

Granice eksploatacji utworów w ramach uczenia maszynowego. Naruszenie praw autorskich a dozwolony użytek w zakresie TDM

**Limitations of Exploitation of Works in Machine Learning.
Copyright Infringement vs. Permitted Use in the Scope of TDM**

Abstract

The main subject of analysis undertaken in this paper is the question of whether the current state of the law leads to an appropriate balance between the interests of creators and the economic development of countries. To this end, an in-depth analysis of Articles 3 and 4 of the DSM directive on the training of artificial intelligence systems was carried out. As part of the paper, the question of the requirement for legal access was raised and the meaning of the opt-out clause was discussed. The final section of the paper presents a comparative legal study of the United States' legal system. It answers the question of whether previous Polish regulations relating to the DSM Directive sufficiently protected creators before the amendment.

SŁOWA KLUCZOWE: TDM, sztuczna
inteligencja, dozwolony użytek,
naruszenie praw autorskich

KEYWORDS: TDM, artificial
intelligence, fair use,
copyright infringement

ALEKSANDRA BUKOWIECKA – magister prawa, Uniwersytet Komisji Edukacji
Narodowej w Krakowie, ORCID – 0009-0005-5219-0568,
e-mail: aleksandra.bukowiecka00@gmail.com

ALEKSANDRA FARON – magister prawa, Uniwersytet Komisji Edukacji Narodowej
w Krakowie, ORCID – 0009-0000-8638-9843,
e-mail: aleksandra.faron00@gmail.com

1 | Wprowadzenie

Na gruncie prawa autorskiego ochroną może być objęty wyłącznie sposób wyrażenia. Nie podlegają jej zatem odkrycia, idee, procedury, metody i zasady działania oraz koncepcje matematyczne, w oparciu o które może przebiegać proces szkolenia systemów generatywnej sztucznej inteligencji (*artificial intelligence* – dalej: „AI”), jak i które mogą stanowić niezbędny element do wytworzenia przez te systemy nowych treści. Niemniej jednak wskazany proces musi zostać poprzedzony wprowadzeniem do danego modelu AI danych, które mogą być przedmiotem ochrony prawa autorskiego. Niezależnie od faktu, że finalnie dany system nie powinien powielać w swoich końcowych wynikach utworów, z których zaczerpnął idee, procedury czy metody, to działanie polegające na kopiowaniu materiału podczas trenowania AI może zostać uznane za naruszenie praw autorskich^[1].

Ostatnia nowelizacja dokonana przez polskiego prawodawcę na gruncie dyrektywy DSM^[2] skłania do wielu pytań związanych z odpowiednim wyważeniem praw twórcy oraz wzrostu znaczenia Unii Europejskiej w kontekście dynamicznego rozwoju sztucznej inteligencji. Stąd też próba odpowiedzi na powyższy problem na łamach tejże pracy w oparciu o wybrane zagadnienia przedmiotowe.

¹ „Aby rozpocząć proces uczenia algorytmów, twórcy sztucznej inteligencji zmuszeni są do wykonywania czynności, które stanowią ingerencję w monopol prawny posiadaczy praw, co wiąże się z pobieraniem, nagrywaniem i cyfrowym kopiowaniem materiałów chronionych prawem autorskim, nawet jeśli dzieje się to tylko w pamięci operacyjnej systemu i ma charakter tymczasowy”. Grzegorz Tylec, Sebastian Kwiecień, Łukasz Sarowski, Adam Sadowski, „Is It Possible to License Works Used in the Learning Process of Artificial Intelligence Algorithms?”, 5. Szczególny problemem przy ustaleniu naruszenia pojawia się w przypadku korzystania z dużych zbiorów danych pozyskanych z *shadow libraries* (zwanymi też „pirackimi bibliotekami”) takimi jak *Sci-Hub*, *Z-Library* czy *Library Genesis* (*LibGen*). Terminem tym określa się internetowe bazy danych z trudno dostępnymi materiałami, np. ze względu na systemy płatnego dostępu (*paywall*) lub inne zabezpieczenia praw autorskich. Fauzan Eka Kusuma, Rahmi Rahmi, „The Use of Shadow Libraries at Universitas Indonesia” *First Monday*, nr 10 (2023): 1.

² 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, s. 92) – dalej: „dyrektywa DSM”.

2 | Dane treningowe AI – problem „dostępu zgodnego z prawem” i zakresu opt-out

2.1. Trenowanie generatywnej sztucznej inteligencji

Problem naruszenia praw autorskich w związku z wykorzystaniem systemów sztucznej inteligencji należy rozpocząć od przedstawienia długo wyczekiwanej regulacji Aktu w sprawie o sztucznej inteligencji^[3], który wszedł w życie 1 sierpnia 2024 r. Rozporządzenie wprowadza definicję systemu AI^[4], lecz pomija kwestię scharakteryzowania samego uczenia maszynowego^[5]. Wobec tego powstaje problem, jaki zakres działań podlega pod trenowanie AI oraz jaki jest jego związek z eksploracją tekstów i danych wyrażoną w dyrektywie DSM.

Proces trenowania AI, w szczególności modeli generatywnej sztucznej inteligencji^[6], przebiega na głębokim uczeniu, opartym na sztucznych sie-

³ 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) (Tekst mający znaczenie dla EOG), (Dz. Urz. UE L 2024/1689 z 12.07.2024 r.) – dalej: „rozporządzenie” lub „akt o AI”. Rozporządzenie UE jest pierwszym na świecie kompleksowym zbiorem przepisów, którego celem jest objęcie w ramy prawne postępującego rozwoju technologii – systemów i modeli sztucznej inteligencji.

⁴ Przez system AI należy rozumieć „system maszynowy, który został zaprojektowany do działania z różnym poziomem autonomii po jego wdrożeniu oraz który może wykazywać zdolność adaptacji po jego wdrożeniu, a także który – na potrzeby wyraźnych lub dorozumianych celów – wnioskuje, jak generować na podstawie otrzymanych danych wejściowych wyniki, takie jak predykcje, treści, zalecenia lub decyzje, które mogą wpływać na środowisko fizyczne lub wirtualne” (art. 3 pkt 1 rozporządzenia). Akt o AI w art. 3 pkt 29 definiuje również dane treningowe jako „dane wykorzystywane do trenowania systemu AI poprzez dopasowanie jego parametrów podlegających uczeniu”.

⁵ W literaturze podaje się, że „uczenie maszynowe polega na trenowaniu algorytmów w taki sposób, aby samodzielnie (lub przy wsparciu człowieka) odkrywały one zależności, wzorce i korelacje w zestawach danych, które są im przedstawiane. W rezultacie takie algorytmy i modele uczenia maszynowego mogą w sposób bardziej efektywny – niż człowiek – generować określone rezultaty”. Michał Nowakowski, *Sztuczna inteligencja. Praktyczny przewodnik dla sektora innowacji finansowych* (Warszawa: Wolters Kluwer, 2023), 29-31.

