

KATARZYNA JASIŃSKA

Trenowanie sztucznej inteligencji a naruszenie praw autorskich. Aspekty dowodowe

Artificial Intelligence Training and Copyright Infringement: Evidentiary Aspects

The inspiration for this paper was a lawsuit against OpenAI and related companies filed in 2023 in the United States District Court of the Northern District of California, San Francisco Division. The plaintiffs (authors) alleged, among other things, copyright infringement by using their novels to train AI without their consent. The purpose of this article is to assess, through the prism of Polish law, the plaintiffs' chances of success in a similar dispute hypothetically taking place in Poland. However, due to the limited scope of this paper, its scope was limited to evidentiary issues. This is an interesting aspect, which has been neglected in the doctrine so far, and at the same time it is undoubtedly crucial from a practical point of view. The conclusions of the considerations made will be related to the postulates presented regarding the solution at the legal level of some copyright problems related to the training of AI.

KATARZYNA JASIŃSKA, doktor nauk prawnych, Krakowska Akademia im. A.F. Modrzewskiego w Krakowie

ORCID – 0000-0002-1746-1095, e-mail: kjasinska@afm.edu.pl

SŁOWA KLUCZOWE: sztuczna inteligencja, branża kreatywna, trenowanie sztucznej inteligencji, naruszenie praw autorskich, postępowanie dowodowe, domniemania faktyczne, postępowania pomocnicze, wnioski o udzielenie informacji, wnioski o zabezpieczenie środka dowodowego, wnioski o wydanie lub wyjawienie środka dowodowego

KEYWORDS: artificial intelligence, creative industry, artificial intelligence training, copyright infringement, evidentiary proceedings, factual presumptions, ancillary proceedings, request for information, application for securing evidence, application for issuance or disclosure of evidence

1 | Wprowadzenie

Ostatnie lata to lata niezwykle dynamicznego rozwoju tzw. sztucznej inteligencji^[1]. Jest ona obecna niemal w każdej dziedzinie naszego życia, prywatnego i zawodowego, a nawet publicznego. Narzędzia wykorzystujące się są wykorzystywane m. in. w medycynie^[2], w wymiarze sprawiedliwości^[3]

¹ Termin „sztuczna inteligencja” stworzył John McCarthy w 1956 r. Andreas Kaplan i Michael Haenlein zdefiniowali to pojęcie jako zdolność systemu do prawidłowego interpretowania danych pochodzących z zewnątrz, nauki na ich podstawie oraz wykorzystywania tej wiedzy, aby wykonywać określone zadania i osiągać cele poprzez elastyczne dostosowanie – Andreas Kaplan, Michael Haenlein, „Siri, Siri in my Hand, who’s the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence” *Business Horizons*, nr 1 (2019): 15–25.

² *AI in Clinical Medicine. A Practical Guide for Healthcare Professionals*, red. Michael F. Byrne (Hoboken: John Wiley and Sons Ltd, 2023).

³ Katarzyna Jasińska, „Judge’s E-Assistant in Common Courts”, [w:] *Artificial Intelligence and Human Rights*, red. Laura Miraut Martin, Mariusz Załucki (Madryt: Dykinson Editorial, 2021), <https://www.dykinson.com/libros/artificial-intelligence-and-human-rights/9788413778174/>), 77–90; Katarzyna Jasińska, „Czy »referendarza z Lublina« może zastąpić »sędzia robot«?” *Studia Prawnicze. Rozprawy i materiały*, nr 1 (2020), 89–101. <https://sp.ka.edu.pl/numery/2020-1/studia-prawnicze-rim-2020-1-jasinska.pdf>.

i w sądownictwie polubownym^[4], w architekturze^[5], w mass mediach^[6] (w tym w prasie, radio i telewizji^[7], a także filmie^[8] i książkach^[9]) oraz w innych dziedzinach sztuki^[10], w tym w sztuce użytkowej^[11], w muzyce^[12], a także w modzie^[13].

Aktualnie oczy nie tylko specjalistów z wielu branż, ale także przeciętnych obywateli zwrócone są w stronę ChatGPT. Jest to narzędzie opracowane przez firmę OpenAI, korzystające z technologii SI, a konkretnie z sieci neuronowych i uczenia maszynowego. Jest to generatywny system

⁴ Katarzyna Szczudlik, *Rozwiązywanie sporów (arbitraż) a Sztuczna Inteligencja, Założenia do strategii AI w Polsce. Plan działań Ministerstwa Cyfryzacji* (Warszawa: Ministerstwo Cyfryzacji, 2018). <http://cli.re/LNXz9N> [dostęp: 23.09.2020]; Beata Więzowska-Czepiel, „Sztuczna inteligencja w arbitrażu – wsparcie dla arbitrów w procesie podejmowania decyzji” *Studia Prawnicze. Rozprawy i Materiały*, nr 2 (2019): 57–65. <http://hdl.handle.net/11315/28194>.

⁵ Henriette Bier, „Implications of Robotics and AI in Architecture”, [w:] *Disruptive Technologies: The Convergence of New Paradigms in Architecture*, red. Philippe Morel, Henriette Bier (Cham: Springer International Publishing, 2023), 3–5.

⁶ Wedle Encyklopedii pwn środki masowego przekazu, mass media, to m. in. prasa, radio, telewizja, także film, książki, nagrania muzyczne czy Internet, <https://encyklopedia.pwn.pl/encyklopedia/mass%20media.html>. [dostęp: 11.08.2023].

⁷ Michael Graßl, Jonas Schützeneder, Klaus Meier, „Artificial intelligence as a tool of assistance A scientific and practical perspective on AI in journalism” *Journalism Research*, nr 1 (2022): 3–24.

