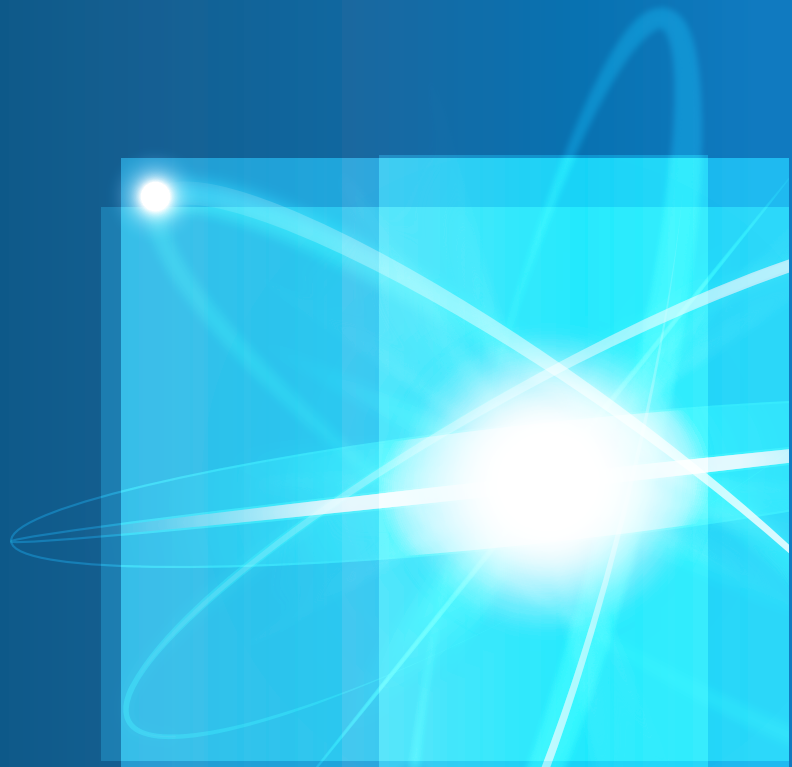


PME

2/2024

Prawo Mediów Elektronicznych

Kwartalnik Naukowy



CBiKE



Dostęp online:

<https://bibliotekacyfrowa.pl/publication/157980>

<https://repozytorium.uni.wroc.pl/publication/150803>

<https://www.pme.uni.wroc.pl>

DOI: 10.34616/150803

Prawo Mediów Elektronicznych

Kwartalnik Naukowy

pod redakcją prof. Jacka Gołaczyńskiego

Uniwersytet Wrocławski

<https://orcid.org/0000-0002-3295-7099>

i pod redakcją dr Marii Dymitruk (redaktor numeru 2)

Uniwersytet Wrocławski

<https://orcid.org/0000-0003-1003-9083>

2/2024

Wrocław 2024

Redakcja

Redaktor naczelny: prof. dr hab. Jacek Gołaczyński, Uniwersytet Wrocławski

Sekretarz redakcji: dr Rafał Skibicki, Uniwersytet Wrocławski

Członek redakcji: Agata Jałowicka, Uniwersytet Wrocławski

Rada Programowa

Przewodniczący: dr hab. Wojciech Wiewiórowski, Uniwersytet Gdański, EIOD

Członkowie:

r.pr. Włodzimierz Chróścik

sędzia Jacek Czaja, Naczelny Sąd Administracyjny

adw. Rafał Dębowski

dr hab. Kinga Flaga-Gieruszyńska, prof. nadzw. USz

dr hab. Włodzimierz Gromski, emerytowany prof. UWrocław

adw. Xawery Konarski

prof. Avv. Michele Angelo Lupoi, University of Bologna

prof. dr hab. Jacek Mazurkiewicz, Uniwersytet Zielonogórski

dr hab. Grzegorz Sibiga, prof. nadzw. Instytut Nauk Prawnych PAN

dr hab. Joanna Studzińska, prof. nadzw. Akademia Leona Koźmińskiego

prof. dr hab. Grażyna Szpor, Uniwersytet Kardynała Stefana Wyszyńskiego

prof. dr Andreas Wiebe, University of Goettingen

prof. dr hab. Krzysztof Wójtowicz, Uniwersytet Wrocławski

dr hab. Dariusz Szostek, prof. nadzw. Uniwersytet Śląski w Katowicach

dr hab. Piotr Stec, prof. nadzw. Uniwersytet Opolski

dr hab. Radim Polcak, prof. Uniwersytetu w Brnie

dr hab. Svetlana Fursa, Taras Shevchenko, National University of Kyiv

dr hab. Marlena Jankowska, prof. nadzw. UŚ

Recenzenci

prof. hab. dr Vytautas Nekrošius, Vilnius University

dr hab. Andrzej Adamski, prof. nadzw. UMK

prof. Zsolt Balogh, Corvinus University

prof. dr hab. Sławomir Cieślak, Uniwersytet Łódzki

prof. dr hab. Jacek Górecki, Uniwersytet Śląski

prof. em. dr Wolfgang Kilian, University of Hannover

prof. dr hab. Ryszard Markiewicz, Uniwersytet Jagielloński

dr hab. Marek Świerczyński, Uniwersytet Kardynała Stefana Wyszyńskiego

prof. Richard Warner Ph.D, Kent College of Law, Chicago

dr hab. Kazimierz Zgrzyzek, prof. nadz. UŚ

© Copyright by Wydział Prawa, Administracji i Ekonomii Uniwersytetu Wrocławskiego 2024

Czasopismo dofinansowane ze środków Ministerstwa Edukacji i Nauki

w ramach programu „Rozwój Czasopism Naukowych” (projekt nr RCN/SP/0553/2021/1).

Korekta (język polski): Sebastian Surrendra

Projekt i wykonanie okładki: Agata Jałowicka

Skład i opracowanie techniczne: Munda Maciej Torz

Wydawca

E-Wydawnictwo. Prawnicza i Ekonomiczna Biblioteka Cyfrowa.

Wydział Prawa, Administracji i Ekonomii Uniwersytetu Wrocławskiego

ISSN 2082-100X

Adres redakcji

Uniwersytet Wrocławski, Wydział Prawa, Administracji i Ekonomii,

Centrum Badań Problemów Prawnych i Ekonomicznych Komunikacji Elektronicznej

ul. Uniwersytecka 22/26, 51-145 Wrocław

e-mail: jacek.golaczynski@uwr.edu.pl

Wersją pierwotną (referencyjną) czasopisma jest wydanie elektroniczne

Spis treści

Jakub Jurowicz, Małgorzata Paja

Wykorzystywanie w biznesie treści generowanych przez sztuczną inteligencję
– praktyczne *case study* zarządzania ryzykiem prawnoautorskim 7

Jakub Jurowicz, Małgorzata Paja

Zwielokrotnianie utworów w procesach tworzenia i funkcjonowania dużych
modeli językowych 23

Daniel Kula

Web scraping na potrzeby szkolenia sztucznej inteligencji a sytuacja autora 41

Joanna Matczuk, Adrianna Pilecka

Systemy sztucznej inteligencji w procesach administracyjnych i w pracy urzędów patentowych na świecie – przegląd wybranych inicjatyw oraz analiza dalszego ich rozwoju w kontekście zasad etycznych dla sztucznej inteligencji.
Część 1 59

Jakub Jurowicz

Uniwersytet Jagielloński
Wydział Prawa i Administracji
ORCID: 0009-0002-7113-5992

Małgorzata Paja

Uniwersytet Jagielloński
Wydział Prawa i Administracji
ORCID: 0009-0003-0779-9636

Wykorzystywanie w biznesie treści generowanych przez sztuczną inteligencję – praktyczne *case study* zarządzania ryzykiem prawnoautorskim

Streszczenie

Tekst został poświęcony biznesowemu ryzykom wynikającym ze stosowania w przedsiębiorstwach sztucznej inteligencji (dalej: „AI”) – związanym z brakiem powstania lub brakiem skutecznego nabycia praw autorskich do generacji AI. Zawarta w tekście analiza przedstawia genezę powstawania takich ryzyk, związaną ze stopniem przygotowania przedsiębiorców do wdrożenia narzędzi AI w swoich przedsiębiorstwach, a następnie omawia wybrane przykłady takich ryzyk. Autorzy tekstu skoncentrowali się na omówieniu zagrożeń wynikających ze spornego statusu prawnego generacji AI na gruncie prawa autorskiego (w szczególności w branży IT) i praktycznych konsekwencjach braku nabycia praw autorskich do dzieł tworzonych przy wykorzystaniu AI – z uwzględnieniem orzecznictwa sądów krajów członkowskich UE w tym zakresie. Analiza ryzyk biznesowych została omówiona w szczególności w kontekście relacji pracownik–pracodawca oraz w kontekście relacji pracodawca–klient. Zawarta w tekście analiza obejmuje również omówienie obszarów zarządzania ryzykiem związanym ze stosowaniem narzędzi AI. W tekście omówiono również stosowane przez praktyków prawa rozwiązania mające na celu zabezpieczenie omówionych ryzyk, przedstawiając i analizując przykłady rozwiązań na gruncie prawa umów oraz przykłady działań zmierzających do odpowiedniego przygotowania przedsiębiorców oraz pracowników do odpowiedzialnego stosowania narzędzi AI.

Słowa kluczowe:

AI, utwór, ryzyko, przedsiębiorca, zarządzanie, biznes

1. Wstęp

Narzędzia AI są już powszechnie stosowane w biznesie, w tym w branży kreatywnej przez takie podmioty jak software house’y, agencje marketingowe lub biura

architektoniczne. Środowisko biznesowe poszukuje jak najprostszych rozwiązań, aby określić i zminimalizować ryzyka związane ze stosowaniem narzędzi AI i potencjalnym naruszeniem cudzych lub brakiem ochrony swoich praw autorskich.

Czy pomimo trwającej dyskusji na temat statusu generacji AI (outputu uzyskanego po wprowadzeniu promptu) można korzystać z AI w sposób bezpieczny, tj. niepowodujący zbyt dużego ryzyka naruszenia cudzych praw autorskich lub braku nabycia takich praw od swoich pracowników lub współpracowników? Uważamy, że tak, i w naszym artykule skupimy się na krótkiej analizie dostępnych aktualnie rozwiązań prawnych, pozwalających przedsiębiorcom świadomie zarządzać ryzykiem, jakie powoduje stosowanie narzędzi AI w sferze praw autorskich.

2. Brak powszechnej świadomości ryzyka związanego ze stosowaniem narzędzi AI

Perspektywa optymalizacji czasu, kosztów, zasobów w związku z korzystaniem z narzędzi AI zachęca przedsiębiorców do wdrażania takich rozwiązań. Z obserwacji, jakich dokonujemy w naszej praktyce zawodowej, wynika, że w większości przypadków w kalkulacji zysków i strat korzystania z narzędzi AI przedsiębiorcy w ogóle nie uwzględniają analizy ryzyka prawnoautorskiego lub całkowicie je marginalizują. Często wśród przedsiębiorców pojawia się przekonanie, że korzyści, jakie dają narzędzia AI, są tak duże, że kwestię praw autorskich można zignorować.

Dobrze rozumiemy biznesowe benefity, jakie wynikają z korzystania z narzędzi AI. Strategicznym błędem, z perspektywy prowadzenia biznesu, jest jednak korzystanie z nich w sposób przypadkowy i nieskoordynowany. Zarządzanie ryzykiem to kluczowa cecha decydująca o rynkowym sukcesie. W taki sam sposób przedsiębiorca powinien podchodzić do korzystania z narzędzi AI.

W obserwowanej przez nas praktyce obrotu gospodarczego pojawiają się twierdzenia, że skoro racjonalni przedsiębiorcy powszechnie korzystają z narzędzi AI, to być może ochrona prawnoautorska generacji AI nie jest konieczna. Ta sama praktyka pokazuje nam jednak, że wyzwania i problemy powstałe w wyniku nieuporządkowanego korzystania z narzędzi AI pojawią się dopiero kilka lub kilkanaście lat później, kiedy świadomość prawna w tym zakresie stanie się bardziej powszechna. W naszej ocenie im szybciej zostaną rozstrzygnięte kwestie prawnoautorskie związane ze stosowaniem AI, tym większa korzyść dla gospodarki i mniej strat dla przedsiębiorców.

Błądność wnioskowania o nieistotności zagadnień prawnoautorskich w kontekście narzędzi AI dobitnie pokazują aktualne spory dotyczące korzystania z komponentów oprogramowania na licencjach *open source*. Wśród *software house*'ów korzystanie z komponentów na licencjach *open source* jest powszechne, natomiast świadomość konsekwencji prawnych z tym związanych jest znikoma. Powszechne są sytuacje, w których przedsiębiorca jest przekonany, że posiada prawa autorskie do kodu źródłowego i dystrybucję swojego produktu opiera na modelu licencyjnym, pomimo tego, że w skład kodu źródłowego takiego produktu wchodzi komponenty udostępniane na tzw. wirusujących licencjach *open source* (tzw. licencjach *copyleft*), które często uniemożliwiają komercyjne licencjonowanie kodu. Jak bowiem przyjmuje się w doktrynie, implementacja do oprogramowania elementów kodu udostępnianych na licencjach *open source*, zawierających klauzule wirusujące / klauzule *copyleft* (np. *GNU General Public License v.3*¹), skutkuje wystąpieniem tzw. efektu wirusowego, polegającego na „infekowaniu” całości oprogramowania, które konsekwentnie musi być w całości udostępniane na zasadach *open source*².

W naszej ocenie podobne wyzwania i spory będą miały miejsce na gruncie narzędzi AI – stosowanych obecnie tak samo powszechnie jak dotychczas używano (i wciąż jest tak) oprogramowania typu *open source*. Dlatego naszym zawodowym celem powinno być wspieranie przedsiębiorców w świadomym korzystaniu z narzędzi AI.

3. Portfolio IP przedsiębiorstwa a narzędzia AI.

Wybrane obszary zarządzania ryzykiem związanym ze stosowaniem narzędzi AI

Przedmiotem niniejszego opracowania są kluczowe aspekty zarządzania przez przedsiębiorców ryzykiem prawnym związanym z zagadnieniami prawnoautorskimi i wykorzystywaniem AI w przedsiębiorstwie. Przedstawione rozważania dotyczą problematyki:

1) kwalifikacji treści generowanych przez AI jako utworów chronionych prawem autorskim lub jako wytworów, które nie podlegają takiej ochronie;

¹ Treść licencji GNU General Public License v.3, <https://www.gnu.org/licenses/gpl-3.0.html> [dostęp 4.06.2024].

² P. Wasilewski, *Open source w umowach IT*, [w:] *Licencjonowanie i prawa autorskie w umowach IT – aspekty praktyczne*, LEX/el. 2021.

2) zarządzania portfolio własności intelektualnej przedsiębiorstwa powstającej z wykorzystaniem narzędzi AI – co obejmuje:

- a) analizę relacji prawnych pomiędzy pracownikami oraz współpracownikami a pracodawcami w kontekście korzystania przez tych pierwszych z narzędzi AI (w tym w kontekście tworzenia wewnętrznych polityk i formułowania w umowach o pracę klauzul umownych skutecznie chroniących interes pracodawcy),
- b) stosowanie klauzul umownych zabezpieczających interes pracodawcy, którego pracownicy korzystają z narzędzi AI w umowach zawieranych przez takiego pracodawcę z klientami.

Istotnym elementem analizy jest również ocena wpływu ewentualnego zakwestionowania ochrony prawnoautorskiej generacji AI na wycenę własności intelektualnej przedsiębiorstwa.

3.1. Zagadnienia wstępne – utwór czy wytwór

Nie sposób omawiać zagadnień dotyczących zarządzania ryzykiem wynikającym ze stosowania narzędzi AI, nie rozstrzygając w sposób chociażby uproszczony kwestii ochrony prawnoautorskiej generacji AI.

Zgodnie z brzmieniem art. 1 ust. 1 z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych³ przedmiotem prawa autorskiego (tj. utworem) jest każdy przejaw działalności twórczej o indywidualnym charakterze, ustalony w jakiejkolwiek postaci, niezależnie od wartości, przeznaczenia i sposobu wyrażenia. Jak wskazuje się w doktrynie, utwór może stanowić jedynie efekt twórczości człowieka – aby dane dzieło uznać za utwór chroniony prawem autorskim, musi on być wynikiem twórczej pracy osoby fizycznej⁴. Wytwory generowane przez narzędzia naśladujące ludzki proces twórczy nie są uznawane za utwory⁵. Kluczowe dla uzyskania statutu utworu jest, aby człowiek miał decydujący wpływ na powstanie dzieła, nadając mu twórczy i indywidualny charakter⁶.

³ Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych, Dz. U. z 2022 r. poz. 2509 (dalej: „pr. aut.”).

⁴ E. Laskowska-Litak, [w:] R. Markiewicz (red.), *Ustawy autorskie. Komentarze, Tom I. Komentarz do ustawy o prawie autorskim i prawach pokrewnych*, Warszawa 2021, komentarz do art. 1 pr. aut., teza 49.

⁵ A. Niewęglowski, [w:] *idem* (red.), *Prawo autorskie. Komentarz*, Warszawa 2021, komentarz do art. 1 pr. aut., teza 1.

⁶ *Ibidem*.

Rezultat „samodzielnej” pracy AI nie spełnia cech utworu i nie przysługuje mu ochrona prawnoautorska⁷. Generacja ta jest jedynie wytworem należącym do szeroko rozumianej domeny publicznej⁸. Jednak już twórczy dobór fragmentów kodu wygenerowanego przez AI lub twórcza modyfikacja projektu lub materiału marketingowego stworzonego przez AI będą podlegały ochronie prawnoautorskiej. Kwestia ta jest jednak w rzeczywistości znacznie bardziej zniuansowana. Naszym zdaniem korzystanie z narzędzi AI w celu tworzenia tekstów, grafik, dźwięków itp. może odbywać się według trzech scenariuszy⁹ – tak jak dotychczas rozważano to w kontekście twórczości z wykorzystaniem komputerów:

1) treści są generowane przez AI „samodzielnie”, a użytkownik narzędzia ogranicza swoją rolę do wprowadzenia prostego promptu i nie modyfikuje tak powstałej generacji – mamy wtedy do czynienia z tzw. *purely AI-generated works*¹⁰ lub inaczej – *machine-generated work*¹¹;

2) treści są tworzone przez człowieka ze wsparciem narzędzi AI (tzw. *computer assisted creativity*¹² lub inaczej – *machine-enhanced works*¹³), jak w przypadku korzystania z narzędzi do obróbki grafiki, w których moduły AI są wykorzystywane np. do zwiększenia dokładności zaznaczania i wycinania elementów;

3) treści są „współtworzone” przez człowieka i AI, jak choćby w przypadku uzupełniania przez AI wyciętych i pustych miejsc w grafice lub na zdjęciu.

Największe ryzyko (w zasadzie graniczące z pewnością) braku powstania utworu i chroniących go praw autorskich powstaje w pierwszej ze wskazanych sytuacji. Ograniczenie się przez użytkownika do wprowadzenia krótkiego i nieskomplikowa-

⁷ R. Markiewicz, *ChatGPT i prawo autorskie Unii Europejskiej*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2023, nr 2(160), s. 146.

⁸ A. Stanek, *Kto jest twórcą, a kto zarobi na wytworach AI? Prawa autorskie a sztuczna inteligencja*, <https://www.forbes.pl/opinie/prawa-autorskie-a-sztuczna-inteligencja-kto-jest-tworca/n8tlx9b> [dostęp 4.06.2024]; R. Rozwadowski, *Uregulowanie zasad działania sztucznej inteligencji jest konieczne*, <https://www.prawo.pl/prawo/sztuczna-inteligencja-a-skutki-braku-uregulowan-prawnych,522083.html> [dostęp 4.06.2024]; P.P. Juściński, *Prawo autorskie w obliczu rozwoju sztucznej inteligencji*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2019, nr 1.

⁹ D. Flisak, *Odpowiedzi na pytania ze szkolenia pt. „Sztuczna inteligencja: zrozumieć, opanować, zastosować. Prawne aspekty funkcjonowania”, które odbyło się 5 lutego 2024 r.*, LEX/el. 2024.

¹⁰ A. Trapova, *Copyright for AI-generated works: a task for the internal market?*, https://discovery.ucl.ac.uk/id/eprint/10166871/1/Trapova_Trapova%20-%20Copyright%20for%20AI-generated%20works%20a%20task%20for%20the%20internal%20market%20-%20reworked_05122022_clean_final.pdf [dostęp 4.06.2024].

¹¹ D. Flisak, *Odpowiedzi...*

¹² D. Flisak, *Komentarz do wybranych przepisów ustawy o prawie autorskim i prawach pokrewnych*, LEX/el. 2018, komentarz do art. 1 pr. aut., teza 4.

¹³ D. Flisak, *Odpowiedzi...*

nego promptu w celu nakazania AI realizacji określonego zamysłu na kształt generacji i dokonania przez nią „samodzielnego” doboru parametrów tworzonej treści nie spełnia bowiem przesłanek uznania za utwór¹⁴.

Ryzyko ulega znacznemu zmniejszeniu w drugim ze wskazanych scenariuszy. Jeśli AI jest wykorzystywane jedynie w formie narzędzia ułatwiającego pracę twórcy (np. pomaga wycinać, obrabiać lub modyfikować elementy powstającego dzieła, justować tekst itp.), to co do zasady przemawia to za uznaniem takiego dzieła za twórczość człowieka chronioną prawem autorskim¹⁵ – innymi słowy, sytuacja ta nie będzie znacząco różnić się od tworzenia przez człowieka utworów w postaci animacji, grafik komputerowych itp.

Najwięcej kontrowersji towarzyszy jednak trzeciej z przedstawionych sytuacji, tj. powstawaniu treści „współtworzonych” przez człowieka i AI. Mowa tutaj o tworzeniu dzieła zarówno przez człowieka, jak i maszynę – w sposób wielopoziomowy, tj. gdy wkłady AI i człowieka tworzą jedną całość¹⁶. Do takiej sytuacji dochodzi np., gdy fotografia stworzona i obrobiona przez człowieka jest następnie modyfikowana przez AI poprzez np. wygenerowanie zupełnie innego tła niż ujęte przez człowieka, dodawanie nowych elementów na fotografii, wypełnianie luk powstających przez wycięcie przez fotografa niektórych elementów zdjęcia.

W ostatnim z omówionych przykładów, aby zwiększyć prawdopodobieństwo uznania, że użytkownik AI wniósł w powstanie generacji wkład twórczy pozwalający na powstanie ochrony prawnoautorskiej, taki użytkownik nie może ograniczać się do wprowadzania prostych promptów, a powinien maksymalizować swój wysiłek twórczy poprzez np. wielokrotne promptowanie, wprowadzanie promptów zawierających znaczną liczbę szczegółów wpływających na kształt generacji, a nawet powinien podjąć się następczej obróbki generacji¹⁷.

Należy stwierdzić, że o powstaniu praw autorskich do generacji AI będzie decydowała indywidualna ocena każdej z nich. Nie jest możliwe sztywne (np. procentowe) określenie stopnia swobody wyborów dokonanych przez użytkownika AI, tak samo jak nie jest to możliwe w przypadku oceniania utworów nie będących generacją AI. Innymi słowy, ograniczenie się przez użytkownika AI do zdawkowego i niefinezyjnego określenia w prompcie jedynie tego, co powinna przedstawiać ge-

¹⁴ *Ibidem.*

¹⁵ *Ibidem.*

¹⁶ *Ibidem.*

¹⁷ Por. szerzej o możliwościach twórczego wykorzystywania AI: Z. Okoń, *Czy w Midjourney można stworzyć utwór?*, <https://www.linkedin.com/pulse/czy-w-midjourney-mo%C5%BCna-stworzy%C4%97-utw%C3%B3r-zbigniew-oko%C5%84-cvywf/> [dostęp 20.05.2024]; D. Flisak, *Odpowiedzi...*

neracja, z pewnością nie wystarczy do uznania takiej generacji za utwór. Im bardziej jednak użytkownik, za pomocą promptów, będzie intensywniej wpływał na kształt generacji (np. coraz bardziej szczegółowo określając styl, atmosferę, układ, rozmycie tła, sposób badania światła itd.), tym większa szansa na powstanie utworu¹⁸.

3.2. Jak w praktyce wykazać, że generacja AI jest utworem?

Powyższa kwestia wydaje się jasna, jednak w praktyce kluczowe może okazać się udokumentowanie twórczego wkładu w pracownika/współpracownika w powstanie utworu. Dla przykładu, sąd czeski, rozstrzygając spór pomiędzy grafikami a praską kancelarią adwokacką, oddalił powództwo o ochronę prawnoautorską, wytoczone przez grafika, który przy wykorzystaniu AI wykonał na rzecz pozwanej kancelarii zamówienie w postaci identyfikacji graficznej. Powód domagał się ustalenia, że jest autorem spornej grafiki, a także zobowiązania pozwanej kancelarii do usunięcia grafik ze strony internetowej oraz zakazania jakiegokolwiek innego postępowania, mogącego zagrozić lub naruszyć prawa powoda jako autora grafik. Naruszenia swoich praw powód upatrywał w upublicznieniu grafik bez jego zgody na stronie internetowej kancelarii.

Powództwo zostało oddalone z uwagi na fakt, że powód nie sprostował ciężarowi dowodu w zakresie sposobu powstania grafiki (w tym przede wszystkim, kto i jakich promptów użył w celu stworzenia grafiki). Powód argumentował, że AI wytworzyła grafiki na podstawie jego poleceń, dlatego to jemu przysługują prawa autorskie. Grafik nie był jednak w stanie wykazać, jak przebiegał ten proces – w szczególności nie przedstawił treści poleceń (promptów) jakie wydawał, aby uzyskać docelowe generacje (*output*). Z tym stanowiskiem nie zgodził się sąd (wyrok Sądu Miejskiego w Pradze z dnia 11 października 2023 r.¹⁹).

3.3. Narzędzia AI w relacji pracownik–pracodawca

Ryzyka związane z korzystaniem z narzędzi AI w biznesie są najwyższe, kiedy AI jest wykorzystywana w sposób zupełnie dowolny przez pracowników przedsiębiorstwa, bez jakiegokolwiek planu i kontroli. Dla przejrzystości wyводу, w tej części opracowania, przedsiębiorcę, którego pracownicy wykorzystują AI, będziemy nazywać zbiorczo „pracodawcą”.

¹⁸ *Ibidem*.

¹⁹ Wyrok Sądu Miejskiego w Pradze z 11 października 2023 r., 10 C 13/2023, https://justice.cz/documents/14569/1865919/10C_13_2023_10/108cad3e-d9e8-454f-bfac-d58e1253c83a [dostęp 21.05.2024].

Zgodnie z brzmieniem art. 12 ust. 1 pr. aut., jeżeli ustawa lub umowa o pracę nie stanowią inaczej, pracodawca, którego pracownik stworzył utwór w wyniku wykonywania obowiązków ze stosunku pracy, nabywa z chwilą przyjęcia utworu autorskie prawa majątkowe w granicach wynikających z celu umowy o pracę i zgodnie z zamiarem stron. Zgodnie z przyjmowanym w doktrynie poglądem nabycie przez pracodawcę praw autorskich do utworu stworzonego przez pracownika ma charakter pochodny²⁰. **Oznacza to, że pracodawca nabywa prawa autorskie do utworu dopiero, jeśli uprzednio nabędzie je pracownik-autor.**

W kontekście ryzyka braku nabycia praw autorskich przez pracodawcę należy również odnieść się do art. 74 ust. 3 pr. aut., zgodnie z którym prawa majątkowe do programu komputerowego stworzonego przez pracownika w wyniku wykonywania obowiązków ze stosunku pracy przysługują pracodawcy, o ile umowa nie stanowi inaczej. W odróżnieniu od nabycia praw przez pracodawcę praw do utworu na podstawie art. 12 pr. aut. nabycie przez pracodawcę praw do programu komputerowego stworzonego przez jego pracownika ma charakter pierwotny. Oznacza to, że pracodawca nabywa prawa autorskie do stworzonego przez pracownika programu komputerowego, nie będąc następcą prawnym pracownika-autora. Takie nabycie jest konstytutywne – pracodawca uzyskuje nowe, nieistniejące wcześniej prawo²¹.

Oba omówione przypadki nabycia przez pracodawcę praw autorskich do pracowniczych utworów (lub programów komputerowych) mogą nie ziścić się w przypadku nieodpowiedniego korzystania przez pracowników z narzędzi AI. Jeśli bowiem pracownik będzie korzystał z narzędzi AI w taki sposób, że:

1) dostarczone pracodawcy treści (teksty, grafiki, programy komputerowe itp.) będą stanowić zaledwie wspomniany wyżej *purely AI-generated content* lub

2) treści będą „współtworzone” przez pracownika i AI bez wniesienia przez pracownika wkładu twórczego umożliwiającego uznanie generacji za utwór, – to w takiej sytuacji, odpowiednio w przypadku utworów pracowniczych określonych w art. 12 pr. aut. – pracownik nie będzie mógł z mocy prawa przenieść praw autorskich na pracodawcę, a w przypadku programów komputerowych, o których mowa w art. 74 pr. aut. – pracodawca nie nabędzie takich praw w sposób pierwotny. W obu bowiem przypadkach nie będziemy mieli do czynienia z utworami, a zatem nie powstaną prawa autorskie, które może nabyć pracodawca.