⁶ Takich jak np. generatory obrazów DALL-E, Stable Diffusion, Midjourney i ChatGPT.

ciach neuronowych^[7]. Można w nim wyróżnić dwie istotne fazy: szkolenia (tzw. faza „wejściowa” – *input*) i generowania treści (tzw. faza „wyjściowa” – *output*). W przypadku pierwszej z nich kluczową rolę odgrywa TDM, gdyż stanowi podstawowe narzędzie szkolenia systemów generatywnej sztucznej inteligencji^[8].

Duże generatywne modele nie tylko stanowią element rozwoju technologicznego, przyczyniając się do masowego i łatwego tworzenia nowych treści^[9], ale jednocześnie stwarzają ryzyko nie tylko dla przyszłości sztuki *per se*, ale przede wszystkim dla praw twórców, gdy trenowanie tych modeli będzie naruszać prawa wyłączne, chyba że zastosowanie znajdą odpowiednie wyjątki i ograniczenia. Jednym z nich jest dozwolony użytek zawarty w dyrektywie DSM, na mocy którego korzystanie z utworów na potrzeby TDM (a zatem również uczenia maszynowego^[10]) jest dozwolone, jeżeli zostaną spełnione odpowiednie warunki.

⁷ Zdaniem Michała Nowakowskiego głębokie uczenie można uznać za podzbiór uczenia maszynowego, lecz nie wyklucza odrębnego traktowania tej techniki. Podaje się, że głębokie uczenie „wykorzystuje wiele warstw przetwarzania, aby odkryć wzorce i strukturę w bardzo dużych zbiorach danych. Każda warstwa uczy się koncepcji z danych, na których opierają się kolejne warstwy; im wyższy poziom, tym bardziej abstrakcyjne koncepcje, które są uczone; uczenie głębokie nie zależy od wcześniejszego przetwarzania danych i automatycznie wydobywa określone cechy”. Nowakowski, *Sztuczna inteligencja. Praktyczny przewodnik dla sektora innowacji finansowych*, 34.

⁸ Systemy te „umożliwiają elastyczne generowanie treści, np. w postaci tekstu, dźwięku, obrazów lub materiałów wideo, i mogą z łatwością wykonywać szeroki zakres różnych zadań” (motyw 99 rozporządzenia). Maryna Manteghi, „Can text and data mining exceptions and synthetic data training mitigate copyright-related concerns in generative AI?” *Law, Innovation and Technology*, nr 2 (2024): 2.

⁹ W przypadku generatywnego modelu sztucznej inteligencji jakim jest ChatGPT wskazuje się, że „algorytm głębokiego uczenia się może rozpoznawać, podsumowywać, tłumaczyć, przewidywać i generować tekst oraz inne treści na podstawie wiedzy uzyskanej z ogromnych zbiorów danych”, a co więcej „może on tworzyć nową zawartość, a nie tylko analizować i działać na dostarczonych danych”. Ryszard Markiewicz, „ChatGPT a prawo Unii Europejskiej” *Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace z Prawa Własności Intelektualnej*, nr 2 (2023): 144.

¹⁰ Akt o AI zarówno w treści przepisu art. 53 ust. 1 lit. c (jak i w motywie 104 i 106) „wyraźnie łączy korzystanie z utworów chronionych prawem autorskim w celu szkolenia modeli sztucznej inteligencji z wyjątkiem dotyczącym eksploracji tekstu i danych (TDM) zawartym w art. 4 dyrektywy CDSM”. Paul Keller, „A First Look at the Copyright Relevant Parts in the Final AI Act Compromise” *Kluwer Copyright Blog*, 11 grudzień, 2023. <https://copyrightblog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/>. Zdaniem jednak części autorów „regulacja prawna dotycząca TDM

2.2. Wyjątek TDM – użytek badawczy i użytek komercyjny

Dyrektywa DSM w art. 3 (użytek badawczy)^[11] oraz art. 4 (użytek komercyjny)^[12] wprowadza obligatoryjny dozwolony użytek w zakresie TDM. Terminem tym określono zautomatyzowaną technikę analityczną służącą do analizowania tekstów i danych w postaci cyfrowej (w tym dźwięki i obrazy) w celu wygenerowania informacji, obejmujących między innymi wzorce, tendencje i korelacje^[13]. Na potrzeby TDM mogą być wykorzystywane w całości lub części m.in. utwory, artystyczne wykonania, fonogramy,

nie rozwiązuje problemu legalności uczenia maszynowego opartego na utworach chronionych prawem autorskim”. Wskazują, że algorytmy uczenia się AI wykraczają poza TDM, w związku z tym uczenie maszynowe powinno być traktowane jako nowe nieuregulowane pole eksploatacji. W swojej argumentacji podkreślają, że „jednym z celów sztucznej inteligencji jest naśladowanie ludzkiego intelektu, co prowadzi do tworzenia nowych utworów (takich jak teksty, obrazy, dźwięki). Tak więc informacje uzyskane w procesie TDM są dalej przetwarzane przez AI w celu utworzenia nowych dóbr intelektualnych”. Dlatego też „jeśli zatem działalność AI sprowadza się jedynie do tworzenia nowych informacji, reguł, wzorców czy korelacji, to może ona ograniczać się do pola eksploatacji, jakim jest TDM. Jeśli jednak wygenerowane dane przybierają określony sposób ekspresji (np. grafika komputerowa, utwór muzyczny lub tekst literacki), taka działalność AI powinna wykraczać poza pole eksploatacji, czyli poza TDM”. W konsekwencji nawet skuteczne dokonanie przez twórców zastrzeżenia z art. 4 dyrektywy nie wyłączałyby obowiązku uzyskania ich zgody na eksploatację utworów w celu trenowania sztucznej inteligencji, jak również prawa żądania z tego tytułu odpowiedniego wynagrodzenia. Tylec, Kwiecień, Sarowski, Sadowski, „Is It Possible to License Works Used in the Learning Process of Artificial Intelligence Algorithms?”, 7-10.

¹¹ Art. 3 ust. 1 dyrektywy: Państwa członkowskie wprowadzają przepisy przewidujące wyjątek od praw przewidzianych w art. 5 lit. a) i art. 7 ust. 1 dyrektywy 96/9/WE, w art. 2 dyrektywy 2001/29/WE oraz w art. 15 ust. 1 niniejszej dyrektywy w odniesieniu 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.

¹² Art. 4 ust. 1 dyrektywy: Państwa członkowskie wprowadzają przepisy przewidujące wyjątek lub ograniczenie dotyczące praw przewidzianych w art. 5 lit. a) i art. 7 ust. 1 dyrektywy 96/9/WE, art. 2 dyrektywy 2001/29/WE, art. 4 ust. 1 lit. a) i b) dyrektywy 2009/24/WE oraz art. 15 ust. 1 niniejszej dyrektywy w odniesieniu do zwielokrotnień i pobrań dostępnych zgodnie z prawem utworów i innych przedmiotów objętych ochroną do celów eksploracji tekstów i danych.

¹³ Art. 2 pkt 2 dyrektywy oraz motyw 8 dyrektywy.

wideogramy oraz bazy danych chronione prawem *sui generis*^[14] z wyłączeniem programów komputerowych w przypadku użytku badawczego^[15].

