ te obrazy są tworzone przez komputerowe modele, które uczą się na podstawie danych i niekoniecznie odzwierciedlają rzeczywistość. Jednak istnieją pewne cechy i techniki, które można

[3] *AI or Not – AI Detection for Truth Seekers*, [on-line:] https://www.aiornot.com – 14.11.2025.

[4] N. Khanna, G.T.C. Chiu, J.P. Allebach, E.J. Delp, *Forensic techniques for classifying scanner, computer generated and digital camera images*, [w:] *IEEE. International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2008, https://doi.org/10.1109/ICASSP.2008.4517944.

[5] N. Kamali, K. Nakamura, A. Chatzimparmpas, J. Hullman, M. Groh, *How to Distinguish AI-Generated Images from Authentic Photographs*, Northwestern University 2024, https://doi.org/10.48550/arXiv.2406.08651; D.M. Patel, *Artificial Intelligence & Generative AI for Beginners: The Complete Guide (Generative AI & Chat GPT Mastery Series)*, Independently published, 2023.

[6] J. Jacques, *How to Identify AI-generated and Fake Images*, 2023, [on-line:] https://jacquesjulien.com/identify-fake-images/ – 10.06.2025; *How to spot AI-generated images*, Techoist,, 24.01.2013, [on-line:] https://www.youtube.com/watch?v=zqRcjbft3zg – 16.05.2025.

wykorzystać do próby rozpoznania takich obrazów", a następnie wylistował kilka z nich:

> 1. Niespójność lub nierealność: Obrazy generowane przez SI mogą często zawierać niespójne lub nierealne elementy. Na przykład, mogą zawierać surrealistyczne sceny, niemożliwe perspektywy lub nietypowe połączenia przedmiotów.
> 2. Brak szczegółów lub szczegóły nadmiernie dopracowane: Obrazy SI mogą zawierać brakujące szczegóły, które zwykle występują w naturalnych zdjęciach, lub nadmiernie dopracowane szczegóły, które wydają się nieproporcjonalne do reszty obrazu.
> 3. Powtarzalność lub symetria: Niektóre algorytmy SI mogą wykazywać powtarzalność lub symetrię w swoich obrazach, co jest mniej typowe dla naturalnych scen.
> 4. Błędy generatywne: Algorytmy SI czasami popełniają błędy generatywne, które mogą objawiać się jako artefakty lub nieprawidłowości w obrazie. Mogą to być np. zniekształcenia, smugi, pęknięcia lub inne nieoczekiwane efekty wizualne.
> 5. Analiza metadanych: można również spróbować przeanalizować metadane obrazu, takie jak data utworzenia, autor lub źródło. Obrazy SI mogą nie mieć realnych metadanych lub zawierać informacje, które sugerują ich automatyczne generowanie.
> 6. Użycie narzędzi do rozpoznawania obrazów, takich jak algorytmy analizy treści, aby ocenić zawartość i charakter obrazu. Jednak te narzędzia mogą być ograniczone, jeśli obraz jest unikalny lub trudny do sklasyfikowania.
> Ekspertyza ludzka: W niektórych przypadkach trudno jest dokładnie stwierdzić, czy obraz został wygenerowany przez SI, dlatego ekspertyza ludzka może być konieczna do dokładnej oceny[7].

Cele i pytania badawcze

Badanie przedstawione w niniejszym artykule stanowi pierwszą część serii eksperymentów mających na celu ocenę zdolności uczestników do rozróżniania obrazów autentycznych od generowanych przez sztuczną inteligencję oraz analizę czynników wpływających na rozpoznawalność obrazów sztucznych o różnorodnej tematyce. Do realizacji założeń badawczych zastosowano technikę eye-tracking (ET) czyli śledzenia ruchu gałek ocznych. Ludzkie oko wykonuje 3–4 ruchy na sekundę (sakkady), a uwaga wzrokowa przemieszcza się między elementami otoczenia, zwykle poza

[7] OpenAI, *ChatGPT (wersja GPT-4)* [Model językowy], 2025, [on-line:] https://chat.openai.com/ – 16.05.2025.

świadomością. Eyetracking to metoda badawcza, która śledzi ruchy oczu i punkty skupienia wzroku (fiksacje), pozwalając analizować to, jak postrzegamy bodźce wizualne[8].

Jako pierwsze w planowanym cyklu, badanie to pełni również funkcję pilotażową, pozwalającą na weryfikację i dopracowanie proponowanej metodologii. Grupę badawczą stanowiła próba studentów pochodzenia polskiego. Sformułowano więc następujące pytania badawcze:

- W jaki sposób użytkownicy rozpoznają obrazy wygenerowane przez SI i jakie czynniki mają na to decydujący wpływ?
- Czy wstępna wiedza uczestników o możliwościach generatywnych algorytmów wpływa na ich zdolność do rozróżniania obrazów SI od fotografii?
- Czy różnice w poziomie wiedzy na temat generowanych obrazów SI przekładają się na rozpoznawalność tych obrazów przez badanych?
- Czy analiza ruchu gałek ocznych (eye-tracking) dostarcza dodatkowych informacji, które poszerzają obserwacje płynące z ankiet i przyczyniają się do wnioskowania w zakresie rozpoznawalności sztucznych obrazów?
- Czy wskazówki generowane przez SI (np. ChatGPT) dotyczące percepcji obrazów odpowiadają rzeczywistemu sposobowi, w jaki ludzie postrzegają obrazy SI, czy też istnieją między nimi istotne różnice?
- Czy ludzie interpretują obrazy SI w sposób zgodny z regułami „ustalanymi" przez algorytmy, czy też tworzą własne schematy postrzegania?

Przegląd literatury przedmiotu

Jeszcze kilkadziesiąt lat temu nie zastanawiano się powszechnie, czy to, co widzimy, istnieje naprawdę. Przemiany związane z erą cyfrową sprawiły, że zjawisko fałszywych multimediów stało się coraz bardziej powszechne. Technologie GAN generujące realistyczne obrazy, w tym twarze, rodzą wyzwania związane z identyfikacją autentyczności mediów. Przykładowo, Facebook w 2019 r. poinformował o tworzeniu fałszywych kont z generowanymi przez SI zdjęciami profilowymi[9].

W ostatnich dekadach przeprowadzono liczne badania dotyczące relacji człowieka ze sztuczną inteligencją, w tym percepcji oraz zdolności

[8] D. Smołucha, *Eye-tracking in Cultural Studies*, „Perspektywy Kultury" 2019, t. 27, nr 4, s. 169–183, https://doi.org/10.35765/pk.2019.2704.12.

[9] D. Fallis, *The epistemic threat of deepfakes*, „Philosophy & Technology" 2021, vol. 34, s. 623–643, https://doi.org/10.1007/s13347-020-00419-2.

rozróżniania obrazów generowanych komputerowo od rzeczywistych[10]. Uczestnicy badania mieli za zadanie rozpoznawać różnorodne obiekty, takie jak samochody, krajobrazy, budynki czy wnętrza. W analizach statystycznych wykorzystywano m.in. metody Bayesa, pozwalające na określenie, czy zebrane dane wspierają hipotezę alternatywną (istnienie efektu) czy hipotezę zerową (jej brak).