²⁰ J. Ożegalska-Trybalska, [w:] R. Markiewicz (red.), *Ustawy autorskie. Komentarze, Tom I. Komentarz do ustawy o prawie autorskim i prawach pokrewnych*, Warszawa 2021, komentarz do art. 12 pr. aut., teza 13.

²¹ A. Niewęglowski, *op. cit.*, komentarz do art. 74 pr. aut., teza 5.

Dobrym przykładem sytuacji, w których zobrazowane ryzyko braku nabycia praw autorskich może się zmaterializować, jest branża IT. Aktualnie powszechne jest korzystanie przez programistów z narzędzi AI podczas prac nad kodem programu komputerowego. Jako przykładowe narzędzie może posłużyć GitHub Copilot, który może być w pracy pracownika-programisty wykorzystywany na różne sposoby. Może np. jedynie weryfikować poprawność kodu stworzonego przez pracownika, tworzyć mniejsze elementy kodu, ale może również generować prawie całość funkcjonującego programu komputerowego lub generować mniejsze fragmenty kodu na podstawie instrukcji pracownika²² (tzw. *snippets*)²³. W tym drugim przypadku potencjalnie możliwa jest sytuacja, w której pracownik nie wniesie swojego wkładu twórczego do programu komputerowego i w konsekwencji nie będzie możliwe nabycie praw autorskich przez pracodawcę. W obecnym stanie techniki korzystanie z takich narzędzi AI co do zasady i tak wymaga twórczego wkładu człowieka, a narzędzia AI mają jedynie charakter wspomagający. Nie zmienia to jednak faktu, że korzystanie z narzędzi AI może sprawić, że w procesie tworzenia oprogramowania będzie mniej elementów twórczych niż w przypadku pracy bez takich „wspomagaczy”. Jeśli wkład ludzki potrzebny do osiągnięcia końcowego efektu stanie się zbyt mały, aby uznać go za rezultat pracy człowieka, to takie dzieło nie będzie chronione prawem autorskim²⁴.

Ponadto, warto w tym miejscu wspomnieć o krzyżującym się ryzyku korzystania przez pracowników ze „stockowego” kodu komputerowego udostępnianego na wspomnianych już wcześniej „wirusujących” licencjach typu *open source* oraz wspomnianego powszechnego korzystania w branży IT z narzędzi AI typu GitHub Copilot. W takim stanie faktycznym potencjalnie możliwa jest sytuacja, w której pracownik, tworząc program komputerowy, wykorzysta do stworzenia znacznej części kodu znalezione przez siebie w sieci elementy wirusującego oprogramowania typu *open source*, a następnie elementy te zostaną scalone lub zmodyfikowane przez ww. narzędzie AI²⁵. W takiej sytuacji powstaje ryzyko braku nabycia przez pracodawcę

²² A. Wachowska, M. Ręgorowicz, K. Dymek, *Programisto, czy nadal „tworzysz”?* Prawa autorskie a korzystanie z Copilot i innych narzędzi AI wspierających tworzenie oprogramowania, <https://www.traple.pl/programisto-czy-nadal-tworzysz-prawa-autorskie-a-korzystanie-z-copilot-i-innych-narzedzi-ai-wspierajacych-tworzenie-oprogramowania/> [dostęp 21.05.2024].

²³ F. Rabbi, A. Champa, M. Zibran, R. Islam, *AI Writes, We Analyze: The ChatGPT Python Code Saga*, [w:] *21st International Conference on Mining Software Repositories (MSR '24)*, April 15–16, 2024, Lisbon, Portugal, s. 177.

²⁴ *Ibidem*.

²⁵ Przykład takiego zastosowania narzędzi AI wskazano w B. Berabi *et al.*, *B. Prompts for GPT-3.5 & GPT-4*, [w:] *DeepCode AI Fix: Fixing Security Vulnerabilities with Large Language Models*, Zurich 2024, <https://arxiv.org/pdf/2402.13291> [dostęp 4.06.2024].

praw autorskich do kodu dostarczonego przez programistę. Ponadto, pozostała część oprogramowania, która została połączona ze zmodyfikowanymi przez AI (ale wciąż wirusującymi elementami oprogramowania *open source*), będzie podlegać „zainfekowaniu” licencją *open source* i zgodnie z zasadami takiej licencji możliwości jego komercjalizacji będą ograniczone (brak możliwości odpłatnego licencjonowania, konieczność udostępnienia kodu źródłowego).

Skutki wskazanych zagrożeń ujawniają się na wielu płaszczyznach. Przykładowo:

1) brak nabycia praw autorskich do kodu programu komputerowego w podanych scenariuszach może uniemożliwić odliczenie 50% kosztów uzyskania przychodów (KUP)²⁶. Koszty uzyskania przychodu to wydatki, jakie ponosi się, aby osiągnąć przychód. Dla pracownika, który w ramach swoich obowiązków pracowniczych tworzy utwory, koszty te wynoszą 50% uzyskanego przychodu, pomniejszonego o składki ZUS (zazwyczaj 20%). Aby skorzystać z możliwości rozliczenia 50% KUP, zgodnie z brzmieniem art. 22 ust. 9 pkt 3) ustawy z dnia 26 lipca 1991 r. o podatku dochodowym od osób fizycznych²⁷, wynagrodzenie uzyskane przez pracownika musi zostać wypłacone z tytułu przeniesienia praw autorskich przez twórcę utworu (a więc także programu komputerowego). Zaistnienie sytuacji, w której w jednym z przedstawionych scenariuszy wykorzystania przez pracownika narzędzi AI nie dochodzi do powstania praw autorskich, siłą rzeczy nie jest możliwe nabycie takich praw przez pracodawcę, a co za tym idzie – w naszej opinii nie dochodzi do spełnienia wymogów przywołanego art. 22 ust. 9 pkt 3) ustawy o PIT i brak jest możliwości odliczenia 50% KUP;

2) ziszczenie się jednego z opisanych scenariuszy wykorzystania przez programistę narzędzi AI i konsekwentny brak nabycia przez pracodawcę praw autorskich będą skutkować tym, że pracodawca, w myśl zasady *nemo plus iuris*, nie będzie w stanie skutecznie licencjonować lub przenieść praw do elementów powstałych w taki sposób na swoich klientów – a tym samym prawidłowo wykonać umowy z takimi klientami. Nie dość, że w realiach kontraktowych taka sytuacja może prowadzić do obciążenia pracodawcy karami umownymi, to jeszcze powstaje ryzyko, że z dostarczonych klientowi elementów programu komputerowego,

²⁶ Z zastrzeżeniem, że nie dotyczy to sytuacji omówionego wyżej pierwotnego nabycia praw autorskich przez pracodawcę programisty, tj. gdy pracodawca i programista nie zmodyfikują umownie reguły z art. 74 pr. aut. – por. Interpretacja Ogólna Ministra Finansów z dnia 15 września 2020 r., Nr DD3.8201.1.2018, w sprawie zastosowania 50% kosztów uzyskania przychodów do honorarium autorskiego, Dz. Urz. MF z 2020 r. poz. 10.

²⁷ Ustawa z dnia 26 lipca 1991 r. o podatku dochodowym od osób fizycznych, Dz. U. z 2024 r. poz. 226 (dalej: „ustawa o PIT”).

z uwagi na brak ochrony prawnoautorskiej, będą mogli korzystać np. jego konkurenci²⁸. Alternatywna ochrona praw do oprogramowania może być co prawda realizowana na podstawie przepisów o zwalczaniu nieuczciwej konkurencji, ale jest to w praktyce znacznie trudniejsza podstawa dochodzenia roszczeń niż przepisy ustawy prawnoautorskiej.

3.4. Generacje AI – wybrane zabezpieczenia umowne

Przygotowanie wewnętrznej polityki bezpiecznego korzystania z narzędzi AI to jednak nie wszystko. W celu zabezpieczenia interesu przedsiębiorcy korzystanie z narzędzi AI powinno mieć swoje odzwierciedlenie w odpowiednich postanowieniach umownych. Tyczy się to w szczególności uwzględnienia w umowach okoliczności wykorzystania modeli AI w kontekście postanowień dotyczących transferu praw autorskich. Rekomendowane jest zadbanie o umieszczenie stosownych klauzul zarówno w umowach zawieranych przez przedsiębiorców z ich pracownikami lub współpracownikami, jak i w umowach, które tacy przedsiębiorcy zawierają ze swoimi klientami.

Dla przejrzystości wyводу, w tej części opracowania przedsiębiorcę, którego pracownicy lub współpracownicy wykorzystują narzędzia AI, będziemy nazywać zbiorczo „przedsiębiorcą”, jego pracowników lub współpracowników – zbiorczo mianem „współpracownika”, a kontrahentów takiego przedsiębiorcy będziemy zbiorczo określać pojęciem „zamawiający”.

3.4.1. Precyzyjne wskazanie, że narzędzia AI będą wykorzystywane do wykonania umowy

W przypadku gdy przedsiębiorca dostarcza usługi lub wykonuje dzieła z użyciem narzędzi AI, kwestia ta powinna być wyraźnie wskazana w umowie, a zakres podziału ryzyka związany z korzystaniem z narzędzi AI powinien być jasno ustalony w umowie.

Tutaj przykładowa, przygotowana przez autorów klauzula informująca o korzystaniu przez przedsiębiorcę z narzędzi AI: „Zamawiający przyjmuje do wiadomości i akceptuje fakt, że niektóre elementy Dzieła dostarczane zgodnie z Umową mogą zostać opracowane przy użyciu «systemów AI» (tj. systemów maszynowych, które zostały zaprojektowane do działania z różnym poziomem autonomii po ich wdrożeniu oraz które mogą wykazywać zdolność adaptacji po ich wdrożeniu, a także które – na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na

²⁸ *Ibidem.*

podstawie otrzymanych danych wejściowych wyniki, takie jak predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne)”²⁹.

To przykładowe postanowienie jest dość ogólne i zabezpiecza bardziej przedsiębiorcę niż zamawiającego. W negocjacjach zamawiający mógłby oczekiwać sprecyzowania przez przedsiębiorcę narzędzi AI, jakie będą wykorzystywane przy wykonywaniu umowy, tak aby bardziej kompleksowo ocenić ryzyka. Jednocześnie częstym błędem popełnianym przez przedsiębiorcę jest bezrefleksyjne składanie w umowie oświadczeń, że usługi lub dzieło zostaną wykonane przez niego samodzielnie, będą stanowić utwór i nie będą posiadały wad prawnych. W przypadku składania takich oświadczeń przedsiębiorca bierze na siebie nieświadomie wszelkie ryzyka związane z korzystaniem z narzędzi AI, nawet jeśli biznesowo uzasadnione mogłoby być, aby część tego ryzyka ponosił zamawiający.

3.4.2. Zobowiązania gwarancyjne oraz klauzule indemnifikacyjne

Kolejną kwestią umowną jest ustalenie zakresu odpowiedzialności przedsiębiorcy i ryzyka ponoszonego przez zamawiającego w związku z korzystaniem z narzędzi AI poprzez odpowiednie określenie oświadczeń gwarancyjnych: „Wykorzystanie elementów Dzieła stworzonych przy wykorzystaniu systemów AI (i) nie spowoduje konieczności udzielenia podmiotom trzecim licencji do kodu źródłowego przekazanego przez zamawiającego w celu wykonania Dzieła przez przedsiębiorcę, ani kodu źródłowego opracowanego przez przedsiębiorcę dla zamawiającego w ramach lub w związku z wykonywaniem Dzieła oraz (ii) nie będzie ograniczać zamawiającego w zgodnym z Umową korzystaniu lub rozporządzaniu Dziełem”.

Powyższe przygotowane przez autorów postanowienie można uznać za kontraktowe minimum. Przedsiębiorca, jako profesjonalista, powinien zadbać o to, czy wykorzystanie generacji AI nie spowoduje konieczności licencjonowania kodu źródłowego, co może mieć miejsce w przypadku, gdy pracownicy lub współpracownicy przedsiębiorcy wykonujący zamówienie skorzystają z tego typu systemów AI, które są udostępniane (najczęściej w wersjach nieodpłatnych) przez dostawców wskazujących w swoich warunkach umów (*terms and conditions*), że treści wprowadzane (*input*) lub uzyskiwane wskutek procesu generacji (*output*) są objęte licencją udzielaną na rzecz takiego dostawcy.

²⁹ Za definicją z art. 3 pkt 1 rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) (Dz. Urz. L 2024/1689, z 12.7.2024).

Często spotykanym postanowieniem umownym jest też zobowiązanie przedsiębiorcy do współdziałania z zamawiającym w przypadku pojawienia się roszczeń osób trzecich względem zamawiającego, wstąpienie do sporu i zaspokojenie uzasadnionych roszczeń osób trzecich. Za przykład może posłużyć klauzula przygotowana przez autorów: „W przypadku zgłaszania względem zamawiającego przez osoby trzecie jakichkolwiek roszczeń związanych z naruszeniem praw własności intelektualnej do elementów Dzieła, przedsiębiorca zobowiązuje się wstąpić na wezwanie zamawiającego do toczącego się sporu oraz zaspokoić uzasadnione roszczenia osób trzecich, stwierdzone prawomocnym rozstrzygnięciem właściwego organu lub sądu lub objętych ugodą sądową lub pozasądową, pod warunkiem, że treść takiej ugody została zaakceptowana przed jej zawarciem przez przedsiębiorcę. Ponadto, przedsiębiorca zobowiązuje się zwolnić zamawiającego od obowiązku świadczenia z tego tytułu, zaś w przypadku niewykonania tego zobowiązania będzie on odpowiedzialny za szkodę zamawiającego, na którą składać się będą w szczególności zasądzone od zamawiającego koszty zastępstwa procesowego, koszty sądowe oraz inne zasądzone świadczenia”.

4. Wnioski

Wyzwania prawne związane z wykorzystywaniem w biznesie treści generowanych przez AI wynikają ze ścisłego powiązania zagadnień prawnoautorskich (wątpliwości co do zakresu pojęcia utworu i wynikającej z niego ochrony) ze skomplikowanymi zagadnieniami natury technologicznej (procesy technologiczne związane z narzędziami AI, w których wykorzystywane lub tworzone są utwory). Brak świadomego korzystania z narzędzi AI prowadzi do eskalacji ryzyka przedsiębiorstwa, wpływa na jego wycenę i powoduje ryzyko odpowiedzialności kontraktowej względem kontrahentów.

Ryzyka związane z komercyjnym wykorzystaniem treści generowanych przez AI nie są iluzoryczne, czego dowodzi coraz częstsze kwestionowanie ochrony prawnoautorskiej treści generowanych przez AI w trakcie sporów sądowych (np. dotyczących autorstwa treści stworzonych przez AI) – zarówno w krajach tradycji *common law* i w krajach azjatyckich³⁰, jak i w ramach naszego kręgu „prawno-kulturowego”

³⁰ Por. np. wyrok Sądu ds. Internetu w Pekinie z dnia 27 listopada 2023 r., Jing 0491 Min Chu No. 11279 (2023), <https://english.bjinternetcourt.gov.cn/pdf/BeijingInternetCourtCivilJudgment112792023.pdf> [dostęp 4.06.2024] oraz wyrok United States District Court – District of Columbia z dnia 18.08.2023 r., 22-1564 (BAH), https://ecf.dcd.uscourts.gov/cgi-bin/show_public_doc?2022cv1564-24 [dostęp 4.06.2024].

(por. przywołany w niniejszej analizie wyrok Sądu Miejskiego w Pradze³¹), co wskazuje na rosnące znaczenie tego zagadnienia w praktyce prawniczej.

Jednak, jak wynika z naszej analizy, pomimo że prawo nie jest jeszcze w pełni dostosowane do wyzwań związanych z AI, istnieją narzędzia prawne pozwalające umiejętnie ograniczać ryzyko związane z korzystaniem z AI w biznesie. Możliwe jest elastyczne stosowanie już istniejących instytucji prawnych do zabezpieczania ryzyk biznesowych wynikających z korzystania z AI – np. na gruncie prawa cywilnego, a w szczególności na gruncie prawa umów. Konieczne jest jednak ciągle monitorowanie zmian prawnych i technologicznych i dostosowywanie strategii zarządzania ryzykiem do nowych realiów.

Bibliografia

Akty prawne

Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) (Dz. Urz. L 2024/1689, z 12.7.2024).

Ustawa z dnia 26 lipca 1991 r. o podatku dochodowym od osób fizycznych (Dz. U. z 2024 r. poz. 226).

Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz. U. z 2022 r. poz. 2509).

Źródła

Berabi B., *Prompts for GPT-3.5 & GPT-4*, [w:] *DeepCode AI Fix: Fixing Security Vulnerabilities with Large Language Models*. Zurych 2024, <https://arxiv.org/pdf/2402.13291> [dostęp 4.06.2024].

Flisak D., *Odpowiedzi na pytania ze szkolenia pt. „Sztuczna inteligencja: zrozumieć, opanować, zastosować. Prawne aspekty funkcjonowania”*, które odbyło się 5 lutego 2024 r., LEX/el. 2024.

Interpretacja Ogólna Ministra Finansów z dnia 15 września 2020 r., Nr DD3.8201.1.2018, w sprawie zastosowania 50% kosztów uzyskania przychodów do honorarium autorskiego (Dz. Urz. MF z 2020 r. poz. 10).

Licencja GNU General Public License v.3, <https://www.gnu.org/licenses/gpl-3.0.html> [dostęp 4.06.2024].

Okoń Z., *Czy w Midjourney można stworzyć utwór?*, <https://www.linkedin.com/pulse/czy-w-midjourney-mo%C5%BCna-stworzy%C4%87-utw%C3%B3r-zbigniew-oko%C5%84-cvywf/> [dostęp 20.05.2024].

³¹ Wyrok Sądu Miejskiego w Pradze z dnia 11 października 2023 r., 10 C 13/2023, https://justice.cz/documents/14569/1865919/10C_13_2023_10/108cad3e-d9e8-454f-bfac-d58e1253c83a, [dostęp 21.05.2024].

- Rozwadowski R., *Uregulowanie zasad działania sztucznej inteligencji jest konieczne*, <https://www.prawo.pl/prawo/sztuczna-inteligencja-a-skutki-braku-uregulowan-prawnych,522083.html> [dostęp 4.06.2024].
- Stanek A., *Kto jest twórcą, a kto zarobi na wytworach AI? Prawa autorskie a sztuczna inteligencja*, <https://www.forbes.pl/opinie/prawa-autorskie-a-sztuczna-inteligencja-kto-jest-tworca/n8tlx9b> [dostęp 4.06.2024].
- Trapova A., *Copyright for AI-generated works: a task for the internal market?*, https://discovery.ucl.ac.uk/id/eprint/10166871/1/Trapova_Trapova%20-%20Copyright%20for%20AI-generated%20works%20a%20task%20for%20the%20internal%20market%20-%20reworked_05122022_clean_final.pdf [dostęp 4.06.2024].
- Wachowska A., Ręgorowicz M., Dymek K., *Programisto, czy nadal „tworzysz”? Prawa autorskie a korzystanie z Copilot i innych narzędzi AI wspierających tworzenie oprogramowania*, <https://www.traple.pl/programisto-czy-nadal-tworzysz-prawa-autorskie-a-korzystanie-z-copilot-i-innych-narzedzi-ai-wspierajacych-tworzenie-oprogramowania/> [dostęp 21.05.2024].
- Wyrok Sądu ds. Internetu w Pekinie z dnia 27 listopada 2023 r., Jing 0491 Min Chu No. 11279 (2023), <https://english.bjinternetcourt.gov.cn/pdf/BeijingInternetCourtCivilJudgment112792023.pdf> [dostęp 4.06.2024].
- Wyrok Sądu Miejskiego w Pradze z dnia 11 października 2023 r., 10 C 13/2023, https://justice.cz/documents/14569/1865919/10C_13_2023_10/108cad3e-d9e8-454f-bfac-d58e1253e83a [dostęp 21.05.2024].
- Wyrok United States District Court – District of Columbia z dnia 18.08.2023, 22-1564 (BAH), https://ecf.dcd.uscourts.gov/cgi-bin/show_public_doc?2022cv1564-24 [dostęp 4.06.2024].

Literatura

- Flisak D., *Komentarz do wybranych przepisów ustawy o prawie autorskim i prawach pokrewnych*, LEX/el. 2018.
- Juściński P.P., *Prawo autorskie w obliczu rozwoju sztucznej inteligencji*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2019, nr 1.
- Laskowska-Litak E., [w:] R. Markiewicz (red.), *Ustawy autorskie. Komentarze*, Tom I. *Komentarz do ustawy o prawie autorskim i prawach pokrewnych*, Warszawa 2021.
- Markiewicz R., *ChatGPT i prawo autorskie Unii Europejskiej*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2023, nr 2(160).
- Niewęglowski A., [w:] *idem* (red.), *Prawo autorskie. Komentarz*, Warszawa 2021.
- Ożegalska-Trybalska J., [w:] R. Markiewicz (red.), *Ustawy autorskie. Komentarze*, Tom I. *Komentarz do ustawy o prawie autorskim i prawach pokrewnych*, Warszawa 2021
- Rabbi F., Champa A., Zibran M., Islam R., *AI Writes, We Analyze: The ChatGPT Python Code Saga*, [w:] *21st International Conference on Mining Software Repositories (MSR '24)*, April 15–16, 2024, Lisbon, Portugal.
- Wasilewski P., *Open source w umowach IT*, [w:] *Licencjonowanie i prawa autorskie w umowach IT – aspekty praktyczne*, LEX/el. 2021.

Business exploitation of content generated by ai – a practical case study of copyright risk management

Abstract

The text is focused on the business risks arising from the use of AI tools in enterprises – related to the lack of emergence of copyrights to AI generations or related to the lack of effective acquisition of copyrights to AI generations rights. The analysis contained in the text presents the genesis of the emergence of such risks, related to the level of preparation of entrepreneurs for the implementation of AI tools in their companies, and then discusses selected examples of such risks. The authors of the text focused on discussing the risks arising from the disputed legal status of AI generation under copyright law (in particular in the IT industry) and the practical consequences of not acquiring copyright in works created using AI – taking into account the jurisprudence of EU member state courts in this regard. The analysis of business risks is discussed in particular in the context of the employee-employer relationship and in the context of the employer-client relationship. The analysis in the text also includes a review of the areas of risk management related to the use of AI tools. The text also discusses solutions used by legal practitioners to secure the discussed risks, presenting and analysing examples of solutions on the grounds of contract law and examples of appropriate preparation of entrepreneurs and employees for responsible use of AI tools.

Keywords

AI, work, risk, entrepreneur, management, business

Jakub Jurowicz

Uniwersytet Jagielloński
Wydział Prawa i Administracji
ORCID: 0009-0002-7113-5992

Małgorzata Paja

Uniwersytet Jagielloński
Wydział Prawa i Administracji
ORCID: 0009-0003-0779-9636

Zwielokrotnianie utworów w procesach tworzenia i funkcjonowania dużych modeli językowych

Streszczenie

W odniesieniu do tzw. dużych modeli językowych (ang. *large language models*, LLM) jednym z głównych zagadnień jest zagrożenie dla wyłącznych praw twórców, a w szczególności – praw autorskich. Wspomniane zagrożenie materializuje się zarówno na etapie powstawania LLM, jak i późniejszego funkcjonowania takiego modelu. W przedmiocie korzystania z cudzej twórczości spory w doktrynie prawa toczą się przede wszystkim w odniesieniu do zakresu dozwolonego użytku utworów w ramach eksploracji tekstów i danych, w odniesieniu do tzw. zjawiska zapamiętywania danych treningowych przez LLM oraz wobec zjawiska tzw. regurgitacji, tj. „zwracania” fragmentów chronionych danych treningowych przez LLM w generowanych treściach. Wskazane spory podsycało procedowanie najnowszej legislacji dotyczącej sztucznej inteligencji, tj. przede wszystkim Aktu o sztucznej inteligencji, którego kolejne wersje pojawiały się w I kwartale 2024 r. i którego ostateczna wersja została uchwalona 13 czerwca 2024 r. – uwzględniając odesłania do przepisów o eksploracji tekstów i danych na gruncie Dyrektywy o prawie autorskim na jednolitym rynku cyfrowym. Celem niniejszego tekstu jest analiza poszczególnych etapów powstawania i funkcjonowania LLM oraz weryfikacja zachodzącego w ich trakcie zwielokrotniania chronionej twórczości w kontekście ewentualnego stosowania przepisów o dozwolonym użytku.

Słowa kluczowe

AI, utwór, eksploatacja, dozwolony użytek, eksploracja, tekstów, danych, TDM

Wstęp

Dynamiczny rozwój sztucznej inteligencji (dalej: „AI”) spowodował również gwałtowny wzrost liczby problemów natury prawnej. W odniesieniu do modeli AI

ogólnego przeznaczenia¹ jednym z głównych zagadnień jest zagrożenie dla wyłącznych praw twórców, a w szczególności majątkowych praw autorskich. Wspomniane zagrożenie materializuje się co do zasady na dwóch płaszczyznach, tj. na etapie trenowania modelu AI oraz na etapie tworzenia generacji przez AI. Autorzy, dla zwiększenia przejrzystości dalszej analizy, wskazują w tym miejscu, że na potrzeby niniejszego artykułu, mając na myśli modele AI ogólnego przeznaczenia, będą posługiwać się terminem „LLM” (ang. *large language models*)² – jako że rozważania poczynione w takcie analizy odnoszą się, jak wskazuje tytuł, właśnie do dużych modeli językowych.

Powyższe zagadnienie wróciło na agendę doktrynalnej dyskusji z nową mocą, gdy wraz z kolejnymi uchwalonymi wersjami Aktu o sztucznej inteligencji 13 marca 2024 r. Parlament Europejski przyjął swoją wersję tekstu Aktu o sztucznej inteligencji³, następnie 14 maja 2024 r. swoją wersję uchwaliła Rada Europejska⁴, a końcowe brzmienie AI Act ukształtowano w wersji z 13 czerwca 2024 r. W toku prac nad kolejnymi wersjami AI Act prawodawca europejski zdecydował się zaadresować problemy dotyczące trenowania LLM na dziełach objętych cudzymi prawami autorskimi (por. ostateczna treść art. 53 ust. 1 lit. c) AI Act), ujmując pośród obowiązków dostawców takich modeli obowiązek wprowadzenia polityki zapewnienia zgodności

¹ Model AI, w tym model AI trenowany dużą ilością danych z wykorzystaniem nadzoru własnego na dużą skalę, który wykazuje znaczną ogólność i jest w stanie kompetentnie wykonywać szeroki zakres różnych zadań, niezależnie od sposobu, w jaki model ten jest wprowadzany do obrotu, i który można zintegrować z różnymi systemami lub aplikacjami niższego szczebla – z wyłączeniem modeli AI, które są wykorzystywane na potrzeby działań w zakresie badań, rozwoju i tworzenia prototypów przed wprowadzeniem ich do obrotu – za definicją z art. 3 pkt 63 rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) (Dz.U. L 2024/1689, z 12.7.2024) (dalej: „AI Act”).

² I. Emanuilov, T. Margoni, *Memorisation in generative models and EU copyright law: an interdisciplinary view*, Kluwer Copyright Blog, 26.03.2024, <https://copyrightblog.kluweriplaw.com/2024/03/26/memorisation-in-generative-models-and-eu-copyright-law-an-interdisciplinary-view/> [dostęp 21.05.2024].

³ Rezolucja ustawodawcza Parlamentu Europejskiego z dnia 13 marca 2024 r. w sprawie wniosku dotyczącego rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (akt w sprawie sztucznej inteligencji) i zmieniającego niektóre akty ustawodawcze Unii (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_PL.pdf [dostęp 17.05.2024].

⁴ Propozycja Rozporządzenia Parlamentu Europejskiego i Rady w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) w wersji przyjętej 14 maja 2024 r., 2021/0106(COD), PE-CONS 24/24, <https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/pl/pdf> (dalej: „AI Act”) [dostęp 21.05.2024].

z unijnym prawem autorskim, w szczególności z myślą o identyfikacji i przestrzeganiu, w tym poprzez najnowocześniejsze technologie, zastrzeżenia praw wyrażonego zgodnie z art. 4 ust. 3 dyrektywy (UE) 2019/790⁵.

W toku tworzenia LLM do zwielokrotniania utworów chronionych prawem autorskim może dochodzić na kilku etapach:

1) na etapie tworzenia LLM, tj.:

- a) w trakcie pozyskiwania danych zawierających utwory – aby wytrenowanie LLM było możliwe, konieczne jest zgromadzenie znacznej ilości danych. Odbywa się to na wiele sposobów – np. poprzez wewnętrzne gromadzenie danych (wydobywanie ich z zasobów twórcy modelu), *crowdsourcing* (pozyskiwania danych poprzez przekazywanie ich przez inne osoby) lub poprzez zautomatyzowane gromadzenie danych (np. w drodze tzw. *web scrapingu*, polegającego na wydobywaniu danych ze stron internetowych za pomocą różnych narzędzi informatycznych)⁶;
- b) podczas przetwarzania takich utworów w celu uzyskania odpowiedniej jakości danych treningowych dla trenowanego LLM – czynności te obejmują sprawdzanie danych zgromadzonych do treningu pod kątem adekwatności i kompletności, a następnie oczyszczanie danych i formatowanie ich na potrzeby szkolenia⁷;

2) w ramach procesu zwanego zapamiętywaniem lub memoryzacją (ang. *memorisation*) odbywającego się w niektórych LLM – z prowadzonych badań⁸ wynika, że niektóre LLM budowane w oparciu o mechanizmy *deep learning* są podatne na zapamiętywanie (tj. przechowywanie) fragmentów danych treningowych. Przyjmuje się, że tak jak ludzki mózg zapamiętuje fragmenty informacji, aby się uczyć, tak samo robią też LLM;

3) w wyniku generowania treści przez AI po wprowadzeniu przez użytkownika promptu lub promptów – w procesie korzystania z LLM możliwa jest

⁵ Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790 z dnia 17 kwietnia 2019 r. w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym oraz zmiany dyrektyw 96/9/WE i 2001/29/WE, Dz. Urz. UE L 130 z 17.05.2019 r., s. 92–125 (dalej: „dyrektywa DSM”).