⁸ L Kavitha, „Copyright Challenges in the Artificial Intelligence Revolution: Transforming the Film Industry from Script to Screen” *Trinity Law Review*, nr 1 (2024): 2–8; Abhijit Ghosh, „Artificial Intelligence as an Innovation in the Film Industry” *Research Inspiration*, nr 3 (2023): 12–16; Md Monirul Islam, „Artificial Intelligence in Indian Films: Anukul and AI Ethics” *Short Film Studies*, nr 2 (2022): 183–192.

⁹ Książka napisana przez komputer prawie zdobyła nagrodę literacką. <http://booknews.pl/?p=4169>. [dostęp: 11.08.2023].

¹⁰ James Hutson, Martin Lang, „Content Creation or Interpolation: AI Generative Digital Art in the Classroom” *Metaverse*, nr 1 (2023): 1–13.

¹¹ Yu-Hsu Lee, Chun-Yao Chiu, „The Impact of AI Text-to-Image Generator on Product Styling Design” [w:] *Human Interface and Management of Information*, red. Yumi Ashaki, Hirohiko Mori (Cham: Springer International Publishing, 2021), 502–515.

¹² „Le Monde: Mają miliony fanów. Jest jeden haczyk. Nie istnieją”, oprac. M. Kucharczyk. <https://wydarzenia.interia.pl/raport-media-zagraniczne/news-le-monde-maja-miliony-fanow-jest-jeden-haczyk-nie-istnieja,nId,6952972>. [dostęp: 11.08.2023].

¹³ Heidi Härkönen, „Fashion Piracy and Artificial Intelligence—does the New Creative Environment Come with New Copyright Issues” *Journal of Intellectual Property Law & Practice*, nr 3 (2020): 163–172.

SI, czyli taki, który umożliwia tworzenie nowych treści, a nie tylko analizuje istniejące dane lub działa na ich podstawie^[14]. Aktualnie oparte jest o model multimodalny GPT-4 (multimodalność oznacza, że oprócz tekstu może przyjmować inne typy danych – obrazy, dźwięki, animacje czy filmy video, co odróżnia go od poprzednika, który jest modelem językowym). Ten triumfalny pochód SI zdają się jednak zakłócać problemy z prawami autorskimi. Mają one różne odsłony, a jedna z nich to problem z możliwym naruszeniem praw autorskich podmiotów, których utwory są wykorzystywane do trenowania (nawet jeśli efekt pracy SI sam w sobie nie przejmuje elementów twórczych utworów wykorzystanych do trenowania).

Przyczynkiem do napisania artykułu stał się pozew pisarzy przeciwko OpenAI i powiązanim z nią firmom, który wpłynął do Sądu Okręgowego Stanów Zjednoczonych Północnego Dystryktu Kalifornii, Oddział San Francisco w dniu 28.06.2023 r. (dalej „pозew”). Powodowie – pisarze Paul Tremblay, Mona Awad i inni – zarzucili m.in. naruszenie praw autorskich poprzez użycie do trenowania SI ich powieści, bez ich zgody^[15]. Warto zauważyć, że wykorzystanie efektów swojej pracy zakwestionowali także programiści i artyści wizualni^[16].

Sprzeciw artystów oraz podejmowanie prób walki z twórcami i dystrybutorami nowoczesnych narzędzi SI jest zrozumiałe z praktycznego punktu widzenia. Nie jest to tylko walka o idee i o poszanowanie praw autorskich, czy o jakość tego, co w przyszłości będziemy czytać czy oglądać. Dzięki swojej wspomnianej wyżej zdolności do kreowania zupełnie nowych treści, SI staje się zagrożeniem dla wielu twórców, już teraz pozbawiając ich pracy (mimo, że aktualne narzędzia SI nie stanowią jeszcze tzw. silnej sztucznej inteligencji, czyli posiadającej wszystkie atrybuty ludzkiego umysłu, jak m. in. kreatywności czy własna inicjatywa^[17]). Skalę zagro-

¹⁴ Danica Lo, „What is generative AI? Here’s how ChatGPT and artificial intelligence will change business” FAST COMPANY. <https://www.fastcompany.com/90826178/generative-ai>. [dostęp: 11.08.2023].

¹⁵ <https://lmlitigation.com/pdf/03223/tremblay-openai-complaint.pdf>. [dostęp: 11.08.2023]. Analogiczny pozew został złożony w tym sądzie w dniu 07.07.2023 r., <https://lmlitigation.com/pdf/03416/silverman-openai-complaint.pdf>. [dostęp: 11.08.2023].

¹⁶ Blake Brittain, „Lawsuits Accuse AI Content Creators of Misusing Copyrighted Work” REUTERS. <https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/>, 17 stycznia 2023 r. [dostęp: 11.08.2023].

¹⁷ W przedmiocie tzw. silnej SI, zwanej też ogólną, por. Martin V. Butz, „Towards Strong AI” *KI – Kunst Intelligenz*, 35 (2021): 91–101.

żenie dobrze ilustruje przykład opowiadania *Dzień, w którym komputer napisze powieść* (*Konpyuta ga shosetsu wo kaku hi*), które walcząc o nagrodę literacką Nikkei Hoshi Shinichi zakwalifikowało się do rywalizacji i przeszło też pierwszy etap selekcji. Równie wymowny jest fakt, że w ramach *self-publishingu* pojawia się tak dużo książek generowanych przez SI, że Amazon wprowadził ograniczenie do maksymalnie trzech zgłoszeń książek dziennie na jedno (!) konto^[18]. W lutym bieżącego roku poinformowano, że na Amazonie pojawiło się ponad 200 książek, które były tworzone albo współtworzone przez z ChatGPT^[19]. Problem został zauważony także przez niektórych tradycyjnych wydawców, którzy w umowach wydawniczych wymagają od twórców zobowiązań, że podczas pracy nie będą wspomagali się narzędziami SI.