Zakresem zastosowania art. 3 objęto wyłącznie instytucje dziedzictwa kulturowego^[16] oraz organizacje badawcze^[17], które prowadzą badania

¹⁴ Ponadto wskazuje się, że wobec tak sformułowanej definicji TDM nie jest wykluczone uwzględnienie tego wyjątku również w przypadku zdigitalizowanych utworów, które były pierwotnie utrwalone w postaci analogowej. Ryszard Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790* (Warszawa: Wolters Kluwer, 2021), 51-52. Wyjątki z art. 3 i 4 dyrektywy „dotyczą prawa czasowego lub trwałego zwielokrotniania w całości lub w części autorskich baz danych, prawa producenta baz danych chronionych prawem *sui generis* co do wtórnego wykorzystania bazy w całości lub istotnej części co do ilości lub jakości (art. 5 lit. a i art. 7 ust. 1 dyrektywy 96/9/WE), autorskiego wyłącznego prawa zwielokrotniania (art. 2 dyrektywy InfoSoc) oraz prawa wydawców prasowych w przedmiocie publikacji online (art. 15 dyrektywy DSM)”. Elżbieta Traple, „Granice eksploracji tekstów i danych na potrzeby maszynowego uczenia się przez systemy sztucznej inteligencji”, [w:] *Prawo sztucznej inteligencji i nowych technologii*, red. Bogdan Fischer, Adam Pązik, Marek Świerczyński (Warszawa: Wolters Kluwer, 2021), 24-25.

¹⁵ Użytek badawczy nie obejmuje programów komputerowych z uwagi na brak odwołania w treści przepisu (w przeciwieństwie do użytku komercyjnego) do dyrektywy Parlamentu Europejskiego i Rady 2009/24/WE z dnia 23 kwietnia 2009 r. w sprawie ochrony prawnej programów komputerowych (Dz. Urz. UE L 111, s. 16). Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 51-52; Aleksandra Bagińska-Masiota, „Dozwolony użytek w zakresie eksploracji tekstów i danych w świetle Dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/790” *Annales Universitatis Paedagogicae Cracoviensis / Studia De Cultura*, nr 1 (2022): 121.

¹⁶ Przez instytucje dziedzictwa kulturowego rozumie się ogólnodostępną bibliotekę lub muzeum, archiwum lub instytucję dziedzictwa filmowego lub instytucję dziedzictwa dźwiękowego (art. 2 pkt 3 dyrektywy).

¹⁷ Organizacje badawcze w rozumieniu art. 2 pkt 1 dyrektywy oznaczają uczelnię, łącznie z należącymi do niej bibliotekami, instytut badawczy lub inny podmiot, których głównym celem jest prowadzenie badań naukowych lub działalności edukacyjnej obejmującej także prowadzenie badań naukowych w sposób nienastawiony na zysk lub poprzez ponowne inwestowanie całości zysków w swoje badania naukowe lub zgodnie z uznaną przez dane państwo członkowskie misją realizowania interesu publicznego; w taki sposób, że z dostępu do wyników takich badań naukowych nie może korzystać na preferencyjnych warunkach przedsiębiorstwo mające decydujący wpływ na taką organizację (zatem ze względu na uwarunkowania strukturalne wynikające przykładowo z posiadanych akcji lub udziałów w takiej organizacji). Możliwości powołania się na art. 3 dyrektywy odebrano natomiast indywidualnym badaczom, doktorantom i studentom, którzy chcą prowadzić badania niezależnie, jak również szkołom. Markiewicz, *Prawo*

w sposób nienastawiony na zysk (lub inwestują go ponownie w badania^[18]). Dodatkowo dyrektywa w art. 7 wprowadza zakaz umownego wyłączania eksploracji tekstów i danych w przypadku użytku badawczego, wobec tego każde postanowienie umowne sprzeczne z art. 3 będzie bezskuteczne^[19]. W przypadku art. 4 nie zdecydowano się na jego ograniczenie podmiotowe i może być wykorzystany do dowolnego celu^[20] (w tym zarobkowego^[21]). Wskazany użytek wykazuje jednak mniejsze znaczenie praktyczne z uwagi na wprowadzenie możliwości zastrzeżenia przez twórców w odpowiedni sposób praw do zwielokrotniania i pobierania do celów TDM (tzw. *opt-out*).

Specyfika dozwolonego użytku ustanowionego w dyrektywie sprowadza się do tego, że zasadniczo TDM nie wkracza w zakres praw autorskich, gdyż jego rezultaty obejmują wyłącznie odnalezione zależności i tendencje^[22]. Stąd też w literaturze wprowadzenie tego wyjątku spotkało się z krytyką, uznając go za zbyt restrykcyjny mechanizm, który w założeniu sprzyjający innowacji paradoksalnie doprowadzi do jej zahamowania na terenie Unii Europejskiej^[23]. Niemniej jednak podczas procesu eksploracji tekstów

autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790, 49.

¹⁸ Wskazuje się, że organizacje badawcze mogą czerpać zysk odpłatnie udostępniając prowadzone badania TDM podmiotom trzecim pod warunkiem, że zostanie on przeznaczony na dalsze badania. Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 50.

¹⁹ „Wydaje się więc, że o ile utwór lub inny przedmiot praw pokrewnych został publicznie rozpowszechniony, o tyle zastosowanie znajdzie przedmiotowy wyjątek. Tym samym w stosunku do utworów lub innych przedmiotów praw pokrewnych, które zostały publicznie rozpowszechnione, podmioty uprawnione majątkowo nie mogą umownie wykluczyć eksploracji tekstów i danych”. Bagieńska-Masiota, „Dozwolony użytek w zakresie eksploracji tekstów i danych w świetle Dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/790”, 122.

²⁰ Jak wskazuje motyw 18 dyrektywy „oprócz znaczenia dla badań naukowych, techniki eksploracji tekstów i danych są szeroko wykorzystywane także przez podmioty prywatne i publiczne 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”.

²¹ Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 62.

²² Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 43.

²³ Ze szczególną krytyką spotkało się przyznanie podmiotom uprawnionym w art. 4 dyrektywy możliwości dokonania zastrzeżenia, dzięki któremu mogą wyłączyć korzystanie z ich treści na potrzeby TDM. Tym samym wprowadzony

i danych dochodzi do zwielokrotniania utworów, nawet jeśli w końcowej fazie uzyskiwane są niechronione prawem autorskim elementy^[24], dlatego nie sposób pominąć w tym aspekcie zabezpieczenia interesów twórców.

2.3. Legalność dostępu a legalność źródła

Jednym z problemów formującym się zarówno na gruncie użytku badawczego, jak i użytku komercyjnego TDM jest wymóg zgodnego z prawem dostępu do utworów, który nie został zdefiniowany w dyrektywie DSM^[25]. Jego istota sprowadza się do ochrony interesów twórców i uprawnionych podmiotów przed bezprawnym korzystaniem z ich utworów. Problemem natomiast pozostaje zakres tego ograniczenia i jakie skutki przyniesie dla rozwoju AI.