⁶ J. Tan, *What Is AI Model Training?*, 30.04.2024, <https://engage-ai.co/what-ai-model-training/> [dostęp 21.05.2024].

⁷ D. O’Keefe, *How Does AI Model Training Work?*, apian, 24.01.2014, <https://appian.com/blog/acp/ai/how-does-ai-model-training-work.html> [dostęp 21.05.2024]; AILabs, *Data Cleaning and AI Model Training in Algorithmic Training*, <https://www.aillabs.global/blog/data-cleaning-and-ai-model-training-in-algorithmic-training> [dostęp 21.05.2024].

⁸ N. Carlini *et al.*, *Extracting Training Data from Large Language Models*, <https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting> [dostęp 21.05.2024].

sytuacja, w której model „zwraca” w swoich generacjach (tj. w outputcie) mniejsze lub większe fragmenty danych treningowych – a więc także utworów (np. tekstów lub grafik)⁹.

Celem niniejszego tekstu jest omówienie zagadnienia zwielokrotniania utworów w kontekście tworzenia i funkcjonowania LLM oraz próba rozstrzygnięcia, czy *de lege lata* takie korzystanie z cudzych utworów wymaga zgody podmiotu uprawnionego, czy też nie wymaga takiej zgody.

1. Kwestie techniczne i terminologiczne

1.1. Trenowanie LLM

Przed przystąpieniem do szczegółowych rozważań w przedmiocie zwielokrotniania utworów (1) w procesie tworzenia LLM, (2) dokonującego się w LLM poprzez ich zapamiętywanie oraz (3) zwielokrotniania odbywającego się w trakcie generowania treści przez AI, a będącego wynikiem zapamiętywania chronionych i w konsekwencji – tzw. regurgitacji treści wchodzącej w skład danych treningowych LLM¹⁰, należy wyjaśnić zagadnienia wstępne, których zrozumienie jest niezbędne dla poprawnego śledzenia toku dalszego wywodu.

LLM potrafią generować np. teksty, obrazy i inne treści. Opracowanie i trenowanie takich modeli wymagają jednak dostępu do ogromnych ilości danych. Dane te są przez algorytmy oparte na uczeniu maszynowym (a konkretniej – jego podkategorii, tj. uczeniu głębokim, ang. *deep learning*) analizowane w procesie zwanym trenowaniem lub uczeniem. Wspomniane algorytmy robią to konsekwentnie poprzez uogólnianie dostarczonych danych i wychwytywanie występujących w dostarczonych danych wzorców i relacji¹¹. Dzięki temu trenowany model może następnie stosować wyuczone na wprowadzonych danych wzorce i zależności, analizując nowe dane, które wcześniej nie były modelowi znane¹² (np. dane wejściowe wprowadzane przez użytkownika modelu).

⁹ Z. Okoń, https://www.linkedin.com/posts/zbigniew-okon_openai-nyt-prawoautorskie-activity-7152621678215278592-jLnF/?originalSubdomain=pl [dostęp 21.05.2024].

¹⁰ *Ibidem*; OpenAI, *OpenAI and journalism*, <https://openai.com/index/openai-and-journalism/> [dostęp 21.05.2024].

¹¹ J. Patterson, A. Gibson, *Deep Learning. Praktyczne wprowadzenie*, Gliwice 2018, s. 19–20.

¹² K. Kunc-Urbańczyk, *Uczenie maszynowe jako pole eksploatacji utworów – 2. Uczenie maszynowe – wyjaśnienie podstawowych pojęć*, [w:] B. Fischer, A. Pązik, M. Świerczyński (red.), *Prawo sztucznej inteligencji i nowych technologii*, Warszawa 2021.

1.2. Wykorzystanie utworów w procesie tworzenia i funkcjonowania LLM jako pole eksploatacji w postaci zwielokrotniania utworów

Omówienie kwestii wstępnych należy zakończyć na kontekstowym omówieniu zagadnień prawnych. Dla zrozumienia analizy przeprowadzonej w ramach niniejszego tekstu kluczowe jest zrozumienie dwóch pojęć charakterystycznych dla prawa autorskiego, tj. „pola eksploatacji” oraz jednego z jego rodzajów – „zwielokrotniania utworów”.

W polskim porządku prawnym pole eksploatacji utworu to sposób korzystania z tego utworu. Ustawowe przykłady pól eksploatacji zostały wskazane w art. 50 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych¹³ i obejmują poszczególne rodzaje (1) utrwalania i zwielokrotniania utworu, (2) obrotu oryginałem albo egzemplarzami oraz (3) rozpowszechniania utworu w sposób inny niż poprzez obrót utworem. Kluczowe z punktu widzenia niniejszej analizy pole eksploatacji (tj. zwielokrotnianie utworu) oznacza wytwarzanie określoną techniką kolejnego lub kolejnych egzemplarzy utworu (względem pierwszego egzemplarza utworu, stworzonego wskutek utrwalenia) – przy zastosowaniu dowolnych technik (np. druku)¹⁴.

Trzeba jednak zaznaczyć, że o ile uznanie pola eksploatacji za sposób korzystania z utworu można uznać za bezsporne, o tyle w doktrynie toczy się dyskusja co do kryteriów wyodrębniania pól eksploatacji innych niż wymienione w art. 50 pr. aut. – jako że przepis ten zawiera ich katalog otwarty¹⁵. W kontekście analizowanego zagadnienia jest to kwestia o tyle istotna, że wyodrębnienie nowego pola eksploatacji skutkuje koniecznością stosowania do niego przepisów art. 41 ust. 2 pr. aut. (umowy prawnoautorskie obejmują pola eksploatacji wyraźnie w nich wymienione) oraz art. 41 ust. 4 pr. aut. (umowa może dotyczyć tylko pól eksploatacji, które są znane w chwili jej zawarcia). W odniesieniu do zarysowanego sporu o wyodrębnianie pól eksploatacji innych niż określone w art. 50 pr. aut. K. Klafkowska-Waśniowska za Elżbietą Traple wskazuje, że pojęcie „pola eksploatacji” wiąże się z formą wykorzystania utworu – odrębną pod względem ekonomicznym lub technicznym¹⁶.

¹³ Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych, Dz. U. z 2022 r. poz. 2509 (dalej: „pr. aut.”).

¹⁴ K. Klafkowska-Waśniowska, [w:] R. Markiewicz (red.), *Ustawy autorskie. Komentarze. Tom I. Komentarz do ustawy o prawie autorskim i prawach pokrewnych*, Warszawa 2021, komentarz do art. 50 pr. aut., teza 25.

¹⁵ K. Gliściński, [w:] A. Michalak (red.), *Ustawa o prawie autorskim i prawach pokrewnych. Komentarz*, Warszawa 2019, komentarz do art. 50 pr. aut., nb 1.

¹⁶ K. Klafkowska-Waśniowska, *op. cit.*, teza 15.

Zdaniem autorów, procesy korzystania z utworów w trakcie tworzenia lub funkcjonowania LLM odbywają się w ramach pola eksploatacji w postaci zwielokrotniania. Należy jednak odnotować, że w doktrynie pojawiły się rozważania co do tego, czy jeden z takich procesów, tj. *machine learning* należy traktować jako swego rodzaju subpole eksploatacji¹⁷. Autorzy przychylają się do tezy, zgodnie z którą, choć z technicznego punktu widzenia możliwe jest uznanie *machine learning'u* za odrębne pole eksploatacji¹⁸, to nie zmienia to jednak faktu, że takie wyodrębnienie nie jest uzasadnione z perspektywy kryterium ekonomicznego. Można zgodzić się ze stwierdzeniem, że *machine learning* pochłania znaczne ilości zasobów i wymaga dużych nakładów finansowych. Należy jednak podkreślić, że *machine learning* jest tylko etapem wytwarzania określonego produktu lub usługi (np. określonego LLM) – autorzy zgadzają się, że dopiero takie produkty lub usługi można oceniać jako posiadające znaczący potencjał ekonomiczny. Sam *machine learning* jako jeden z procesów prowadzących do ich wytworzenia nie ma potencjału wystarczającego, aby uznać go za odrębne pole eksploatacji w rozumieniu art. 50 pr. aut.¹⁹

W konsekwencji poczynionych ustaleń, na potrzeby dalszych rozważań, korzystanie z utworów w ramach omawianych procesów będzie przez autorów postrzegane przez pryzmat pola eksploatacji, jakim jest zwielokrotnianie, dlatego też w dalszej części artykułu przeanalizowane zostanie, kiedy tworzenie i funkcjonowanie LLM prowadzi do zwielokrotniania utworów i czy w danym przypadku zwielokrotnianie to jest objęte dozwolonym użytkowaniem. Dodatkowo, na marginesie powyższych rozważań:

– należy jednak stwierdzić, że uznanie jednego z przejawów korzystania z utworów w procesie tworzenia i funkcjonowania LLM za odrębne od zwielokrotniania pole eksploatacji mogłoby wobec takiego odrębnego pola eksploatacji rodzić wątpliwości co do stosowania do niego przepisów o dozwolonym użytku – jak np. art. 4 dyrektywy DSM (więcej szczegółowych rozważań w dalszej części tekstu);

– w orzecznictwie TSUE, bez względu na legislacyjną siatkę terminologiczną korzystania z utworu przyjętą w prawodawstwach krajów członkowskich Unii Europejskiej, niektóre z pól eksploatacji należy uznać za zharmonizowane – w tym

¹⁷ K. Kunc-Urbańczyk, *Uczenie maszynowe jako pole eksploatacji utworów – 6. Możliwość uznania uczenia maszynowego za odrębne subpole zwielokrotniania metodą cyfrową*, [w:] B. Fischer, A. Pązik, M. Świerczyński (red.), *op. cit.*

¹⁸ *Ibidem.*

¹⁹ *Ibidem.*

właśnie prawo do zwielokrotniania²⁰ (obejmujące również utrwalanie²¹) oraz prawo do komunikowania utworu publicznie²². Dzięki temu rozważania poczynione w niniejszej analizie w oparciu o przepisy polskiego porządku prawnego są przynajmniej w pewnym zakresie aktualne w odniesieniu do porządków prawnych pozostałych krajów członkowskich UE.

2. Zwielokrotnianie utworów w związku z tworzeniem LLM

2.1. Przykłady zwielokrotniania utworów w procesie tworzenia LLM

Jak wspomniano, do zwielokrotniania utworów dochodzi na etapie tworzenia LLM. Aby model został wytrenowany na określonym (i ogromnym) zbiorze danych treningowych, dane te muszą zostać najpierw zgromadzone, a następnie wstępnie przetworzone. Z technicznych aspektów czynności gromadzenia danych wynika, że aby możliwe było „nakarmienie” nimi trenowanego LLM, niezbędne jest uprzednie zwielokrotnienie takich danych (a więc również danych stanowiących utwory chronione prawem autorskim) – choćby poprzez zapisanie ich w pamięci komputera.

Pewne wątpliwości może nasuwać to, czy również przetwarzanie zgromadzonych danych, aby uzyskać pożądaną jakość dane treningowe, powinno być kwalifikowane jako zwielokrotnienie utworów. Surowe dane zgromadzone przed rozpoczęciem trenowania LLM (a więc np. pozyskane z różnych źródeł teksty, grafiki, zdjęcia, pliki wideo itp.) muszą przejść tzw. proces wektoryzacji, polegający na konwersji danych na dane liczbowe, tj. do formatu wektorowego umożliwiającego trenowanie algorytmu²³. W procesie wektoryzacji danych dochodzi np. do oczyszczenia tekstów poprzez usunięcie z nich wielkich liter, znaków specjalnych, symboli, cyfr, znaków interpunkcyjnych, spójników oraz do zapisania słów w podstawowej formie gramatycznej²⁴, przetworzenia obrazów poprzez standaryzację wymiarów, zmniejszenie jakości plików, pominięcie jednolitych fragmentów tła oraz ujednolicenie obszarów podobnych barw i normalizację kolorów²⁵. Istnieją wątpliwości, czy

²⁰ Wyrok Trybunału Sprawiedliwości z dnia 29 lipca 2019 r., C-469/17, *Funke Medien*, Dz. Urz. UE C 319/5 z 2019 r.

²¹ K. Klafkowska-Waśniowska, *op. cit.*, teza 23.

²² Wyrok Trybunału Sprawiedliwości z dnia 13 lutego 2014 r., C-466/12, *Svensson*, ZOTSiS 2014/2/I-76.

²³ K. Kunc-Urbańczyk, *Uczenie maszynowe jako pole eksploatacji utworów – 4.2. Wektoryzacja danych*, [w:] B. Fischer, A. Pązik, M. Świerczyński (red.), *op. cit.*

²⁴ *Ibidem*.

²⁵ *Ibidem*.

tego typu operacje na zgromadzonych danych (w tym danych stanowiących utwory) zawsze muszą wiązać się z ponownym zwielokrotnieniem utworów. Zdaniem autorów możliwe są sytuacje, w których w trakcie procesu wektoryzacji danych będzie dochodziło do zwielokrotniania utworów, ale możliwe też są przypadki, w których wektoryzacja danych nie będzie prowadziła do zwielokrotniania – pośrednio wskazuje na to R. Markiewicz, zauważając, że w szerszym kontekście eksploracji tekstów i danych mogą wystąpić przypadki eksploracji, które nie stanowią zwielokrotniania, a taka eksploracja odpowiada „czytaniu utworu” przez człowieka i nie wkracza w monopol prawnoautorski²⁶.

2.2. Zwielokrotnianie utworów w procesie tworzenia LLM a dozwolony użytek utworów

Aby zwielokrotnianie danych (w tym utworów) w ramach szeroko rozumianego etapu tworzenia LLM (a więc obejmującego gromadzenie danych, ich wektoryzację oraz trenowanie modelu) nie naruszało praw podmiotów uprawnionych, twórcy modelu muszą mieć podstawę prawną do takiego korzystania. Podstawę taką mogą stanowić umowy licencyjne zawarte przez twórców LLM z podmiotami uprawnionymi lub przepisy powszechnie obowiązującego prawa. Przedmiotem dalszych rozważań będzie drugi ze wskazanych przypadków.

Wobec tego, że zakres uprawnienia do zwielokrotniania utworów (prawa twórców do zezwalania lub zabrania takiego zwielokrotniania) jest w prawie unijnym (a konsekwentnie także w polskim) ujmowany maksymalnie szeroko²⁷ (por. brzmienie art. 2 dyrektywy 2001/29/WE²⁸), nie ulega wątpliwości, że zwielokrotnianie utworów w związku z tworzeniem LLM co do zasady wkracza w monopol prawnoautorski podmiotów uprawnionych. W doktrynie prawa rozważano, czy podstawą prawną do takiego zwielokrotniania utworów bez zgody podmiotów uprawnionych mogą być przepisy dyrektywy DSM²⁹. Na ten moment, tj. po zmianach w proponowanej treści AI Act (w wersji przyjętej 13 marca 2024 r. przez

²⁶ R. Markiewicz, 2.2. *Eksploracja tekstów i danych (art. 3 i 4)*, [w:] *idem* (red.), *Prawo autorskie na jednolitym rynku cyfrowym. Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, Warszawa 2021.

²⁷ Z. Okoń, *Wykorzystanie utworów dla potrzeb głębokiego uczenia w świetle europejskiego prawa autorskiego* [w:] K. Flaga-Gieruszyńska, J. Gołaczyński, D. Szostek (red.), *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, Warszawa 2019.

²⁸ Dyrektywa 2001/29/WE Parlamentu Europejskiego i Rady z dnia 22 maja 2001 r. w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym, Dz. Urz. UE L 167 z 22.06.2001 r., s. 10–19 (dalej: „dyrektywa InfoSoc”).

²⁹ Z. Okoń, *op. cit.*

Parlament Europejski i w wersji przyjętej 14 maja 2024 r. przez Radę Europejską), a następnie na gruncie ostatecznie przyjętej wersji tekstu AI Act, zdaniem autorów istnieje możliwość objęcia zwielokrotniania dokonującego się w procesie tworzenia LLM (od etapu gromadzenia danych treningowych do etapu samego trenowania modelu) zakresem dozwolonego użytku utworów z art. 3 i art. 4 dyrektywy DSM w ramach eksploracji tekstów i danych. Wskazuje na to motyw 105 ostatecznej wersji tekstu Aktu o sztucznej inteligencji, który stanowi, że „[...] Rozwój i trenowanie takich modeli wymaga dostępu do ogromnych ilości tekstów, obrazów, materiałów wideo i innych danych. Techniki eksploracji tekstów i danych mogą być w tym kontekście szeroko wykorzystywane do wyszukiwania i analizy takich treści, które mogą być chronione prawem autorskim i prawami pokrewnymi. Każde wykorzystanie treści chronionych prawem autorskim wymaga zezwolenia danego podmiotu praw, chyba że zastosowanie mają odpowiednie wyjątki i ograniczenia dotyczące praw autorskich. Na mocy dyrektywy (UE) 2019/790 wprowadzono wyjątki i ograniczenia umożliwiające, pod pewnymi warunkami, zwielokrotnianie i pobieranie utworów lub innych przedmiotów objętych ochroną do celów eksploracji tekstów i danych. Zgodnie z tymi przepisami podmioty uprawnione mogą zastrzec swoje prawa do swoich utworów lub innych przedmiotów objętych ochroną, aby zapobiec eksploracji tekstów i danych, chyba że odbywa się to do celów badań naukowych. W przypadku gdy prawo do wyłączenia z eksploracji zostało w odpowiedni sposób wyraźnie zastrzeżone, dostawcy LLM muszą uzyskać zezwolenie od podmiotów uprawnionych, jeżeli chcą dokonywać eksploracji tekstów i danych odnośnie do takich utworów”. Jest to również akceptowane przez polską doktrynę prawa, w której wskazuje się, że w ramach eksploracji tekstów i danych mieszczą się takie czynności (etapy) jak:

- 1) identyfikacja odpowiednich danych i uzyskanie do nich dostępu;
- 2) kopiowanie danych w celu ich późniejszej analizy;
- 3) przetworzenie danych do odpowiedniego formatu;
- 4) ekstrakcja tekstów i danych;
- 5) analiza danych w celu identyfikacji związków, zależności, wzorów lub tendencji;
- 6) wykorzystanie rezultatów analizy, tj. użycie danych, przedstawienie rezultatów analizy, weryfikacja wyników (w tym wraz z fragmentami analizowanych tekstów)³⁰.

³⁰ R. Markiewicz, *op. cit.*

Zanim objęcie tworzenia LLM zakresem dozwolonego użytku z dyrektywy DSM zostało szerzej zaakceptowane, w doktrynie rozważano, czy podstawą do zwielokrotniania utworów w takim procesie może być przepis art. 5 dyrektywy InfoSoc. Zgodnie z tym przepisem tymczasowe zwielokrotnianie, mające charakter przejściowy lub dodatkowy, stanowiące integralną i podstawową część procesu technologicznego, którego jedynym celem jest umożliwienie legalnego korzystania z utworu i które nie ma odrębnego znaczenia ekonomicznego, jest wyłączone z zakresu przysługującego twórcy wyłącznego prawa do zwielokrotniania. Za dopuszczalne uznane zostało jednak zwielokrotnianie dokonywane w określonym przebiegu procesu technicznego, tj. pobieranie utworów stanowiących surowe dane, ich przetwarzanie do postaci danych treningowych, a następnie ich przetwarzanie przez sieć neuronową w procesie trenowania i usuwanie po zakończeniu tego procesu, które przebiegało w sposób zautomatyzowany³¹. Rozważania te można jednak uznać obecnie za drugorzędne w związku z przyjęciem dyrektywy DSM oraz brzmieniem ostatecznej wersji tekstu AI Act.

Na marginesie warto wspomnieć, że zarysowane wątpliwości towarzyszyły również polskiemu ustawodawcy, który w 2024 r. podjął kolejną próbę implementacji dyrektywy DSM do polskiego porządku prawnego. W projekcie ustawy z dnia 14 lutego 2024 r.³² proponowane brzmienie art. 26³ ust. 1 pr. aut. zakładało, że nie wolno zwielokrotniać rozpowszechnionych utworów w celu eksploracji tekstów i danych na potrzeby tworzenia generatywnych modeli sztucznej inteligencji, chyba że uprawniony wyraźnie wyraził na to zgodę. Z kolei projekt ustawy przekazany do prac legislacyjnych 16 maja 2024 r.³³ i ostateczna treść uchwalonej ustawy³⁴ dopuszczają możliwość eksploracji tekstów i danych w celu tworzenia generatywnych modeli sztucznej inteligencji – chyba że uprawniony wyraźnie sprzeciwi się takiemu korzystaniu z jego utworów. Zmiana brzmienia propozycji i ostatecznej wersji polskich przepisów implementujących dyrektywę DSM jest w naszej ocenie konsekwencją wskazanych wyżej zmian, jakim podlegały kolejne wersje projektu AI Act.

³¹ Z. Okoń, *op. cit.*

³² Projekt ustawy z dnia 14 lutego 2024 r. o zmianie ustawy o prawie autorskim i prawach pokrewnych oraz niektórych innych ustaw, <https://legislacja.rcl.gov.pl/docs//2/12382002/13037394/13037395/dokument656778.pdf> [dostęp 22.05.2024].

³³ Rządowy projekt ustawy o zmianie ustawy o prawie autorskim i prawach pokrewnych oraz niektórych innych ustaw, wniesiony do Sejmu 16 maja 2024, <https://www.sejm.gov.pl/sejm10.nsf/agent.xsp?symbol=RPL&Id=RM-0610-36-24> [dostęp 22.05.2024].

³⁴ Ustawa z dnia 26 lipca 2024 r. o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi, Dz. U. z 2024 r. poz. 1254.

3. Zwielokrotnianie w ramach zjawiska zapamiętywania

3.1. Przykłady zwielokrotniania w ramach zjawiska zapamiętywania

Jak wskazują Ivo Emanuilov i Thomas Margoni, powołując się na badania przeprowadzone przez naukowców zrzeszonych w USENIX Association, zostało empirycznie wykazane, że LLM mogą wykazywać zachowanie podobne do tego, co w odniesieniu do pracy ludzkiego mózgu nazywamy pamięcią fotograficzną lub ejdetyczną, tj. zdolnością do przywoływania informacji po ich jednokrotnym obejrzeniu³⁵. Należy w tym miejscu podkreślić za ww. autorami, że nie każdy algorytm uczenia maszynowego zachowuje się w taki sam sposób i istnieją przypadki, w których zapamiętywanie jest zamierzoną cechą procesu uczenia modelu³⁶. Przedmiotem niniejszej analizy jest jednak takie zjawisko zapamiętywania, które stanowi niezamierzoną cechę tego procesu. Innymi słowy, przedmiotem analizy jest zapamiętywanie przez LLM danych treningowych niezbędne do tego, aby LLM dokonywały bardziej efektywnego uogólniania (tj. zdolności LLM do adaptacji i prawidłowego reagowania na nieznanne mu wcześniej dane, które nie wchodziły w zakres danych treningowych)³⁷.

W związku z poczynionym spostrzeżeniem, że wynikający z art. 2 dyrektywy InfoSoc zakres prawa twórców do zezwalania lub zabrania zwielokrotniania utworów jest w prawie unijnym ujmowany szeroko, należy uznać, że zwielokrotnianie utworów w toku procesu zapamiętywania wkracza w monopol prawnoautorski. Po raz kolejny pojawia się zatem pytanie, czy takie zwielokrotnianie może odbywać się w granicach dozwolonego użytku wynikającego z przepisów prawa.

3.2. Zwielokrotnianie utworów w ramach zjawiska zapamiętywania a dozwolony użytek utworów

W pierwszej kolejności należy zastanowić się, czy zwielokrotnianie utworów w procesie zapamiętywania jest objęte zakresem dozwolonego użytku wynikającego z dyrektywy DSM. Jak wynika z przywołanych wcześniej art. 3 i art. 4 dyrektywy DSM, przepisy te obejmują zwielokrotnianie i pobieranie utworów do celów

³⁵ I. Emanuilov, T. Margoni, *Forget me not: memorisation in generative sequence models trained on open source licensed code*, 27.02.2024, s. 11, <https://zenodo.org/records/10635479>; [dostęp 20.05.2024].

³⁶ *Ibidem*.

³⁷ L. Goetz, N. Seedat, R. Vandersluis, *Generalization – a key challenge for responsible AI in patient-facing clinical applications*, „Nature. npj Digital Medicine” 2024, 7, 126, <https://www.nature.com/articles/s41746-024-01127-3> [dostęp 22.05.2024].

eksploracji tekstów i danych. Do rozstrzygnięcia pozostaje zatem, czy zwielokrotnianie utworów w procesie ich zapamiętywania przez LLM (ale przed ich zwielokrotnieniem w formie outputu – o czym niżej) stanowi zwielokrotnienie mieszczące się w sformułowaniu „do celów eksploracji tekstów i danych”.

Zdaniem autorów, zwielokrotnianie utworów odbywające się od momentu gromadzenia surowych danych treningowych, przez proces wektoryzacji danych, aż do momentu zakończenia trenowania LLM (na określonych danych – utworach) wchodzi w zakres dozwolonego użytku związanego z eksploracją tekstów i danych. Autorzy twierdzą jednak, że zwielokrotnianie utworów odbywające się przez to, że wytrenowany LLM zapamiętał określone elementy danych treningowych, takim dozwolonym użytkowaniem nie jest już objęte. Tezę tę autorzy opierają przede wszystkim na założeniu, że dozwolone zwielokrotnianie wynikające z art. 3 i art. 4 dyrektywy DSM stanowi wyjątek, który zgodnie z zasadą *exceptiones non sunt extendendae* nie powinien być wykładany rozszerzająco. Takim nadużyciem byłoby z kolei stwierdzenie, że zwielokrotnianie odbywające się faktycznie już po procesie konwersji danych treningowych do formatu odczytywalnego przez sieć neuronową i po procesie trenowania takiej sieci nadal jest objęte zakresem dozwolonego użytku wynikającego z omawianych przepisów.

Podobne spostrzeżenie należy poczynić względem „wyrwy” tworzonej w monopolu prawnoautorskim przez art. 5 ust. 1 lit. b) dyrektywy InfoSoc. Jak wspomniano, przewiduje on, że tymczasowe zwielokrotnianie, mające charakter przejściowy lub dodatkowy, które stanowi integralną i podstawową część procesu technologicznego i którego jedynym celem jest umożliwienie legalnego korzystania z utworu i które nie ma odrębnego znaczenia ekonomicznego, jest wyłączone z zakresu przysługującego twórcy wyłącznego prawa do zwielokrotniania. Nie sposób jednak uznać, że zwielokrotnianie odbywające się wskutek zapamiętania danych treningowych przez LLM ma charakter przejściowy lub dodatkowy. Ma to związek ze stwierdzeniem, że zapamiętywanie mniejszych lub większych fragmentów danych treningowych po pierwsze – może stanowić trwałe zakodowanie danych w LLM i po drugie – ma wpływ na zwiększenie efektywności zdolności uogólniania, tj. na zmniejszenie niedokładności modelu w przewidywaniu wartości wyników na nieznanym mu dotąd danych. Konsekwentnie nie można stwierdzić, że zwielokrotnianie nietymczasowe i nieobjęte automatycznego usuwania utworów nie jest objęte zakresem dozwolonego użytku przewidzianym w art. 5 ust. 1 lit. b) dyrektywy InfoSoc³⁸.

³⁸ R. Markiewicz, *ChatGPT i prawo autorskie Unii Europejskiej*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2023, nr 2(160), s. 152.

4. Zwielokrotnianie w ramach zjawiska regurgitacji treści przez LLM

4.1. Przykłady zwielokrotniania w ramach zjawiska regurgitacji treści przez LLM

Konsekwencją opisanego zjawiska zapamiętywania utworów wchodzących w skład danych treningowych są przypadki, gdy po wprowadzeniu odpowiedniego promptu lub promptów generacja tworzona przez LLM zawiera całe fragmenty danych, na których model został wytrenowany.

Na gruncie sporu „New York Times” przeciwko OpenAI, zjawisko takie zostało przez pozwanego dostawcę ChataGPT³⁹ nazwane „regurgitacją”⁴⁰ – znanym z medycyny określeniem zjawiska niekontrolowanego cofania się treści pokarmowej z żołądka do przełyku⁴¹. W toku sporu „New York Times” wykazuje, że empirycznie zweryfikował, że po wprowadzeniu do promptu początkowych fragmentów swoich artykułów uzyskiwał w outpucie generowanym przez ChatGPT praktycznie identyczne fragmenty swoich tekstów. Według wydawcy oznacza to, że – po pierwsze – LLM stworzony przez OpenAI został wytrenowany na tekstach, do których prawa autorskie przysługują gazecie i – po drugie – że możliwe są przypadki zwielokrotniania tych tekstów w treściach generowanych przez sztuczną inteligencję⁴².