Szczególne znaczenie w tym obszarze ma rozpoznawanie twarzy (ang. Face Recognition, FR), które aktywuje wyspecjalizowane obszary mózgu, w tym wrzecionowaty obszar twarzowy (Fusiform Face Area, FFA). Neuronalne mechanizmy FFA odpowiadają za ekstrakcję cech rysów twarzy oraz ich układu przestrzennego, co umożliwia identyfikację tożsamości. Badania z wykorzystaniem elektroencefalografii (EEG) wykazały charakterystyczną aktywność mózgu podczas rozpoznawania twarzy[11], a analiza aktywności neuronalnej pozwala na niezawodne dekodowanie twarzy generowanych przez sztuczną inteligencję[12].

W badaniach wykorzystywano również technikę ET[13]. Co istotne, twarze generowane przez SI często były postrzegane jako bardziej wiarygodne niż ich prawdziwe odpowiedniki. Uczestnicy zwracali uwagę na „nieregularne zęby" czy „asymetryczne oczy", które badacze określili jako artefakty charakterystyczne dla obrazów syntetycznych. Porównując twarze ludzkie z twarzami lalek, wykazano, że te pierwsze wywołują trwalsze reakcje neuronalne.

[10] S. Lyu, H. Farid, *How realistic is photorealistic?*, „IEEE Transactions on Signal Processing" 2005, vol. 53, issue 2, s. 845–850, https://doi.org/10.1109/TSP.2004.839896; Y. Wang, P. Moulin, *On discrimination between photorealistic and photographic images*, [w:] *IEEE. International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2006, https://doi.org/10.1109/ICASSP.2006.1660304; J.F. Lalonde, A.A. Efros, *Using color compatibility for assessing image realism*, [w:] *IEEE. International Conference on Computer Vision*, IEEE, 2007, https://doi.org/10.1109/ICCV.2007.4409107.

[11] J.P. Tauscher, S. Castillo, S. Bossey, M. Magno, *EEG-Based Analysis of the Impact of Familiarity in the Perception of Deepfake Videos*, [w:] *IEEE. International Conference on Image Processing (ICIP)*, IEEE, 2021, https://doi.org/10.1109/ICIP42928.2021.9506082; E. Twardoch-Raś, *Co widzą sieci neuronowe? Strategie widzenia maszynowego w projektach artystycznych opartych na technikach rozpoznawania i analizy twarzy*, „Kultura Współczesna" 2023, nr 1 (121), https://doi.org/10.26112/kw.2023.121.03.

[12] M.L. Moshel, A.K. Robinson, T.A. Carlson, T. Grootswagers, *Are you for real? Decoding realistic AI-generated faces from neural activity*, „Vision Research" 2022, vol. 199, https://doi.org/10.1016/j.visres.2022.108079.

[13] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, J. van de Weijer, *Eye-tracking: a comprehensive guide to methods and measures*, Oxford University Press, 2015; R. Wawer, M. Wawer, *Wykorzystanie nowoczesnych technik komputerowych do pomiaru emocji na podstawie badania fotografii*, „Annales Universitatis Paedagogicae Cracoviensis. Studia De Cultura" 2011, t. 2, s. 49–57, [on-line:] https://studiadecultura.uken.krakow.pl/article/view/1572 – 22.11.2025.

Wprowadzano też czynniki motywacyjne, np. nagrody za poprawne rozpoznanie obrazów czy rywalizację między grupami. Inne badania analizowały różnice w percepcji twarzy między dziećmi a dorosłymi – dzieci częściej skupiały wzrok na ustach robotów (mechanicznych i humanoidalnych), dorośli natomiast koncentrowali się na oczach[14].

Przeprowadzenie badań

Eksperyment został przeprowadzony w specjalnie przygotowanym laboratorium. W badaniu wykorzystano narzędzia: GazePoint GP3 HD Eye Tracker 150 Hz, dedykowanego oprogramowania GazePoint Assistant oraz aplikację Ogama w wersji 5.1 do przygotowania projektu eksperymentu, gromadzenia danych i podstawowej analizy wyników. Użyto komputera stacjonarnego z dwoma ekranami o przekątnej 27 cali (4K, 60Hz). Badacz i respondent zajmowali przeciwległe miejsca przy biurku, aby ograniczyć kontakt wzrokowy. Odległość badanego od monitora oraz eye trackera wynosiła około 1,5 metra[15]. Aby zapobiec ewentualnemu rozproszeniu uwagi, w laboratorium przebywali tylko prowadzący i badany[16].

Ze względu na wymogi aplikacyjne, badani poddawani byli podwójnej dziesięciopunktowej kalibracji. Oprogramowanie użyte w badaniu, zsynchronizowane z urządzeniem GazePoint prezentuje wynik kalibracji w punktach, gdzie niższa wartość oznacza dokładniejszą kalibrację. Po licznych testach przeprowadzonych przed eksperymentem ustalono, że wartość poniżej 100 punktów jest prawidłowa dla dokładnego pomiaru (wartości większe niż 2000 punktów występowały we wcześniejszych eksperymentach oraz testach w przypadku znaczących dysfunkcji wzroku, słabego oświetlenia lub zaburzonego procesu kalibracji).

Na początku eksperymentu respondentom wyświetlano krótką instrukcję, przypominającą o konieczności unikania zbędnych ruchów ciała, zwłaszcza poruszania głową. Ekspozycja slajdu trwała 5 sekund. Przed każdym wyświetlonym obrazem wybranym do eksperymentu, przez 5 sekund pojawiał się czarny slajd z białym punktem (50 x 50 pikseli), co miało ułatwić rozpoczęcie obserwacji kolejnego obrazu mniej więcej z tego samego obszaru ekranu (tzw. slajd zerujący). W tym czasie badacz zadawał

[14] E. Park, K.J. Kim, A.P. del Pobil, *Facial Recognition Patterns of Children and Adults Looking at Robotic Faces*, „International Journal of Advanced Robotic Systems" 2012, vol. 9, issue 1, https://doi.org/10.5772/47836.

[15] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, J. van de Weijer, dz. cyt.

[16] A.T. Duchowski, *Eye-tracking methodology*, Springer, Berlin 2017.

pytanie „Czy obraz, który zaraz zobaczysz jest prawdziwy czy wygenerowany przez sztuczną inteligencję?"

Po zakończeniu badania ET, respondent proszony był o wypełnienie elektronicznej ankiety, zawierającej wszystkie zdjęcia w wersji monochromatycznej i w innej kolejności niż w eksperymencie, w której mieli krótko uzasadnić, dlaczego dany obraz uznali za prawdziwy lub wygenerowany przez SI.

Przedmiot badań

W badaniu wykorzystano 11 (zob. Aneks) obrazów z 4 kategorii: twarze, koty, budynki i samochody. Podział ten wynikał z poziomu zaawansowania algorytmów GAN, które generują obrazy w wysokiej rozdzielczości, często trudne do odróżnienia od fotografii[17]. Autorzy, mający doświadczenie w pracy ze sztuką generatywną i dydaktyką, uznali taki wybór za trafny. Przyjęto następujące kryteria:

- Twarze: Nowoczesne modele SI skutecznie tworzą realistyczne wizerunki ludzi, przydatne w marketingu i projektowaniu bez naruszania praw wizerunkowych. Jednocześnie stanowią zagrożenie, m.in. jako narzędzie manipulacji (deep fake'i).
- Koty: SI ma trudności z realistycznym odwzorowaniem złożonych struktur, takich jak futro czy ruch zwierząt, co skutkuje artefaktami.
- Budynki i samochody: Grafika komputerowa w grach i filmach bywa trudna do odróżnienia od rzeczywistości. Dla osób obeznanych z takimi wizualizacjami zadanie rozpoznania obrazu może być łatwiejsze.