Zdaniem Ryszarda Markiewicza warunek dostępu zgodnego z prawem ma „charakter obiektywny w tym sensie, że dla jego stwierdzenia nie ma znaczenia usprawiedliwiona lub nieusprawiedliwiona świadomość danego podmiotu w tej kwestii”^[26]. Pewne wskazówki dotyczące interpretacji przedmiotowego wymogu można odnaleźć w motywie 14 dyrektywy, który wyróżnia cztery rodzaje zgodnego z prawem dostępu do treści:

dozwolony użytek jest iluzoryczny z uwagi na łatwość w zablokowaniu dostępu do szerokiego grona materiałów. Co więcej, organizacje badawcze i inne przedsiębiorstwa wykorzystujące TDM w celu rozwoju sztucznej inteligencji znajdują się w niekorzystnej pozycji konkurencyjnej w porównaniu z podmiotami działającymi w państwach takich jak Stany Zjednoczone, gdzie klauzula *fair use* zapewnia większą elastyczność w ocenie czy doszło do naruszenia praw chronionych prawem autorskim. Traple, „Granice eksploracji tekstów i danych na potrzeby maszynowego uczenia się przez systemy sztucznej inteligencji”, 26; Gina Maria Ziaja, „The Text and Data Mining Opt-Out in Article 4(3) CDSMD: Adequate Veto Right for Rightholders or a Suffocating Blanket for European Artificial Intelligence Innovations?” *Journal of Intellectual Property Law & Practice*, nr 5 (2024): 455.

²⁴ Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 43.

²⁵ Wskazany warunek pojawił się już wcześniej w treści innych aktów prawnych tj. art. 5 ust. 1 dyrektywy InfoSoc oraz art. 6 i 9 dyrektywy 96/6/WE, lecz jego interpretacja w przypadku dyrektywy 2019/790 budzi wiele wątpliwości.

²⁶ Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 52.

1. na gruncie polityki otwartego dostępu^[27],
2. poprzez ustalenia umowne między podmiotami uprawnionymi i organizacjami badawczymi lub instytucjami dziedzictwa kulturowego, takich jak abonamenty^[28],
3. na podstawie innych zgodnych z prawem środków^[29],
4. swobodnie dostępnych w internecie^[30].

Pojawia się zatem wątpliwość, czy dyrektywa wprowadza wymóg legalności źródła (kopii danego utworu), czy za relewantne należy przyjąć korzystanie z treści otwarcie dostępnych w internecie, czyli niezależnie od zamieszczenia ich tam w poszanowaniu praw twórcy. W drugim przypadku można w istocie mówić o legalnym dostępie do nielegalnego źródła. W treści dyrektywy brak jest wyraźnego odniesienia do konieczności sięgania do „źródła” danego utworu. Prowadzi to do możliwości przyjęcia różnych interpretacji tego wymogu, które w zależności od swobodniejszego

²⁷ Creative Commons lub podobne. Juan Carlos Fernández-Molina, Fernando Esteban de la Rosa, „Copyright and Text and Data Mining: Is the Current Legislation Sufficient and Adequate?” *Portal. Libraries and the Academy* 24, nr 3 (2024): 662.

²⁸ Stosownie do treści motywu 14 dyrektywy „w przypadku abonamentów posiadanych przez organizacje badawcze lub instytucje dziedzictwa kulturowego, związane z nimi osoby objęte tymi abonamentami należy uznawać za posiadające zgodny z prawem dostęp”.

²⁹ Przykładowo nabycie egzemplarza utworu, korzystanie online ze zbiorów bibliotecznych czy też uzyskanie do niego dostępu w ramach użyczenia. Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 53; Fernández-Molina, Esteban de la Rosa, „Copyright and Text and Data Mining: Is the Current Legislation Sufficient and Adequate?”, 662.

³⁰ Jak wskazuje dyrektywa: „zgodny z prawem dostęp powinien również obejmować dostęp do treści, które są swobodnie dostępne w internecie”, które mogą stanowić np. zawartość strony internetowej lub platform społecznościowych. Fernández-Molina, Esteban de la Rosa, „Copyright and Text and Data Mining: Is the Current Legislation Sufficient and Adequate?”, 662. Sformułowanie „zgodny z prawem dostęp powinien również obejmować [...]” zdaje się tworzyć kolejną kategorię legalnego dostępu. Jednocześnie zwraca się uwagę, aby „treści swobodnie dostępnych w internecie” nie utożsamiać z „treściami podanymi do publicznej wiadomości w internecie” (motyw 14 oraz art. 4 ust. 3 dyrektywy). Różne określenia przyjęte przez ustawodawcę unijnego w dyrektywie skłaniają do wniosku, że są to „formy zgodnego z prawem dostępu do różnych rodzajów treści”. Thomas Margoni, „Saving Research: Lawful Access to Unlawful Sources Under Art. 3 CDSM Directive?” *Kluwer Copyright Blog*, 22 grudzień, 2023. <https://copyrightblog.kluweriplaw.com/2023/12/22/saving-research-lawful-access-to-unlawful-sources-under-art-3-cdsm-directive/>.

czy bardziej restrykcyjnego podejścia mogą doprowadzić do rozszerzenia zastosowania omawianych wyjątków lub też stworzenia iluzoryczności ochrony innowacji.

Elżbieta Traple obowiązek legalnego dostępu zrównuje z legalnością źródła, wskazując, że pomimo szeregu trudności z jego identyfikowaniem, stanowi to w istocie „podstawową gwarancję legalności samego procesu TDM”^[31]. Natomiast literalne brzmienie omawianych przepisów wskazuje na samą „legalność dostępu”, będąc zatem przychylniejszym warunkiem, gdyż ogrom kopii danego utworu utrudnia ustalenie, czy utwór został udostępniony w internecie za zgodą uprawnionego. Nie oznacza to tym samym, że prowadzący TDM mają pełną swobodę w dostępie do wszelkich treści obchodząc zabezpieczenia^[32].

Skupiając się na szerokiej interpretacji, należy rozważyć jej konsekwencje i czy powinna być zarezerwowana wyłącznie dla instytucji badawczych. Treść motywu 14 dyrektywy może skłaniać do wniosku, że legalny dostęp dotyczy wyłącznie art. 3, gdyż nie pada w nim żadne odniesienie do użytku komercyjnego.

Zdaniem Thomas Margoni łagodniejszy wymóg powinien obejmować jedynie użytek badawczy, ponieważ „badania naukowe zasługują na specjalne traktowanie ze względu na ich niezwykle znaczenie edukacyjne, naukowe, społeczne, gospodarcze i kulturowe”^[33], co równoważy interesy

³¹ Traple, „Granice eksploracji tekstów i danych na potrzeby maszynowego uczenia się przez systemy sztucznej inteligencji”, 29.