4.2. Zwielokrotnianie utworów w ramach zjawiska regurgitacji treści przez LLM

Zdaniem autorów przykłady zwielokrotniania utworów wskazane w pkt 5.1. pozostawiają najmniej wątpliwości co do ich nieuprawnionego charakteru (zakładając oczywiście, że w danym przypadku nie udzielono odpowiedniej licencji). Z pewnością bowiem takie zwielokrotnianie nie może być objęte zakresem dozwolonego użytku, czy to z art. 3 lub art. 4 dyrektywy DSM (generowanie treści nie stanowi bo-

³⁹ R. Metz, *OpenAI Claims New York Times Isn't 'Telling the Full Story' in Suit*, Bloomberg, 8.01.2024, <https://www.bloomberg.com/news/articles/2024-01-08/openai-says-new-york-times-not-telling-the-full-story-in-suit?embedded-checkout=true> [dostęp 22.05.2024].

⁴⁰ Tak też w referacie *Naruszenie prawa autorskiego w wytworach generatywnej sztucznej inteligencji*, ogłoszonym przez Z. Okonia 22.03.2024 r. podczas 10. Forum Prawa Mediów Elektronicznych.

⁴¹ J. Ryzko, *Gastroenterologia pediatryczna – pytania i odpowiedzi*, „Gastroenterologia Praktyczna” 2012, nr 1, s. 81–82.

⁴² M. Bastian, *OpenAI claims New York Times' prompting strategy violates its terms of service*, 6.01.2024, <https://the-decoder.com/openai-claims-new-york-times-prompting-strategy-violates-its-terms-of-service/> [dostęp 22.05.2024].

wiem eksploracji tekstów i danych), czy to z art. 5 ust. 1 lit. b) dyrektywy InfoSoc – nie jest to wszakże zwielokrotnianie, które ma charakter przejściowy lub dodatkowy.

Na dodatek generowanie treści przez LLM wskutek tzw. regurgitacji, nawet jeśli stanowiłoby dozwolone zwielokrotnianie (o czym, jak wspomniano, w opinii autorów nie może być mowy), to i tak stanowiłoby to wykraczające poza ramy dozwolonego użytku (zarówno z dyrektywy DSM, jak i z dyrektywy InfoSoc) korzystanie z utworu poprzez ich publiczne udostępnianie⁴³.

5. Wnioski

Podsumowując, rozpatrywanie zagadnienia zwielokrotniania utworów w procesach tworzenia i funkcjonowania LLM prowadzi do następujących wniosków:

- do zwielokrotniania danych stanowiących utwory może dochodzić na różnych etapach tych procesów, tj. (1) podczas zbierania danych treningowych, (2) w trakcie procesu wektoryzacji danych, (3) w toku dalszego trenowania LLM, (4) wskutek zapamiętywania utworów zawartych w danych treningowych, aż po (4) możliwą regurgitację utworów w treściach generowanych przez LLM;

- nie każdy z ww. przypadków zwielokrotniania jest objęty zakresem dozwolonego użytku, a więc potencjalnie może naruszać prawa podmiotów uprawnionych;

- zdaniem autorów zwielokrotnienie utworów odbywające się od momentu gromadzenia surowych danych treningowych, przez proces wektoryzacji danych, aż do momentu zakończenia trenowania LLM na określonych danych (utworach) wchodzi w zakres dozwolonego użytku związanego z eksploracją tekstów i danych;

- w ocenie autorów zwielokrotnianie utworów odbywające się przez to, że wytrenowany LLM zapamiętał określone elementy danych treningowych, nie jest objęte dozwolonym użytkowaniem wynikającym z art. 3 i art. 4 dyrektywy DSM (i konsekwentnie – art. 26² oraz art. 26³ polskiej ustawy o prawie autorskim i prawach pokrewnych znowelizowanej w lipcu 2024 r. ustawą implementującą dyrektywę DSM do polskiego porządku prawnego⁴⁴). Przepisy te ustanawiają wyjątek od monopolu prawnoautorskiego, który nie powinien być wykładany rozszerzająco;

- zdaniem autorów zwielokrotnianie danych stanowiących utwory, odbywające się wskutek zapamiętania danych treningowych przez LLM, nie ma charakteru

⁴³ I. Emanuilov, T. Margoni, *Memorisation...*

⁴⁴ Ustawa z dnia 26 lipca 2024 r. o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi, Dz. U. z 2024 r. poz. 1254.

przejściowego lub dodatkowego, a także nie obejmuje automatycznego usuwania utworów – dlatego też nie jest objęte zakresem dozwolonego użytku przewidzianym w art. 5 ust. 1 lit. b) dyrektywy InfoSoc;

– w ocenie autorów zwielokrotnianie utworów w ramach zjawiska regurgitacji treści przez LLM nie może być objęte zakresem dozwolonego użytku ani z art. 3 lub art. 4 dyrektywy DSM (generowanie treści nie stanowi bowiem eksploracji tekstów i danych), ani z art. 5 ust. 1 lit. b) dyrektywy InfoSoc (nie jest to zwielokrotnianie, które ma charakter przejściowy lub dodatkowy).

Należy również dodać, że wnioski płynące z analizy mają charakter wyjątkowo praktyczny – choćby z uwagi na znaczną liczbę sporów sądowych, w których dostawcy LLM są pozywani przez kolejnych twórców (por. przywołany w tekście przykład sporu „New York Times” przeciwko OpenAI). Autorzy pragną jednak przypomnieć, że są to spory sądowe odbywające się na gruncie prawa amerykańskiego i w tradycji prawa *common law*. Konsekwentnie, ewentualne rozstrzygnięcia nie wpłyną na wykładnię omawianych w tekście przepisów dyrektywy DSM i dyrektywy InfoSoc. Z takiego założenia wychodzą również najprawdopodobniej dostawcy LLM, którzy nie chcąc ryzykować zmniejszenia popytu na ich usługi (do czego mogłoby dojść, jeśli użytkownicy narzędzi AI obawialiby się, że korzystanie przez nich z tych narzędzi mogłoby uzasadniać kierowanie przeciwko nim roszczeń z tytułu praw autorskich), wprowadzają w swoich warunkach umów tzw. klauzule indemnifikacyjne, w których, pod pewnymi warunkami, zobowiązują się zwolnić użytkownika z odpowiedzialności zarówno za ewentualne naruszenia cudzych praw w związku z danymi wykorzystanymi do trenowania ich LLM, jak i w związku z naruszeniami związanymi z generacją treści przez AI⁴⁵.

Bibliografia

Akty prawne

Dyrektywa 2001/29/WE Parlamentu Europejskiego i Rady z dnia 22 maja 2001 r. w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym (Dz. Urz. UE L 167 z 22.06.2001 r., s. 10–19).

⁴⁵ Por. np. Copilot Copyright Commitment firmy Microsoft, <https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/> [dostęp 22.05.2024] lub Szczegółowe warunki korzystania z usługi Google Workspace, <https://workspace.google.com/terms/service-terms/> [dostęp 22.05.2024].

- Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790 z dnia 17 kwietnia 2019 r. w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym oraz zmiany dyrektyw 96/9/WE i 2001/29/WE (Dz. Urz. UE L 130 z 17.05.2019 r., s. 92–125).
- Projekt ustawy z dnia 14 lutego 2024 r. o zmianie ustawy o prawie autorskim i prawach pokrewnych oraz niektórych innych ustaw z dnia 14 lutego 2024 r., <https://legislacja.rcl.gov.pl/docs//2/12382002/13037394/13037395/dokument656778.pdf> [dostęp 22.05.2024 r.].
- Propozycja Rozporządzenia Parlamentu Europejskiego i Rady w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji), tekst w wersji przyjętej w dniu 14 maja 2024 r., 2021/0106(COD), PE-CONS 24/24, <https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/pl/pdf> [dostęp 21.05.2024].
- Rezolucja ustawodawcza Parlamentu Europejskiego z dnia 13 marca 2024 r. w sprawie wniosku dotyczącego rozporządzenia Parlamentu Europejskiego i Rady ustanawiającego zharmonizowane przepisy dotyczące sztucznej inteligencji (akt w sprawie sztucznej inteligencji) i zmieniającego niektóre akty ustawodawcze Unii (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_PL.pdf [dostęp 17.05.2024].
- Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) (Dz. U. L 2024/1689 z 12.07.2024 r.).
- Rządowy projekt ustawy o zmianie ustawy o prawie autorskim i prawach pokrewnych oraz niektórych innych ustaw, wniesiony do Sejmu w dniu 16 maja 2024 r., <https://www.sejm.gov.pl/sejm10.nsf/agent.xsp?symbol=RPL&Id=RM-0610-36-24>, [dostęp 22.05.2024].
- Ustawa z dnia 26 lipca 2024 r. o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi (Dz. U. z 2024 r. poz. 1254).
- Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz. U. z 2022 r. poz. 2509, t.j. z dnia 6.12.2022 r.).

Orzecznictwo

- Wyrok Trybunału Sprawiedliwości z dnia 13 lutego 2014 r., C-466/12, *Svensson*, ZOTSiS 2014/2/I-76
- Wyrok Trybunału Sprawiedliwości z dnia 29 lipca 2019 r., C-469/17, *Funke Medien* (Dz. Urz. UE C 319/5 z 23.09.2019 r.).

Źródła

- Bastian M., *OpenAI claims New York Times' prompting strategy violates its terms of service*, 6.01.2024, <https://the-decoder.com/openai-claims-new-york-times-prompting-strategy-violates-its-terms-of-service/> [dostęp 22.05.2024].
- Carlini N, Tramèr F, Wallace E., Jagielski M., Herbert-Voss A., Lee K., Roberts A., Brown T., Erlingsson D.U., Oprea A., Raffel C, *Extracting Training Data from Large Language Models*, <https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting> [dostęp 21.05.2024].

- Copilot Copyright Commitment firmy Microsoft, <https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/> [dostęp 22.05.2024].
- Emanuilov I., Margoni T., *Forget me not: memorisation in generative sequence models trained on open source licensed code*, 27.02.2024, <https://zenodo.org/records/10635479> [dostęp 20.05.2024].
- Emanuilov I., Margoni T., *Memorisation in generative models and EU copyright law: an interdisciplinary view*, Kluwer Copyright Blog, 26.03.2024, <https://copyrightblog.kluweriplaw.com/2024/03/26/memorisation-in-generative-models-and-eu-copyright-law-an-interdisciplinary-view/> [dostęp 21.05.2024].
- Goetz L., Seedat N., Vandersluis R., *Generalization – a key challenge for responsible AI in patient-facing clinical applications*, „Nature. npj Digital Medicine” 2024, 7, 126, <https://www.nature.com/articles/s41746-024-01127-3> [dostęp 22.05.2024].
- Metz R., *OpenAI Claims New York Times Isn't 'Telling the Full Story' in Suit*, Bloomberg, 8.01.2024, <https://www.bloomberg.com/news/articles/2024-01-08/openai-says-new-york-times-not-telling-the-full-story-in-suit?embedded-checkout=true>, [dostęp: 22.05.2024].
- O’Keefe D., *How Does AI Model Training Work?*, appian, 24.01.2024, <https://appian.com/blog/acp/ai/how-does-ai-model-training-work.html> [dostęp 21.05.2024].
- Okoń Z., https://www.linkedin.com/posts/zbigniew-okon_openai-nyt-prawoautorskie-activity-7152621678215278592-jLnF/?originalSubdomain=pl [dostęp 21.05.2024].
- Okoń Z., *Naruszenie prawa autorskiego w wytworach generatywnej sztucznej inteligencji*, referat wygłoszony podczas 10. Forum Prawa Mediów Elektronicznych.
- Szczegółowe warunki korzystania z usługi Google Workspace, <https://workspace.google.com/terms/service-terms/> [dostęp 22.05.2024].
- Tan J., *What Is AI Model Training?*, 30.04.2024, <https://engage-ai.co/what-ai-model-training/> [dostęp 21.05.2024].

Literatura

- Glücksniński K., [w:] A. Michalak (red.), *Ustawa o prawie autorskim i prawach pokrewnych. Komentarz*, Warszawa 2019, komentarz do art. 50 pr. aut.
- Klaffkowska-Waśniowska K., [w:] R. Markiewicz (red.), *Ustawy autorskie. Komentarze. Tom I. Komentarz do ustawy o prawie autorskim i prawach pokrewnych*, Warszawa 2021, komentarz do art. 50 pr. aut.
- Kunc-Urbańczyk K., *Uczenie maszynowe jako pole eksploatacji utworów – 2. Uczenie maszynowe – wyjaśnienie podstawowych pojęć*, [w:] B. Fischer, A. Pązik, M. Świerczyński (red.), *Prawo sztucznej inteligencji i nowych technologii*, Warszawa 2021.
- Kunc-Urbańczyk K., *Uczenie maszynowe jako pole eksploatacji utworów – 4.2. Wektoryzacja danych*, [w:] B. Fischer, A. Pązik, M. Świerczyński (red.), *Prawo sztucznej inteligencji i nowych technologii*, Warszawa 2021.
- Kunc-Urbańczyk K., *Uczenie maszynowe jako pole eksploatacji utworów – 6. Możliwość uznania uczenia maszynowego za odrębne subpole zwielokrotniania metodą cyfrową*, [w:] B. Fischer, A. Pązik, M. Świerczyński (red.), *Prawo sztucznej inteligencji i nowych technologii*, Warszawa 2021.
- Markiewicz R., 2.2. *Eksploracja tekstów i danych (art. 3 i 4)*, [w:] R. Markiewicz (red.), *Prawo autorskie na jednolitym rynku cyfrowym. Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, Warszawa 2021.

- Markiewicz R., *ChatGPT i prawo autorskie Unii Europejskiej*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2023, nr 2(160).
- Okoń Z., *Wykorzystanie utworów dla potrzeb głębokiego uczenia w świetle europejskiego prawa autorskiego*, [w:] K. Flaga-Gieruszyńska, J. Gołaczyński, D. Szostek (red.), *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, Warszawa 2019.
- Patterson J., Gibson A., *Deep Learning. Praktyczne wprowadzenie*, Gliwice 2018.
- Ryżko J., *Gastroenterologia pediatryczna – pytania i odpowiedzi*, „Gastroenterologia Praktyczna” 2012, nr 1.

Reproduction of works in the processes of creation and functioning of large language models

Abstract

With regard to so-called large language models, one of the main issues is the threat to the exclusive rights of creators – and in particular – copyright. The aforementioned threat materialises both at the stage of the creation of the LLM and at the stage of the subsequent functioning of such model. On the subject of the use of others’ works, the disputes in the legal doctrine are mainly related to the scope of fair use of works within the framework of text and data mining, to the so-called phenomenon of memorisation of training data by LLM and to the phenomenon of so-called regurgitation, i.e. the “return” of fragments of protected training data by LLM in generated content. The disputes indicated were fuelled by the procedure of the latest AI legislation, i.e. primarily the Artificial Intelligence Act, of which subsequent versions appeared in Q1 2024 – and the latest version of which was adopted on 13th of June 2024 – including references to provisions on text and data mining under the Digital Single Market Copyright Directive. The purpose of this text is to analyse the different stages of the creation and operation of LLM and to verify the reproduction of protected works that takes place during these stages in the context of the possible application of the fair use provisions.

Keywords

AI, work, exploitation, fair use, text, data, mining, TDM

Daniel Kula

Uniwersytet Wrocławski

Wydział Prawa, Administracji i Ekonomii

ORCID: 0009-0007-4911-0488

Web scraping na potrzeby szkolenia sztucznej inteligencji a sytuacja autora

Streszczenie

W dobie sztucznej inteligencji (AI) i powszechnego web scrapingu dzieła dostępne w internecie są codziennie kopiowane na potrzeby zbierania danych treningowych do szkolenia modeli AI. W USA, skąd wiele przedsiębiorstw dostarcza rozwiązania oparte na AI, obserwujemy wzrost liczby postępowań sądowych związanych z tym problemem. Dostawcy tych rozwiązań bronią się, powołując doktrynę *fair use*. W tożsamych przypadkach w Europie argumentacja najprawdopodobniej skupi się na wyjątkach od praw wyłącznych wprowadzonych przez prawodawcę unijnego w dyrektywie Parlamentu Europejskiego i Rady (UE) 2019/790. Analiza wprowadzonych dyrektywą wyjątków prowadzi jednak do wniosku, że regulacje te w istocie nie pozostają optymalne ani z perspektywy podmiotów uprawnionych, ani scrapujących.

Słowa kluczowe

web scraping, eksploracja tekstów i danych, dyrektywa DSM, dozwolony użytek, sztuczna inteligencja, uczenie maszynowe, głębokie uczenie, wielkie modele językowe, zastrzeżenie korzystania, uprawnienia autora

Web scraping – odpowiedź na potrzeby społeczeństwa informacyjnego?

Niewątpliwie społeczeństwo informacyjne cechuje się wysokim zapotrzebowaniem na dostęp do danych. Wraz z rozwojem technologii informacyjnych wytwarzamy i publikujemy w internecie więcej danych niż kiedykolwiek, a rok do roku tempo wytwarzania i publikacji wzrasta. Według [WorldWideWebSize.com](https://www.worldwide-web-size.com) na dzień 30 kwietnia 2024 r. w ramach sieci WWW (World Wide Web) indeksuje się ponad 5 mld stron internetowych¹.

¹ WorldWideWebSize.com. Daily estimated size of the world wide web, [https://www.worldwide-web-size.com/#:~:text=Web \(The Internet\)](https://www.worldwide-web-size.com/#:~:text=Web%20(The%20Internet)) [dostęp 30.04.2024].

Zettabajty danych dostępne w internecie stanowią obiekt zainteresowań wielu podmiotów trudniących się agregacją i analizą tych ogromnych zestawów informacji w każdym możliwym celu. Od statystyki i badań naukowych, przez internetowe porównywarki cen biletów lotniczych, po producentów rozwiązań opartych na AI – wszyscy chcą mieć dostęp do najświeższych i najdokładniejszych danych partykularnych kategorii. Nierzadko są to dane prawnie chronione – dane osobowe, utwory w rozumieniu prawa autorskiego czy bazy danych będących przedmiotem praw *sui generis*. Współcześnie najpopularniejsze metody „wydobycia” danych z internetu określa się terminem web scrapingu.

Web scraping (z ang. skrobanie sieci) to proces, który na potrzeby niniejszego artykułu w uproszczeniu można zdefiniować jako technikę ekstrakcji danych z internetu i zapisywanie ich w systemie plików lub bazie danych dla późniejszego odtwarzania i analizy². Z uwagi na objętość, różnorodność oraz tempo publikacji nowych treści praktycznie niemożliwe jest ręczne pozyskiwanie danych przez człowieka, zwłaszcza jeśli dane podlegają ciągłej aktualizacji³. Wymusza to zastosowanie odpowiednich rozwiązań technicznych w postaci botów scrapingowych – skryptów, których zadaniem jest eksploracja stron internetowych i ekstrakcja określonej kategorii danych lub całej zawartości stron internetowych.

Rola web scrapingu w funkcjonowaniu wielu działów nowoczesnej gospodarki nie może zostać przeceniona. Jeden z najbardziej dynamicznie rozwijających się w ostatnich latach sektorów jest immanentnie związany z tą metodą ekstrakcji danych z internetu – AI, w szczególności rozwiązania oparte na ML (*Machine Learning* – z ang. uczenie maszynowe) oraz jego dalszych iteracjach – DL (*Deep Learning* – z ang. uczenie głębokie) czy LLM (*Large Language Models* – z ang. wielkie modele językowe). Te zaawansowane algorytmy szkoli się w celu osiągnięcia zamierzonej funkcjonalności przy użyciu danych treningowych. Rozwiązania oparte na ML wymagają odpowiedniej ilości danych do pomyślnego szkolenia.

Pokłosie beztroskiego web scrapingu jest widoczne w licznych postępowaniach sądowych toczących się przeciwko przedsiębiorstwom wprowadzającym na rynek produkty oparte na AI, których głównym elementem pozostaje kwestia naruszenia

² B. Zhao, *Web Scraping*, [w:] L.A. Schintler, C.L. McNeely (red.), *Encyclopedia of Big Data*, Springer International Publishing, 2018, s. 1, https://www.researchgate.net/publication/317177787_Web_Scraping [dostęp 30.04.2024].

³ V. Krotov, L. Silva, *Legality and ethics of web scraping*, [w:] *Twenty-fourth Americas Conference on Information Systems*, New Orleans 2018, s. 2, https://www.researchgate.net/publication/324907302_Legality_and_Ethics_of_Web_Scraping [dostęp 30.04.2024].

praw autorskich do utworów wykorzystywanych do szkolenia modeli⁴. Głównym argumentem pozwanych w tych postępowaniach jest odwołanie do doktryny *fair use*, której najbliższym, ale jednak nietożsamym odpowiednikiem pozostaje funkcjonujący w prawie kontynentalnym dozwolony użytek⁵.

Niniejszy artykuł ma na celu zbadanie, w jakim zakresie praktyka web scrapingu, obejmująca utwory oraz inne przedmioty objęte ochroną na gruncie prawa własności intelektualnej, może naruszać zastrzeżone prawa podmiotów uprawnionych. Analiza koncentruje się na ocenie, w jaki sposób wyjątki i ograniczenia wprowadzone przez dyrektywę Parlamentu Europejskiego i Rady (UE) 2019/790 z 17.04.2019 r. w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym oraz zmiany dyrektyw 96/9/WE oraz 2001/29/WE⁶ wpływają na zdolność podmiotów udostępniających swoje utwory lub inne przedmioty objęte ochroną w internecie do skutecznego zastrzeżenia swoich praw wobec praktyk web scrapingu. Artykuł dąży do ukazania, w jakim stopniu istniejące regulacje prawne uniemożliwiają podmiotom uprawnionym ochronę swoich praw przed nieautoryzowanym pozyskiwaniem i wykorzystywaniem utworów i innych przedmiotów objętych ochroną, w szczególności wprowadzenie zastrzeżenia za pomocą środków nadających się do odczytu maszynowego.

Dozwolony użytek w prawie polskim a *web scraping*

Istota dozwolonego użytku sprowadza się do możliwości korzystania z chronionego utworu bez konieczności uzyskania odrębnej zgody uprawnionego. Eksploatacja utworu w zakresie dozwolonego użytku wyłącza bowiem możliwość dochodzenia przez uprawnionego roszczeń z tytułu naruszenia praw autorskich⁷. W przypadku polskich uregulowań ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych⁸ dozwolony użytek został uregulowany w Rozdziale 3 Odstępie 3. Jest on co do zasady nieodpłatną formą korzystania z utworów przez użytkowników. W niektórych przypadkach twórcy przysługuje prawo do wynagrodzenia

⁴ Zob. pozew Getty Images Inc. v. Stability AI inc., https://fingfx.thomsonreuters.com/gfx/legal-docs/byvrlkmwnve/GETTY_IMAGES_AI_LAWSUIT_complaint.pdf [dostęp 30.04.2024].

⁵ K. Grzybczyk, *Fair use jako przykład ograniczenia monopolu autorskiego klauzulą generalną*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2020, nr 1, s. 24.

⁶ Dz. Urz. UE L 160/62 z 17.05.2024 r.; dalej: „dyrektywa DSM”

⁷ W. Machała, *Dozwolony użytek chronionych utworów w polskim prawie autorskim w świetle cywilistycznej koncepcji prawa*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2001, nr 78, s. 59.

⁸ Dz. U. z 1994 r. Nr 24, poz. 83 ze zm.; dalej: „pr. aut.”

(np. rozpowszechnianie aktualnych wypowiedzi i fotografii reporterskich w prasie, radiu i telewizji – art. 25 ust. 2 w zw. z ust. 1 pkt 1 lit. b pr. aut.). W doktrynie różni się dozwolony użytek osobisty (prywatny) oraz publiczny.

Postaci dozwolonego użytku zostały uregulowane w modelu kazuistycznym. Zgodnie z dominującym poglądem doktryny enumeratywnie wskazanie postaci dozwolonego użytku wyłącza stosowanie wykładni rozszerzającej, zwłaszcza przyjmowanie nowych postaci dozwolonego użytku *per analogiam*⁹. Co więcej, korzystanie z utworów w granicach dozwolonego użytku powinno odbywać się pod warunkiem wymienienia imienia i nazwiska twórcy oraz źródła przy uwzględnieniu istniejących możliwości¹⁰ (art. 34 pr. aut.), a jednocześnie taki dozwolony użytek nie może naruszać normalnego korzystania z utworu lub godzić w słusne interesy twórcy (art. 35 pr. aut.). Dla korzystania z utworu w ramach dozwolonego użytku wymagane jest więc spełnienie łącznie tych trzech warunków.

Przenosząc poczynione rozważania na grunt web scrapingu, należy zauważyć w ślad za doktryną, że „(e)tap polegający na skopiowaniu danych, o ile mamy do czynienia z utworem chronionym prawem autorskim i nie możemy oprzeć się na ustawowym wyjątku, zawsze wkracza w wyłączne prawo do zwielokrotniania, gwarantowane zarówno traktatami międzynarodowymi, jak i ustawami krajowymi”¹¹. Przed uchwaleniem dyrektywy DSM słusznie wskazywało się również, że nie istniały podstawy do zakwalifikowaniem zbierania danych dla potrzeb wygenerowania danych treningowych poza zakresem monopolu autorskiego, nawet jeśli dane te były pobierane i przechowywane przez krótki czas, a co więcej – sam proces przetworzenia (wektoryzacji) do postaci czytelnej dla sieci neuronowej stanowi zwielokrotnianie¹². Ten stan prawny wykluczał dopuszczalność takiego zwielokrotniania bez odpowiedniej zgody uprawnionego podmiotu.

Przyjmując, że nowa postać dozwolonego użytku zostałaby wprowadzona do porządku prawnego przez ustawodawcę (przykładowo w postaci implementacji wyjątków zawartych w dyrektywie DSM jako nowych postaci dozwolonego użyt-

⁹ J. Barta, R. Markiewicz, *Prawo autorskie i prawa pokrewne*, Warszawa 2021, s. 81.

¹⁰ Uwzględnienie istniejących możliwości powinno jednak być interpretowane restrykcyjnie, z uwagi na treść art. 5 ust. 3 lit. a) dyrektywy 2001/29/WE Parlamentu Europejskiego i Rady z dnia 22 maja 2001 r. w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym (Dz. Urz. UE L 167/44 z 22.06.2001 r.), gdzie kształt tej przesłanki został określony zwrotem „poza przypadkami, w których okaże się to niemożliwe” – zob. *ibidem*, s. 84.

¹¹ E. Traple, [w:] B. Fischer, A. Pązik, M. Świerczyński (red.), *Prawo sztucznej inteligencji i nowych technologii*, Warszawa 2021, s. 22.

¹² Z. Okoń, [w:] K. Flaga-Gieruszyńska, J. Gołaczyński, D. Szostek (red.), *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, Warszawa 2019, s. 109–110.

ku¹³), nadal pozostawałaby konieczność kumulatywnego wypełnienia pozostałych warunków ustawowych w postaci wymienienia imienia i nazwiska twórcy oraz źródła, a ponadto niegodzenia w słusne interesy twórcy i normalne korzystanie przez niego z utworu. Możliwość spełnienia obu z tych warunków należy podać w wątpliwość. Mówiąc o warunku niegodzenia w słusne interesy twórców, jest to wątpliwe tym bardziej w przypadku web scrapingu dokonywanego na potrzeby szkolenia narzędzi generatywnej AI, które wraz ze wzrostem popularności stanowią realne zagrożenie dla interesów uprawnionych. W literaturze problem nie pozostaje obojętny. Postuluje się nawet jego potencjalne rozwiązania, np. w postaci ustanowienia systemu wynagradzania podmiotów uprawnionych, których utwory były wykorzystywane do szkolenia generatywnej AI, a jej wytwory zastępują w obrocie te, które zostały stworzone przez ludzkich twórców¹⁴.

De lege lata nie ma możliwości wykonywania web scrapingu w ramach dozwolonego użytku w rozumieniu pr. aut., ponieważ nie wchodzi on w zakres żadnej z ustawowo uregulowanych jego form. W obliczu obiektywnej potrzeby zapewnienia możliwości rozwoju technologii wymagających web scrapingu (w tym rozwoju AI) prawodawca unijny zawarł w ramach dyrektywy DSM dwa wyjątki od funkcjonujących w unijnym systemie prawnym praw wyłącznych (art. 3 i 4 dyrektywy DSM), w odniesieniu do utworów, które są dostępne zgodnie z prawem na potrzeby eksploracji tekstów i danych (dalej: „TDM”).