Każda kategoria zawierała jedno zdjęcie autentyczne i jedno lub dwa wygenerowane. Zdjęcia samochodów, budynków i kotów wygenerowano w serwisie PIXLR.com (na podstawie prostych haseł, np. „Audi, RS6, gray"). Obrazy twarzy pochodziły z ThisPersonDoesNotExist.com. Autentyczne zdjęcia twarzy i budynków pobrano z kolekcji Google (Creative Commons), zdjęcie samochodu – ze strony racecars24.pl, a kota – z archiwum własnego.

Respondenci

W badaniu wzięło udział 32 uczestników: 8 kobiet i 24 mężczyzn, w wieku 18–35 lat (średnia wieku 22,3; najmłodszy 19, najstarszy 33). Poziom wiedzy

[17] A. Tsagaris, A. Pampoukkas, *Create Stunning AI Art Using Craiyon, DALL-E and Midjourney: A Guide to AI-Generated Art for Everyone Using Craiyon*, DALL-E and Midjourney, Independently published, 2022; S. Lyu, H. Farid, dz. cyt., s. 845–850.

na temat obrazów generowanych przez SI oceniono następująco: 10 osób – podstawowy (brak praktyki w użytkowaniu GAN), 21 – średnio zaawansowany (sporadyczne generowanie grafik), 1 osoba – zaawansowany (częste użytkowanie i pogłębiona wiedza).

Wyniki badań

Rozpoznawalność

W trakcie przeprowadzanego eksperymentu uczestnikom zaprezentowano łącznie 11 kolorowych obrazów w przypadkowej kolejności, były to wizerunki: 3 twarzy, 3 kotów, 3 samochodów i 2 budynków. Spośród tego zbioru, 8 stanowiły grafiki wygenerowane przez SI (oznaczone jako *false* – F), natomiast 3 reprezentowały autentyczne fotografie (*true* – T), były to: twarz, kot i samochód.

W ogólnej analizie wyników, uczestnicy udzielili odpowiedzi F łącznie 157 razy, co stanowi odchylenie od oczekiwanego wyniku na poziomie 99 razy. Jednocześnie, odpowiedzi T zarejestrowano 86 razy, co również różni się od spodziewanej wartości, która wynosi 10 (tab. 1). W 3 przypadkach nie udzielono odpowiedzi na skutek niepewności badanych.

Tabela 1. Liczba rozpoznanych realnych i wygenerowanych fotografii

	Rozpoznano	Oczekiwano
T	86	96
F	157	256
Razem	243	352

Źródło: opracowanie własne.

Test niezależności chi-kwadrat wykazał istotny statystycznie wpływ rodzaju fotografii (sztuczna czy prawdziwa) na rozpoznawalność ($p < 0{,}035$, $a = 0{,}05$). W identyfikacji autentycznych zdjęć pomylono się 10 razy na 96 odpowiedzi (10%), a w przypadku wygenerowanych grafik błędne odpowiedzi stanowiły ponad jedną trzecią: 99 na 256 odpowiedzi (39%), co ilustruje tabela 1.

W analizie danych dotyczących identyfikacji autentyczności obrazów przez różne płcie (tab. 2) stwierdzono za pomocą testu Chi-kwadrat, że płeć nie ma wpływu na rozpoznawalność: dla obrazów realnych i sztucznych $p = 0{,}657$ oraz $p = 0{,}476$ odpowiednio dla poziomu istotności $a = 0{,}05$.

Tabela 2. Liczba rozpoznanych realnych i wygenerowanych fotografii przez kobiety i mężczyzn

	Rozpoznano		Oczekiwano	
	K	M	K	M
T	19	67	24	75
F	43	114	64	200
Razem	62	181	88	275

Źródło: opracowanie własne.

Jak już wspomniano, 10 z badanych osób określiło swój poziom wiedzy na temat obrazów generowanych przez SI jako podstawowy, 21 jako średnio-zaawansowany i jedna jako zaawansowany, którą ostatecznie przypisano do grupy średniej. Sprawdzono zależność rozpoznawalności obrazów od poziomu wiedzy wstępnej i również w tym przypadku nie zaobserwowano statystycznie istotnego związku pomiędzy tymi zmiennymi: $p = 0{,}554$ dla sztucznej grafiki $p = 0{,}359$ dla prawdziwej ($a = 0{,}05$). Jak wynika z tabeli 3, przedstawiającej średnią rozpoznawalności fotografii, najtrudniejszą kategorią w identyfikacji były twarze, najłatwiejszymi – samochody i budynki.

Tabela 3. Średnia rozpoznawalności według kategorii zdjęcia

Kategoria	samochody	budynki	Twarze	Koty
% rozpoznawalności	0,92	0,83	0,43	0,61

Źródło: opracowanie własne.

Kolejnym etapem (weryfikacyjnym), po wyświetleniu kolorowych obrazów, było zaprezentowanie uczestnikom czarno-białych miniatur tych samych grafik z pytaniami, które miały na celu zrozumienie przyczyn wyboru. Oceniano, czy uczestnicy byli w stanie skutecznie zapamiętać swoje wcześniejsze odpowiedzi. Mimo że z pozoru wydawało się, że 11 różnych obrazów może być zbyt dużym zbiorem aby weryfikacja była adekwatna, okazało się, że respondenci w większości zapamiętali swoje odpowiedzi. Błąd popełniono jedynie 37 na 315 razy (12%). Obejmował on przypadki udzielenia innej odpowiedzi niż za pierwszym razem oraz sytuacje, gdy uczestnicy skomentowali, że nie pamiętają swojej wcześniejszej oceny. W tej części badania z 33 uczestników, 28 oddało poprawnie wypełnione kwestionariusze.

Jeśli chodzi o realne fotografie, to podczas weryfikacji wszystkie 3 zostały w większości rozpoznane poprawnie, co w arkuszu kodowano jako „t-t". Jako najczęstsze powody takiej oceny, respondenci podawali m.in.:

- Dla twarzy: „ostre szczegóły"; „detale, gra świateł"; „oczy ludzkie"; „cera naturalna o strukturze"; „widać dobrze cienie, niedoskonałości".
- Dla kota: „tło naturalne, sierść"; „normalne zdjęcie, są niedoskonałości"; „nie widziałem nic nienaturalnego"; „sceneria wokoło, bardzo naturalne"; „tło, zbyt dużo elementów".
- Dla samochodu: „kołpak, normalne krawędzie"; „felga realna"; „odbicia, refleksy realne"; „światło"; „brak defektów, gra cieni naturalna".

Z analizy odpowiedzi w drugiej serii dotyczących zdjęć wygenerowanych przez SI (kodowane jako „f-f"), najczęściej poprawnie oceniano grafiki przedstawiające samochody – obie w ok. 89%. Jako powód wyboru wskazywano wygląd tablicy rejestracyjnej, logotypu, tła (linie budynku, trybun, proporcje) oraz brak kierowcy w ruchomym pojeździe.