³² Na przykład w postaci paywall, hasła, technologicznego środka ochrony lub – wyłącznie w przypadku użytku komercyjnego – ważnego ograniczenia umownego. Takie działanie bezsprzecznie wychodziłoby poza zakres legalnego dostępu, a więc byłoby bezprawne i nie podlegałoby pod dozwolony użytek TDM. Margoni, „Saving Research: Lawful Access to Unlawful Sources Under Art. 3 CDSM Directive?”. Brak jest jednak odpowiedzi na pytanie jakimi środkami można ograniczać TDM, a które stanowią już naruszenie obligatoryjnego dozwolonego użytku. Tym samym nie jest jasne „w jakim stopniu obejście środków technicznych, naruszenie warunków umowy licencyjnej lub innej umowy lub unikanie środków geoblokowania wpływa na legalny dostęp na mocy dyrektywy DSM”. Juha Vesala, „Developing Artificial Intelligence-Based Content Creation: Are EU Copyright and Antitrust Law Fit for Purpose?” *IIC – International Review of Intellectual Property and Competition Law*, nr 3 (2023): 358.

³³ Margoni stoi na stanowisku, że „ograniczenia dotyczące beneficjentów, celu i wykorzystania wyraźnie określone w art. 3 odzwierciedlają równowagę interesów między misją publiczną tych organizacji a ochroną interesów posiadaczy praw”. Margoni, „Saving Research: Lawful Access to Unlawful Sources Under Art. 3 CDSM Directive?”.

twórców z rozwojem technologii nienastawionym na zysk. Wobec tego autor uznaje za uzasadnione, aby TDM obejmował również nielegalne treści, lecz wyłącznie w przypadku, gdy sam dostęp do nich jest zgodny z prawem (np. zabezpieczenia nie są obchodzone w bezprawny sposób). Odrzucenie koncepcji legalnego źródła na rzecz legalnego dostępu w przypadku art. 3 ma kilka istotnych zalet, gdyż uprawnia do korzystania z ogromu treści bez konieczności ponoszenia kosztów przez organizacje badawcze oraz eliminuje niepewność prawną związaną z wprowadzonym wyjątkiem, tym samym nadając mu znaczenie^[34]. Taka faworyzacja badań naukowych nie znajduje jednak wyraźnego odzwierciedlenia w części normatywnej dyrektywy. Skoro obie normy stanowią o legalności dostępu należałoby przyjąć dla nich jednolitą interpretację w celu zachowania spójności.

W komentarzu European Copyright Society (dalej: „ECS”) trafnie wskazano, że „nierozsądne byłoby oczekiwanie, że zautomatyzowane narzędzia wydobywcze będą przeprowadzać kontrole legalności wydobywanych treści”^[35]. Dlatego też za mieszczące się w granicach dostępu zgodnego z prawem uznano korzystanie z treści dostępnych swobodnie w internecie bez technicznych ograniczeń, a zatem pochodzących z nielegalnych źródeł^[36].

Restrykcyjne rozumienie wymogu legalnego dostępu prowadzi do pozbawienia praktycznego znaczenia dozwolonego użytku z art. 3 i 4 dyrektywy i zwiększa ryzyko procesów sądowych z tytułu naruszenia praw autorskich, ponieważ „weryfikacja pochodzenia materiałów źródłowych jest niemożliwa w przypadku dużych zbiorów danych”^[37]. W związku z czym postuluje się, aby zamiast tego wymagać „wystarczającej staranności od

³⁴ Ibidem.

³⁵ Jonathan Griffiths, Tatiana Eleni Synodinou, Raquel Xalabarder, „Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market” *European Copyright Society*, 3 maja (2022): 12. https://europeancopyright-society.org/2022/05/03/https-europeancopyrightsocietydotorg-files-wordpress-com-2022-05-ecs_exceptions_final-1-pdf/.

³⁶ Griffiths, Synodinou, Xalabarder, „Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market” 12.

³⁷ „Podejście to jest również problematyczne pod względem koncepcyjnym, ponieważ zakłada, że kopia źródłowa, która ma być wykorzystana do TDM, ma możliwy do określenia status praw autorskich (naruszający w swoim pochodzeniu lub wynikający z bezprawnego obejścia), co często nie jest prawdą”. Vesala, „Developing Artificial Intelligence-Based Content Creation: Are EU Copyright and Antitrust Law Fit for Purpose?”, 359.

osób angażujących się w TDM – biorąc pod uwagę m.in. charakter źródeł wykorzystywanych danych”, a zatem czy korzystano z wiarygodnych platform, czy też pozyskane dane znajdowały się na pirackich stronach internetowych^[38]. Takie podejście wydaje się lepiej wyważać sporne interesy, lecz może być trudne do zastosowania w praktyce, m.in. jak ustalić granicę między wiarygodną a piracką stroną. Niemniej jednak zasadnym zdaje się przyjęcie interpretacji, która nie stanie się barierą dla trenowania systemów AI, jednocześnie też nie liberalizującej nadmiernie warunku legalnego dostępu poprzez nieograniczone wykorzystywanie nielegalnych źródeł.

Na uwagę zasługuje też inna kwestia, a mianowicie dozwolony użytek TDM jest nieodpłatny w tych sytuacjach, gdy nie zawarto stosownej umowy. Z uwagi jednak na treść motywu 17 dyrektywy w przypadku użytku badawczego można wywieść wniosek o zakazie wprowadzania wynagrodzenia dla podmiotów uprawnionych^[39]. Jakkolwiek samo prowadzenie TDM przez organizacje badawcze nie może być związane z zapłatą wynagrodzenia na rzecz twórców wykorzystywanych w tym celu utworów, to pozyskanie tych utworów w legalny sposób może wiązać się z wysokimi kosztami^[40]. Wobec tego art. 3 dyrektywy w istocie chroni jedynie przed warunkowym

³⁸ Ibidem.

³⁹ Zgodnie z motywem 17 dyrektywy: „ze względu na charakter i zakres wyjątku, który jest ograniczony do jednostek prowadzących badania naukowe, wszelkie ewentualne szkody wyrządzone podmiotom uprawnionym w związku ze stosowaniem tego wyjątku powinny być minimalne. W związku z tym państwa członkowskie nie powinny wprowadzać rekompensaty dla podmiotów uprawnionych w odniesieniu do korzystania w ramach wyjątków dotyczących eksploracji tekstów i danych wprowadzonych na podstawie niniejszej dyrektywy”. Wobec tego nie jest wykluczone zawarcie odpłatnych licencji, lecz wyłącznie na rzecz podmiotów prowadzących TDM w celach komercyjnych. Wskazuje się, że „skoro dopuszczalne jest całkowite zakazanie przez podmiot praw wyłącznych korzystania z użytku komercyjnego TDM, to jest również legalne uzależnienie takiego działania od umownie ustalonego wynagrodzenia”. Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 63-64; tak też: Traple, „Granice eksploracji tekstów i danych na potrzeby maszynowego uczenia się przez systemy sztucznej inteligencji”, 26.