W związku z realnym zagrożeniem w obronę praw twórców przed sztuczną inteligencją zaangażowało się wiele organizacji zrzeszających twórców^[20]. Jednak, póki co, najwięcej fermentu walka artystów ze sztuczną inteligencją wywołała w branży filmowej, a nie literackiej. To tam, ze względu na ogromne koszty produkcji filmowej i jej czasochłonność, oszczędność pieniędzy i czasu jest na wagę złota. „Zatrudnienie” SI jest tańsze niż zatrudnienie człowieka, który dodatkowo pracuje zdecydowanie wolniej. W związku z tym, że w branży filmowej sztuczną inteligencję

¹⁸ Grzegorz Kubera, „Ludzie nie czytają książek, za to piszą ich bardzo dużo. »Zalewają nas powieści generowane przez AI«” *Business Insider*, 1 października 2023 r. <https://businessinsider.com.pl/technologie/nowe-technologie/ksiazki-pisane-przez-ai-trafiaja-hurtowo-do-amazona-jest-reakcja-sklepu/xc74v2e>. [dostęp: 17.10.2023].

¹⁹ Ibidem.

²⁰ Twórcy na wojnie z AI. Pilna potrzeba regulacji. https://www.rp.pl/biznes/art38928151-tworcy-na-wojnie-z-ai-pilnie-potrzeba-regulacji?fbclid=IwAR0YcOrHxjdifspaD5wzDism399-dtXqxYjhLu65kZ1y7vN7_FAONO-PLQA. [dostęp: 12.08.2023]; An Open Letter to policy makers on Artificial Intelligence, <https://www.cisac.org/Newsroom/articles/global-creators-and-performers-demand-creative-rights-ai-proliferation>. [dostęp: 12.08.2023]; Artificial Intelligence and Creative Work, <https://www.creatorsrightsalliance.org/ai-and-creative-work>. [dostęp: 12.08.2023]; Artificial intelligence: practical steps for members. <https://www2.societyofauthors.org/2023/06/07/artificial-intelligence-practical-steps-for-members/>. [dostęp: 12.08.2023]; Recent developments in artificial intelligence have created a wave of excitement about its possible applications, but also genuine concern about the potential for writers' works to be exploited without recognition or compensation, <https://www.alcs.co.uk/news/our-principles-for-ai-and-authors#:~:text=We%2obelieve%20that%20AI%20has,effective%20and%20appropriate%20policy%20framework>. [dostęp: 12.08.2023].

można wykorzystać na każdym etapie produkcji, a nawet jeszcze przed jej rozpoczęciem (na etapie oceniania szans na sukces rynkowy planowanego przedsięwzięcia^[21], np. wykorzystując ScriptBook^[22]), to każdy twórca filmowy może czuć się zagrożony. Bunt przeciwko sztucznej inteligencji jako pierwsi wszczęli scenarzyści. Już zdarza się, że zamiast tworzyć scenariusz „od zera”, dostają jedynie do doszlifowania wstępny draft scenariusza, przygotowany przez si. Poza urażoną dumą często wybitnych scenarzystów, w oczywisty sposób negatywnie przekłada się to na wynagrodzenie za ich pracę^[23].

Celem artykułu jest ocena przez pryzmat polskiego prawa szans powodów na sukces w analogicznym sporze, hipotetycznie toczącym w się Polsce, jednak – w uwagi na ograniczone ramy artykułu – w zakresie ograniczonym tylko do aspektów dowodowych. Jest to aspekt niezwykle ciekawy, pomijany dotychczas w doktrynie, a jednocześnie bez wątpienia kluczowy z praktycznego punktu widzenia. Nawet jeśli powodom udałoby się przekonać orzekający w sprawie sąd w kluczowych kwestiach materialnych, a więc że trenowanie si stanowi wkroczenie w monopol praw autorskich, że nie znajdują w danym wypadku żadne przepisy o dozwolonym użytku, a może nawet, że trenowanie stanowi naruszenie praw autorskich osobistych, to nie będzie to miało żadnego znaczenia, jeśli nie uda im się udowodnić, że to właśnie ich konkretne utwory zostały w danym przypadku wykorzystane do trenowania konkretnego narzędzia wykorzystującego si.

Główny problem stanowi to, że o ile wydaje się możliwe zidentyfikowanie, który tekst został stworzony przez si, a który nie^[24], to już ustalenie, jakie dane zostały wykorzystane na etapie uczenia narzędzia, które dane tekst stworzyło, nie jest możliwe^[25].

²¹ Kiedy AI Zastępuje producenta, a algorytmy przewidują sukces filmu, <https://papaya.rocks/pl/news/kiedy-ai-zastepuje-producenta-a-algorytmy-przewidujaja-sukces>. [dostęp: 13.08.2023]; Jacek Samborski, *AI w produkcji filmowej*. <https://filmpro.com.pl/ai-w-produkcji-filmowej/>. [dostęp: 13.08.2023].

²² <https://www.scriptbook.io/#/>. [dostęp: 13.08.2023].

²³ Andrzej Krajewski, Najpopularniejsze seriale mogą zniknąć z winy sztucznej inteligencji. <https://gospodarka.dziennik.pl/praca/artykuly/8710449,ai-chatgpt-sztuczna-inteligencja-hollywood-strajk-praca-pracownik.html>. [dostęp: 12.08.2023].

²⁴ Por. na ten temat np. Chaka Chaka, „Detecting AI Content in Responses Generated by ChatGPT, YouChat, and Chatsonic: The Case of Five AI Content Detection Tools” *Journal of Applied Learning & Teaching*, 6 (2023): 1–11.