Niewiele mniej, bo ok. 82% poprawnych odpowiedzi plus ok. 7% przypadków, gdzie F zmieniono na T oraz ok. 4% F przy ocenie czarno-białej grafiki, otrzymały ilustracje budynków. Respondenci komentowali rozmazane, nierówne kontury, grafiki wyglądające jak tapety stworzone w edytorach graficznych oraz nienaturalne cienie i roślinność.

Ilustracje kotów wygenerowane przez SI uzyskały ok. 46% i ok. 43% poprawnych odpowiedzi, a w obu przypadkach aż po ok. 28% osób oceniło wersję czarno-białą poprawnie, po błędnej ocenie wersji kolorowej. Jako powody podawano niesymetryczne oczy, *unoszące się* łapy, zbyt idealną sierść oraz tło. Niektórzy zwracali uwagę na niepewność, gdyż część elementów wyglądała naturalnie, a część nie, a także na możliwość zastosowania filtrów.

Najgorzej wypadły zdjęcia twarzy generowane przez SI (tab. 3). Pierwsze zdjęcie otrzymało 25% ocen poprawnych („f-f"), które wzrosły do 32% po pokazaniu wersji czarno-białej. Drugie zdjęcie miało ok. 21% poprawnych ocen, wzrastających do ok. 29% po drugiej ocenie. Z komentarzy wynika, że cechy uznawane przez niektórych za dowód prawdziwości, dla innych były sygnałem twórczości SI.

Respondenci, którzy oznaczyli grafiki jako T, komentowali, że zrobili to m.in. ze względu na: „ilość szczegółów"; „niedoskonałe"; „nie jest idealny"; „zęby nierówne"; „żadnych zniekształceń"; „wyraźne detale, światło naturalne"; i „zmarszczki, niedoskonałości". Podobne cechy wskazywały osoby, które oznaczyły zdjęcia jako F. Były to m.in. takie komentarze jak: „coś nie pasuje z oczami, usta"; „cera, odbicie światła w oczach"; „zęby, oczy"; „źrenice"; „prawa strona włosy po prawej stronie, końcówki zębów"; „kolczyki". Co ciekawe, w kilku przypadkach poprawnej odpowiedzi udzielonej na temat zdjęcia czarno-białego, treść komentarzy powtarzała się, np.: „oczy zbyt idealne"; „odbicie światła, kolczyki, brwi nieregularne – kobieta sobie na to nie pozwoli"; „refleksy w oczach nie są prawdziwe".

Z analizy udzielonych odpowiedzi na etapie weryfikacji wynika, że najwięcej błędów respondenci popełnili przy ocenie obu zdjęć twarzy wygenerowanych przez SI – kolejno 25% i ok. 21%. Jeśli chodzi o komentarze, to pojawiły się m.in. takie jak: „coś nie pasuje z oczami, usta, kolczyki"; „zęby"; „krzywa lewa źrenica, zarys ust"; „prawa strona, włosy po prawej stronie, końcówki zębów"; „cera, odbicie światła w oczach"; „źrenice".

Percepcja wizualna

Ścieżki wzroku dobrze ilustrują sposób, w jaki poszukiwano artefaktów na zdjęciach wyświetlanych przez 5 sekund. Ponieważ skanowanie wzrokiem zazwyczaj obejmowało całą powierzchnię grafiki, mapa uwagi, działająca na zasadzie spłaszczenia (algorytm Gaussa) różnic pomiędzy danymi pomiarowymi czasu lub liczby fiksacji w newralgicznych miejscach, powodowała znaczne zredukowanie informacji i nie nadawała się do wizualnej analizy percepcji[18]. Podstawowe charakterystyki liczbowe ET obliczano poprzez uśrednianie liczby fiksacji na serii slajdów dla badanych zgrupowanych najpierw według płci, a następnie poziomu zaawansowania. Testy Manna-Whitneya (na skutek nierównolicznych grup) nie wykazały jednak statystycznie istotnych różnic dla płci (μ_K = 15,6 i μ_M = 12,9, p = 0,175, a = 0,05), jak i poziomu wiedzy (μ_p = 13,8 i μ_z = 13,5, p = 0,675, a = 0,05, jedną osobę z wiedzą zaawansowaną w celu uproszczenia obliczeń ostatecznie dołączono do grupy średniozaawansowanych).

Poniżej przedstawiono przykłady ścieżek wzroku będących wizualizacją danych z eye-trackera, ukazujących, gdzie i w jakiej kolejności uczestnicy badania kierowali wzrok na ekran. Zaprezentowane przykłady odnoszą się do całej grupy badawczej i najlepiej ilustrują strategie poszukiwawczo-analityczne respondentów. Jak można zauważyć w tabeli 4, analiza wizualna zdjęć budynków zawiera sprawdzenie linii perspektywy, głównych osi konstrukcyjnych (np. pionów kolumn).

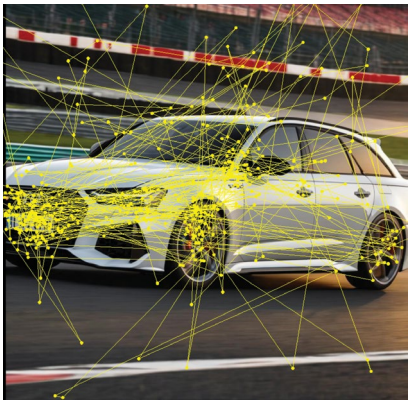
Klucze percepcyjne z etapu weryfikacji są wyraźnie widoczne na zdjęciach samochodów (tab. 5). Badani najpierw zwracają uwagę na elementy identyfikujące model auta, takie jak logotyp, reflektory, lusterko oraz główne atrybuty techniczne – koła, klamki, maska. Wyodrębnienie obszarów, np. kół i maski, oraz analiza statystyk lokalnych pozwoliły zauważyć różnice w percepcji osób z wiedzą podstawową i zaawansowaną. Pierwsi dłużej skupiają się na kołach, drudzy – na masce, gdzie znajdują się kluczowe identyfikatory pojazdu.

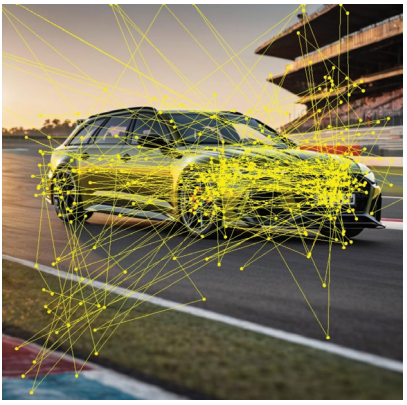
[18] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, J. van de Weijer, dz. cyt.

Tabela 4. Ścieżki wzroku zdjęć budynków (zdjęcia wygenerowane przez SI)

Źródło: opracowanie własne.

Tabela 5. Ścieżki wzroku zdjęć samochodów (trzecie zdjęcie jest realne)





Źródło: opracowanie własne.

Żeby udowodnić, iż różnice percepcji wzrokowej prezentowanych obrazów pomiędzy grupami nie są przypadkowe należało przeprowadzić test Studenta. Pokazał on, że różnice średnich długości trwania fiksacji dla obu grup są statystycznie istotne: $t(2) = 9{,}802$, $p = 0{,}010$, $a = 0{,}05$.