⁴⁰ Należy zaznaczyć, że „podmioty praw autorskich nie są zobowiązane do udostępniania tego bogactwa danych. Mogą zdecydować, czy to zrobić i na jakich warunkach. Jeśli jednak to robią, nie mogą ograniczać – lub w kategoriach ekonomicznych – segmentować tej oferty”. Martin Kretschmer, Thomas Margoni, Pinar Oruçs, „Copyright Law, and the Lifecycle of Machine Learning Models” *IIC – International Review of Intellectual Property and Competition Law*, nr 1 (2024): 127.

dostępem (czyli zakazującego TDM lub prowadzeniem go za dodatkową opłatą), nie zaś przed nieuzyskaniem tego dostępu^[41].

Brak dysponowania wystarczającymi środkami w celu pozyskania legalnej kopii każdego z utworów w gruncie rzeczy uniemożliwi prowadzenie TDM^[42]. W konsekwencji sztuczna inteligencja będzie trenowana na znacznie ograniczonej ilości danych, co pozbawi ją przymiotu konkurencyjności na światowym rynku.

W tym przedmiocie należy wspomnieć również o transpozycji art. 3 i 4 dyrektywy do polskiego porządku prawnego. W art. 26² oraz art. 26³ pr. aut.^[43] zrezygnowano z wymienienia *explicite* w tekście ustawy o wymogu „dostępu zgodnego z prawem”. Brak szerszego objaśnienia tej kwestii należy ocenić krytycznie, gdyż brzmienie znowelizowanych przepisów skłania do wniosku, że eksploracja tekstów i danych może przebiegać bez względu na charakter źródła – legalnego czy też nielegalnego.

2.4. Wyłączenie TDM przez uprawnione podmioty

Użytek komercyjny TDM wyróżnia się od użytku badawczego wyposażeniem podmiotów uprawnionych w instrument *opt-out*, który mogą zastosować poprzez wyrażnie zastrzeżenie korzystania z utworów i innych przedmiotów objętych ochroną w odpowiedni sposób, np. za pomocą środków nadających się do odczytu maszynowego w przypadku treści, które zostały podane do publicznej wiadomości w internecie^[44]. Regulacja

⁴¹ Warto nadmienić, że „usługi często umożliwiają dostęp do swoich zbiorów danych za pośrednictwem interfejsów programowania aplikacji”, zaś „z technicznego punktu widzenia dość łatwo jest zaprojektować API, które zezwala tylko na określoną liczbę żądań, określoną długość lub złożoność zapytań, lub też określoną funkcję wyszukiwania i pobierania”. Wobec tego niejasna jest granica, kiedy stosowane warunki dostępu „staną się formą (niedozwolonego) ograniczenia praw ustanowionych w art. 3”. Kretschmer, Margoni, Oruśc, „Copyright Law, and the Lifecycle of Machine Learning Models”, 127-128.

⁴² Fernández-Molina, Esteban de la Rosa, „Copyright and Text and Data Mining: Is the Current Legislation Sufficient and Adequate?”, 657.

⁴³ Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz. U. z 2022 r., poz. 2509) – dalej: „pr. aut.”.

⁴⁴ Art. 4 ust. 3 dyrektywy: Wyjątek lub ograniczenie przewidziane w ust. 1 ma zastosowanie, pod warunkiem że korzystanie z utworów i innych przedmiotów objętych ochroną, o których mowa w tym ustępie, nie zostało wyrażnie zastrzeżone przez podmioty uprawnione w odpowiedni sposób, na przykład za pomocą

ta spotkała się z żywą krytyką, wskazując, że rozwiązania przyjęte w dyrektywie DSM prowadzą do załamania rozwoju sztucznej inteligencji na obszarze Unii Europejskiej^[45].

Wskazać należy, że w razie skutecznego zastrzeżenia dozwolony użytek komercyjny dyrektywy przestaje mieć zastosowanie. Powstaje zatem pytanie jakie działania twórców mogą zostać uznane za wystarczające, a które nie mieszczą się w zakresie *opt-out*.

Wymogiem skuteczności omawianego zastrzeżenia jest dokonanie go wyraźnie i w odpowiedni sposób, np. za pomocą środków nadających się do odczytu maszynowego łącznie z metadanymi i warunkami korzystania ze strony internetowej lub usługi, a zatem środków, które są „rozpoznawalne” przez techniki informatyczne^[46] – w przypadku treści, które zostały podane do publicznej wiadomości w internecie. Natomiast w innych przypadkach odpowiednie może być zastrzeżenie praw w drodze środków, takich jak umowy czy jednostronne oświadczenia^[47].

środków nadających się do odczytu maszynowego w przypadku treści, które zostały podane do publicznej wiadomości w internecie.

⁴⁵ Znikome praktyczne zastosowanie użytku komercyjnego TDM wynika z możliwości wyłączenia tego wyjątku przez uprawnione podmioty, podczas gdy państwa takie jak Stany Zjednoczone, Chiny czy Japonia będą przodować w dziedzinie AI, gdyż TDM jest uznawany w nich za dozwolony użytek niezależnie od osiąganego zysku. Traple, „Granice eksploracji tekstów i danych na potrzeby maszynowego uczenia się przez systemy sztucznej inteligencji”, 26-27; Griffiths, Synodinou, Xalabarder, „Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market”, 17; Mateusz Kupiec, „Amerykański fair use i japoński non-enjoyment – drogowskazy dla europejskich rozwiązań prawnych w zakresie TDM?” *Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace z Prawa Własności Intelektualnej*, nr 2 (2021): 136.

⁴⁶ Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 63. Podobne określenie zostało zawarte również w dyrektywie 2019/1024, przez gdzie w art. 2 pkt 13 „format nadający się do odczytu maszynowego” zdefiniowano jako „format pliku ustrukturyzowany tak, aby aplikacje mogły łatwo zidentyfikować, rozpoznać i pozyskać określone dane, w tym poszczególne stwierdzenia faktów, i ich wewnętrzną strukturę”.

⁴⁷ We wskazanych pozostałych przypadkach „zastrzeżenie powinno być dostatecznie wyraźne i łatwe do dostrzeżenia przez zamierzającego prowadzić TDM”. Markiewicz, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*, 63. Jak podaje dalej motyw 18 dyrektywy: „zastrzeżenie praw do celów eksploracji tekstów i danych nie powinno dotyczyć innych sposobów korzystania”, jednocześnie też „podmioty uprawnione powinny mieć możliwość stosowania środków, które zapewniają poszanowanie poczynionych przez nie w tym względzie zastrzeżeń”.

Opt-out może być skutecznie zrealizowany poprzez dodanie metadanych typu `robots.txt`^[48] do swoich treści online, wobec tego zapewniono „stosunkowo niewielką swobodę podmiotom zajmującym się komercyjnie eksploracją tekstów i danych”^[49]. Z uwagi na niejasną regulację *opt-out* nie jest wiadome jakie konkretne działania nie będą uznane za wyraźne i odpowiednie zastrzeżenie, stąd też należy krytycznie ocenić niedoprecyzowanie tego wyjątku. Brak jest zatem odpowiedzi czy zastrzeżenia niewyrażone w kodzie, np. w plikach PDF lub obrazach^[50] będą uznane za wystarczające.