²⁵ Karina Kunc-Urbańczyk, „Uczenie maszynowe jako pole eksploatacji utworów”, [w:] *Prawo sztucznej inteligencji i nowych technologii*, red. Bogdan Fischer, Adam

2 | Stan faktyczny i kwestie dowodowe

2.1. Stan faktyczny

Duże modele językowe uczą się na podstawie ogromnych ilości danych, które są niezbędne dla ich trenowania i dalszego funkcjonowania. Dzięki nim potrafią podsumowywać, tłumaczyć, przewidywać i generować tekst oraz inne formy wyrazu^[26]. Do trenowania dużych modeli językowych używa się wielu rodzajów materiałów, a książki stanowią kluczową część zestawów danych szkoleniowych, ponieważ zawierają długie przykłady wysokiej jakości tekstów^[27]. Aktualnie twierdzi się, że duże modele językowe są zwykle szkolone na zbiorach danych wystarczająco dużych, aby uwzględnić prawie wszystko, co zostało zamieszczone w Internecie^[28]. Niestety, z punktu widzenia twórcy, dochodzącego swoich praw z tytułu naruszenia praw autorskich poprzez użycie jego utworu do trenowania si, informacja ta nie ma dużej wartości. Jednocześnie bardziej precyzyjne dane nie są powszechnie dostępne. Autorzy wymienionych pozwów złożonych w imieniu pisarzy powołali się na informacje ujawnione przez OpenAI w 2018 r. i 2020 r. w dwóch artykułach. W pierwszym z nich wskazano, że do uczenia modelu językowego GPT-1 użyto zestawu danych BooksCorpus, który zawiera ponad 7000 unikalnych, niepublikowanych książek z różnych gatunków, w tym przygodowych, fantastyka i romanse^[29]. W drugim wskazano enigmatycznie, że 17% danych szkoleniowych GPT3 to zbiory nazwane Books1 i Books2 oraz anglojęzyczna Wikipedia^[30]. Przede wszystkim jednak są to zbiory nazwane Common Crawl (60%) i WebText2 (22%)^[31]. OpenAI nigdy nie ujawniła, jakie książki wchodziły w skład wymienionych

Pązik, Marek Świerczyński (Warszawa, 2021, LEX/el), pkt 4.1.

²⁶ Angie Lee, *What Are Large Language Models Used For?*. <https://blogs.nvidia.com/blog/2023/01/26/what-are-large-language-models-used-for/>. [dostęp: 13.08.2023].

²⁷ Alec Radford et. al., *Improving Language Understanding by Generative Pre-Training*, (2018): 3, 4. <https://www.mikecaptain.com/resources/pdf/GPT-1.pdf>. [dostęp: 13.08.2023].

²⁸ Lee, *What Are Large Language Models Used For?*

²⁹ Radford et. al., *Improving Language Understanding by Generative Pre-Training*, (2018): 4.

³⁰ Tom B. Brown et. al., *Language Models are Few-Shot Learners*, 2020 r., 9, <https://arxiv.org/pdf/2005.14165.pdf>. [dostęp: 13.08.2023].

³¹ Ibidem.

zbiorów, ani nie podała nawet ogólnej informacji na temat bazy materiałów użytych do trenowania GPT-4^[32]. Autorzy pozwu szacują, że Books1 zawiera ok. 63,000 tytułów, a Books2 ok. 294,000 tytułów^[33].

Również i te dane okazałyby się niewystraszająca w przypadku pozwu twórcy przekonanego, że dany tekst został wytworzony przez SI, która była trenowana z wykorzystaniem jego utworu.

W stanie faktycznym, który stał się inspiracją do napisania artykułu, sytuacja twórców wydaje się być lepsza. Zakwestionowali oni bowiem przygotowanie przez ChatGPT podsumowania książek ich autorstwa. Przykładowe polecenie brzmiało: „Summarize in detail the first part of »The Cabin at the End of the World« by Paul Tremblay”. Na skutek postawienia tego rodzaju zadań SI sporządziła szczegółowe, choć nie pozbawione błędów, posumowanie książek^[34]. Pozornie zasadny wydaje się więc wniosek, że skoro SI była w stanie przygotować szczegółowe streszczenie utworu, to wcześniej musiała zostać zasilona jego treścią. Ze względu jednak na fakt, że ChatGPT zasilany jest wieloma treściami, które obejmują na przykład wypowiedzi internautów dyskutujących o książkach, czy choćby ich streszczenia przygotowane przez ludzi, informujące o ich treści^[35], to z punktu widzenia dowodowego sytuacja powodów może nie być tak komfortowa, jak się wydaje.

2.2. Domniemanie faktyczne

Moim zdaniem na gruncie polskiej procedury cywilnej z pomocą powodom mógłby przyjść art. 231 k.p.c. dotyczący domniemań faktycznych.

³² OpenAI (2023), *GPT-4 Technical Report*, 2. <https://cdn.openai.com/papers/gpt-4.pdf>. [dostęp: 13.08.2023]; ogólnie w przedmiocie zasobów GPT-4 por. Brian John Aboze, *GPT-3.5 vs. GPT-4: Unveiling the Power of the Next-Generation Language Models*. <https://deepchecks.com/gpt-3-5-vs-gpt-4-unveiling-the-power-of-the-next-generation-language-models/>. [dostęp 17.10.2023].

³³ Pkt 31 pozwu.

³⁴ Pkt 31 pozwu.

³⁵ <https://qz.com/openai-lawsuit-copyright-books-chatgpt-generative-ai-1850609334>. [dostęp: 14.08.2023]. W zasadzie na tę samą okoliczność zwrócono uwagę w pkt 41 pozwu („The summaries get some details wrong. These details are highlighted in the summaries. This is expected, since a large language model mixes together expressive material derived from many sources”).