Web scraping jako zwielokrotnianie utworów na potrzeby TDM

Idąc dalej, należy zadać sobie pytanie, czy *web scraping* dokonywany na potrzeby uzyskania danych treningowych do szkolenia AI może być traktowany jako przejaw „zwielokrotniania i pobrań do celów eksploracji tekstów i danych” zgodnie z zawartą w art. 2 pkt 2) dyrektywy DSM definicją TDM. W rozumieniu dyrektywy mianem eksploracji tekstów i danych określa się „zautomatyzowaną technikę analityczną służącą do analizowania tekstów i danych w postaci cyfrowej w celu wygenerowania informacji, obejmujących między innymi wzorce, tendencje i korelacje”. Szkolenie modeli AI zawiera w sobie element TDM, a faktem jest, że prawodawca unijny jednoznacznie

¹³ Najnowsza dostępna wersja projektu nowelizacji pr. aut. zakłada implementację tych wyjątków właśnie w ramach sekcji poświęconej dozwolonemu użytkowi, w projektowanych art. 26² oraz 26³, <https://legislacja.rcl.gov.pl/docs//2/12382002/13037430/dokument664634.zip> [dostęp 30.04.2024].

¹⁴ Zob. M. Senftleben, *Generative AI and Author Remuneration*, „International Review of Intellectual Property and Law” 2023, nr 54, s. 1535–1560, *passim*, <https://doi.org/10.1007/s40319-023-01399-4>.

uznaje, że *web scraping* dokonywany na potrzeby szkolenia AI mieści się w zakresie tej definicji. Ewidentny tego przykład stanowi wskazanie na potrzebę zachowania zgodności tych działań z dyrektywą DSM (w szczególności w zakresie zastrzeżenia, o którym mowa w art. 4 ust. 3 dyrektywy DSM) w motywach¹⁵ rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji)¹⁶.

Wspólną cechą obu wyjątków (tj. wyjątków z art. 3 i 4 dyrektywy DSM) jest istnienie warunku „zgodnego z prawem dostępu” do utworów. Prawodawca unijny wyjaśnia w motywie 14 dyrektywy DSM, odnoszącym się do wyjątku zawartego w art. 3, rozumienie dostępu zgodnego z prawem jako „dostępu do treści opartego na polityce otwartego dostępu lub na ustaleniach umownych pomiędzy podmiotami uprawnionymi i organizacjami badawczymi lub instytucjami dziedzictwa kulturowego, takich jak abonamenty lub na innych zgodnych z prawem środkach. [...] Zgodny z prawem dostęp powinien również obejmować dostęp do treści, które są swobodnie dostępne w internecie”. Natomiast w przypadku art. 4 dyrektywy DSM zgodnie z motywem 18 dyrektywy DSM w szczególności chodzi o sytuacje, gdy utwór lub inny przedmiot objęty ochroną „został podany do publicznej wiadomości w internecie”, o ile podmioty uprawnione nie zastrzegły w odpowiedni sposób praw do zwielokrotniania i pobierania do celów TDM. Dokonanie tej dystynkcji w motywach dyrektywy prowadzi do wniosków, że w przypadku wyjątku z art. 4 dyrektywy DSM samo nabycie abonamentu na korzystanie z utworów nie będzie warunkowało samo w sobie wystąpienia przesłanki „zgodnego z prawem dostępu” na potrzeby TDM. Przy tym warto zaznaczyć, że w przypadku wyjątku zawartego w art. 4 dyrektywy DSM prawodawca unijny nie wyłączył możliwości zawierania umów sprzecznych z jego postanowieniami tak, jak uczynił to w przypadku art. 3, 5 i 6. Nic nie stoi więc na przeszkodzie, aby uprawnieni na mocy umowy ustanowili wynagrodzenie za korzystanie z utworów i innych przedmiotów objętych ochroną, do których przysługują im prawa wyłączne¹⁷.

¹⁵ W szczególności należy wskazać tu na treść motywu 105, który w ostatnim zdaniu jednoznacznie wskazuje: „W przypadku gdy prawo do wyłączenia z eksploracji zostało w odpowiedni sposób wyraźnie zastrzeżone, dostawcy modeli AI ogólnego przeznaczenia muszą uzyskać zezwolenie od podmiotów uprawnionych, jeżeli chcą dokonywać eksploracji tekstów i danych odnośnie do takich utworów”.

¹⁶ Dz. Urz. UE L z 12.07.2024 r.

¹⁷ R. Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym. Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, Warszawa 2021, s. 64.

Wyjątek z art. 3 dyrektywy DSM

Pierwszy wyjątek umożliwiający skorzystanie z dozwolonego użytku na potrzeby web scrapingu uregulowany jest w art. 3 ust. 1 dyrektywy DSM i odnosi się do zwielokrotnień i pobrań dokonywanych przez organizacje badawcze i instytucje dziedzictwa kulturowego „w celu przeprowadzenia, do celów badań naukowych, eksploracji tekstów i danych na utworach lub innych przedmiotach objętych ochroną, do których mają zgodny z prawem dostęp”. Na potrzeby TDM podmioty te mogą skorzystać z wyjątku od praw wyłącznych skonkretyzowanych w: a) art. 5 lit. a oraz art. 7 ust. 1 dyrektywy 96/9/WE Parlamentu Europejskiego i Rady z dnia 11 marca 1996 r. w sprawie ochrony prawnej baz danych¹⁸ i b) art. 2 dyrektywy 2001/29/WE Parlamentu Europejskiego i Rady z dnia 22 maja 2001 r. w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym¹⁹.

W ocenie autora wprowadzenie wyjątku na rzecz organizacji badawczych i instytucji dziedzictwa kulturowego nie budzi większych kontrowersji oraz pozostaje w pełni uzasadnione z uwagi na potrzebę rozwoju technologii i nauki właśnie przez te podmioty. Z tego powodu rozważania na temat wyjątku wprowadzonego przez art. 3 dyrektywy DSM należy ograniczyć do krótkiej charakterystyki przedstawionej powyżej, a dalsza część niniejszego artykułu zostanie poświęcona wyjątkowi przewidzianemu w art. 4.

Wyjątek z art. 4 dyrektywy DSM

Wprowadzony w art. 4 dyrektywy DSM wyjątek, który nie posiada tak skonkretyzowanego zakresu podmiotowego jak w art. 3, przy jednoczesnym rozszerzeniu zakresu przedmiotowego (o prawa wyłączne zawarte w art. 4 ust. 1 lit. a) i b) dyrektywy 2009/24/WE z dnia 23 kwietnia 2009 r. w sprawie ochrony prawnej programów komputerowych²⁰), budzi zdecydowanie większe kontrowersje. W art. 4 prawodawca unijny nie zdefiniował bowiem adresata normy, którą państwa członkowskie mają zaimplementować, wskazując jedynie na „znaczenie technik eksploracji tekstów i danych dla podmiotów prywatnych i publicznych do analizowania dużych ilości danych w różnych sferach życia codziennego i do różnych celów, w tym przez służby państwowe, do podejmowania złożonych decyzji biznesowych oraz do rozwijania nowych aplikacji i technologii” (motyw 18 dyrektywy DSM).

¹⁸ Dz. Urz. UE L 77/39 z 27.03.1996 r.

¹⁹ Dz. Urz. UE L 167/44 z 22.06.2001 r.

²⁰ Dz. Urz. UE L 111/52 z 5.05.2009 r.

Analizując dalej motyw 18, potrzeba takiego uregulowania wyniku również z konieczności zniwelowania niepewności co do możliwości zastosowania art. 5 ust. 1 dyrektywy 2001/29/WE w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym do procesu TDM (tak jak w przypadku web scrapingu, zakładającego potrzebę trwałego zwielokrotnienia danych, w celu ich następczej obróbki dla szkolenia modeli AI). Niewątpliwie prawodawca unijny miał na celu stworzenie przyjaznego środowiska dla podmiotów prywatnych do rozwijania technologii AI na terenie UE, rozpoznając wyzwania natury prawnej związane z pozyskiwaniem danych treningowych.

Temporalna granica przechowywania zwielokrotnionych utworów i innych pobranych przedmiotów objętych ochroną została zakreślona w ust. 2 omawianego artykułu i obejmuje przechowywanie „[...] tak długo, jak jest to konieczne do celów eksploracji tekstów i danych”. Dane powinny więc zostać usunięte w momencie, w którym zostały już wygenerowane informacje obejmujące wzorce, tendencje i korelacje, o których mowa w art. 2 ust. 2 dyrektywy DSM. Takie rozwiązanie z jednej strony pozwala na ograniczenie skutków sanowanego naruszenia praw wyłącznych poprzez minimalizację czasu, w którym będą one przechowywane przez podmiot dokonujący TDM, z drugiej jednak może stanowić istotne ograniczenie możliwości następczego zbadania plików na potrzeby przeprowadzenia dowodu w przypadku powstania sporu pomiędzy uprawnionym z tytułu praw autorskich a dokonującym TDM.

Prawodawca unijny w art. 4 ust. 3 dyrektywy DSM wprowadza warunek, który musi zostać spełniony dla skorzystania z wyjątku zawartego w art. 4 – brak wyraźnego zastrzeżenia korzystania z utworów i innych przedmiotów objętych ochroną przez podmiot uprawniony. Zastrzeżenie to musi być wyrażone „[...] w odpowiedni sposób, na przykład za pomocą środków nadających się do odczytu maszynowego w przypadku treści, które zostały podane do publicznej wiadomości w internecie”. Prawodawca unijny wprowadza więc domniemanie zgody uprawnionego na korzystanie z utworów i innych przedmiotów objętych ochroną do celów TDM udostępnionych w internecie. Możliwość zastrzeżenia przez podmioty uprawnione korzystania z utworów i innych przedmiotów objętych ochroną jest – zgodnie z dominującymi poglądami – niedostateczna dla realizowania uzasadnionych potrzeb korzystających z TDM oraz stawia te podmioty z krajów UE w gorszej pozycji w stosunku do tych, które operują chociażby w jurysdykcji USA, Japonii czy Chin²¹.

²¹ Tak m.in. E. Traple, *op. cit.*, s. 26 oraz R. Markiewicz, *op. cit.*, s. 64.

W opinii autora domyślna zgoda uprawnionego wprowadzona w art. 4 ust. 3 dyrektywy DSM pozostaje problematyczna. Jest tak z uwagi na fakt, że podmioty uprawnione nie zawsze udostępniają swoje utwory i inne przedmioty objęte ochroną w sieci samodzielnie, a wydaje się, że częstszym zjawiskiem jest udostępnianie utworów za pośrednictwem zewnętrznych podmiotów takich jak dostawcy platform internetowych, którym uprawnieni udzielają licencji na rozpowszechnienie utworu w sieci lub przenoszą na nich autorskie prawa majątkowe do utworu. To właśnie te zewnętrzne podmioty mają faktyczną (techniczną) możliwość zastrzeżenia korzystania z utworów w odpowiedni sposób, w tym z użyciem środków nadających się do odczytu maszynowego, o których mowa w ust. 3 omawianego artykułu. Z tego względu przed analizą skutków, jakie art. 4 dyrektywy DSM może wyrzucić w sferze interesów podmiotów uprawnionych, należy przeanalizować, w jaki sposób utwory są udostępniane w sieci i jak te przypadki rzutują na możliwość wprowadzenia zastrzeżenia.

Podmiot uprawniony udostępnia samodzielnie

Przypadki, gdy utwór lub inny przedmiot objęty ochroną jest udostępniony bezpośrednio przez podmiot uprawniony (w witrynie operowanej osobiście przez niego) lub inny podmiot²² (w taki sposób, że podmiot uprawniony ustala warunki korzystania przez użytkowników), pozostają raczej rzadkością w przypadku niezależnych twórców, jednak zdarzają się²³. Często można natomiast zauważyć udostępnianie w sieci utworów przez wydawców, którym przysługują autorskie prawa majątkowe – prowadzących działalność gospodarczą opartą na publikacji treści w internecie. W takim układzie korzystanie z utworów przez użytkowników serwisu może odbywać się zarówno nieodpłatnie, jak i za wynagrodzeniem,

²² Z uwagi na niniejszy śródtytuł celowo pominięte zostały rozważania dotyczące dostawców usług pośrednich innych niż platformy internetowe, w rozumieniu rozporządzenia Parlamentu Europejskiego i Rady (UE) 2022/2065 z dnia 19 października 2022 r. w sprawie jednolitego rynku usług cyfrowych oraz zmiany dyrektywy 2000/31/WE (Dz. Urz. UE L 277/65 z 27.10.2022 r.), (dalej: „DSA”), ponieważ rola tych podmiotów pozostaje marginalna z perspektywy ustalania warunków korzystania z utworów udostępnianych w witrynie operowanej bezpośrednio przez podmiot uprawniony.

²³ Np. polski artysta Taco Hemingway udostępnia swoją twórczość (albumy muzyczne) do pobrania w formie skompresowanych folderów w formacie .zip na stronie <https://pobierz.1-800-oswiecenie.com/> [dostęp 30.04.2024], a adres strony internetowej zmienia się wraz z wydaniem każdego kolejnego albumu. Na stronie nie sposób szukać jednak jakichkolwiek warunków umowy licencyjnej, której autor udziela swoim słuchaczom, należy więc przyjąć, że mogą z nich korzystać w ramach dozwolonego użytku osobistego.

często w postaci subskrypcji lub jednorazowej opłaty za dostęp do utworu lub ich grupy, a warunki i pola eksploatacji, na jakich użytkownicy mogą korzystać z utworu w ramach licencji²⁴ będą ustalane bezpośrednio przez odpowiedni podmiot uprawniony.

Przenosząc te rozważania na grunt web scrapingu dokonywanego w ramach art. 4 dyrektywy DSM, należy przyjąć, że konsekwencje dla legalności TDM w zależności od warunków korzystania z utworu narzuconych przez podmiot uprawniony mogą przyjmować zróżnicowane postaci. Przede wszystkim warunki korzystania z utworu narzucone przez podmiot uprawniony mogą całkowicie wyłączać możliwość zwielokrotnienia utworu na potrzeby TDM. Podmiot uprawniony posiada więc faktyczną możliwość zastrzeżenia korzystania, w tym za pomocą środków możliwych do odczytu maszynowego. W świetle treści art. 7 ust. 1 dyrektywy DSM nie stoi również na przeszkodzie, aby podmiot uprawniony, udostępniając utwór lub inny przedmiot objęty ochroną, zastrzegł możliwość korzystania z utworu na potrzeby TDM, jednocześnie uzależniając zgodę na takie korzystanie od zapłaty wynagrodzenia – np. w postaci zwiększonej opłaty za dostęp czy indywidualnie negocjowanych warunków²⁵.

Podmiot trzeci

Powszechnym zjawiskiem jest udostępnianie utworów i innych przedmiotów objętych ochroną za pośrednictwem podmiotów trzecich, w szczególności platform internetowych w rozumieniu DSA czy dostawców usług udostępniania treści online w rozumieniu dyrektywy DSM. W tym zakresie na potrzeby udostępnienia utworu przez podmiot trzeci, autor lub inny podmiot uprawniony udziela administratorowi witryny licencji, często również w zakresie udzielania sublicencji (w dyspozycji art. 68 ust. 3 pr. aut.) lub przenosi na niego autorskie prawa majątkowe do utworu.

Przypadek przeniesienia autorskich praw majątkowych wydaje się zdecydowanie łatwiejszy do zakwalifikowania w zakresie ewentualnego zastrzeżenia korzy-

²⁴ Mając na względzie sposób zawarcia umowy licencyjnej w tym modelu, tj. w sposób dorozumiany i z użyciem wzorca umownego, należy wykluczyć możliwość przeniesienia autorskich praw majątkowych do utworu zgodnie z regulacją art. 53 pr. aut. Oczywiście zakres licencji może zostać określony przez podmiot uprawniony w bardzo zróżnicowany sposób, w tym np. z użyciem powszechnych licencji typu Creative Commons.

²⁵ Zob. np. GETTY IMAGES CONTENT LICENSE AGREEMENT, Restricted Uses, lit. k, <https://www.gettyimages.com/eula> [dostęp 21.05.2024]

stania z utworów na potrzeby TDM. W tym przypadku bowiem to administrator serwisu staje się podmiotem uprawnionym w znaczeniu art. 4 i jego pozycja prawna zostaje praktycznie zrównana z tą podmiotu uprawnionego opisywaną w poprzedniej jednostce redakcyjnej²⁶.

W pierwszym przypadku główną rolę będzie grała treść umowy zawieranej przez podmiot uprawniony z administratorem witryny. To właśnie jej warunki będą w istocie określać dalsze losy utworu lub innego przedmiotu objętego ochroną na potrzeby TDM. W opinii autora w jej treści powinny znaleźć się postanowienia dotyczące ewentualnego wprowadzenia zastrzeżenia do struktury strony internetowej przez podmiot trzeci. Niestety, w większości przypadków jej treść będzie kształtowana przez nienegocjowalne wzorce umów narzucane przez platformy internetowe i innych administratorów witryn w postaci np. regulaminu usługi. Wydaje się, że sens wprowadzenia prawa do zastrzeżenia, o którym mowa w art. 4 ust. 3 dyrektywy DSM, pozostaje wypaczony właśnie z uwagi na tę okoliczność.

Wniosek taki wynika z wykładni literalnej ust. 3 – wskazania *explicite* podmiotu uprawnionego jako zdolnego do poczynienia zastrzeżenia przez unijnego prawodawcę. Natomiast patrząc na uwarunkowania rynkowe, chociażby analizując warunki korzystania jednej z popularniejszych platform internetowych – YouTube – podmiot uprawniony, udostępniając na niej swój utwór lub inny przedmiot objęty ochroną, będzie udzielał „[...] każdemu innemu użytkownikowi Usługi ogólnoświatowej, niewyłącznej i nieodpłatnej licencji na korzystanie z jego Treści w ramach Usługi (w tym na zwielokrotnianie, rozpowszechnianie, modyfikowanie, wyświetlanie i wykonywanie) wyłącznie w zakresie dopuszczonym w ramach Usługi”²⁷. Dopuszczalny zakres, o którym mowa powyżej, jest kształtowany przez wolę samego dostawcy platformy internetowej. Ten musi co prawda uwzględnić powszechnie obowiązujące przepisy prawa, które jednak na ten moment nie wskazują *explicite* obowiązku uwzględnienia woli podmiotu uprawnionego do dokonania zastrzeżenia. Pozostaje to niezrozumiałe w obliczu tej obiektywnej przewagi kontraktowej i faktu, że to podmiot trzeci ma faktyczną (techniczną) możliwość wprowadzenia zastrzeżenia. Takiego obowiązku nie nałożono bowiem na platformy internetowe w DSA ani na dostawców usług udostępniania treści online w dyrektywie DSM.

²⁶ „Praktycznie”, ponieważ twórcy w dalszym ciągu będą przysługiwać niezbywalne autorskie prawa osobiste, o których mowa w art. 16 pr. aut.

²⁷ *Licencja dla innych użytkowników*, <https://www.youtube.com/t/terms>, [dostęp 30.04.2024].

Udostępnienie przez osobę trzecią nieuprawnioną

Równie powszechnym zjawiskiem jest udostępnianie utworów i innych przedmiotów objętych ochroną przez osoby nieuprawnione do takiego udostępnienia. Już w momencie udostępnienia przez ww. podmiot trzeci dochodzi do naruszenia monopolu prawnoautorskiego i podmiot dokonujący takiego udostępnienia jest odpowiedzialny za naruszenie. Słusznie wskazuje się bowiem w literaturze, że „[...] jeżeli osoba na własnej stronie internetowej chce zamieścić utwory osób trzecich, które są objęte ochroną prawnoautorską, musi uzyskać zezwolenie twórcy na korzystanie z utworu w ramach tego pola eksploatacji, aby nie narazić się na zarzut naruszenia praw autorskich, niezależnie od tego, czy robi to w ramach działalności komercyjnej czy niekomercyjnej”²⁸. Każdy przejaw udostępniania utworów i innych przedmiotów podlegających ochronie może być rozpatrywany w kategoriach naruszenia autorskich praw majątkowych, niezależnie od tego, czy takie udostępnienie odbywa się na własnej stronie internetowej udostępniającego, za pośrednictwem dostawców usług pośrednich, czy np. za pośrednictwem programów *peer-to-peer*.

Nasuwa się więc pytanie, czy podmiot dokonujący TDM ze źródeł, które pierwotnie już zostały obciążone naruszeniem monopolu autorskiego, może skutecznie powołać się na wyjątek zawarty w art. 4 dyrektywy DSM. Przywołany już motyw 18 wskazywałby, że utwór lub inny przedmiot objęty ochroną, który „został podany do publicznej wiadomości w internecie”, *prima facie* samoistnie gwarantuje spełnienie przesłanki dostępu zgodnego z prawem. Takie stwierdzenie istotnie mogłoby zostać podniesione przez dokonującego *web scraping*. Patrząc na regulacje dyrektywy DSM, istnieje uzasadnione przypuszczenie, że utwór lub inny przedmiot objęty ochroną podany do publicznej wiadomości w internecie może być przedmiotem zwielokrotnienia w zakresie art. 4, o ile nie zostało dokonane zastrzeżenie. Kolejnym argumentem za wskazaną tezę jest ocena możliwości, jakie posiada taki podmiot w zakresie analizy uwarunkowań prawnoautorskich, które są przedmiotem umowy pomiędzy podmiotami od niego niezależnymi. Konstrukcja ww. argumentów oparta jest jednak na niewiedzy podmiotu dokonującego *web scrapingu* o nielegalności źródła, natomiast wydaje się, że należałoby jednoznacznie oddzielić przypadki, gdzie podmioty świadomie scrapują z nielegalnych źródeł. Na ten moment, z uwagi na stosunkowo krótki okres po uchwaleniu dyrektywy DSM, ciężko udzielić jednoznacznej odpowiedzi – nie wykształciła się jeszcze praktyka orzecznicza. W opinii autora więcej argumentów

²⁸ M. Nowikowska, [w:] J. Sieńczyło-Chlabicz (red.), *Prawo własności intelektualnej. Teoria i praktyka*, Warszawa 2021, s. 374.

przemawia jednak za uznaniem zwielokrotnienia obciążonego pierwotnym naruszeniem monopoli autorskiego za wykraczające poza zakres art. 4 dyrektywy DSM. Biorąc pod uwagę obowiązek zawężającej interpretacji wyjątków oraz dotychczasowe orzecznictwo TSUE w sprawach dotyczących wyjątków od praw wyłącznych, zdecydowanie bardziej uprawnione jest przyjęcie kryterium legalności źródła – tak jak TSUE przyjął, odpowiadając na pytanie prejudycjalne w odniesieniu do dozwolonego użytku osobistego w sprawie *ACI Adam BV i in. v. Stichting de Thuiskopie i Stichting Onderhandeligen Thuiskopie vergoeding*²⁹. Kwestia wykorzystywania utworów i innych przedmiotów objętych ochroną pochodzących z nielegalnych źródeł dla TDM niewątpliwie pozostaje problematyczna. Praktyka pokazuje bowiem, że podmioty dokonujące web scrapingu na potrzeby szkolenia AI korzystają również z nich³⁰.

Ułomność „możliwości odczytania maszynowego”

Powracając jednak do sytuacji, w których podmiot uprawniony ma możliwość dokonania zastrzeżenia, o którym mowa w art. 4 ust. 3 dyrektywy DSM, warto zwrócić uwagę na nieprecyzyjność prawodawcy unijnego w formułowaniu wymogu co do formy, w jakiej zastrzeżenie może być poczynione. Użycie sformułowania „na przykład za pomocą środków nadających się do odczytu maszynowego” wymaga uzupełnienia o treść motywu 18 dyrektywy DSM. Prawodawca unijny wyjaśnia w nim, że „w przypadku treści, które zostały podane do publicznej wiadomości w internecie za stosowne powinno się wyłącznie uznać zastrzeżenie tych praw za pomocą środków nadających się do odczytu maszynowego, łącznie z metadanymi i warunkami korzystania ze strony internetowej lub usługi”. Prawodawca dopuszcza zatem możliwość dokonania zastrzeżenia w warunkach korzystania ze strony internetowej lub usługi, wyraźnie nie zdając sobie sprawy, że boty scrapingowe wykorzystywane do dokonywania zwielokrotniania i pobierania treści na potrzeby TDM to w istocie proste programy, które zazwyczaj nie dokonują analizy warunków czy regulaminów zamieszczonych w witrynach, których zawartość pobierają.

²⁹ Wyrok Trybunału z dnia 10 kwietnia 2014 r. w sprawie C-435/12, *ACI Adam BV i in. v. Stichting de Thuiskopie i Stichting Onderhandeligen Thuiskopie vergoeding*, EU:C:2014:254.

³⁰ Np. podczas szkolenia swojego LLM Meta użyła i przyznała się do użycia zestawu danych zawierającego książki z nielegalnego źródła „the Books3 section of ThePile”. Według agencji Reuters osoba, która stworzyła ten zestaw danych, podaje, że sporządziła go z ponad 196 640 książek, K. Paul, *Meta used copyrighted books for AI training despite its own lawyers' warnings, authors allege*, 13.12.2023, <https://www.reuters.com/technology/meta-used-copyrighted-books-ai-training-despite-it-s-own-lawyers-warnings-authors-2023-12-12/> [dostęp 30.04.2024].

Brak regulacji odnoszącej się bezpośrednio do tego, co to znaczy, że zastrzeżenie „jest możliwe do odczytu maszynowego”, przy jednoczesnym wymienieniu warunków strony internetowej lub usługi jako jeden ze sposobów zastrzeżenia, prowadzi do powstania dysonansu, który w ocenie autora prawodawca unijny powinien uzupełnić o wprowadzenie jednolitego sposobu wprowadzenia zastrzeżenia, a to z kilku powodów.

Po pierwsze, na rynku wewnętrznym warunki korzystania ze stron internetowych występują w znacznej liczbie języków i często w ramach jednej witryny nie są nawet dostępne dodatkowo w języku angielskim, z uwagi na fakt, że taki warunek nie w każdym przypadku³¹ przewidują akty prawa unijnego, natomiast akty prawa krajowego często wymuszają na administratorach jedynie udostępnianie warunków w języku narodowym (np. art. 7 ustawy z dnia 7 października 1999 r. o języku polskim [Dz. U. z 1999 r. Nr 90, poz. 999 ze zm.] w odniesieniu do obrotu konsumenckiego). Prowadzi to do sytuacji, w której podmiot planujący wykorzystać bota scrapingowego na terenie UE musi go zaprogramować w sposób, który umożliwi mu co najmniej samodzielną analizę warunków korzystania we wszystkich językach urzędowych UE, co znacząco zwiększa poziom skomplikowania takiego bota, który z założenia ma być prostym środkiem do pozyskania treści z internetu. Nietrudno wywnioskować również, że z uwagi na takie utrudnienia w zakresie stworzenia bota scrapingowego podmioty będą: a) rezygnować z web scrapingu, pozostając w jurysdykcji UE, lub b) scrapować, nie zwracając uwagi na warunki korzystania z witryn.

Po drugie, wykształcona praktyka pozwala na przyjęcie innego środka w postaci pliku *robots.txt* zamieszczanego w metadanych witryny (co również stara się zauważyć prawodawca unijny w motywie 18, odnosząc się do „metadanych”), w tym zakresie jednak poszczególni scrapujący nie posiadają jednolitych warunków wyłączenia. Przykładowo, OpenAI w odniesieniu do swojego GPTBot’a scrapującego dane na potrzeby działania modelu GPT-4 może zostać wyłączony w odniesieniu do strony, w której w pliku *robots.txt* znajdują się dwie linijki tekstu³²:

User-agent: GPTBot

Disallow: /

³¹ Np. w art. 14 ust. 6 DSA prawodawca unijny wprowadził obowiązek dla bardzo dużych platform internetowych i bardzo dużych wyszukiwarek do publikowania swoich warunków korzystania z usług w językach urzędowych wszystkich państw członkowskich, w których oferują usługi, jednak zakres podmiotowy tego przepisu pozostaje istotnie ograniczony.

³² GPTBot, OpenAI Platform, <https://platform.openai.com/docs/gptbot> [dostęp 30.04.2024].

Nie wszyscy producenci popularnych modeli AI są tak transparentni w kwestii blokowania swoich botów scrapingowych. Poszukując informacji, można natrafić chociażby na artykuły popularnonaukowe, które zawierają zbiory takich komend³³, które muszą zostać zaimplementowane do pliku *robots.txt*, aby ten potencjalnie ograniczył dostęp dla poszczególnych botów. Są one jednak najczęściej wskazywane przez podmioty zewnętrzne, niezwiązane z działalnością poszczególnych przedsiębiorstw zajmujących się web scrapingiem. Analiza tych zbiorów prowadzi do wniosku, że każdy bot wymaga osobnej komendy w pliku *robots.txt*. Wydaje się to kuriozalne w obliczu konieczności zapewnienia respektowania prerogatywy podmiotów uprawnionych do zastrzeżenia korzystania, o której mowa w art. 4 ust. 3 dyrektywy DSM.

Z uwagi na to należy postulować ujednoczenie przez prawodawcę unijnego metod zastrzeżenia, tak aby umożliwić korzystanie z uprawnień przez podmioty uprawnione oraz pewność co do prawa dla podmiotów dokonujących TDM na podstawie art. 4 ust. 3.