W sytuacji gdy pozyskanie utworów w celu wykorzystania ich jako danych wejściowych odbywa się bez zgody uprawnionych podmiotów może dojść do naruszenia wyłącznych praw autorskich do reprodukcji^[51]. W jaki jednak sposób twórcy mogą pozyskać informacje o tym, że ich utwory zostały wykorzystane do trenowania sztucznej inteligencji? Na uwagę zasługuje wprowadzony w akcie o AI obowiązek przejrzystości, zgodnie z którym dostawcy modeli sztucznej inteligencji są zobowiązani do podania do publicznej wiadomości wystarczająco szczegółowego streszczenia na temat treści wykorzystanych do trenowania^[52]. Warunek ten jest pewnym

⁴⁸ „Plik `robots.txt` przekazuje robotom wyszukiwarek oraz programom crawlującym wytyczne, które strony i pliki mogą odwiedzać”. Nie jest to jednak rozwiązanie, które całkowicie wykluczy indeksowanie URL, ale pozwala na zablokowanie crawlowania i indeksowania samej zawartości strony internetowej. „Robots.txt – do czego służy i jakich błędów unikać przy jego tworzeniu?” *Whitepress*. <https://www.whitepress.com/pl/baza-wiedzy/642/robots-txt-do-czego-sluzy-i-jakich-bledow-unikac-przy-jego-tworzeniu>. [dostęp: 20.10.2024].

⁴⁹ Bagieńska-Masiota, „Dozwolony użytek w zakresie eksploracji tekstów i danych w świetle Dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/790”, 124; Griffiths, Synodinou, Xalabarder, „Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market”, 17.

⁵⁰ Griffiths, Synodinou, Xalabarder, „Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market”, 18.

⁵¹ Manteghi, „Can text and data mining exceptions and synthetic data training mitigate copyright-related concerns in generative AI?”, 3.

⁵² Art. 53 ust. 1 lit. d rozporządzenia. Obowiązek przejrzystości został doprecyzowany w motywie 107 rozporządzenia, zgodnie z którym „streszczenie to powinno więc na przykład wymieniać główne zbiory danych, które wykorzystano do trenowania modelu, takie jak duże prywatne lub publiczne bazy lub archiwa danych, oraz powinno zawierać opisowe wyjaśnienie innych wykorzystanych źródeł danych. Urząd ds. AI powinien zapewnić wzór streszczenia, który powinien być prosty i skuteczny oraz umożliwiać dostawcy przedstawienie wymaganego streszczenia w formie opisowej”.

krokiem w kierunku przejrzystości danych szkoleniowych, jednak nie sprawdzi się w każdym przypadku z faktu braku konieczności podawania przez dostawców wszystkich użytych danych^[53].

Warto zaznaczyć, że opt-out jest mechanizmem *ex ante*^[54]. Wobec tego podmioty prowadzące TDM nie są zobowiązane do usuwania ze swoich algorytmów treści, na których szkolono modele AI zanim nastąpiło wyraźne zastrzeżenie. W przypadku jednak wytrenowania systemu AI z pominięciem skutecznie dokonanego zastrzeżenia dochodzi do naruszenia. Dlatego też przedsiębiorstwa technologiczne zaczęły rozwijać technologie, które pozwalają „oduczyć” danych informacji z dużych modeli AI, jednak zauważa się, że jest to proces bardziej kosztowny i czasochłonny niż zapłacenie twórcy wynagrodzenia^[55].

W ramach omawianego wyjątku należy również przedstawić spór LAION e.V. vs. Kneschke wytoczony przed Sądem Okręgowym w Hamburgu^[56].

⁵³ Natomiast innym sposobem na pozyskanie stosownych informacji o grafikach użytych do trenowania AI może być skorzystanie z narzędzia „Have I Been Trained?”, które pozwala na przeszukiwanie danych użytych do szkolenia sztucznej inteligencji. Platforma ta daje możliwość opt-out za pomocą rejestru „Do Not Train”, dzięki któremu twórcy mają możliwość oznaczenia swoich utworów w celu wyłączenia ich ze zbioru danych do trenowania przyszłych modeli AI. Rejestr ten jest respektowany przez takie przedsiębiorstwa jak HuggingFace i Stability. „Frequently Asked Questions” *Have I Been Trained?*. <https://haveibeentrained.com>. [dostęp: 20. 10.2024]. Podobne narzędzie opracowuje OpenAI o nazwie Media Manager, które ma umożliwić twórcom określenie, w jaki sposób chcą, aby ich prace były uwzględniane lub wykluczane w zakresie uczenia maszynowego. „OpenAI will Release a Media Manager Tool in 2025, Allowing Creators to Block AI Training”, VentureBeat. <https://venturebeat.com/ai/openai-will-release-a-media-manager-tool-in-2025-allowing-creators-to-block-ai-training/>. [dostęp: 20.10.2024].

⁵⁴ Wskazuje się, że „w przeciwieństwie do niego, obowiązek przejrzystości w ustawie o sztucznej inteligencji odnosi się do treści, które zostały już wykorzystane”, dlatego też „aby ten obowiązek przejrzystości miał zauważalny wpływ na realizację mechanizmu opt-out dla podmiotów praw autorskich, opt-out musiałby być wykonywany *ex post*”. Ziaja, „The Text and Data Mining Opt-Out in Article 4(3) CDSMD: Adequate Veto Right for Rightholders or a Suffocating Blanket for European Artificial Intelligence Innovations?”, 457.

⁵⁵ Ziaja, „The Text and Data Mining Opt-Out in Article 4(3) CDSMD: Adequate Veto Right for Rightholders or a Suffocating Blanket for European Artificial Intelligence Innovations?”, 457.

⁵⁶ Wyrok Sądu Okręgowego w Hamburgu z dnia 27.09.2024 (310 O 227/23). Proces ten jako pierwszy w Europie podjął problem legalności wykorzystywania utworów chronionych prawem autorskim w celu szkolenia generatywnych modeli sztucznej inteligencji. Przyczyną sporu było pobranie przez LAION e.V (niemiecka organizacją non-profit, która tworzy szeroko stosowane szkoleniowe

Rozprawa skoncentrowała się wokół ewentualnego skutecznego zastosowania *opt-out*, ponieważ w podsekcji warunków korzystania z serwisu *Bigstock* dokonano zastrzeżenia w języku angielskim (sformatowanego w HTML) o zakazie używania m.in. zautomatyzowanych programów, botów „w jakimkolwiek celu”, w tym do pobierania treści^[57].

W orzeczeniu z 27 września 2024 r. uznano, że LAION może powołać się na dozwolony użytek badawczy, ponieważ udostępnia swoje zbiory danych za darmo – działanie to jest niekomercyjne i służące badaniom naukowym, wobec tego nie doszło do naruszenia. Niemniej jednak w orzeczeniu zawarto kilka niewiążących przemyśleń na temat *opt-out*, a mianowicie że zastrzeżenie sformułowane w „języku naturalnym” może być uznane za „zrozumiałe dla maszyn”, przy czym ocena ta „zawsze będzie musiała być udzielona w odniesieniu do rozwoju technicznego istniejącego w odpowiednim czasie korzystania z utworu”. Sąd nie odniósł się szerzej w kwestii legalności szkolenia modeli sztucznej inteligencji ani generowanych przez nie wyników. Możliwe, że orzecznictwo powstałe na gruncie przyszłych sporów dostarczy odpowiedzi na wiele nurtujących pytań względem wskazanych wyżej problemów.