Stanowi on, że sąd może uznać za ustalone fakty mające istotne znaczenie dla rozstrzygnięcia sprawy, jeżeli wniosek taki można wyprowadzić z innych ustalonych faktów (domniemanie faktyczne). Przepis ten pozwala na dokonanie ustaleń faktycznych polegających na uznaniu faktów mających istotne znaczenie dla rozstrzygnięcia sprawy, jeżeli wniosek taki można wyprowadzić z innych ustalonych faktów^[36].

W analizowanym przypadku podstawą domniemania byłby fakt dokonania szczegółowego streszczenia powieści przez ChatGPT, a faktem domniemanym byłoby zapoznanie się z powieściami przez ChatGPT. Uważam, że analogiczne podejście można by zastosować w przypadku, gdyby dane narzędzie poproszone o stworzenie tekstu w stylu konkretnego autora przygotowało wytwór, który faktycznie stylowi temu, by odpowiadał (choć i tutaj zapewne w praktyce będzie można spotkać się z zarzutem, że wiedzę o danym stylu si pozyskała w wyniku analizy innych treści opisujących ten styl).

Jest to domniemanie wzruszalne, które może być wzruszone przez wykazanie nieprawidłowości przyjętego rozumowania: brak ustalenia faktu będącego podstawą wniosku, lub niezasadność – w świetle wiedzy i doświadczenia życiowego – wyprowadzonego wniosku (np. z powołanie się na wspomniany wyżej argument, że si z innych źródeł niż sama powieść mogła pozyskać wiedzę o jej treści). Domniemanie faktyczne nie prowadzi do zmiany ciężaru dowodu w procesie i do jego podważenia wystarczy dowód przeciwny, a zbędne jest przeprowadzenie dowodu przeciwności^[37]. Rozwiązanie takie wydaje się być uczciwym w sytuacji, kiedy to pozwany dysponuje wszelkimi danymi co do wykorzystanej do trenowania bazy materiałów, a powodowie nie mają do nich dostępu.

³⁶ Wyrok SN z 19.02.2002 r., IV CKN 718/00, LEX nr 54362.

³⁷ Lech Morawski, *Domniemania a dowody prawnicze* (Toruń: Wydawnictwo UMK, 1981), 104; wyrok SN z 17.05.2007 r., III CSK 429/06, LEX nr 274129; wyrok SN z 30.03.2000 r., III CKN 811/98, LEX nr 51364; Małgorzata Sieńko, „Komentarz do artykułu 231”, [w:] *Kodeks postępowania cywilnego. Komentarz aktualizowany. Tom I. Art. 1-477(16)*, red. Małgorzata Manowska (LEX/el., 2022), teza 3.

2.3. Tzw. postępowania pomocnicze

Rozważając możliwości wykorzystania w tego typu sprawach tzw. postępowań pomocniczych przewidzianych w postępowaniu odrębnym w sprawach własności intelektualnej, przede wszystkim należy wskazać, że twórcom nie przyjdzie w takich przypadkach z pomocą instytucja wezwania do udzielenia informacji. Zgodnie bowiem z art. 479¹³§1 k.p.c. uprawniony może się domagać udzielenia określonej kategorii informacji, a mianowicie informacji o pochodzeniu i sieciach dystrybucji towarów lub usług, szczególnie wskazanych w art. 479¹⁵§1 k.p.c. Z pewnością nie należy do nich informacja o danych wykorzystanych do trenowania si. Teoretycznie natomiast właściwym środkiem prawnym wydaje się być wniosek o zabezpieczenie dowodu z art. 479⁹⁸ k.p.c. oraz wniosek wydanie lub wyjawienie dowodu z art. 479¹⁰⁶ k.p.c.

Zgodnie art. 479⁹⁸ k.p.c. sąd udziela zabezpieczenia środka dowodowego na wniosek uprawnionego, który uprawdopodobnił zarówno roszczenie, jak i interes prawny w zabezpieczeniu, przy czym interes prawny w zabezpieczeniu środka dowodowego istnieje, gdy brak żądanego zabezpieczenia uniemożliwia lub poważnie utrudnia przytoczenie lub udowodnienie istotnych faktów, jak również, gdy zachodzi ryzyko zniszczenia środka dowodowego lub opóźnienie w uzyskaniu środka dowodowego może uniemożliwić lub poważnie utrudnić osiągnięcie celu postępowania dowodowego, lub gdy z innych przyczyn zachodzi potrzeba stwierdzenia istniejącego stanu rzeczy. Co ważne, ta instytucja prawna ma nie tylko cel zachowawczy, ale także wydobywczy. Oznacza to, że – inaczej niż zabezpieczenie dowodu z art. 309 k.p.c. – może zmierzać do uzyskania środków dowodowych, potwierdzających fakty istotne dla rozstrzygnięcia sprawy, a więc pozwala uzyskać dostęp do informacji o faktach dotyczących naruszenia i umożliwić ocenę zasadności i zakresu roszczeń, których zamierza dochodzić poszkodowany (wniosek ten może zostać złożony także przed wszczęciem postępowania)^[38].