Podsumowanie i postulaty

Wyjątek zawarty w art. 4 dyrektywy DSM, pozostaje niewystarczający z uwagi na dalszy brak pewności co do prawa z perspektywy podmiotów dokonujących web scrapingu danych szkoleniowych dla AI oraz z perspektywy podmiotów uprawnionych. Dla uprawnionych w praktyce wprowadzenie zastrzeżenia, o którym mowa w art. 4 ust. 3, do witryny, w której udostępniają oni swoje utwory, może okazać się nadmiernie utrudnione lub niemożliwe. Pozostaje niepewność co do znaczenia terminu „możliwości odczytania maszynowego”, na którego interpretacji orzecznictwo zapewne „połamie sobie zęby” z uwagi na fakt, że prawodawca unijny nie ujednotliił ram dla zastrzeżeń w metadanych. Dostawcy rozwiązań opartych na AI aktualnie ewidentnie stawiają tę metodę przed zastrzeżeniem dokonany w warunkach strony internetowej, a prawodawca unijny powinien rozważyć wprowadzenie jednolitej metody zastrzeżenia, do której to dostawcy rozwiązań będą musieli się dostosować.

W istocie projektując rozwiązania, które miały na celu wywarzyć zapewnienie rozwoju technologii z przeciwnymi interesami podmiotów uprawnionych, w opinii autora prawodawca unijny powinien mieć na względzie najbardziej interesujący dla obu stron aspekt – finansowy.

³³ N. Clarke, *Block the Bots that Feed “AI” Models by Scraping Your Website*, <https://neil-clarke.com/block-the-bots-that-feed-ai-models-by-scraping-your-website/> [dostęp 30.04.2024].

Wydaje się, że zdecydowanie bardziej praktycznym rozwiązaniem, które mogłoby ograniczyć przyszłe spory wynikające z poszczególnych nieścisłości dotyczących legalnego przeprowadzania procedury TDM, byłoby upoważnienie organizacji zbiorowego zarządzania do negocjowania zbiorowych umów licencyjnych z podmiotami prywatnymi, które dokonują web scrapingu, tak jak prawodawca unijny uczynił to w przypadku korzystania przez instytucje dziedzictwa kulturowego z niedostępnych w obrocie handlowych utworów i innych przedmiotów objętych ochroną w art. 8 dyrektywy DSM.

Co więcej, niejednolita praktyka podmiotów rynkowych co do plików *robots.txt* oraz brak realnego wpływu na te metadane przez podmioty uprawnione udostępniające swoje utwory za pośrednictwem podmiotów zewnętrznych zdecydowanie warunkują potrzebę wprowadzenia jednolitych ram dla zastrzeżenia, o którym mowa w art. 4 ust 3 dyrektywy DSM, oraz obowiązek uwzględnienia woli podmiotów uprawnionych w zakresie dokonania zastrzeżenia. W tym zakresie wydaje się, że wystarczające byłoby wprowadzenie wymogu uzyskania stanowiska podmiotu uprawnionego (np. w postaci prostego *checkboxu*).

Bibliografia

Źródła

- Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790 z 17.04.2019 r. w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym oraz zmiany dyrektyw 96/9/WE oraz 2001/29/WE (Dz. Urz. UE L 160/62 z 17.05.2024 r).
- Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz. U. z 1994 r. Nr 24, poz. 83 ze zm.).
- Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2024/1689 z dnia 13 czerwca 2024 r. w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektyw 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji) (Dz. Urz. UE L z 12.07.2024 r.).
- Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2022/2065 z dnia 19 października 2022 r. w sprawie jednolitego rynku usług cyfrowych oraz zmiany dyrektywy 2000/31/WE (akt o usługach cyfrowych) (Dz. Urz. UE L 277/65 z 27.10.2022 r.).
- Dyrektywa 96/9/WE Parlamentu Europejskiego i Rady z dnia 11 marca 1996 r. w sprawie ochrony prawnej baz danych (Dz. Urz. UE L 77/39 z 27.03.1996 r.).
- Dyrektywa 2001/29/WE Parlamentu Europejskiego i Rady z dnia 22 maja 2001 r. w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym (Dz. Urz. UE L 167/44 z 22.06.2001 r.).
- Dyrektywa Parlamentu Europejskiego i Rady 2009/24/WE z dnia 23 kwietnia 2009 r. w sprawie ochrony prawnej programów komputerowych (Dz. Urz. UE L 111/52 z 5.05.2009 r).
- Ustawa z dnia 7 października 1999 r. o języku polskim (Dz. U. z 1999 r. Nr 90, poz. 999 ze zm.).

Orzecznictwo

Wyrok Trybunału Sprawiedliwości Unii Europejskiej z dnia 10 kwietnia 2014 r. w sprawie C-435/12, *ACI Adam BV i in. v. Stichting de ThuisKopie i Stichting Onderhandeligen ThuisKopie vergoeding*, EU:C:2014:254.

Literatura

- Barta J., Markiewicz R., *Prawo autorskie i prawa pokrewne*, Warszawa 2021.
- Fischer B., Pązik A., Świerczyński M., *Prawo sztucznej inteligencji i nowych technologii*, Warszawa 2021.
- Flaga-Gieruszyńska K., Gołaczyński J., Szostek D., *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, Warszawa 2019.
- Grzybczyk K., *Fair use jako przykład ograniczenia monopolu autorskiego klauzulą generalną*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2020, nr 1.
- Krotov V., Silva L. et. al., *Twenty-fourth Americas Conference on Information Systems*, New Orleans 2018.
- Machała W., *Dozwolony użytek chronionych utworów w polskim prawie autorskim w świetle cywilistycznej koncepcji prawa*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2001, nr 78.
- Markiewicz R., *Prawo autorskie na jednolitym rynku cyfrowym. Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, Warszawa 2021.
- Senftleben M., *Generative AI and Author Remuneration*, „International Review of Intellectual Property and Law” 2023, nr 54.
- Schintler A., McNeely C.L., *Encyclopedia of Big Data*, „Springers International Publishing” 2017,
- Sieńczyło-Chłabicz J., *Prawo własności intelektualnej. Teoria i praktyka*, Warszawa 2021.

Internet

- Clarke N., *Block the Bots that Feed “AI” Models by Scraping Your Website*, neil-clarke.com/category/technical, 23.8.2023.
- GETTY IMAGES CONTENT LICENSE AGREEMENT, Restricted Uses, lit. k, <https://www.gettyimages.com/eula>
- GPTBot, OpenAI Platform, <https://platform.openai.com/docs/gptbot>
- <https://neil-clarke.com/block-the-bots-that-feed-ai-models-by-scraping-your-website/>
- <https://pobierz.1-800-oswiecenie.com/>
- Licencja dla innych użytkowników*, <https://www.youtube.com/t/terms>
- Paul K., *Meta used copyrighted books for AI training despite its own lawyers’ warnings, authors allege*, <https://www.reuters.com/technology/meta-used-copyrighted-books-ai-training-despite-its-own-lawyers-warnings-authors-2023-12-12/>
- Pozew Getty Images Inc. v. Stability AI inc., <https://fingfx.thomsonreuters.com/gfx/legaldocs/byvrlkmwnve/GETTY IMAGES AI LAWSUIT complaint.pdf>
- Projekt z dnia 18 kwietnia 2024 r. ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych oraz niektórych innych ustaw, <https://legislacja.rcl.gov.pl/docs//2/12382002/13037430/dokument664634.zip>

Reuters, Technology, [reuters.com/technology](https://www.reuters.com/technology), 13.12.2023.

WorldWideWebSize.com. Daily estimated size of the world wide web, <https://www.worldwide-websize.com/#:~:text=Web> (The Internet)

Web scraping for artificial intelligence training v. author's position

Abstract

In the age of artificial intelligence (AI) and widespread web scraping, artworks available on the internet are copied daily for the purpose of obtaining training data for AI models. In the U.S., where many companies are providing AI-based solutions, we are seeing an increase in the number of lawsuits related to this issue. The providers of these solutions defend themselves by invoking the doctrine of fair use. In the same cases in Europe, the arguments will most likely focus on the exceptions to the exclusive rights introduced by the EU legislator in directive (EU) 2019/790 of the European Parliament and of the Council. However, an analysis of the exceptions introduced by the directive leads to the conclusion that these regulations do not in fact remain optimal from the perspective of rightholders or scrapers.

Keywords

webscraping, text and data mining, DSM Directive, fair use, artificial intelligence, machine learning, deep learning, large language models, use restriction, author's rights

dr Joanna Matczuk
ORCID: 0009-0009-5083-7566

Adrianna Pilecka
ORCID: 0009-0008-3242-5333

Systemy sztucznej inteligencji w procesach administracyjnych i w pracy urzędów patentowych na świecie – przegląd wybranych inicjatyw oraz analiza dalszego ich rozwoju w kontekście zasad etycznych dla sztucznej inteligencji. Część 1¹

Streszczenie

Zaawansowane technologie, takie jak sztuczna inteligencja (dalej: „AI”), zyskują coraz większe znaczenie w działalności urzędów na całym świecie, w tym urzędów własności intelektualnej. Artykuł omawia rozwój oraz różne sposoby wykorzystania AI przez urzędy, tj. EUIPO, EPO, USPTO i WIPO, w tym konkretnych narzędzi stosowanych przez urzędy i udostępnianych użytkownikom, automatyzację procesów związanych z rejestracją praw własności przemysłowej, analizą adekwatnych zbiorów danych w kontekście prowadzonych poszukiwań znaków czy badań patentowych. Przedstawia również wyzwania i korzyści wynikające z integracji nowoczesnych technologii w obszarze ochrony własności intelektualnej.

Słowa kluczowe

AI, SI, sztuczna inteligencja, urzędy ds. własności intelektualnej, własność intelektualna, własność przemysłowa, postępowanie administracyjne

Uwagi wstępne

Od kilku lat toczy się debata o możliwościach i prawno-etycznych ograniczeniach stosowania rozwiązań opartych na sztucznej inteligencji w praktyce są-

¹ Publikacja jest pierwszą częścią cyklu artykułów analizujących obecny stan wykorzystania technologii AI w pracy urzędów patentowych oraz ukazujących perspektywę jej dalszego wykorzystania przez pryzmat zasad etycznych dotyczących godnej zaufania sztucznej inteligencji przyjmowanych globalnie i na szczeblu regionalnym.

downiczej². O ile w tej sferze proces implementacji takich rozwiązań wciąż nie wydaje się zbyt zaawansowany, o tyle kwestia ta inaczej prezentuje się w działalności urzędów ds. własności intelektualnej, gdzie w ciągu ostatnich lat narzędzia AI znalazły zastosowanie we wsparciu procedur administracyjnych, zwłaszcza w zakresie badań i udzielania praw własności intelektualnej. Widoczne tempo ich rozwoju i potencjał technologicznego postępu stwarza realne możliwości niemal całkowitej automatyzacji takich procedur i teoretyczną możliwość wykluczenia z nich tzw. czynnika ludzkiego. To z kolei, na co często zwraca się uwagę, komentując tego typu rozwiązania, może powodować rozmaite ryzyka (np. związane chociażby z dyskryminacyjnym aspektem AI). Ochroną przed takimi ryzykami, ale jednocześnie i barierą stojącą na drodze ekspansji takich rozwiązań mogą okazać się narzędzia tzw. *soft law* w postaci zbiorów zasad etycznych przyjmowane na różnych szczeblach.

W niniejszym opracowaniu, które podzielone jest na trzy artykuły, autorki w formie sprawozdawczej przedstawiają rozwiązania SI wprowadzone do praktyki regionalnych i krajowych urzędów ds. własności intelektualnej, ukazując tempo i skalę zachodzących zmian oraz prezentując dostępne rozwiązania. W pierwszej (niniejszej) części omówiono szczerzej projekty prowadzone przez wiodące światowe organizacje, tj. Urząd Unii Europejskiej ds. Własności Intelektualnej (dalej: EUIPO), Międzynarodową Organizację Własności Intelektualnej (dalej: WIPO), Europejski Urząd Patentowy (dalej: EPO) oraz Urząd Patentów i Znaków Towarowych Stanów Zjednoczonych (ang. United States Patent and Trademark Office) – USPTO, które z perspektywy praktyków prawa wydają się kluczowe i mające największe oddziaływanie na pozostałe jurysdykcje. W części drugiej autorki dokonają przeglądu rozwiązań zaimplementowanych na szczeblu krajowym w wybranych urzędach. Natomiast w części trzeciej cyklu poruszone zostaną zagadnienia z zakresu etyki dotyczące godnej zaufania sztucznej inteligencji. Autorki spojrzą również na prezentowane rozwiązania biur ds. własności intelektualnej przez pryzmat przyjętych przez te instytucje celów zastosowań AI i ich zgodności z dyrektywami etycznymi. Przyjrzą się też potencjalnym mechanizmom odnoszącym się do roli człowieka w ramach procedur administracyj-

² M. Górski, *Dziesiąty element. Sztuczna inteligencja jako sędzia a prawo do sądu*, PME 2022, nr 3, i przywołana tam literatura; P.M. Nowotko, *AI in judicial application of law and the right to a court*, „Procedia Computer Science” 2021, vol. 192, <https://www.sciencedirect.com/science/article/pii/S1877050921017324>, [dostęp 15.07.2024]; M. Bartoszek, *Zastosowanie sztucznej inteligencji w sądownictwie w świetle zasady skutecznej ochrony sądowej*, „Folia Iuridica Universitatis Wratislaviensis” 2022, vol. 11, s. 8–29.

nych, w tym mechanizmom regulującym kwestię zredukowania tej roli w procesach decyzyjnych.

Obserwując wydarzenia ostatnich kilkunastu miesięcy, a zwłaszcza rozgłos, jaki zyskało rozwiązanie ChatGPT stworzone przez OpenAi, wydawać by się mogło, że narzędzia sztucznej inteligencji (dalej: „AI” lub „SI”) pojawiły się w gospodarce na istotną skalę dopiero w listopadzie 2022 r.³ Skutkiem takiego przekonania jest również utożsamienie AI z tzw. generatywną sztuczną inteligencją, na której modelu ChatGPT został oparty. Nie kwestionując roli, jaką rozwiązanie wspierane przez Microsoft odegrało w upowszechnieniu wykorzystania narzędzi AI, podkreślić należy, że termin AI jest znacznie szerszy, a rozwój i wdrażanie narzędzi, także w zastosowaniach publicznych, opartych na różnorodnych systemach AI to perspektywa co najmniej kilkunastu lat.

Pod pojęciem sztucznej inteligencji mieści się wiele rozmaitych systemów i rozwiązań. Szeroko termin ten definiuje WIPO, wskazując, że to „dyscyplina informatyki, której celem jest opracowywanie maszyn i systemów, które mogą wykonywać zadania uważane za wymagające ludzkiej inteligencji, przy ograniczonej lub zerowej ingerencji człowieka”⁴. W konsekwencji w tym szerokim pojęciu pomieszczają się różnorodne techniki uczenia maszynowego (ang. *machine learning* – ML)⁵, w tym nadzorowanego (ang. *supervised machine learning*)⁶, nienadzorowanego (ang. *unsupervised machine learning*)⁷ czy uczenia głębokiego (ang. *deep learning*) wykorzystywanego do przyuczania głębokich sieci neuronowych (ang. *Deep Neural Network* – DNN)⁸, procesy przetwarzania języka naturalnego (ang. *Natural Lan-*

³ 30 listopada 2022 r. miało miejsce upublicznienie ChatGPT.

⁴ Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence; Document Code, WIPO/IP/AI/2/GE/20/1 REV, 29 maja 2020 r., https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=499504 [dostęp 9.05.2024].

⁵ Uczenie maszynowe to zbiór technik polegających na przetwarzaniu danych. Dane te mogą stanowić przygotowany wcześniej zbiór lub może być to strumień danych przetwarzanych w czasie rzeczywistym; J. Popper, J. Hermann, S. Bergweiler, *Artificial intelligence across industries*, IEC White Paper, XI 2018, s. 35 i n.

⁶ Sposób uczenia, w którym zbiór danych treningowych, na których uczy się algorytm, zawiera dołączone rozwiązanie problemu, tzw. etykiety albo klasy, <https://www.ibm.com/topics/machine-learning> [dostęp 1.06.2024].

⁷ Sposób uczenia modelu AI, w którym dane uczące są nieoznakowane tzn. nie posiadają etykiet. IBM, *What is machine learning (ML)?*, <https://www.ibm.com/topics/machine-learning> [dostęp 1.06.2024].

⁸ Podkategoria uczenia maszynowego oparta na głębokich sieciach neuronowych, które pozwalają maszynie na samodzielne uczenie. Struktura głębokich sieci neuronowych składa się z wielu warstw, których jednostki przekształcają dane wejściowe w informacje, wykorzystywane następnie przez kolejne warstwy. Dzięki tej strukturze maszyna może uczyć się poprzez własne przetwarzanie danych. Microsoft, <https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2> [dostęp 1.06.2024].

guage Processing – NLP)⁹, a także powstałe przy ich zastosowaniu modele, w tym modele generatywnej sztucznej inteligencji (GenAI)¹⁰.

Precyzyjne określenie pojęcia systemu AI stanowiło jedno z wyzwań dla twórców AI Act¹¹, unijnego rozporządzenia, którego celem jest kompleksowa regulacja kwestii związanych z AI, czego wyrazem były kilkukrotne modyfikacje definicji systemu AI w kolejnych jego projektach¹². Ostatecznie przyjęta definicja tego pojęcia wskazuje, że system AI oznacza system oparty na maszynach zaprojektowany do działania na różnych poziomach autonomii, który może wykazywać zdolność adaptacji po wdrożeniu i który, w celach jawnych lub ukrytych, wnioskuje na podstawie otrzymywanych danych wejściowych, w jaki sposób generować dane wyjściowe, takie jak prognozy, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne.

Zainteresowanie możliwościami wykorzystywania AI w różnych aspektach działalności człowieka jest powszechne. W styczniu 2019 r. WIPO opublikowało raport z badań dotyczących trendów technologicznych, w którym przeanalizowano dane zawarte w zgłoszeniach patentowych i publikacjach naukowych¹³. Przeanalizo-

⁹ Techniki i metody umożliwiające urządzeniom rozpoznawanie, rozumienie i generowanie tekstu i mowy poprzez połączenie lingwistyki komputerowej – opartej na regułach modelowania ludzkiego języka – z modelowaniem statystycznym, uczeniem maszynowym (ML) i głębokim uczeniem. Za: J. Holdsworth, *What is NLP (natural language processing)?*, IBM, 6.06.2024, <https://www.ibm.com/topics/natural-language-processing> [dostęp 7.05.2024].

¹⁰ Wielkoskalowe modele uczenia maszynowego, wytrenowane na ogromnych zbiorach danych, wykorzystujące sieci neuronowe do identyfikowania wzorców w istniejących danych w celu generowania nowych treści. wytrenowane na obszernych zbiorach danych. Najpopularniejszymi z nich są duże modele językowe (ang. *large language models*, LLM), ale modele takie mogą być oparte na obrazach, kodzie oprogramowania, danych chemicznych czy finansowych. Modele te generują na podstawie tzw. promptów (komend) nowe treści w postaci tekstu pisanego, dźwięku, obrazów lub filmów. Za: T. Doliński, *Sztuczna inteligencja: od wąskiej specjalizacji do twórczej wszechstronności*, „Gazeta SGH”, 2.02.2024, <https://gazeta.sgh.waw.pl/meritum/sztuczna-inteligencja-od-waskiej-specjalizacji-do-tworczej-wszechstronnosci> [dostęp 1.06.2024].

¹¹ Rozporządzenie Parlamentu Europejskiego w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektywy 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji). Projekt aktu prawnego aktualny na dzień 19.04.2024 r., https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_PL.pdf [dostęp 7.05.2024].

¹² L. White, M. Evans, *The AI Act – A step closer to the first law on Artificial Intelligence*, The Norton Rose Fulbright Data Protection Report, 18.05.2023, <https://www.dataprotectionreport.com/2023/05/the-ai-act-a-step-closer-to-the-first-law-on-artificial-intelligence/> [dostęp 8.05.2024].

¹³ Patent Trends – Assistive Technology, WIPO 2021, https://www.wipo.int/tech_trends/en/ [dostęp 06.05.2024]; WIPO Launches State-of-the-Art Artificial Intelligence-Based Image Search Tool for Brands, Geneva, PR/2019/831, 1.09.2019, https://www.wipo.int/pressroom/en/articles/2019/article_0005.html [dostęp 6.05.2024].

wano poziom innowacji w zakresie AI i ustalono, że od czasu pierwszego pojawienia się AI w latach 50. miało miejsce ponad 340 tys. zgłoszeń patentowych związanych z AI (z czego większość po 2013 r.)¹⁴ oraz 1,6 mln publikacji naukowych związanych z AI¹⁵. Tylko pomiędzy rokiem 2016 a 2022 wzrost zgłoszeń patentowych obejmujących systemy AI wzrósł o 718%¹⁶.

Narzędzia opierające się na AI od lat zyskują stale na popularności również w obszarze własności intelektualnej, dając nadzieję na usprawnienie przyspieszenie i usprawnienie procesów rejestracji czy zarządzania prawami. I trudno się temu dziwić, gdyż liczba zgłoszeń przedmiotów praw własności intelektualnej rośnie z roku na rok. Aby zrozumieć skalę zjawiska, warto spojrzeć na niektóre tylko statystyki. Przykładowo do EUIO w latach 2021–2023 wpłynęło ok. 300 tys. zgłoszeń wzorów wspólnotowych oraz wpływało rocznie ok. 175 tys. zgłoszeń unijnych znaków towarowych; liczba złożonych odwołań w sprawach *inter partes* wzrosła z 1 559 w 2021 r. do 1 855 w 2023 r.¹⁷ Zgodnie z szacunkami WIPO w 2020 r. w 149 urzędach na świecie zarejestrowanych było ok. 64,4 mln znaków towarowych. W tym samym roku zgłoszono ich ok. 13,4 mln, co stanowi 1,9 mln więcej niż w roku poprzedzającym¹⁸. Liczby te pokazują skalę potencjalnych wyzwań, z którymi muszą się mierzyć urzędy, oceniając wpływające do nich zgłoszenia i rozstrzygając w sprawach spornych. Jako obszary zastosowań AI można wskazać m.in. wsparcie zgłaszających przy doborze odpowiednich klas towarów i usług w trakcie rejestracji znaków, sygnalizowanie przeszkód rejestracji, identyfikację oczywistych błędów, zastosowanie interaktywnych formularzy. Algorytmy AI wykorzystywane mogą być również przez urzędy m.in. do klasyfikacji zgłoszeń, do poszukiwania wcześniejszych dokumentów ze stanu techniki w odniesieniu do patentów, czy też do identyfikowania kolizji ze wcześniejszymi prawami ze znaków towarowych¹⁹.

¹⁴ W 2020 r. liczba zgłoszeń patentowych związanych z AI szacuje się na ponad 93 tys., Annual patent applications related to AI, by status, World, 2020, <https://ourworldindata.org/grapher/ai-related-patents-applications-and-patents-granted?time=latest> [dostęp 8.05.2024].

¹⁵ *Ibidem*.

¹⁶ WIPO, AI Inventions, <https://tind.wipo.int/record/49113?v=pdf> [dostęp 8.05.2024].

¹⁷ T. Gawliczek, *EUPO otwiera szeroko drzwi dla alternatywnych metod rozwiązywania sporów*, 27.03.2024, <https://www.prawo.pl/biznes/alternatywne-metody-rozwiazywania-sporow-w-euipo,526120.html> [dostęp 29.04.2024].

¹⁸ WIPO, *World Intellectual Property Indicators*, 2021, https://www.wipo.int/edocs/pubdocs/en/wipo_pub_941_2021.pdf [dostęp 30.04.2024].

¹⁹ D.S. Gangjee, *A quotidian revolution: artificial intelligence, and trade mark law*, [w:] R. Abbott (red.), *Research Handbook on Intellectual Property and Artificial Intelligence*, Edward Elgar, 2022, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4081317 [dostęp 3.05.2024].

Nic dziwnego zatem, że szybki rozwój algorytmów AI w ostatnich latach sprawił, że zaczęto coraz bardziej wykorzystywać technologię AI do usprawniania pracy w obszarze IP (z ang. *Intellectual Property*), przechodząc tym samym od teoretycznych badań do praktycznego jej zastosowania. AI stopniowo wkraczała w coraz to kolejne obszary działalności praktyków IP i rewolucjonizowała ich podejście do pracy, zmieniając i usprawniając procesy jej wykonywania²⁰. Nic dziwnego więc, że wykorzystanie systemów AI w obszarze własności intelektualnej nie ograniczyło się do celów komercyjnych, ale zaczęło się odbywać na poziomie organizacji państwowych i międzynarodowych realizujących zadania w obszarze ochrony własności intelektualnej. Mimo szeregu trudności związanych z jej stosowaniem badają one i wdrażają systemy AI, aby ułatwić pracę urzędnikom.

Biura ds. własności intelektualnej na całym świecie dostrzegają potencjał sztucznej inteligencji oraz to, jak ta technologia zmienia obecny krajobraz procesów i procedur okołoprawnych. Eksploracja, badanie AI w obszarze IP i wykorzystywanie jej aspektów mają wymiar globalny. Szereg inicjatyw na poziomie krajowym, regionalnym i międzynarodowym warte jest odnotowania.

Specjalistów IP na całym świecie nurtują jednak obawy dotyczące sztucznej inteligencji i związanych z nią problemów, mimo że jest ona ujmowana i przedstawiana przez zainteresowane instytucje jako narzędzie ułatwiające pracę ludziom, zwiększające efektywność czy zapewniające większą jednolitość (koherencję) praktyk w poszczególnych obszarach. Nie zmienia to jednak faktu, że zastosowania AI, a w szczególności generatywnej sztucznej inteligencji, budzi niepokój związany z etyką postępowania, w tym transparentnością działania tych systemów oraz faktycznym zakresem zastępowania tzw. czynnika ludzkiego w ramach procesów związanych z IP oraz przyszłymi tego implikacjami, zwłaszcza jeśli mowa o ich zastosowaniu w instytucjach publicznych.

Temat ten ze względu na jego przekrojowość autorki zdecydowały podzielić się na trzy części, w każdej z nich analizując nieco inny aspekt problemu.

W niniejszym artykule, będącym pierwszą częścią cyklu, omówiono szczerzej projekty prowadzone przez wiodące światowe organizacje, tj. EUIPO, WIPO, EPO oraz USPTO, które z perspektywy praktyków prawa wydają się kluczowe i mają największe oddziaływanie na pozostałe jurysdykcje. W części drugiej dokonano przeglądu rozwiązań zaimplementowanych na szczeblu krajowym. Natomiast w części

²⁰ S. Katyal, A. Kesari, *Trademark Search. Artificial Intelligence and the Role of the Private Sector*, „Berkeley Technology Law Journal”, 4.01.2021, opublikowany 6.02.2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3760112 [dostęp 3.05.2024].

trzeciej cyklu poruszono zagadnienia z zakresu etyki dotyczące godnej zaufania AI, starając się spojrzeć na prezentowane rozwiązania biur ds. własności intelektualnej przez pryzmat przyjętych przez te instytucje celów zastosowań AI i ich zgodności z dyrektywami etycznymi.

Systemy AI urzędach IP na świecie – przegląd inicjatyw

1) WIPO²¹

WIPO przyjęło na siebie rolę lidera w zakresie różnorodnych inicjatyw dotyczących wykorzystania AI w różnych aspektach własności intelektualnej. Poza działalnością w obszarze analizy i promocji rozwiązań legislacyjnych²² jest to widoczne także w zastosowaniach administracyjnych. WIPO prowadzi m.in. aktualizowane zestawienia inicjatyw związanych z narzędziami AI w praktykach urzędów krajowych i regionalnych, m.in. Index of AI initiatives in IP offices²³. Na uwagę zasługuje fakt, że WIPO stworzyło forum służące dyskusji i promocji różnorodnych inicjatyw zarówno na szczeblu międzynarodowym, jak i lokalnym. Od 2020 r. organizowane są cykliczne spotkania w ramach „WIPO Conversation on IP and Frontier Technologies”, podczas których przedstawiciele polityki, administracji, biznesu i nauki omawiają kwestie związane z rozwojem technologicznym i jego wpływem na własność intelektualną. Piąta z tych sesji, która odbyła się w kwietniu 2022 r., dotyczyła wykorzystania narzędzi AI przez urzędy patentowe²⁴.

Warto zauważyć, że od początku w dyskusjach toczących się w ramach powyższych inicjatyw adresowano kwestie etyczne związane z wykorzystaniem AI w administracji i dopuszczalności stosowania AI w procesie wydawania decyzji. W dokumencie pokonferencyjnym WIPO z 21 maja 2020 r.²⁵ postanowiono szereg pytań

²¹ WIPO, *AI tools for common use*, 7.12.2023, https://www.wipo.int/edocs/mdocs/cws/en/cws_11/cws_11_9b_ai_ib.pdf [dostęp 5.05.2024].

²² WIPO stworzyło i prowadzi indeks inicjatyw legislacyjnych w obszarze AI i własności intelektualnej: WIPO, *Artificial Intelligence and Intellectual Property Strategy Clearing House*, https://www.wipo.int/about-ip/en/artificial_intelligence/strategy-search.jsp?territory [dostęp 2.05.2024].

²³ WIPO, *Index of AI initiatives in IP offices*, https://www.wipo.int/about-ip/en/artificial_intelligence [dostęp 30.04.2024].

²⁴ WIPO, *WIPO Conversation on Intellectual Property (IP) and Frontier Technologies: Fifth Session*, https://www.wipo.int/meetings/en/details.jsp?meeting_id=68588 [dostęp 3.05.2024].