Zaobserwowano również, iż kobiety i mężczyźni poświęcają uwagę w różnym stopniu tym dwóm obszarom. Dlatego przeanalizowano je za pomocą narzędzia AOI (Area Of Interest). Migracje spojrzeń koło–maska mężczyzn były stanowczo bardziej intensywne i częstsze niż kobiet. Test Studenta wykazał statystycznie istotne różnice pomiędzy tymi podgrupami na podstawie znormalizowanej liczby fiksacji: $t(5) = 3{,}708$, $p = 0{,}014$, $a = 0{,}05$.

Klasyczny wzór percepcji twarzy zdefiniowany przez A.L. Yarbusa[19] jako odwrócony trójkąt w kontekście grafik generowanych przez SI wymaga rozszerzenia. Oprócz oczu, nosa i ust, wzrok badanych zahacza także o uszy, często zawierające ważne szczegóły – zamiast trójkąta powstaje wielokąt (tab. 6).

Zauważono różnice w analizie oczu między użytkownikami z wiedzą podstawową a (średnio-)zaawansowaną. Pierwsi równomiernie rozkładali uwagę na oboje oczu, drudzy skupiali się głównie na jednym, zwykle lewym oku twarzy wygenerowanej. Przy rozpoznawaniu twarzy realnej badani patrzyli na oboje oczu, niezależnie od poziomu wiedzy. Można przypuszczać, że jedno wybrane oko może dostarczyć informacji o realności obiektu. Statystycznie istotne różnice (potwierdzone testem Studenta) zaobserwowano także w sposobie skanowania twarzy przez kobiety i mężczyzn. Dominującym kierunkiem analizy twarzy przez mężczyzn był dół–góra (usta–czoło–włosy), czego u kobiet nie zaobserwowano. Średnie liczby relatywnych migracji wzroku w pionie wyniosły: $\mu_K = 8{,}0$ i $\mu_M = 13{,}8$, $p = 0{,}043$, $a = 0{,}05$.

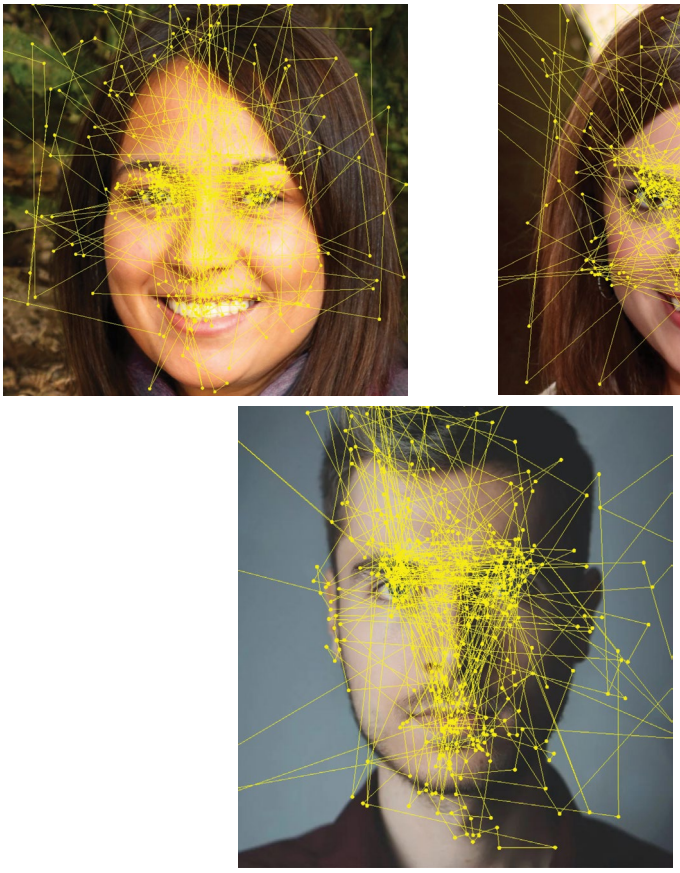
Podobnie postrzegane są wizerunki kotów: uwaga skupia się na pyszczku i uszach, a łapy stają się ważnym wyznacznikiem realizmu, pokazując pośrednio ciężar zwierzęcia. Badani zwracają też uwagę na krawędzie sylwetki, ponieważ w grafice komputerowej najtrudniej wyrenderować (in. zwizualizować) sierść, a artefakty najłatwiej dostrzec na granicy sierść – otoczenie. Jeśli kot nie zajmuje całego kadru, analizowane jest też otoczenie w poszukiwaniu realnych detali, np. wzorków na ceracie.

Choć wygenerowane wizerunki kotów zidentyfikowano poprawnie w 40%, a prawdziwe zdjęcie w 100%, nie zaobserwowano istotnych różnic w ścieżkach i mapach ET obu grafik. Natomiast wykryto różnice percepcyjne między kobietami a mężczyznami. Kobiety wykonywały mniej fiksacji, równomiernie rozłożonych w różnych obszarach zdjęcia, a u mężczyzn

[19] A.L. Yarbus, *Eye Movements and Vision*, Springer, Boston 1967.

stwierdzono większy rozrzut liczby fiksacji, potwierdzony testem wariancji Fishera: $p = 0{,}043$, $a = 0{,}05$.

Tabela 6. Gazeploty zdjęć twarzy (trzecie zdjęcie jest realne)



Źródło: opracowanie własne.

Podsumowanie

Celem badania była ocena postrzeganej autentyczności fotografii generowanych przez sztuczną inteligencję poprzez ich porównanie z rzeczywistymi zdjęciami. Analizie poddano cztery kategorie obiektów: twarze ludzkie, koty, samochody oraz budynki. Uczestnicy proszeni byli o ocenę obrazów pod względem ich realności, wskazując cechy, które miały decydujący wpływ na ich werdykt.

Największy poziom trudności odnotowano w przypadku kategorii „twarze". Przy ocenie obrazów syntetycznych respondenci identyfikowali typowe artefakty, takie jak nieprawidłowości w obrębie oczu, uszu czy uzębienia. W odniesieniu do fotografii rzeczywistych większość badanych dokonywała trafnej oceny, wskazując na obecność szczegółów, naturalne oświetlenie oraz widoczne niedoskonałości jako wskaźniki autentyczności.

Tabela 7. Gazeploty zdjęć kotów (trzecie zdjęcie jest realne)

Źródło: opracowanie własne.

W drugiej pod względem trudności kategorii – kotów – większość respondentów poprawnie oceniła prawdziwe zdjęcie, wskazując naturalne tło i sierść jako ważne cechy. Niektórzy zwracali uwagę na niesymetryczne oczy i pozycję łap, sugerujące nieprawdziwość. Oceniając ilustracje SI kotów, zauważono trudności – choć część respondentów poprawnie je identyfikowała jako sztuczne, aż 8 osób oceniło poprawnie wersję czarno-białą

po błędnej ocenie kolorowej. Najczęstsze powody oceny to niesymetryczne oczy, łapy „unoszące się w powietrzu", zbyt idealna sierść oraz nienaturalne tło. Niektórzy wyrażali niepewność, zauważając, że niektóre elementy wyglądały naturalnie, inne sztucznie, a także, że mogły być użyte filtry, wpływające na ocenę.