Z kolei na podstawie art. 479¹⁰⁶ k.p.c. powód może domagać się, by pozwany wyjawiał lub wydał środek dowodowy, którym dysponuje, w szczególności (a więc nie jest to katalog zamknięty) dokumenty bankowe,

³⁸ Katarzyna Jasińska, Zbigniew Pinkalski, Justyna Rasiewicz, Anna Sokołowska-Ławniczak, *Metodyka pracy pełnomocnika w sprawach cywilnych z zakresu prawa własności intelektualnej* (Warszawa: Wolters Kluwer, 2021, LEX/el), pkt 2.3.1.

finansowe lub handlowe służące ujawnieniu i udowodnieniu faktów. W założeniu wniosek ten ma umożliwić powodowi zdobycie wiedzy o faktach dotyczących naruszeń jego prawa, których nie jest w stanie uzyskać samodzielnie, co w konsekwencji pozbawia go możliwości ochrony praw^[39]. W związku z tym jednak, że wniosek ten może zostać złożony dopiero wraz z wszczęciem postępowania, to w praktyce będzie on służył raczej zebraniu dowodów obrazujących skalę naruszenia lub rozmiar roszczeń pieniężnych, a nie zebraniu dowodów na sam fakt naruszenia^[40]. Natomiast informacja, jakie dane zostały wykorzystane do trenowania SI, w tym czy był wśród nich utwór pokrzywdzonego, niewątpliwie służyłaby ustaleniu, czy do naruszenia w ogóle doszło. Skorzystanie z wniosku w tym celu jest oczywiście formalnie możliwe, ale wiąże się dla powoda z ryzykiem.

Przesłanką uwzględnienia obu wniosków jest uprawdopodobnienie roszczenia. W sprawach analizowanego typu jedynym argumentem autorów wskazującym na to, że doszło do naruszenia, byłby wspomniany już wyżej fakt, że SI była w stanie przygotować szczegółowe streszczenie utworu, co może (choć nie musi) świadczyć o tym, że wcześniej musiała zostać zasilona jego treścią. Pomocniczo można byłoby odwołać się do wskazanych powyżej danych, z których wynika, jak ogromną ilością danych, w szczególności dostępnych w Internecie, zostało zasilone dane narzędzie, o ile w konkretnym wypadku (w odniesieniu do danego narzędzia) takie dane w ogóle są dostępne. Trudno stwierdzić, czy w praktyce takie rozumowanie przekonałoby orzekający w sprawie sąd. Odnotować należy, że w sprawach innego rodzaju (tzn. gdy w wyniku działania SI nie powstaje streszczenie czy choćby wytwór „w stylu [...]”) nawet o uprawdopodobnienie będzie ciężko; wówczas pokrzywdzonym pozostawałoby odwoływanie się do ewentualnych dostępnych informacji, z których wynikałoby, że dane narzędzie zostało zasilone niemal całą zawartością Internetu.

W przypadku każdego ze wskazanych wyżej wniosków największym wyzwaniem będzie w mojej ocenie określenie środka dowodowego, którego zabezpieczenia albo wyjawienia lub wydania żąda uprawniony/powód i uprawdopodobnienie, że przeciwnik/pozwany dysponuje tym środkiem.

Przede wszystkim wskazać należy, że w doktrynie wyrażony został pogląd, że z istoty omawianych środków prawnych wynika, że zabezpieczeniu

³⁹ Maciej Barczewski, Olga Zinkiewicz-Będźmirowska w: *Postępowania odrębne. System Postępowania Cywilnego*, t. VI, red. Anna Machnikowska (Warszawa: C.H. Beck, 2022), 1012.

⁴⁰ Jasińska et. al., *Metodyka*, pkt 2.4.3.

oraz wyjawieniu lub wydaniu mogą podlegać jedynie rzeczowe środki dowodowe (dokumenty, próbki produktów etc.), w tym także materiały w postaci elektronicznej^[41]. To co w mojej ocenie mogłoby być przydatne powodowi, to dostęp do bazy danych przeciwnika/pozwanego, w której znajdowałyby się wykorzystane do trenowania materiały. Przedstawiciele doktryny nie zamykają drogi do takiego wykorzystania omawianego wniosku, dopuszczając uzyskanie w wyniku jego uwzględnienia dostępu do bazy danych^[42]. Problem podstawowy jednak tkwi w tym, że taka baza danych może powstać w związku z funkcjonowaniem narzędzi SI, ale wcale nie musi. Teoretycznie, dane, którymi dane narzędzie SI jest zasilane, mogą zostać pobrane z sieci, wpuszczone do algorytmu a następnie usunięte. Zapewne gdyby faktycznie miało to miejsce, to byłoby dużym „marnotrawstwem” danych, ale niezależnie od tego, czy ono ma miejsce, czy nie, taki sposób działania mógłby być skutecznym sposobem na unieszkodliwienie omawianych środków procesowych. ChatGPT, którego funkcjonowania dotyczy spór, i który jest inspiracją do niniejszych rozważań, jest złożoną siecią neuronową, której struktury nie jest jawna; OpenAI nie ujawniło jeszcze sposobu, w jaki obsługuje swoją infrastrukturę szkoleniową i wdraża modele^[43]. Tak więc jakakolwiek polemika z twierdzeniami przeciwnika/pozwanego byłaby skazana na niepowodzenie. Dodatkową trudność stanowi fakt, że w typowym przypadku modele nie są tworzone „od zera”, a są „dotrenowywane”^[44].

3 | Podsumowanie

Jak wynika z powyższego, tylko pozornie sytuacja prawna twórców, których utwory zostały streszczone przez SI, rysuje się całkiem dobrze. W praktyce jednak, z dużym prawdopodobieństwem w polskich (i zapewne nie tylko)

⁴¹ Jasińska et. al., *Metodyka*, pkt 2.4.3 i 2.3.2.

⁴² Agnieszka Gołaszewska, „Komentarz do artykułu 479106”, [w:] *Kodeks postępowania cywilnego. Komentarz. Art. 506–1217*, t. 11, red. Agnieszka Gołaszewska, Tomasz Szanciło (Warszawa: C.H. Beck, 2023), Legalis, Nb 7 i 10; Kinga Flaga-Gieraszyńska, Andrzej Zieliński, *Kodeks postępowania cywilnego. Komentarz* (Warszawa, 2022, Legalis), kom. do art. 479106, Nb 3.