²⁵ WIPO, *WIPO Conversation On Intellectual Property (Ip) And Artificial Intelligence (Ai) Second Session Revised Issues Paper On Intellectual Property Policy And Artificial Intelligence*, 21.05.2020, https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1_rev.pdf [dostęp 2.05.2024].

m.in. o dopuszczalność AI w zakresie podejmowania decyzji w sprawie własności intelektualnej:

„(i) Czy sztuczna inteligencja powinna być dopuszczona do podejmowania decyzji w postępowaniu w sprawie wniosków dotyczących własności intelektualnej? Jakie są kwestie prawne związane z wykorzystywaniem aplikacji SI do podejmowania decyzji w procesie udzielania praw własności intelektualnej?

(ii) Jakie rodzaje decyzji mogą być podejmowane przez SI w urzędach ds. własności intelektualnej? Czy istnieją dodatkowe obszary, które urzędy ds. własności intelektualnej powinny zbadać w celu wdrożenia narzędzi SI do ścigania i rejestracji?

(iii) Czy należy podjąć jakieś działania polityczne lub praktyczne w celu zapewnienia odpowiedzialności za decyzje podejmowane w ramach oceny i administrowania wnioskami o ochronę własności intelektualnej, gdy te decyzje są podejmowane przez aplikacje SI? Jakich zasad powinny przestrzegać aplikacje SI w ocenie i administrowaniu zgłoszeniami własności intelektualnej (na przykład zachęcanie do przejrzystości w odniesieniu do korzystania ze sztucznej inteligencji i w odniesieniu do stosowanej technologii)?

(iv) Czy należy przewidzieć jakiegokolwiek zmiany legislacyjne lub regulacyjne w celu ułatwienia lub zaradzić konsekwencjom podejmowania decyzji przez aplikacje SI (na przykład przegląd przepisów prawnych dotyczących uprawnień i swobód decyzyjnych niektórych wyznaczonych urzędników)?

(v) Czy obecne mechanizmy odwoławcze są przygotowane do rozpatrywania odwołań od decyzji podejmowanych przez SI²⁶ [tłum. własne].

Pomimo upływu czterech lat od sformułowania tych kwestii wciąż trudno o jednoznaczne zasady, które mogłyby stanowić wytyczne dla uregulowania ww. kwestii. Nie stanowi to jednak przeszkód w rozwoju i implementacji narzędzi opartych na AI także w praktyce WIPO.

Obecnie dostępnych dla klientów zewnętrznych i ekspertów urzędu jest szereg narzędzi w obszarze klasyfikacji, wstępnej oceny czy tłumaczeń maszynowych.

Jednym z nich jest system Global Goods & Services Terms Explorer²⁷ – eksplorator globalnych terminów dla towarów i usług. Narzędzie to m.in. sugeruje odpowiednie terminy klasyfikacji dla towarów i usług zawierające poszukiwany ciąg znaków. Natomiast funkcja wyszukiwania semantycznego działająca w oparciu o sieci

²⁶ *Ibidem*.

²⁷ WIPO, *Global Goods & Services Terms Explorer*, <https://www.wipo.int/web/ai-tools-services/global-goods-services-terms-explorer> [dostęp 6.05.2024]; WIPO, Prezentacja pt. *Global Goods & Services Terms Explorer* dostępna pod adresem: https://www.wipo.int/edocs/mdocs/madrid/en/mm_id_wg_21/mm_id_wg_21_roundtable_topic_3_ib_2.pdf [dostęp 6.05.2024].

neuronowe wskazuje na rekomendowane podobne terminy. Dodatkowo w tej funkcji dostępny jest ranking semantycznej bliskości i kategorie klasyfikacji nicejskiej. Wskazywane są także ekwiwalenty wyszukiwanych terminów w ośmiu różnych językach. System został wytrenowany na bazach aktywnych i wygasłych znaków m.in. z urzędów w Australii, Kanadzie, Korei, Wielkiej Brytanii, WIPO czy EUIPO.

Kolejnym wartym odnotowania narzędziem jest WIPO Image Similarity Search for Trademarks – wyszukiwarka podobieństwa znaków towarowych skierowana zarówno do ekspertów WIPO, jak i użytkowników zewnętrznych. Narzędzie to działa od 2014 r., natomiast w 2019 r. udoskonalono go, implementując głębokie sieci neuronowe, trenowane przy wykorzystaniu klasyfikacji elementów graficznych Systemu Madryckiego i baz danych dużych urzędów patentowych. Wyniki podobnych znaków prezentowane są na podstawie wyszukiwania łączącego dany koncept (określenie słowne towaru/usługi) z jej przedstawieniem graficznym. W tworzeniu narzędzia uczestniczyły urzędy patentowe z 45 krajów. W 2021 r. wprowadzony został unowocześniony algorytm, który identyfikuje elementy graficzne w znakach towarowych z dużo większą precyzją, za pomocą inteligentnego wykrywania elementów słownych osadzonych w logo podczas jego indeksowania²⁸.

W WIPO działa również automatyczne narzędzie klasyfikacji patentowej (*IPC-CAT- IPC Computer-Assisted Categorization*) oparte na zaawansowanych algorytmach uczenia maszynowego. Pozwala ono na automatyczne przypisywanie zgłoszeń patentowych do odpowiedniej dziedziny technicznej przy wykorzystaniu kodów Międzynarodowej Klasyfikacji Patentowej (IPC). Użytkownik, wpisując tekst, np. streszczenie wynalazku, uzyskuje predykcje symboli IPC, które najlepiej odpowiadają dziedzinie technicznej, do której odnosi się wprowadzony tekst, a także wskazują na poziom trafności proponowanych dopasowań²⁹.

Ponadto wdrożono także Asystenta Klasyfikacji Wiedeńskiej – Vienna Assistant Classification, który pozwala na automatyczne proponowanie kodów klasyfikacji do poszczególnych elementów graficznych znaków towarowych. Podobnie jak w przypadku klasyfikacji patentowej ma to na celu zwiększenie efektywności procesu klasyfikacji zgłoszeń³⁰.

²⁸ WIPO, *Just released: Improved Conceptual Image Similarity algorithm in the Global Brand Database*, 8.01.2021 r., https://www.wipo.int/reference/en/branddb/news/2021/news_0001.html [dostęp 5.05.2024]; WIPO Launches State-of-the-Art Artificial Intelligence-Based Image Search Tool for Brands, Geneva, PR/2019/831, https://www.wipo.int/pressroom/en/articles/2019/article_0005.html [dostęp 6.05.2024].

²⁹ WIPO, *AI-based Classification for IP Data*, <https://www.wipo.int/web/ai-tools-services/classification-assistant> [dostęp 6.05.2024].

³⁰ *Ibidem*.

W WIPO funkcjonuje również oparty na sieciach neuronowych system maszynowego tłumaczenia, oferujący m.in. możliwość tłumaczenia dokumentów patentowych³¹.

Obecnie trwają prace nad wprowadzeniem możliwości wyszukiwania obrazkowego podobnych wzorów przemysłowych w bazie Global Design Database oraz narzędzia pozwalającego użytkownikom na generowanie słów kluczowych ułatwiających wyszukiwanie podobnych znaków towarowych w celu przyspieszenia wypełniania wymaganych opisów zgłaszanych znaków³².

2) EUIPO³³

Inicjatywy związane z możliwością wykorzystania różnych rodzajów narzędzi AI, a także implikacjami ich zastosowania są widoczne na forum Unii Europejskiej od kilku już lat. Efektem tych prac na poziomie legislacyjnym jest AI Act. Niemniej, także w sferze praktycznej instytucje UE podejmują próby wdrożeń elementów AI w celu zautomatyzowania, a przynajmniej usprawnienia części zadań wykonywanych do tej pory przez urzędników. W tym zakresie wyróżnia się niewątpliwie EUIPO³⁴, gdzie wdrożono już kilka narzędzi o przeznaczeniu zarówno wewnętrznym, jak i skierowanych do zewnętrznych odbiorców.

Interesujący przegląd celów, jakie EUIPO stawia sobie w tym obszarze, stanowi dokument Strategic Plan 2025³⁵ (SP 2025) EUIPO przyjęty w 2019 r.³⁶ Oparte na AI

³¹ WIPO, *WIPO Translate – Breaking Language Barriers with AI*, <https://www.wipo.int/web/ai-tools-services/wipo-translate> [dostęp 6.05.2024].

³² WIPO, <https://www.wipo.int/web/ai-tools-services> [dostęp 6.05.2024].

³³ Komisja Europejska, *Artificial Intelligence in the European Commission (AI@EC) A strategic vision to foster the development and use of lawful, safe and trustworthy Artificial Intelligence systems in the European Commission*, dokument zamieszczony na oficjalnej stronie Komisji Europejskiej w dniu 24.01.2024, <https://commission.europa.eu/system/files/2024-01/EN%20Artificial%20Intelligence%20in%20the%20European%20Commission.PDF> [dostęp 1.05.2024]; EUIPO, *Strategic Plan 2025 European Union Intellectual Property Office*, https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/contentPdfs/about_euipo/strategic_plan/SP2025_en.pdf [dostęp 30.04.2024]; T. Lince, *Innovation at the EUIPO: spotlight on digital tools and services*, 20.07.2023, <https://www.worldtrademarkreview.com/data/ip-office-tools-and-services/ip-office-tools-and-services/article/innovation-the-euipo-spotlight-digital-tools-and-services> [dostęp 1.05.2024].

³⁴ W 2021 r. EUIPO obok Koreańskiego Urzędu Patentowego zostało uznane za najbardziej innowacyjny urząd ds. własności intelektualnej na świecie w rankingu „World Trademark Review”, za: S. Bradley, *Blockchain and AI: the EUIPO is top for innovation*, 11.02.2022, <https://www.lexology.com/library/detail.aspx?g=6445f08c-5819-4544-96db-335538f15498> [dostęp 3.05.2024].

³⁵ Obecnie trwa faza konsultacji nad Planem Strategicznym EUIPO do 2030 r., którego jednym z kluczowych filarów ma być optymalizacja wydajności i skuteczności operacyjnej poprzez wdrożenie i wykorzystanie najnowocześniejszych technologii. EUIPO, *Open consultation on the Strategic Plan 2030*, <https://www.euipo.europa.eu/en/the-office/governance/strategic-plan/sp2030> [dostęp 3.05.2024].

³⁶ EUIPO, *Strategic Plan 2025 European Union Intellectual Property Office*, https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/contentPdfs/about_euipo/strategic_plan/SP2025_en.pdf [dostęp 30.04.2024].

mechanizmy już wtedy wspomagały prace urzędników w obszarze porównywania towarów i usług, klasyfikacji, wyszukiwania obrazów czy poprzez funkcje tłumaczenia maszynowego.

Strategic Plan 2025 wyznaczył dalsze ambitne cele dla EUIPO w segmencie rozwiązań technologicznych. Jako jeden z nich (Cel 3.2. pt. Ewolucja w dobie ery cyfrowej) wskazano „zbadanie potencjału rozszerzenia tego typu rozwiązań na inne aspekty cyklu życia znaku towarowego i wzoru w obszarach, takich jak analiza obrazów pod kątem nieprawidłowej treści, tworzenie pism i analiza braków [w aplikacjach zgłoszeniowych – przyp. wł.] w celu ich ograniczenia. Sztandarowe narzędzia, w szczególności TMView, DesignView i zharmonizowana baza danych, zostaną zmodernizowane, aby spełnić oczekiwania użytkowników. Wdrożone zostaną zmodernizowane rozwiązania sieciowe, niektóre z nich będą zasilane przez sztuczną inteligencję i blockchain, z większym naciskiem na poprawę jakości usług i danych, a także na możliwość wyszukiwania nowych rodzajów znaków towarowych”³⁷.

Paradygmatem inicjatyw podejmowanych w ramach EUIPO (tak samo jak w całej UE) jest ukierunkowanie na człowieka (ang. *human centered approach*). Projekty, które EUIPO wdraża lub nad którymi pracuje, z zasady nie mają zatem zastąpić człowieka (tu eksperta EUIPO), ale wspomagać jego pracę. W opisie Kluczowej Inicjatywy 2 Strategic Plan 2025 dotyczącej usprawniania i zwiększania efektywności pracy podkreślono, że cel ten zostanie osiągnięty właśnie przy zwiększeniu „wykorzystania narzędzi pomocniczych, opartych na sztucznej inteligencji i dużych zbiorach danych, które będą oferować wszystkie istotne i kontekstowe informacje w przyjazny dla użytkownika i terminowy sposób, upraszczać bardziej rutynowe operacje oraz dalej usprawniać zarządzanie plikami i inteligentne przydzielanie zadań, co z kolei pozwoli ekspertom skoncentrować swoje wysiłki na zadaniach, które naprawdę wymagają zaangażowania człowieka”³⁸.

Jak pokazuje zawarty w tym opracowaniu przegląd, co najmniej część z założeń przyjętych w ramach Strategic Plan 2025 udało się EUIPO wdrożyć. Jednym z pierwszych narzędzi było powstałe w 2017 r. narzędzie E-translation, pozwalające na tłumaczenie maszynowe w 23 językach. Jest to narzędzie Komisji Europejskiej, z którego korzystają także inne instytucje EU. W przypadku EUIPO wykorzystywane jest głównie dla tłumaczenia decyzji wydawanych w różnych ję-

³⁷ *Ibidem.*

³⁸ *Ibidem.*

zykach postępowania oraz do tłumaczenia terminów określających towary i usługi w zgłoszeniach znaków³⁹.

Powstałe w 2010 r. narzędzie TMView⁴⁰ obok eSearch⁴¹ jest wiodącym systemem wyszukiwania znaków, które oferuje funkcje wyszukiwania obrazkowego. Od 2020 r. jest konsekwentnie rozwijane z użyciem elementów AI i *blockchain*. W 2022 r. zaimplementowany został algorytm umożliwiający natychmiastowe wyszukiwanie znaków towarowych podobnych do innego załadowanego do wyszukiwarki znaku. System został wytrenowany lokalnie przez EUIPO na kilku milionach znaków. Od 2023 r. wyszukiwarka jest zasilana dodatkowo bazami danych największych urzędów patentowych: japońskiego, chińskiego, amerykańskiego i koreańskiego. Razem z EUIPO ich rejestry obejmują 60% wszystkich rejestracji znaków towarowych na świecie. Obecnie w bazie danych narzędzia TMView znajduje się ponad 120 mln znaków⁴².

Podobne narzędzia dostępne są w wyszukiwarce wzorów przemysłowych DesignView, której zmodernizowana wersja, zasilona bazami danych urzędów krajowych została oddana do użytku w maju 2024 r.⁴³

Semantyczna wyszukiwarka towarów i usług Goods and Services semantic search sugeruje kategorie towarów i usług (synonimy, podobne usługi lub towary) w procesie zgłaszania znaku towarowego. Rozwiązanie wbudowane w usługach EUTM e-Filing i Easy Filing⁴⁴ ma zapewnić wydajne wyszukiwanie i walidację terminów wskazywanych przez zgłaszających w dokumentach aplikacyjnych. Oparte jest na modelach *open source*, m.in. Fastex.

Pre-assesment check to narzędzie do wstępnego sprawdzenia zdolności rejestrowej znaku w zakresie niektórych podstaw odmowy (jak np. znak wprowadzający w błąd, sprzeczność z porządkiem publicznym), a także potencjalnego konfliktu z zarejestrowaną domeną z rozszerzeniem EU czy z zarejestrowaną odmianą roślin.

³⁹ EUIPO, *Improved machine translation in eSearch Case Law*, 17.10.2019, https://euipo.europa.eu/ohimportal/en/key-user-newsflash/-/asset_publisher/dIGJZDH66W8B/content/improved-machine-translation-in-esearch-case-law/ [dostęp 30.04.2024].

⁴⁰ Wyszukiwarka znaków towarowych zawartych w bazie EUIPO i innych urzędów patentowych dostępna pod adresem: <https://www.tmdn.org/tmview/#/tmview>.

⁴¹ Wyszukiwarka EUIPO zawierająca m.in. dane o znakach towarowych, wzorach, uprawnionych i pełnomocnikach, dostępna pod adresem: <https://euipo.europa.eu/eSearch/>.

⁴² EUIPO, *AI at EUIPO: In-house image search in TMview extended to all TM5 offices*, informacja na oficjalnej stronie EUIPO, <https://www.euipo.europa.eu/pl/news/ai-at-euipo-in-house-image-search-in-tmview-extended-to-all-tm5-offices> [dostęp 1.05.2024].

⁴³ EUIPN, *DesignView Image Search*, 7.05.2024, <https://www.tmdn.org/publicwebsite/#/news/2580624> [dostęp 7.05.2024].

⁴⁴ Usługi oferowane przez EUIPO w celu dokonywania zgłoszeń znaków towarowych przez internet.

Narzędzie można aktywować w formularzu zgłoszenia znaku unijnego – powiadamia ono o potencjalnych przeszkodach za pomocą komunikatów (alertów)⁴⁵.

Wszystkie wskazane funkcjonalności i systemy są dostępne dla użytkowników zewnętrznych Urzędu. Niemniej EUIPO posiada również narzędzia AI przeznaczone tylko do wewnętrznego użytku ekspertów. Jednym z nich jest porównywarka towarów i usług – AI-based comparison of goods and services. Jej algorytm umożliwia ocenę danej pary towarów lub usług w oparciu o dane historyczne (pochodzące z bazy decyzji wydanych w I instancji EUIPO) oraz Narzędzie Podobieństwa Similarity Tool, które za pomocą wyszukiwania semantycznego wskazuje najbardziej zbliżone terminy. Ten model AI został wytrenowany na podstawie decyzji obejmujących 420 tys. par towarów i usług. Stworzony algorytm pozwala nie tylko na wyszukanie istniejących już przypadków takich samych par towarów i usług, ale zapewnia szersze porównanie oparte na analogicznych terminach. Wyniki są następnie udostępniane ekspertom w celu rozważenia ich relewantności przy podejmowaniu decyzji. System oferuje dodatkową funkcję, dostarczając cytaty z historycznych dokumentów, które eksperci mogą wykorzystać do sporządzenia uzasadnienia decyzji w badanej sprawie⁴⁶.

Kolejnym narzędziem jest Image Search for Absolute Grounds, które wskazuje ekspertom znaki towarowe o podobnych reprezentacjach graficznych zawarte w historycznych zgłoszeniach, posługując się parametrami takimi jak definicje dominujących elementów na obrazie oraz kody klasyfikacji wiedeńskiej. W efekcie pozwala na porównanie praktyki dotyczącej zaistnienia podstaw do odmowy rejestracji znaku w podobnych przypadkach w przeszłości⁴⁷.

Z kolei 6ter Database to narzędzie przeznaczone do wykrywania w zgłoszeniach obrazów przypominających flagi i symbole oraz ich oznaczanie pod kątem przeszkody w rejestracji. Obecnie narzędzie znajduje się w fazie testowej i po jej zakończeniu zostanie udostępnione użytkownikom w ramach Pre-assessment tool⁴⁸.

⁴⁵ EUIPO, *New pre-assessment checks in your EUTM application*, <https://www.euipo.europa.eu/pl/news/new-pre-assessment-checks-in-your-eutm-application> [dostęp 30.04.2024].

⁴⁶ EUIPO & New technologies Connecting IP Offices and Users, *Developing Trade Mark and Industrial Design Tools for Users: Leveraging on New Technologies*, https://ipkey.eu/sites/default/files/ipkey-docs/2023/IPKeySEA_nov2023_Carlos-Luna_Developing-Trade-Mark-and-Industrial-Design-Tools-for-Users-Leveraging-on-New-Technologies.pdf [dostęp 1.05.2024]; EUIPO, *New AI-based comparison of goods and services*, 29.03.2022, <https://euipo.europa.eu/ohimportal/en/-/news/new-ai-based-comparison-of-goods-and-services> [dostęp 30.04.2024].

⁴⁷ EUIPO, *Consolidated annual activity report 2022*, https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/contentPdfs/about_euipo/annual_report/Annual_Activity_Report_2022_en.pdf [dostęp 30.04.2024].

⁴⁸ EUIPO & New technologies....

Narzędzie oparte na AI jest obecnie wdrażane także dla celów badania formalnego zgłoszeń wzorów wspólnotowych Design Deficiency Detection. Narzędzie umożliwia wykrycie wadliwych widoków i wskazanie ekspertowi na potencjalne braki formalne rejestracji, które powinny zostać zakomunikowane zgłaszającemu (jeden widok w danym obrazie, neutralne tło, brak tekstu). Po przeprowadzeniu testów i uzyskaniu wysokiego prawdopodobieństwa poprawnego działania planowane jest udostępnienie narzędzia w ramach usługi *e-Filing* także użytkownikom zewnętrznym⁴⁹.

EUIPO obecnie bada potencjał wykorzystania rozwiązań AI i ich implementacji w sferze asystowania ekspertom w przygotowywaniu uzasadnień wydawanych decyzji, w zakresie opisowości i zdolności odróżniającej znaków towarowych, całościowego znaczenia semantycznego znaku słownego, a także sporządzania podsumowań na podstawie istniejących dokumentów⁵⁰.

Od wielu lat w EUIPO funkcjonuje również, ciągle udoskonalany, *chatbot* odpowiadający na pytania zewnętrznych użytkowników i wspomagający w procesie zgłoszeniowym w usługach *e-Filing* i *Easy Filing*⁵¹.

Obecnie EUIPO nie wykorzystuje modelu generatywnej AI⁵². Funkcjonujące narzędzia zostały oparte na innych, mniej zaawansowanych technologiach ML i NLP. Potencjalne zastosowanie modeli generatywnej AI jest ograniczone przede wszystkim ich dostępnością. Są to bowiem modele oferowane przez dostawców zewnętrznych. Stworzenie takiego modelu wewnątrznie przez EUIPO, a nawet Komisję Europejską, jest mało prawdopodobne, a zastosowanie zewnętrznych rozwiązań niesie zbyt wiele ryzyk, chociażby związanych z brakiem transparentności tych modeli czy ewentualnym naruszeniem praw autorskich. Niemniej, obecnie

⁴⁹ EUIPO, *New EUIPO AI tools, empowering customer services*, https://euiipo.europa.eu/knowledge/pluginfile.php/175303/mod_label/intro/AI%20Webinar%20Feb%202022%20%28SL%20review%29.pdf [dostęp: 1.05.2024].

⁵⁰ *EUIPO & New technologies...*

⁵¹ EUIPO, *New EUIPO AI tools, empowering customer services...*

⁵² Tym niemniej nie można wykluczyć pomocniczego korzystania z dostępnych narzędzi generatywnej inteligencji samodzielnie przez urzędników EUIPO, gdyż organy UE nie wydały generalnego zakazu w tym zakresie. W maju 2023 r. Komisja Europejska przygotowała Wytyczne w zakresie używania rozwiązań osób trzecich (Guidelines for Staff on the use of online available generative artificial intelligence tools), wskazując na ewentualne zagrożenia i zasady korzystania z nich. Z przedstawionych tam zasad jasno wynika reguła ograniczonego zaufania do użycia tych modeli, konieczność krytycznej oceny wszelkich odpowiedzi generowanych przez generatywny model sztucznej inteligencji pod kątem potencjalnych uprzedzeń i niedokładnych informacji, zakaz bezpośredniego powielania wyników generatywnego modelu AI w dokumentach publicznych, zwłaszcza prawnie wiążących, *Guidelines For Staff On The Use Of Online Available Generative Artificial Intelligence Tools*, https://www.asktheeu.org/en/request/13063/response/45877/attach/3/guidelines%20on%20the%20use%20of%20online%20generative%20artificial%20intelligence%20tools.pdf?cookie_passthrough=1 [dostęp 2.05.2024].

w instytucjach EU prowadzone są badania nad możliwością wykorzystania rozwiązań generatywnej AI. Tu wspomnieć przede wszystkim należy o narzędziach GPT Lab, AI Sandbox, JRC@GPT. Te prototypowe rozwiązania zostały opracowane do badania i testowania sposobów wykorzystania modeli takich jak ChatGPT, LLAMA czy Mistral w bezpiecznym środowisku (ang. *sandbox*) w celu tworzenia nowych aplikacji i usług⁵³.

Z dużym prawdopodobieństwem należy założyć, że inicjatywy w ramach EUIPO nie pozostaną bez wpływu na działania krajowych urzędów IP. EUIPO dzieli się bowiem informacjami o wszystkich inicjatywach z innymi urzędami m.in. w ramach EUIPN⁵⁴, a narzędzia AI stworzone przez EUIPO mogą również zostać udostępnione krajowym urzędom patentowym, zamierzającym wdrożyć takie rozwiązania⁵⁵.

3) EPO

Od 2011 r. EPO wdraża tzw. IT roadmap – program informatycznej modernizacji urzędu⁵⁶. W 2019 r. w ramach EPO został powołany specjalny zespół Data Science, którego zadaniem było stworzenie i implementacja AI, w tym systemów ML, aby podnieść szybkość i jakość procedury oceny zgłoszeń patentowych. Zespół ten był i jest wspierany przez ekspertów merytorycznych EPO. Obszary prac zespołu to m.in. zastosowania NLP, Computer Vision oraz ich aplikacja w procesie oceny zgłoszeń patentowych – klasyfikacji, poszukiwań wcześniejszego stanu techniki i tłumaczeń maszynowych⁵⁷.

Strategiczny Plan EPO na 2023 r. jako jeden z trzech głównych celów obrał unowocześnienie systemu IT urzędu⁵⁸. Wśród ujętych w tym celu strategicznym inicjatyw

⁵³ KE, *Artificial Intelligence in the European Commission (AI@EC) A strategic vision to foster the development and use of lawful, safe and trustworthy Artificial Intelligence systems in the European Commission*, dokument zamieszczony na oficjalnej stronie Komisji Europejskiej 24.01.2024, <https://commission.europa.eu/system/files/2024-01/EN%20Artificial%20Intelligence%20in%20the%20European%20Commission.PDF> [dostęp 1.05.2024].

⁵⁴ Sieć Własności Intelektualnej Unii Europejskiej (ang. *European Union Intellectual Property Network*).

⁵⁵ EUIPN, Oficjalna strona European Union Intellectual Property Network, <https://www.tmdn.org/> [dostęp 6.05.2024].

⁵⁶ EPO, *EPO continues IT modernisation with launch of new online filing tool, 1.10.2014*, <https://www.epo.org/en/news-events/press-centre/press-release/2014/451759> [dostęp 2.05.2024].

⁵⁷ R. Tarcu, *AI Initiatives at the EPO*, 29.01.2021, <https://www.oblon.com/ai-initiatives-at-the-epo> [dostęp 2.05.2024]; Yan Tang Demey, D. Golzio, *Search strategies at the European Patent Office*, „World Patent Information” 2020, vol. 63, <https://www.sciencedirect.com/science/article/pii/S0172219020300818?via%3Dihub> [dostęp 6.05.2024].

⁵⁸ EPO, *A vision of sustainability: European Patent Office publishes Strategic Plan 2023 r.*, 27.06.2019 i dostępny tam w formie pdf dokument *EPO – Strategic Plan 2023*, <https://www.epo.org>.

wskazano na implementację narzędzia do przeprowadzenia całościowego procesu oceny zgłoszeń patentowych⁵⁹ oraz udoskonalenie dotychczasowych narzędzi⁶⁰.

Efekt tych wewnętrznych prac stanowi m.in. rozbudowa systemu ANSERA, który jest sukcesywnie uzupełniany nowymi narzędziami i funkcjonalnościami. W połowie 2022 r. wprowadzony został zautomatyzowany system klasyfikacji zgłoszeń patentowych⁶¹. Nowe zgłoszenia są wstępnie przypisywane do jednej z 1400 dziedzin technicznych, a następnie przesyłane do relewantnej grupy ekspertów⁶². Pod koniec 2022 r. dodano nową funkcjonalność pozwalającą na automatyczne przypisywanie kodów CPC (*Cooperative Patent Classification*)⁶³.

W 2022 r. wprowadzony został także oparty w dużym stopniu na technologiach AI moduł Pre Check. Jest on automatycznie włączany wraz z wejściem w aplikację ANSERA. Łączy w sobie różne metody wykorzystujące metadane, słowa kluczowe, klasyfikację. Wyszukuje i pobiera dokumenty ze stanu techniki dla danego zgłoszenia m.in. w oparciu o nazwę zgłaszającego i klasyfikację. Znalezione stan techniki jest pokazywany w formie zoptymalizowanego rankingu pod kątem jego relewantności. Narzędzie dokonuje wyszukiwania wielojęzycznego⁶⁴. W 2023 r. dodana została kolejna istotna funkcjonalność – AI Pre Search – bazująca na zwektoryzowanym przeszukiwaniu korpusu wcześniejszego stanu techniki. Została ona wyszkolona w modelu językowym EP-RoBERTa autorstwa pracowników EPO⁶⁵.

[org/en/news-events/news/vision-sustainability-european-patent-office-publishes-strategic-plan-2023](https://www.epo.org/en/news-events/news/vision-sustainability-european-patent-office-publishes-strategic-plan-2023) [dostęp 2.05.2024].

⁵⁹ Kluczowa inicjatywa 1. Wdrożenie narzędzia patentowego w celu wsparcia kompleksowego elektronicznego procesu przyznawania patentów poprzez modułowe i skalowalne podejście (ang. *Implement a patent tool to support an end-to-end electronic patent granting process through a modular and scalable approach*).