Analiza grafik samochodów (tab. 5) pokazuje, że najczęściej były poprawnie rozpoznawane. Wskazywano na wygląd tablic rejestracyjnych, logotypy, tło (linie budynków, trybuny), proporcje samochodu oraz brak kierowcy w ruchu. Respondenci mieli też niewielkie problemy z oceną grafik budynków. W przypadku zdjęć SI zwracano uwagę na rozmazane, nierówne kontury sugerujące nienaturalność i użycie filtrów. Grafiki często wyglądały jak malowane, a cienie i roślinność były sztuczne.

Różnice w postrzeganiu obrazów przez różne grupy badanych wykryto za pomocą ET. Użytkownicy (średnio-)zaawansowani skupiali uwagę na specjalistycznym szczególe – „kluczu" (np. jednym oku), co zwykle wystarczało do pewnej oceny. Trzy miary ET związane z fiksacjami (czas trwania) i sakkadami (średnia prędkość, długość połączeń) pozwoliły wykryć istotną statystycznie interakcję między percepcją bodźców a poziomem wiedzy (jedną osobę z zaawansowaną wiedzą przypisano do średniozaawansowanych dla uproszczenia) – zob. tabela 8.

Tabela 8. Statystycznie istotna zależność między bodźcem (slajdem) a poziomem wiedzy badanych

Miara ET	SS	DF1	DF2	MS	F	p
Długość połączeń między fiksacjami	76183149	10	320	7618315	3,103	0,000
Średnia prędkość sakkad	50,19	10	320	5,019	1,964	0,037
Średni czas trwania fiksacji	603581,2	10	320	60358,12	2,013	0,032

Źródło: opracowanie własne.

Tabela 9. Statystycznie istotny wpływ płci na strategię percepcji

Miara ET	SS	DF1	DF2	MS	F	p
Długość połączeń między fiksacjami	76183149	10	320	7618315	3,114823	0,001
Prędkość trajektorii spojrzenia	767805,134	10	320	76780,513	3,140	0,001

Źródło: opracowanie własne.

Zaawansowani przeglądali slajdy szybciej, wykonując krótsze przejścia między fiksacjami – „wiedzą, jak patrzeć". Mężczyźni, w odróżnieniu od kobiet, rozkładali uwagę mniej równomiernie. Miary ET, takie jak

prędkość spojrzenia i długość sakkad, różniły się statystycznie istotnie między płciami (tab. 9). Jeśli w procesie sprawdzania różnic pomiędzy grupami badanymi dysponujemy więcej niż jedną zmienną, to można zastosować wielowymiarową analizę wariancji – Manova. Wyniki testu zaprezentowane są w tab. 8–9.

Dla procesu percepcji twarzy zaobserwowano tendencję do wertykalnego skanowania: usta–oczy–włosy. Pełne rozpoznanie zdjęcia rzeczywistego nie wiązało się z nietypowymi zachowaniami percepcyjnymi. Najwięcej trudności sprawiały obrazy kotów, prawdopodobnie z powodu perfekcyjnej sierści („F") i silnych reakcji emocjonalnych. Autorzy postulują, by fotografie kotów traktować jako osobny temat badań.

Konkluzje i dyskusja

Wnioskiem z analizy jest, że zdolność rozpoznawania obrazów generowanych przez SI w dużej mierze zależała od rodzaju obiektu. Grafiki samochodów były najczęściej poprawnie oceniane, podczas gdy twarze ludzkie najczęściej błędnie, co wskazuje na trudności SI w wiarygodnym odwzorowaniu cech ludzkich. Obrazy czarno-białe mogły wpływać na poprawność oceny, sugerując, że pewne cechy łatwiej rozpoznać przy podkreślonym kontraście czy uwydatnieniu detali.

Z ankiety i sondażu wynikało, że respondenci oceniali m.in. realizm obiektów i scenerii, spójność szczegółów względem otoczenia, harmonię i asymetrię. Osoby z wiedzą zaawansowaną w grafice generatywnej poszukiwały artefaktów typowych dla algorytmów GAN. Mimo to zdarzały się nadinterpretacje: cechy uważane przez jednych za dowód autentyczności, dla innych wskazywały na generację SI. Takie zachowania są typowe dla eksperymentów rozpoznawania obrazów SI[20]. Można przypuszczać, że towarzyszy temu nadmierna podejrzliwość, zwłaszcza gdy użytkownicy nie poszerzają praktycznej wiedzy o możliwościach SI, które się ciągle rozwijają. W związku z tym dalsze badania i edukacja w zakresie wykrywania manipulacji cyfrowych są kluczowe dla radzenia sobie z wyzwaniami generowania obrazów przez SI.

Strategia poszukiwania „wpadek SI" analizowana przez ET pokrywała się z głównymi obserwacjami z sondażu. Wzmocniona uwaga dotyczyła kluczowych elementów, takich jak rysy twarzy, pyszczki kotów, logotypy samochodów czy linie budynków, a następnie otoczenia. ET nie ujawnił subtelnych strategii kognitywnych, jak porównywanie ostrości planów czy analiza symetrii, które zaawansowani respondenci stosowali automatycznie. Natomiast metryki ET i testy statystyczne pozwoliły wykryć różnice

[20] M.L. Moshel, A.K. Robinson, T.A. Carlson, T. Grootswagers, dz. cyt.,

w percepcji mężczyzn i kobiet oraz osób o różnym poziomie wiedzy. Ankieta weryfikacyjna okazała się pomocna w badaniu mechanizmów rozpoznawalności zdjęć generatywnych.

Warto podkreślić, że analiza nie wykazała istotnych różnic statystycznych między płciami i poziomem wiedzy w rozpoznawaniu autentyczności obrazów. Konieczne jest jednak powtórzenie eksperymentu na zróżnicowanych grupach wiekowych, płciowych oraz o różnym poziomie zaawansowania w kontekście praktycznego użycia SI. Nowym celem badawczym mogłoby być zrozumienie, jak różnorodność grupy kształtuje wyniki identyfikacji obrazów.

Istotne jest także unikanie nadinterpretacji cech mogących świadczyć o autentyczności lub sztuczności. Przy ocenie obrazów trzeba zachować krytyczne, analityczne podejście, ponieważ istnieje ryzyko błędnej identyfikacji. Użytkownicy powinni być świadomi, że percepcja obrazów generowanych przez SI różni się indywidualnie, np. w zależności od płci, doświadczenia czy wiedzy. To sugeruje odmienne tendencje w rozpoznawaniu autentyczności w różnych grupach demograficznych. W miarę rozwoju SI kluczowe jest, aby użytkownicy regularnie aktualizowali wiedzę o nowych metodach wykrywania i oceny obrazów generowanych przez sztuczną inteligencję, co zwiększy trafność ocen[21].

Ograniczenia

Pierwszym ograniczeniem badania była niewielka liczba uczestników, co mogło wpływać na jakość diagnozy regresji. Jednak próba 30 osób jest uznawana za wystarczającą w badaniach ET prowadzonych w psychologii eksperymentalnej i poznawczej[22].

Uczestnicy tworzyli jednorodną grupę etniczną i byli głównie młodzi. Biorąc pod uwagę rosnącą popularność produktów opartych na sieciach GAN wśród młodych użytkowników (część autorów prowadzi zajęcia z projektowania graficznego i interakcji człowiek–komputer), badanie może przyczynić się do weryfikacji metod oraz wskazania nowych problemów w komunikacji i edukacji związanej z SI.