⁴³ <https://www.hyro.ai/glossary/gpt-3/>. [dostęp: 17.10.2023].

⁴⁴ Ibidem.

realiach prawnych nawet ich szczególna sytuacja wiązałaby się ogromnymi trudnościami dowodowymi. W innych przypadkach ta sytuacja byłaby jeszcze trudniejsza.

Należy więc uznać za zasadne postulaty wprowadzenia jakiegoś rodzaju środka prawnego, który umożliwiłby twórcom przynajmniej częściową rekompensatę szkód, które ponoszą ze względu na działania SI, stanowiącej dla nich poważną konkurencję, a jednocześnie intensywnie czerpiącą z ich pracy^[45]. Ważne jest wprowadzenie regulacji nakazujących oznaczania wytworów SI^[46]. Miałoby to istotne znaczenie w nierównej walce ludzi z algorytmami, jako że wiedza o zaangażowaniu człowieka w proces artystyczny ma pozytywny wpływ na ocenę sztuki przez odbiorców, którzy postrzegają sztukę jako doświadczenie specyficzne dla człowieka^[47].

Natomiast podmioty tworzące i wykorzystujące narzędzia działające w oparciu o SI w swoich działaniach, a przede wszystkim w relacjach z twórcami, powinny uwzględniać ryzyko „podcięcia gałęzi, na której siedzą”. Jak zauważyli prawnicy pisarzy, narzędzia SI opierają się na danych stworzonych przez ludzi, „Ich systemy całkowicie zależą od ludzkiej kreatywności. Jeśli doprowadzą do bankructwa ludzkich twórców, wkrótce doprowadzą do bankructwa samych siebie”^[48]. Bardziej sceptyczny zdaje się być w tej kwestii Ryszard Markiewicz, który nie wyklucza, że „korzystanie z utworów na potrzeby AI w przyszłości stanie się zbędne lub (co bardziej prawdopodobne) istotnie coraz bardziej ograniczone. AI będzie się opierać głównie na własnych wytworach lub utworach, co do których ochrona autorska wygasła”^[49].

Przede wszystkim jednak zawarte w artykule wywody, które niestety nie doprowadziły do żadnego pozytywnego z punktu widzenia twórców czy innych uprawnionych z autorskich praw majątkowych wniosku, są dowodem na ogromne znaczenie regulacji, które będą w stanie zapewnić

⁴⁵ W przedmiocie zgłaszanych w doktrynie propozycji i trudności związanych z ich realizacją por. Ryszard Markiewicz, „ChatGPT i prawo autorskie Unii Europejskiej” *Zeszyty Naukowe Uniwersytetu Jagiellońskiego PPWI*, nr 2 (2023): Lex/el., pkt 4.5.

⁴⁶ Postuluje to m. in. Markiewicz, ChatGPT, pkt 1.7.

⁴⁷ Lucas Bellaiche et. al., „Humans vs. AI: Whether and Why We Prefer Human-Created Compared to AI-Created Artwork” *Cognitive Research: Principles and Implications*, 42 (2023): 22–23.

⁴⁸ Pierwszy taki pozew przeciw ChatGPT. *Zyski ze „skradzionych tekstów i pomysłów*. <https://tvn24.pl/biznes/tech/sztuczna-inteligencja-ai-chat-gpt-pisarze-zlozyli-pozew-przeciw-openai-7212322>. [dostęp: 20.08.2023].

⁴⁹ Markiewicz, ChatGPT, pkt 4.5.

transparentność narzędzi wykorzystujących SI, w tym przede wszystkim w zakresie procesu ich uczenia.

Bibliografia

- Abhijit Ghosh, „Artificial Intelligence as an Innovation in the Film Industry” *Research Inspiration*, nr 3 (2023): 12–16.
- Above Brian John, *GPT-3.5 vs. GPT-4: Unveiling the Power of the Next-Generation Language Models*. <https://deepchecks.com/gpt-3-5-vs-gpt-4-unveiling-the-power-of-the-next-generation-language-models/>.
- Postępowania odrębne. System Postępowania Cywilnego*, t. VI, red. Anna Machnikowska. Warszawa: C.H. Beck, 2022.
- Bellaïche Lucas, Rohin Shahi, Martin Harry Turpin, Anya Ragnhildstveit, Shawn Sprockett, Nathaniel Barr, Alexander Christensen, Paul Seli, „Humans versus AI: Whether and Why we Prefer Human-Created Compared to AI-Created Artwork” *Cognitive Research: Principles and Implications*, 8 (2023): 1–22.
- Bier Henriette, „Implications of Robotics and AI in Architecture”, [w:] *Disruptive Technologies: The Convergence of New Paradigms in Architecture*, red. Morel Philippe, Bier, Henriette. 3–5, Cham: Springer International Publishing, 2023.
- Brittain Blake, „Lawsuits Accuse AI Content Creators of Misusing Copyrighted Work” *REUTERS*.
- Butz Martin V., „Towards Strong AI” *KI – Künst Intelligenz*, (2021): 91–101.
- AI in Clinical Medicine. A Practical Guide for Healthcare Professionals*, red. Michael F. Byrne. Hoboken: John Wiley and Sons Ltd, 2023.
- Chaka Chaka, „Detecting AI Content in Responses Generated by ChatGPT, YouChat, and Chatsonic: The Case of Five AI Content Detection Tools” *Journal of Applied Learning & Teaching*, nr 2 (2023): 1–11.
- Flaga-Gieruszyńska Kinga, Andrzej Zieliński, *Kodeks postępowania cywilnego. Komentarz*. Warszawa: C.H. Beck, 2022, Legalis.
- Gołaszewska Agnieszka, „Komentarz do artykułu 479¹⁰⁶”, [w:] *Kodeks postępowania cywilnego. Komentarz. Art. 506–1217*, t. II, red. Agnieszka Gołaszewska, Tomasz Szanciło. Warszawa: C.H. Beck, 2023, Legalis.
- Graßl Michael, Jonas Schützeneder, Klaus Meier, „Artificial Intelligence as a Tool of Assistance A Scientific and Practical Perspective on AI in Journalism” *Journalism Research*, nr 1 (2022): 3–24.