⁶⁰ Kluczowa inicjatywa 2. Ulepszenie istniejących narzędzi EPO i zarządzanie wysokiej jakości bibliotekami wcześniejszych dokumentów stanu techniki (ang. *Improve existing EPO tools and manage high-quality prior-art libraries*).

⁶¹ WIPO, AI applied to Classification Use Cases at the European Patent Office, https://www.wipo.int/edocs/mdocs/classifications/en/ipc_ce_54/ipc_ce_54_p4.pdf [dostęp 4.05.2024].

⁶² EPO ocenia, że wdrożenie tego systemu podniosło skuteczność poprawnej alokacji zgłoszeń do odpowiednich ekspertów aż do 90%. Tylko kompleksowe zgłoszenia są nadal wstępnie oceniane i kwalifikowane przez grupę specjalistów technicznych EPO, *Quality Report 2022, Annex to the Annual Review*, <https://link.epo.org/web/general/annual-review-2022/en-quality-report-2022.pdf> [dostęp 1.05.2024].

⁶³ EPO, *Quality Report 2022, Annex to the Annual Review*, <https://link.epo.org/web/general/annual-review-2022/en-quality-report-2022.pdf> [dostęp 1.05.2024].

⁶⁴ *Ibidem*.

⁶⁵ A. Klenner Bajaja, *Introducing AI-PreSearch: A Revolutionary AI-Driven Search Tool to support our Patent Examiners*, 28.08.2023, <https://www.linkedin.com/pulse/introducing-ai-presearch-revolutionary-ai-driven-tool-klenner-bajaja/> [dostęp 6.05.2024].

Jak twierdzą przedstawiciele EPO, nowa funkcjonalność pozwala na identyfikację wysokiego odsetka odpowiednich cytatów ze stanu techniki i wyszukiwanie relewantnych dokumentów, których żadna inna istniejąca metoda nie mogła wykryć. AI-PreSearch wypełnia 25% istniejącej luki wyszukiwań, dla których dotychczasowy PreSearch nie był w stanie skutecznie zidentyfikować odpowiednich dokumentów⁶⁶. Podkreślić należy przy tym, że ten etap pracy nad oceną zgłoszenia patentowego jest kluczowy. To bowiem dokumenty ze stanu techniki są podstawą oceny kryterium nowości i poziomu wynalazczego patentu.

4) USPTO⁶⁷

USPTO prowadzi badania nad wykorzystaniem algorytmów sztucznej inteligencji w procesie badania zgłoszeń patentowych i znaków towarowych co najmniej od 2018 r. Już wtedy ówczesny dyrektor USPTO postrzegał te narzędzia jako nieodzowne w pracy urzędu, biorąc pod uwagę wzrastającą liczbę zgłoszeń⁶⁸. W tym aspekcie podkreślane jest jednak podejście „pierwszeństwa człowieka” (ang. *human first approach*)⁶⁹. Technologie AI mają przede wszystkim wspomóc pracowników w dokonywaniu czasochłonnych poszukiwań i skierować ich uwagę na najważniejsze aspekty procesu badania.

Jednym z najistotniejszych narzędzi USPTO wykorzystujących zastosowania AI jest Patents End-to-End search tool (AI)-based, będący systemem skierowanym do ekspertów urzędu i zawierającym zestaw narzędzi do wykorzystania w procesie rejestracyjnym zgłoszeń patentowych. Pierwszym modułem tego systemu korzystającym z AI była funkcjonalność More Like This, a obecnie wdrożona została dodatkowo funkcja Similarity Search, pozwalająca na większą elastyczność eksperta w doborze relewantnych kryteriów⁷⁰.

Funkcja Similarity Search jako dane wejściowe otrzymuje wyselekcjonowane przez eksperta informacje dotyczące zgłoszenia patentowego, a następnie przy wy-

⁶⁶ *Ibidem*.

⁶⁷ USPTO, *Emerging Technologies In USPTO Business Solutions*, 29.05.2018, https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=407118 [dostęp 30.04.2024]. USPTO, *Artificial intelligence tools at the USPTO*, 18.03.2021 r., <https://www.uspto.gov/blog/director/entry/artificial-intelligence-tools-at-the> [dostęp 30.04.2024].

⁶⁸ USPTO, *USPTO Leader: AI is a 'Must-Have' Tool*, 15.11.2019, <https://govciomedia.com/uspto-leader-ai-is-a-must-have-tool/> [dostęp 1.05.2024].

⁶⁹ Scannell T., *USPTO takes human-first approach to AI innovation*, 9.09.2022, <https://www.cio.com/article/406584/uspto-takes-human-first-approach-to-ai-innovation.html> [dostęp 1.05.2024].

⁷⁰ USPTO, *New PE2E Search Tool Using AI Search Features*, <https://www.uspto.gov/sites/default/files/documents/og-aiforpe2e-20211220.pdf> [dostęp 4.05.2024].

korzystaniu odpowiednio wytrenowanego algorytmu tworzy listy krajowych i zagranicznych dokumentów patentowych, które cechują się podobieństwem do ocenianego zgłoszenia patentowego. Ekspert posiada również możliwość elastycznego doprecyzowania zapytań, wybierając fragmenty tekstu lub posługując się kodami CPC⁷¹.

Modele AI wykorzystane w omawianym narzędziu zostały wytrenowane przy użyciu publicznie dostępnych baz danych patentowych, klasyfikacji patentowej, cytatów ze wcześniejszych dokumentów oraz ocen podobieństwa dokonanych przez człowieka. Co istotne, aby zminimalizować ryzyko niekorzystnych odchyżeń (ang. *bias*), dane, na których algorytm jest trenowany, zostały oczyszczone m.in. z informacji o autorze oraz zgłaszającym. Przedmiotowe funkcjonalności zostały zbudowane i działają w zgodności z aktami prawnymi mającymi zapewnić bezpieczne i transparentne wykorzystanie AI⁷².

W najbliższych latach należy się spodziewać dalszych prac nad unowocześnieniem tego systemu, zwłaszcza że pojawiające się w przestrzeni publicznej informacje wskazują na planowane inwestycje w ten obszar na poziomie nawet 70 mln dolarów⁷³.

Ponadto od 2020 r. w USPTO funkcjonuje także narzędzie do automatycznej klasyfikacji, wykorzystujące uczenie maszynowe do klasyfikowania dokumentów patentowych przy użyciu systemu Cooperative Patent Classification dla celów wewnętrznych. System sugeruje odpowiednie symbole CPC i pozwala na zidentyfikowanie zastrzeganego przedmiotu w celu dodatkowego udoskonalenia sugerowanych symboli CPC⁷⁴.

USPTO prowadzi także prace w zakresie narzędzi AI mających wspomagać identyfikację podstaw do odmowy udzielenia patentu i wykrywanie błędów w zgłoszeniach. Narzędzia AI mają przyspieszyć i wspomóc identyfikację np. niespójności w języku zastrzeżeń lub błędów formatowania. Celem implementacji tych narzędzi jest dążenie do jak najszybszego zapewnienia zgłaszającym informacji zwrotnych i przyspieszenia w ten sposób procesu badania zgłoszenia patentowego⁷⁵.

⁷¹ USPTO, *New artificial intelligence functionality in PE2E Search*, <https://www.uspto.gov/sites/default/files/documents/ai-sim-search.pdf> [dostęp 4.05.2024].

⁷² Executive Order 13960, Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, <https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government> [dostęp 8.05.2024].

⁷³ L. Hockley, *USPTO set to award \$70m contract to improve AI-driven patent search*, 20.02.2024, <https://www.worldipreview.com/artificial-intelligence/uspto-set-to-award-dollar70m-contract-to-improve-ai-driven-patent-search> [dostęp 2.05.2024].

⁷⁴ WIPO, *Use of Artificial intelligence (AI) in Auto-Classification*, https://www.wipo.int/edocs/mdocs/classifications/en/ipc_ce_53/ipc_ce_53_p8.pdf [dostęp 2.05.2024].

⁷⁵ R. McCoy, *The State of AI Use by the USPTO: Streamlining Patent Examination*, 5.07.2023, <https://www.mccrus.com/2023/07/05/state-of-ai/> [dostęp 4.05.2024].

Od 2021 r. dostępne dla ekspertów są także narzędzia w zakresie rozpoznawania i wyszukiwania obrazu, pozwalające m.in. na detekcję wielokrotnych zgłoszeń dokonywanych w złej wierze. Narzędzia zostały wytrenowane przy pomocy bazy danych obejmującej historyczne 6-cyfrowe kody przypisywane obrazom umieszczanym w zgłoszeniach, które identyfikują poszczególne elementy grafiki. Wytrenowane na ten sposób systemy ML mogą przewidywać kody nowego obrazu zawartego w znaku towarowym⁷⁶.

W lutym 2024 r. USPTO podpisało kontrakt z prywatnym dostawcą, firmą Clarivate, na dostawę w modelu SaaS oprogramowania do przeszukiwania za pomocą obrazu – DesignVision. Narzędzie ma służyć ekspertom USPTO do wyszukiwania wcześniejszych przykładów ze stanu techniki w celu oceny zgłoszeń wzorów przemysłowych (ang. *design patent*). Nie jest to oprogramowanie dla USPTO, ale funkcjonujące narzędzie stworzone przez Clarivate na bazie własnego algorytmu i bazy danych, oferowane również innym instytucjom publicznym i podmiotom prywatnym. Ma ono stanowić uzupełnienie funkcjonujących już w USPTO narzędzi o podobnym zastosowaniu⁷⁷.

W grudniu 2022 r. wdrożono Wirtualnego Asystenta – USPTO virtual assistant⁷⁸ opartego na systemach ML, przeznaczonego dla klientów urzędu w zakresie znaków towarowych. W 2023 r. jego funkcjonalność rozszerzono na aspekty patentowe⁷⁹. Poza odpowiedziami na podstawowe pytania związane z procesem rejestracji Wirtualny Asystent pozwala np. na szybkie sprawdzenie statusu zgłoszenia.

Podsumowanie

Dokonany przegląd pozwala na postawienie wniosku, że narzędzia oparte na różnorodnych technologiach AI stają się elementem codziennej praktyki wiodących urzędów patentowych. Dążenie do automatyzacji powtarzalnych czynności, przyspieszenia procesów czy koherencji praktyk poprzez zastosowanie tych technologii to jedno z priorytetów tych instytucji. Niewątpliwie przyczynia się do omawianego

⁷⁶ W 2021 r. w USPTO zgłoszono 280 tys. znaków z elementem graficznych, wiele z nich imitujących renomowane graficzne znaki towarowe, Du Hao (Henry), *Unleashing the Power of AI to Fight Bad Faith Trademark Registrations*, 6.08.2022, <https://ipwatchdog.com/2022/08/06/unleashing-power-ai-fight-bad-faith-trademark-registrations/id=150649/> [dostęp 4.05.2024].

⁷⁷ M. Woutersen, *USPTO partners with Clarivate for AI-driven image search tool*, 26.02.2024, <https://www.worldipreview.com/patent/uspto-partners-with-clarivate-for-ai-driven-image-search-tool> [dostęp 4.05.2024].

⁷⁸ USPTO, *USPTO launches new Virtual Assistant*, 15.12.2022, <https://www.uspto.gov/about-us/news-updates/uspto-launches-new-virtual-assistant-0> [dostęp 4.05.2024].

⁷⁹ USPTO, *USPTO Virtual Assistant now available for Patents customers*, 10.10.2023, <https://www.uspto.gov/about-us/news-updates/uspto-virtual-assistant-now-available-patents-customers> [dostęp 4.05.2024].

procesu silnie obecna w tych urzędach digitalizacja procesów i dokumentacji. Pozwala ona na dużych zbiorach danych w pożądanym sposobie trenować modele sztucznej inteligencji. Rosnące możliwości techniczne w niedługiej perspektywie potencjalnie zapewne mogłyby umożliwić niemal całkowitą automatyzację procesów w tych obszarach, minimalizując czynnik ludzki.

Korzystając z tych możliwości, chociażby EUIPO coraz wyraźniej wkracza na ścieżkę algorytmizacji w zakresie nie tylko prezentowania relewantnych danych porównawczych, ale także formułowania argumentacji tego wyboru. Przykładowo, wskazane chociażby wyżej narzędzie Image Search for Absolute Grounds przedstawiające ekspertom orzeczenia zapadłe w przeszłości w analogicznych sprawach potencjalnie mogłoby wpływać na treść finalnej decyzji, poprzez zawężenie ocenianych przez eksperta danych i sugestią rozwiązania (jeśli na nich ekspert miałby poprzestać).

W kolejnych artykułach autorki cyklu przeanalizują, jak postawienie na opisany kierunek rozwoju przez organizacje o największej sile oddziaływania przekłada się na inicjatywy urzędów patentowych w skali krajowej oraz czy i na ile wytyczne w zakresie zasad etycznych regulujących sztuczną inteligencję, które przyjmują perspektywę antropocentryczną, staną się rzeczywistą i skuteczną barierą ograniczającą rozwój tych systemów w omawianych zastosowaniach.

Bibliografia

Źródła

Artificial Intelligence in the European Commission (AI@EC) A strategic vision to foster the development and use of lawful, safe and trustworthy Artificial Intelligence systems in the European Commission, dokument zamieszczony na oficjalnej stronie Komisji Europejskiej w dniu 24.01.2024, <https://commission.europa.eu/system/files/2024-01/EN%20Artificial%20Intelligence%20in%20the%20European%20Commission.PDF> [dostęp 1.05.2024].

Executive Order 13960, "Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government", <https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government> [dostęp 8.05.2024].

National Artificial Intelligence Initiative Act of 2020, (H.R.6216).

Projekt Rozporządzenia Parlamentu Europejskiego w sprawie ustanowienia zharmonizowanych przepisów dotyczących sztucznej inteligencji oraz zmiany rozporządzeń (WE) nr 300/2008, (UE) nr 167/2013, (UE) nr 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 oraz dyrektywy 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (akt w sprawie sztucznej inteligencji), wersja aktualna na dzień 19.04.2024 r., https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_PL.pdf [dostęp 7.05.2024].

Literatura

- Bartoszek M., *Zastosowanie sztucznej inteligencji w sądownictwie w świetle zasady skutecznej ochrony sądowej*, „Folia Iuridica Universitatis Wratislaviensis” 2022, vol. 11.
- Górski M. *Dziesiąty element. Sztuczna inteligencja jako sędzia a prawo do sądu*, „Prawo Mediów Elektronicznych” 2022, nr 3.
- Popper J., Hermann J., Bergweiler S., *Artificial intelligence across industries*, IEC White Paper, XI 2018.

Internet

- Annual patent applications related to AI, by status, World, 2020, <https://ourworldindata.org/grapher/ai-related-patents-applications-and-patents-granted?time=latest> [dostęp 8.05.2024].
- Bradley S., *Blockchain and AI: the EUIPO is top for innovation*, 11.02.2022, <https://www.lexology.com/library/detail.aspx?g=6445f08c-5819-4544-96db-335538f15498> [dostęp 3.05.2024].
- DesignView Image Search, <https://www.tmdn.org/publicwebsite/#/news/2580624> [dostęp 7.05.2024].
- Doligalski T., *Sztuczna inteligencja: od wąskiej specjalizacji do twórczej wszechstronności*, „Gazeta SGH”, 2.02.2024, <https://gazeta.sgh.waw.pl/meritum/sztuczna-inteligencja-od-waskiej-specjalizacji-do-tworczej-wszechstronosci> [dostęp 1.06.2024].
- Du Hao (Henry), *Unleashing the Power of AI to Fight Bad Faith Trademark Registrations*, 6.08.2022, <https://ipwatchdog.com/2022/08/06/unleashing-power-ai-fight-bad-faith-trademark-registrations/id=150649/> [dostęp 4.05.2024]
- EPO, *A vision of sustainability: European Patent Office publishes Strategic Plan 2023 r.*, 27.06.2019 i dostępny tam w formie pdf dokument *EPO – Strategic Plan 2023*, <https://www.epo.org/en/news-events/news/vision-sustainability-european-patent-office-publishes-strategic-plan-2023> [dostęp 2.05.2024].
- EPO, *EPO continues IT modernisation with launch of new online filing tool*, 1.10.2014, <https://www.epo.org/en/news-events/press-centre/press-release/2014/451759> [dostęp 2.05.2024].
- EPO Strategic Plan 2023, https://link.epo.org/web/EPO_Strategic_Plan_2023_en.pdf [dostęp 4.05.2024].
- EPO, Quality Report 2022, Annex to the Annual Review, <https://link.epo.org/web/general/annual-review-2022/en-quality-report-2022.pdf> [dostęp 1.05.2024].
- EUIPN, Oficjalna strona European Union Intellectual Property Network, <https://www.tmdn.org/> [dostęp 6.05.2024].
- EUIPO, *AI at EUIPO: In-house image search in TMview extended to all TM5 offices*, informacja na oficjalnej stronie EUIPO, <https://www.euipo.europa.eu/pl/news/ai-at-euipo-in-house-image-search-in-tmview-extended-to-all-tm5-offices> [dostęp 1.05.2024].
- EUIPO, *EUIPO & New technologies Connecting IP Offices and Users, Developing Trade Mark and Industrial Design Tools for Users: Leveraging on New Technologies*, https://ipkey.eu/sites/default/files/ipkey-docs/2023/IPKeySEA_nov2023_Carlos-Luna_Developing-Trade-Mark-and-Industrial-Design-Tools-for-Users-Leveraging-on-New-Technologies.pdf [dostęp 1.05.2024].
- EUIPO, *Improved machine translation in eSearch Case*, https://euipo.europa.eu/ohimportal/en/key-user-newsflash/-/asset_publisher/dIGJZDH66W8B/content/improved-machine-translation-in-esearch-case-law/ [dostęp 30.04.2024].

- EUIPO, *Open consultation on the Strategic Plan 2030*, <https://www.euipo.europa.eu/en/the-office/governance/strategic-plan/sp2030> [dostęp 3.05.2024].
- EUIPO, *New AI-based comparison of goods and services*, 29.03.2022, <https://euipo.europa.eu/ohim-portal/en/-/news/new-ai-based-comparison-of-goods-and-services> [dostęp 30.04.2024].
- EUIPO, *New EUIPO AI tools, empowering customer services*, https://euipo.europa.eu/knowledge/pluginfile.php/175303/mod_label/intro/AI%20Webinar%20Feb%202022%20%28SL%20review%29.pdf [dostęp 1.05.2024].
- Gangjee D.S., *A quotidian revolution: artificial intelligence, and trade mark law*, [w:] R. Abbott (red.), *Research Handbook on Intellectual Property and Artificial Intelligence*, Edward Elgar, 2022, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4081317 [dostęp 3.05.2024].
- EUIPO, *Strategic Plan 2025 European Union Intellectual Property Office*, https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/contentPdfs/about_euipo/strategic_plan/SP2025_en.pdf [dostęp 30.04.2024].
- Gawliczek T., *EUIPO otwiera szeroko drzwi dla alternatywnych metod rozwiązywania sporów*, publikacja z 27.03.2024, <https://www.prawo.pl/biznes/alternatywne-metody-rozwiazywania-sporow-w-euipo,526120.html> [dostęp 29.04.2024].
- Guidelines For Staff On The Use Of Online Available Generative Artificial Intelligence Tools, https://www.asktheeu.org/en/request/13063/response/45877/attach/3/guidelines%20on%20the%20use%20of%20online%20generative%20artificial%20intelligence%20tools.pdf?cookie_passthrough=1 [dostęp 2.05.2024].
- Hockley L., *USPTO set to award \$70m contract to improve AI-driven patent search*, 20.02.2024, <https://www.worldipreview.com/artificial-intelligence/uspto-set-to-award-dollar70m-contract-to-improve-ai-driven-patent-search> [dostęp 2.05.2024].
- Holdsworth J., *What is NLP (natural language processing)?*, IBM, 6.06.2024, <https://www.ibm.com/topics/natural-language-processing> [dostęp 7.05.2024].
- IBM, *What is machine learning (ML)?*, <https://www.ibm.com/topics/machine-learning> [dostęp 1.06.2024].
- Katyal S., Kesari A., *Trademark Search, Artificial Intelligence and the Role of the Private Sector*, „Berkeley Technology Law Journal”, artykuł z 4.01.2021, opublikowany 6.02.2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3760112 [dostęp 3.05.2024].
- Klenner Bajaja A., *Introducing AI-PreSearch: A Revolutionary AI-Driven Search Tool to support our Patent Examiners*, <https://www.linkedin.com/pulse/introducing-ai-presearch-revolutionary-ai-driven-tool-klenner-bajaja/> [dostęp 6.05.2024].
- Lince T., *Innovation at the EUIPO: spotlight on digital tools and services*, WTR, 20.07.2023, <https://www.worldtrademarkreview.com/data/ip-office-tools-and-services/ip-office-tools-and-services/article/innovation-the-euipo-spotlight-digital-tools-and-services> [dostęp 1.05.2024].
- McCoy R., *The State of AI Use by the USPTO: Streamlining Patent Examination*, 5.07.2023, <https://www.mccrus.com/2023/07/05/state-of-ai/> [dostęp 4.05.2024].
- Micorosoft, <https://learn.microsoft.com/en-us/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2> [dostęp 1.06.2024].
- Nowotko P.M., *AI in judicial application of law and the right to a court*, „Procedia Computer Science” 2021, vol. 192, <https://www.sciencedirect.com/science/article/pii/S1877050921017324>, [dostęp 15.07.2024 r.].

- Oficjalna strona Europejskiego Urzędu Patentowego (EPO), <https://www.epo.org/en/about-us/services-and-activities/quality/enabling-quality/tools> [dostęp 2.05.2024].
- Quality Report 2023, Annex to the Annual Review, <https://link.epo.org/web/general/annual-review-2022/en-quality-report-2022.pdf> [dostęp 4.05.2024].
- Scannell T., *USPTO takes human-first approach to AI innovation*, 9.09.2022, <https://www.cio.com/article/406584/uspto-takes-human-first-approach-to-ai-innovation.html> [dostęp 1.05.2024].
- Tarcu R., *AI Initiatives at the EPO*, 29.01.2021, <https://www.oblon.com/ai-initiatives-at-the-epo> [dostęp 2.05.2024].
- USPTO, *Artificial intelligence tools at the USPTO*, 18.03.2021 r., <https://www.uspto.gov/blog/director/entry/artificial-intelligence-tools-at-the> [dostęp 30.04.2024].
- USPTO, *New artificial intelligence functionality in PE2E Search*, <https://www.uspto.gov/sites/default/files/documents/ai-sim-search.pdf> [dostęp 4.05.2024].
- USPTO, *New PE2E Search Tool Using AI Search Features*, <https://www.uspto.gov/sites/default/files/documents/og-aiforpe2e-20211220.pdf> [dostęp 4.05.2024].
- USPTO, *USPTO launches new Virtual Assistant*, 15.12.2022, <https://www.uspto.gov/about-us/news-updates/uspto-launches-new-virtual-assistant-0> [dostęp 4.05.2024].
- USPTO, *USPTO Leader: AI is a 'Must-Have' Tool*, 15.11.2019, <https://govciomedia.com/uspto-leader-ai-is-a-must-have-tool/> [dostęp 1.05.2024].
- USPTO, *USPTO Virtual Assistant now available for Patents customers*, 10.10.2023, <https://www.uspto.gov/about-us/news-updates/uspto-virtual-assistant-now-available-patents-customers> [dostęp 4.05.2024].
- White L., Evans M., *The AI Act – A step closer to the first law on Artificial Intelligence*, The Norton Rose Fulbright Data Protection Report, 18.05.2023, <https://www.dataprotectionreport.com/2023/05/the-ai-act-a-step-closer-to-the-first-law-on-artificial-intelligence/> [dostęp 8.05.2024].
- WIPO, *AI applied to Classification Use Cases at the European Patent Office*, https://www.wipo.int/edocs/mdocs/classifications/en/ipc_ce_54/ipc_ce_54_p4.pdf [dostęp 4.05.2024].
- WIPO, *AI Inventions*, <https://tind.wipo.int/record/49113?v=pdf> [dostęp 8.05.2024].
- WIPO, *Artificial Intelligence and Intellectual Property Strategy Clearing House*, https://www.wipo.int/about-ip/en/artificial_intelligence/strategy-search.jsp?territory [dostęp 2.05.2024].
- WIPO, *AI tools for common use*, 7.12.2023, https://www.wipo.int/edocs/mdocs/cws/en/cws_11/cws_11_9b_ai_ib.pdf [dostęp 5.05.2024].
- WIPO, *Artificial Intelligence at the EPO*, 23-25 maj, 2018, https://www.wipo.int/edocs/mdocs/globalinfra/en/wipo_ip_itai_ge_18/wipo_ip_itai_ge_18_p8.pdf [dostęp 2.05.2024].
- WIPO, *Emerging Technologies In USPTO Business Solutions*, 29.05.2018, https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=407118 [dostęp 30.04.2024].
- WIPO, *Global Goods & Services Terms Explorer*, <https://www.wipo.int/web/ai-tools-services/global-goods-services-terms-explorer> [dostęp 6.05.2024];
- WIPO, *Index of AI initiatives in IP offices*, https://www.wipo.int/about-ip/en/artificial_intelligence [dostęp 30.04.2024].
- WIPO, *Just released: Improved Conceptual Image Similarity algorithm in the Global Brand Database*, 8.01.2021 r., https://www.wipo.int/reference/en/branddb/news/2021/news_0001.html [dostęp 5.05.2024].

- WIPO, *Patent Trends – Assistive Technology*, WIPO 2021, https://www.wipo.int/tech_trends/en/ [dostęp 6.05.2024].
- WIPO, Prezentacja pt. *Global Goods & Services Terms Explorer* dostępna pod adresem: https://www.wipo.int/edocs/mdocs/madrid/en/mm_ld_wg_21/mm_ld_wg_21_roundtable_topic_3_ib_2.pdf [dostęp 6.05.2024].
- WIPO, *Revised Issues Paper on Intellectual Property Policy and Artificial Intelligence*; Document Code, WIPO/IP/AI/2/GE/20/1 REV, 29.05.2020 r., https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=499504 [dostęp 9.05.2024].
- WIPO, *Similarity algorithm in the Global Brand Database*, 8.01.2021 r., https://www.wipo.int/reference/en/branddb/news/2021/news_0001.html [dostęp 5.05.2024].
- WIPO, *WIPO Conversation On Intellectual Property (Ip) And Artificial Intelligence (Ai) Second Session Revised Issues Paper On Intellectual Property Policy And Artificial Intelligence*, 21.05.2020, https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_2_ge_20/wipo_ip_ai_2_ge_20_1_rev.pdf [dostęp 2.05.2024].
- WIPO, *WIPO Conversation on intellectual property (IP) and artificial intelligence (AI), Third session, summary of second and third sessions 04.11.2020*, wipo/ip/ai/3/ge/20/inf/5, 08.01.2021, https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_ip_ai_3_ge_20/wipo_ip_ai_3_ge_20_inf_5.pdf [dostęp 3.05.2024].
- WIPO, *WIPO Conversation on Intellectual Property (IP) and Frontier Technologies: Fifth Session*, https://www.wipo.int/meetings/en/details.jsp?meeting_id=68588 [dostęp 3.05.2024].
- WIPO, *WIPO Launches State-of-the-Art Artificial Intelligence-Based Image Search Tool for Brands*, Geneva, PR/2019/831, 1.09.2019, https://www.wipo.int/pressroom/en/articles/2019/article_0005.html [dostęp 4.05.2024].
- WIPO, *World Intellectual Property Indicators*, 2021, https://www.wipo.int/edocs/pubdocs/en/wipo_pub_941_2021.pdf [dostęp 30.04.2024].
- WIPO, *WIPO Translate – Breaking Language Barriers with AI*, <https://www.wipo.int/web/ai-tools-services/wipo-translate> [dostęp 6.05.2024].
- WIPO, *Use of Artificial intelligence (AI) in Auto-Classification*, https://www.wipo.int/edocs/mdocs/classifications/en/ipc_ce_53/ipc_ce_53_p8.pdf [dostęp 2.05.2024].
- WIPO, <https://www.wipo.int/web/ai-tools-services> [dostęp 6.05.2024].
- Woutersen M., *USPTO partners with Clarivate for AI-driven image search tool*, 26.02.2024, <https://www.worldipreview.com/patent/uspto-partners-with-clarivate-for-ai-driven-image-search-tool> [dostęp 4.05.2024].
- Yan Tang Demey, Golzio D., *Search strategies at the European Patent Office*, „World Patent Information” 2020, vol. 63, <https://www.sciencedirect.com/science/article/pii/S0172219020300818?via%3Dihub> [dostęp 6.05.2024].

Artificial Intelligence Systems in Administrative Processes and Patent Offices Worldwide – An Overview of Selected Initiatives and Analysis of Development Trends (Part I)

Abstract

Advanced technologies, such as artificial intelligence, are gaining increasing importance in the operations of offices worldwide, including intellectual property offices. This article discusses the development and various uses of artificial intelligence by offices such as EUIPO, USPTO and WIPO, including specific tools employed by these offices and made available to users, automation of processes related to industrial property rights registration, analysis of relevant data sets in the context of trademark searches or patent research. It also presents the challenges and benefits from the integration of modern technologies in the context of intellectual property protection.

Keywords

AI, artificial intelligence, intellectual property offices, intellectual property, industrial property, administrative procedure