Kolejnym ograniczeniem był stosunkowo ograniczony i słabo skategoryzowany zestaw obrazów. Warunki ekspozycji bodźców mogłyby być bardziej zaawansowane, a jednocześnie dostosowane do możliwości

[21] J. van Hees i inn., *Human perception of art in the age of artificial intelligence*, „Frontiers in Psychology" 2025, vol. 15, https://doi.org/10.3389/fpsyg.2024.1497469.

[22] H.Y.Z. Wang, X. Wang, *Expertise differences in cognitive interpreting: A meta-analysis of eye-tracking studies across four decades*, „Wiley Interdisciplinary Reviews: Cognitive Science" 2024, vol. 15, issue 1, https://doi.org/10.1002/wcs.1667.

percepcyjnych człowieka. Autorzy planują w przyszłości zastosować większy i bardziej zróżnicowany zestaw bodźców, aby zapewnić kompleksowe testowanie uczestników.

Rekomendacje

Na podstawie analizy wyników eksperymentu oraz zgromadzonej wiedzy autorzy formułują rekomendacje dla użytkowników końcowych i twórców algorytmów SI w kontekście rozwoju sztuki generatywnej:

- Świadomość użytkowników: kluczowe jest budowanie świadomości dotyczącej typowych błędów w obrazach generowanych przez SI, zwłaszcza w odwzorowaniu ludzkich twarzy i drobnych detali. Użytkownicy powinni rozwijać kompetencje w rozpoznawaniu artefaktów generatywnych oraz stosować metody oceny autentyczności, takie jak analiza symetrii i spójności tła z pierwszym planem.
- Przejrzystość systemów generatywnych: ważne jest wprowadzenie mechanizmów oznaczania treści generowanych przez SI – np. za pomocą tagów, metadanych czy znaków wodnych – aby jasno informować o wykorzystaniu algorytmów w procesie tworzenia lub modyfikacji obrazu.
- Edukacja i zaangażowanie użytkowników: obok zabezpieczeń kluczowy jest komponent edukacyjny, angażujący użytkowników w ocenę generowanych treści. Proponuje się umożliwienie zgłaszania nienaturalnych elementów na obrazach, co wspierałoby dalsze doskonalenie narzędzi SI.
- Dobór poziomu szczegółowości: mimo postępów w sztuce generatywnej wciąż występują problemy z odpowiednim doborem szczegółów. Zarówno zbyt mała, jak i nadmierna liczba elementów, zwłaszcza na drugim i trzecim planie, potęguje nienaturalność obrazów.
- Generowanie elementów chronionych prawnie: w przypadku obiektów rzeczywistych, takich jak logotypy czy znaki towarowe, pojawiają się błędy logiczne i odstępstwa. Maskowanie lub rozmycie tych elementów działa podobnie jak anonimizacja, prowadząc do pikselozy w danych obszarach.
- Priorytety dla twórców SI: należy skupić się na poprawie generowania ludzkich twarzy, gdzie obecne modele mają trudności, np. z dolnym rzędem zębów czy palcami dłoni. Ważne jest zwrócenie uwagi na detale, które użytkownicy uznają za kluczowe, takie jak szczegóły twarzy, subtelne wzory czy linie architektoniczne.

Finansowanie

Badania zostały przeprowadzone w ramach projektu BITSCOPE o nr 2021/03/Y/ST6/00002 finansowanego przez Narodowe Centrum Nauki w ramach programu CHIST-ERA IV, który otrzymał dofinansowanie na podstawie Umowy Finansowej nr 857925 w ramach Programu finansowania badań naukowych i innowacji Unii Europejskiej Horyzont 2020.

Bibliografia

AI or Not – AI Detection for Truth Seekers, [on-line:] https://www.aiornot.com – 14.11.2025.

Duchowski A.T., *Eye-tracking methodology*, Springer, Berlin 2017.

Fallis D., *The epistemic threat of deepfakes*, „Philosophy & Technology" 2021, vol. 34, s. 623–643, https://doi.org/10.1007/s13347-020-00419-2.

Goodfellow I.J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., *Generative Adversarial Networks*, [w:] *Advances in Neural Information Processing Systems*, 2014, vol. 2, https://doi.org/10.48550/arXiv.1406.2661.

Holmqvist K., Nyström M., Andersson R., Dewhurst R., Jarodzka H., van de Weijer J., *Eye-tracking: a comprehensive guide to methods and measures*, Oxford University Press, Oxford, 2015.

Jacques J., *How to Identify AI-generated and Fake Images*, 2023, https://jacquesjulien.com/identify-fake-images/ – 16.05.2025.

Kamali N., Nakamura K., Chatzimparmpas A., Hullman J., Groh M., *How to Distinguish AI-Generated Images from Authentic Photographs*, Northwestern University, 2024, https://doi.org/10.48550/arXiv.2406.08651.

Karras T., Aila T., Laine S., Lehtinen J., *Progressive growing of GANs for improved quality, stability, and variation*, [w:] *International Conference on Learning Representations*, 2018, [on-line:] https://www.researchgate.net/publication/320707565 – 18.11.2025.

Khanna N., Chiu G.T.C., Allebach J.P., Delp E.J., *Forensic techniques for classifying scanner, computer generated and digital camera images*, [w:] *IEEE. International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2008, https://doi.org/10.1109/ICASSP.2008.4517944.

Lalonde J.F., Efros A.A., *Using color compatibility for assessing image realism*, [w:] *IEEE. International Conference on Computer Vision*, IEEE, 2007, https://doi.org/10.1109/ICCV.2007.4409107.

Lyu S., Farid H., *How realistic is photorealistic?*, „IEEE Transactions on Signal Processing" 2005, vol. 53, issue 2, s. 845–850, https://doi.org/10.1109/TSP.2004.839896.

Moshel M.L., Robinson A.K., Carlson T.A., Grootswagers T., *Are you for real? Decoding realistic AI-generated faces from neural activity*, „Vision Research" 2022, vol. 199, https://doi.org/10.1016/j.visres.2022.108079.

OpenAI, *ChatGPT (wersja GPT-4)* [Model językowy], 2025, [on-line:] https://chat.openai.com/ – 16.05.2025.

Patel D.M., *Artificial Intelligence & Generative AI for Beginners: The Complete Guide (Generative AI & Chat GPT Mastery Series)*, Independently published, 2023.

Park E.,. Kim K.J, del Pobil A.P., *Facial Recognition Patterns of Children and Adults Looking at Robotic Faces*, „International Journal of Advanced Robotic Systems" 2012, vol. 9, issue 1, https://doi.org/10.5772/47836.

Smołucha D., *Eye-tracking in Cultural Studies*, „Perspektywy Kultury" 2019, t. 27, nr 4, s. 169–183, https://doi.org/10.35765/pk.2019.2704.12.

Tauscher J.P., Castillo S., Bossey S., Magno M., *EEG-Based Analysis of the Impact of Familiarity in the Perception of Deepfake Videos*, [w:] *IEEE. International Conference on Image Processing (ICIP)*, IEEE, 2021, https://doi.org/10.1109/ICIP42928.2021.9506082.

How to spot AI-generated images, Techoist, 24.01.2013, [on-line:] https://www.youtube.com/watch?v=zqRcjbft3zg – 16.05.2025.