- Härkönen Heidi, „Fashion Piracy and Artificial Intelligence—does the New Creative Environment Come with New Copyright Issues?” *Journal of Intellectual Property Law & Practice*, nr 3 (2020): 163–172. <https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/>.
- Hutson James, „Content Creation or Interpolation: AI Generative Digital Art in the Classroom” *Metaverse*, nr 1 (2023): 1–13.
- Samborski Jacek, *AI w produkcji filmowej*. <https://filmpro.com.pl/ai-w-produkcji-filmowej/>.
- Jasińska Katarzyna, „Judge’s e-assistant in Common Courts”, [w:] *Artificial Intelligence and Human Rights*, red. Martin, Laura Miraut, Załucki, Mariusz. 77–90. Madryt Dykinson Editorial, 2021. <https://www.dykinson.com/libros/artificial-intelligence-and-human-rights/9788413778174/>.
- Jasińska Katarzyna, „Czy »referendarza z Lublina« może zastąpić »sędzia robot«?” *Studia Prawnicze. Rozprawy i materiały*, nr 1 (2020): 89–101.
- Jasińska Katarzyna, Zbigniew Pinkalski, Justyna Rasiewicz, Anna Sokołowska-Ławniczak, *Metodyka pracy pełnomocnika w sprawach cywilnych z zakresu prawa własności intelektualnej*. Warszawa: Wolters Kluwer, 2021.
- Kaplan Andreas, Michael Haenlein, „Siri, Siri in my Hand, who’s the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence” *Business Horizons*, 1 (2019): 15–25.
- Kavitha L. „Copyright Challenges in the Artificial Intelligence Revolution: Transforming the Film Industry from Script to Screen” *Trinity Law Review*, nr 1 (2024): 2–8.
- Kubera Grzegorz, „Ludzie nie czytają książek, za to piszą ich bardzo dużo. »Zalewają nas powieści generowane przez AI«” *Business Insider*, 1 października 2023.
- Kunc-Urbańczyk, Karina, „Uczenie maszynowe jako pole eksploatacji utworów”, [w:] *Prawo sztucznej inteligencji i nowych technologii*, red. Bogdan Fischer, Adam Pązik, Marek Świerczyński. Warszawa: Wolters Kluwer, 2021, Lex.
- Lee Angie, *What Are Large Language Models Used For?*. <https://blogs.nvidia.com/blog/2023/01/26/what-are-large-language-models-used-for/>.
- Lee Yu-Hsu, Chun-Yao Chiu, „The Impact of AI Text-to-Image Generator on Product Styling Design”, [w:] *Human Interface and Management of Information*, red. Yumi Ashaki, Mori Hirohiko. 502–515. Cham: Springer International Publishing 2021.
- Lo Danica, *What is Generative AI? Here’s how ChatGPT and Artificial Intelligence will Change Business*, *Fast Company*, 18 grudnia 2022. <https://www.fastcompany.com/90826178/generative-ai>.
- Postępowania odrębne. System Postępowania Cywilnego*, t. VI, red. Anna Machnikowska. Warszawa: C.H. Beck, 2022.

- Major, Małgorzata, *Koniec strajku amerykańskich scenarzystów. Kto wygrał, a kto przegrał?*. <https://www.wirtualnemedi.pl/artykul/koniec-strajk-scenarzy-sta-seriale-film-zarobki>.
- Markiewicz Ryszard, „ChatGPT i prawo autorskie Unii Europejskiej” *Zeszyty Naukowe Uniwersytetu Jagiellońskiego PPWI*, nr 2 (2023): LEX/el.
- Md Monirul Islam, „Artificial Intelligence in Indian Films: Anukul and AI Ethics” *Short Film Studies*, nr 2 (2022): 183-192.
- Morawski Lech, *Domniemania a dowody prawnicze*. Toruń: Wydawnictwo UMK, 1981.
- OpenAI (2023), *GPT-4 Technical Report*. <https://cdn.openai.com/papers/gpt-4.pdf>.
- Radford Alec, KarthikNarasimhan, TimSalimans, Ilya Sutskever, *Improving Language Understanding by Generative Pre-Training*, 2018. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language_unsupervised/language_understanding_paper.pdf.
- Sieńsko Małgorzata, „Komentarz do artykułu 231”, [w:] *Kodeks postępowania cywilnego. Komentarz aktualizowany*, t. 1, Art. 1-477(16), red. Małgorzata Manowska. Lex/el. 2022.
- Szczudlik Katarzyna, *Rozwiązywanie sporów (arbitraż) a Sztuczna Inteligencja, Założenia do strategii AI w Polsce. Plan działań Ministerstwa Cyfryzacji*. Warszawa: Ministerstwo Cyfryzacji, 2018, <http://cli.re/LNxz9N>.
- Więzowska-Czepiel Beata, „Sztuczna inteligencja w arbitrażu – wsparcie dla arbitrów w procesie podejmowania decyzji” *Studia Prawnicze. Rozprawy i Materiały*, nr 2 (2019): 57-65.