Twardoch-Raś E., *Co widzą sieci neuronowe? Strategie widzenia maszynowego w projektach artystycznych opartych na technikach rozpoznawania i analizy twarzy*, „Kultura Współczesna" 2023, nr 1 (121), https://doi.org/10.26112/kw.2023.121.03.

Tsagaris A., Pampoukkas A., *Create Stunning AI Art Using Craiyon, DALL-E and Midjourney: A Guide to AI-Generated Art for Everyone Using Craiyon*, DALL-E and Midjourney, Independently published, 2022.

Hees J. van, Grootswagers T., Quek G.L., Varlet M., *Human perception of art in the age of artificial intelligence*, „Frontiers in Psychology" 2025, vol. 15, https://doi.org/10.3389/fpsyg.2024.1497469.

Wawer R., Wawer M., *Wykorzystanie nowoczesnych technik komputerowych do pomiaru emocji na podstawie badania fotografii*, „Annales Universitatis Paedagogicae Cracoviensis. Studia De Cultura" 2011, t. 2, s. 49–57, [on-line:] https://studiadecultura.uken.krakow.pl/article/view/1572 – 22.11.2025.

Wang Y., Moulin P., *On discrimination between photorealistic and photographic images*, [w:] *IEEE. International Conference on Acoustics, Speech, and Signal Processing*, IEEE, 2006, https://doi.org/10.1109/ICASSP.2006.1660304.

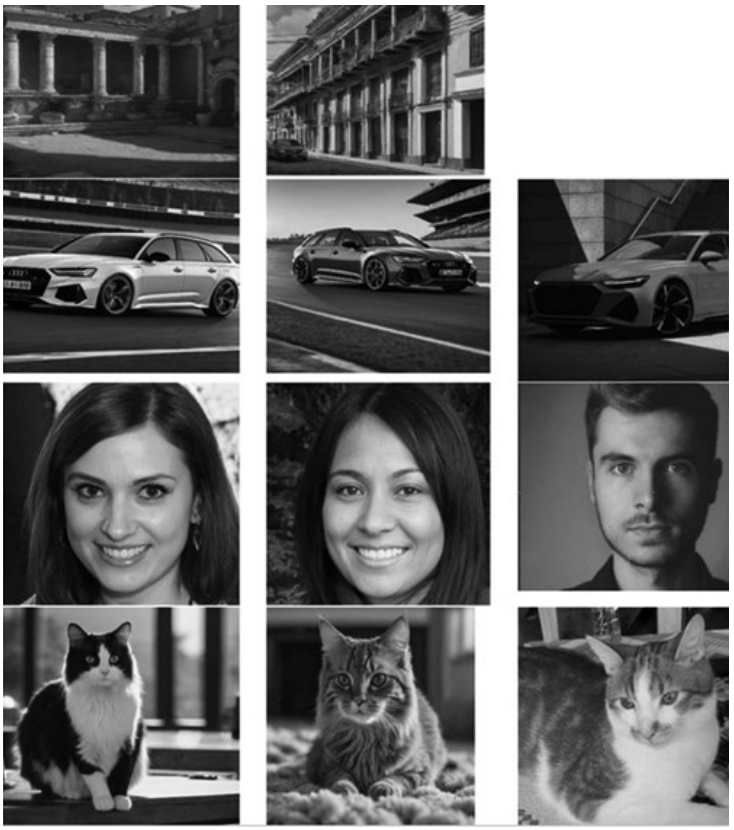
Wang H.Y.Z., Wang X., *Expertise differences in cognitive interpreting: A meta-analysis of eye-tracking studies across four decades*, „Wiley Interdisciplinary Reviews: Cognitive Science" 2024, vol. 15, issue 1, https://doi.org/10.1002/wcs.1667.

Yarbus A.L., *Eye Movements and Vision*, Springer, Boston 1967.

Xiang J., *On generated artistic styles: Image generation experiments with GAN algorithms*, „Visual Informatics" 2023, vol. 7, issue 4, s. 36–40, https://doi.org/10.1016/j.visinf.2023.10.005.

Aneks

Zestaw obrazów wykorzystanych w badaniu (w wersji monochromatycznej). Kategorie tematyczne są ułożone wierszami: (od góry) architektura, samochody, twarze, koty. Pierwsze 2 kolumny – to obrazy SI, 3 – zdjęcia realistyczne.



Streszczenie

Teza i cel: W ostatnich latach zauważalnie wzrósł poziom fotorealizmu obrazów generowanych przez algorytmy sztucznej inteligencji. Celem artykułu jest ocena realizmu tych obrazów poprzez ich porównanie z rzeczywistymi fotografiami czterech kategorii obiektów: twarzy ludzkich, kotów, samochodów i budynków.
Metodologia: Badanie miało charakter eksperymentalny i zostało przeprowadzone z wykorzystaniem techniki eye-trackingu, jako eksperyment wspomagany ankietą oraz testem. Respondenci oceniali autentyczność prezentowanych obrazów, a zebrane dane obejmowały analizy fiksacji, czasu spojrzeń oraz wyniki

kwestionariuszy uzupełniających. Wykorzystano testy statystyczne w celu identyfikacji różnic między grupami uczestników.
Wyniki: Najtrudniejsze do prawidłowego rozpoznania okazały się obrazy przedstawiające twarze ludzkie, natomiast najłatwiejsze – samochody i budynki. Analiza wzorców spojrzeń ujawniła różnice zależne od płci i poziomu wiedzy respondentów. Osoby posiadające większe doświadczenie w grafice generatywnej częściej koncentrowały się na artefaktach typowych dla obrazów tworzonych przez sztuczną inteligencję .
Wnioski: Percepcja realizmu obrazów generowanych przez sztuczną inteligencję jest zależna od doświadczenia odbiorcy oraz rodzaju przedstawionego obiektu. Wyniki badania wskazują na potrzebę dalszej eksploracji procesów poznawczych towarzyszących ocenie autentyczności obrazów oraz formułują rekomendacje dla twórców i użytkowników algorytmów generatywnych w kontekście rozwoju współczesnej sztuki wizualnej.

Słowa kluczowe: obrazy SI, grafiki GAN, eye tracking, okulografia, percepcja wizualna, chat GPT

Perception of artificially generated images: Towards understanding the viewer

Abstract

Thesis and Aim: In recent years, the level of photorealism in images generated by artificial intelligence algorithms has increased significantly. This article aims to assess the realism of such images by comparing them with real photographs across four categories of objects: human faces, cats, cars, and buildings.
Methodology: The study had an experimental character and was conducted using an eye-tracking technique, supported by a questionnaire and a test. Participants evaluated the authenticity of presented images, while collected data included fixation analyses, gaze durations, and complementary questionnaire results. Statistical tests were used to identify differences between participant groups.
Results: Human faces proved to be the most challenging category to distinguish correctly, whereas cars and buildings were the easiest. Analysis of gaze patterns revealed variations related to participants' gender and level of expertise. Individuals with greater experience in generative graphics focused more often on artifacts typical of AI-generated imagery.
Conclusions: The perception of realism in AI-generated images depends on the viewer's experience and the type of object being depicted. The findings underscore the need for further investigation into the cognitive processes involved in evaluating image authenticity and offer recommendations for creators and users of generative algorithms in the context of contemporary visual art development.

Keywords: AI images, GAN graphics, eye tracking, visual perception, chat GPT