3 | Badania prawnoporównawcze na przykładzie prawodawstwa Stanów Zjednoczonych

Technologia TDM i jej właściwe usankcjonowanie w porządkach prawnych państw ma kluczowe znaczenie dla rozwoju i konkurencyjności gospodarek. Pytanie o TDM nie jest więc tylko pytaniem o ochronę praw twórców, ale także o konkurencyjność i możliwość dalszego wzrostu gospodarczego określonych regionów świata. Wyważenie tych dwóch dóbr – praw twórcy i rozwoju państw jest niezwykle trudne, dlatego w ramach dalszej analizy

zbiory danych) zdjęcia niemieckiego fotografa Roberta Kneschke, które było dostępne bez *paywalla* na stronie internetowej *Bigstock*, w celu włączenia go do zbioru danych LAION 5B. Paul Keller, „Machine Readable or not? – Notes on the Hearing in LAION e.v. vs. Kneschke” *Kluwer Copyright Blog*, 22 lipca 2024. <https://copyrightblog.kluweriplaw.com/2024/07/22/machine-readable-or-not-notes-on-the-hearing-in-laion-e-v-vs-kneschke/>.

⁵⁷ Keller, „Machine Readable”.

podjętego przez autorki zagadnienia zbadano inne porządki prawne. W tej części artykułu zostaną przedstawione regulacje amerykańskie.

Choć w Unii Europejskiej wprowadzono osobne regulacje dotyczące naruszenia praw autorskich w związku z wykorzystaniem nowych technologii, w Stanach Zjednoczonych nie zdecydowano się na takie rozwiązanie. Kwestie obejmujące uczenie maszynowe oraz technologię TDM są regulowane na bazie przepisu wprowadzonego do amerykańskiego porządku prawnego już w 1976 r. Zgodnie z § 107 Copyright Act: „zezwała się na korzystanie (*fair use*) z utworów chronionych prawem autorskim [...] dla celów takich jak: krytyka, komentarz, informacja medialna, nauczanie [...], stypendium naukowe, badania”. Ustawa amerykańska określa jednak cztery warunki, w których do takiego dozwolonego użytku (*fair use*) może dojść^[58].

Zacytowany przepis zdaje się nie pozostawiać zbyt wiele przestrzeni w szerokim korzystaniu z dozwolonego użytku. Amerykańskie orzecznictwo doprowadziło jednak do uelastycznienia powyższej regulacji nadając jej w rzeczywistości zupełnie nowy wymiar^[59]. W latach dziewięćdziesiątych ubiegłego wieku sądy amerykańskie sformułowały koncepcję dzieła transformacyjnego (*transformative work*). Koncepcja ta opiera się na założeniu, że autorzy przekształcający dzieło mogą korzystać z *fair use* nawet w celach komercyjnych, w wypadku gdy nowo stworzony przez nich utwór bazujący na innym utworze przekształca go w taki sposób, że tworzy dzieło o nowym znaczeniu lub nowym przekazie^[60]. Sądy Amerykańskie transformacyjność dzieła zestawiały z jego wpływem na rynek (czwarta przesłanka z § 107 Copyright Act). Badano przede wszystkim to, czy oba dzieła – pierwotne oraz to stworzone na podstawie dzieła pierwotnego – trafiają do podobnej lub takiej samej grupy odbiorców, i czy w ten sposób nie naruszono praw majątkowych twórcy dzieła pierwotnego.

Kolejnym krokiem w orzecznictwie amerykańskim była ochrona osób, które wykorzystując dzieło chronione prawem autorskim nie tworzyły

⁵⁸ Zgodnie z regulacją amerykańską do takich przesłanek zalicza się po pierwsze niekomercyjny lub edukacyjny cel i charakter dozwolonego użytku. Po drugie wykorzystywany materiał posiada cechy utworu. Trzecim warunkiem koniecznym do spełnienia jest to, by ilość oraz istotność wykorzystanych elementów utworu była adekwatna. Natomiast czwartym jest zasada by wpływ użycia na potencjalny rynek lub wartość dzieła chronionego prawem autorskim nie naruszały rynku dzieła pierwotnego.

⁵⁹ Jane C. Gisburg, „Fair Use in The United States: Transformed, Deformed, Reformed?” *Singapore Journal of Legal Studies* (2020): 265-294.

⁶⁰ Ibidem, 269-272.

nowego utworu i który byłby niezależny, a zatem funkcjonował na osobnym rynku. Sąd Apelacyjny Stanów Zjednoczonych Szóstego Okręgu orzekł, że z ochrony korzystają użytkownicy, którzy nadali dziełu przez nich wykorzystanemu nowe znaczenie, charakter lub przesłanie^[61] – zasada ta została określona jako *transformative use*. Właśnie ten wyrok stał się podstawą dla ochrony praw osób korzystających z dzieł twórców przy wykorzystaniu nowych technologii.

Jak zaznacza Mateusz Kupiec: bo „wynik wyszukiwanych słów różni się pod względem celu, charakteru, ekspresji, znaczenia i przesłania od strony (oraz książki), z której został zaczerpnięty”^[62]. Tak więc nawet odtworzenie materiału chronionego prawem autorskim ma charakter transformacyjny, gdy takie odtworzenie ma inny niż samo dzieło cel.

Wyrok ten ma szczególne znaczenie w kontekście wciąż toczącego się sporu *Anderson v. Stability AI* dotyczącego zbiorowego pozwu artystów^[63]. Powodowie zarzucili pozwanym naruszenie ich praw autorskich poprzez wykorzystanie ich dzieł w celu trenowania sztucznej inteligencji.

Analizę ewentualnych rozstrzygnięć sądu przeprowadził Matthew Lindberg w swoim artykule^[64]. Lindberg, na łamach toczącej się sprawy, bada § 107 Copyright act oraz cztery wynikające z tegoż przepisu przesłanki, które odpowiadają za zastosowanie w stosunku do dzieła instytucji *fair use* – dozwolonego użytku. Pierwsza z przesłanek dotyczy celu oraz charakteru użytku – odnosi się do tego, czy cel oraz charakter jest niezarobkowy lub służy celom edukacyjnym. To właśnie z tą przesłanką jest powiązana omawiana wcześniej transformacyjność dzieła. Jeśli bowiem utwór posiada nowy cel i charakter, to korzysta z ochrony przewidzianej przez instytucję *fair use*. Jak zaznacza Lindberg, w aż 94% sprawach, w których sąd uznał dzieło za transformacyjne, utwór taki korzystał z ochrony przewidzianej przez „dozwolony użytek”^[65]. Autor wskazuje również, że w analizowanej

⁶¹ Wyrok Sądu Apelacyjnego Stanów Zjednoczonych Szóstego Okręgu z 1994, *Campbell vs. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994).

⁶² Kupiec, „Amerykański *fair use* i japoński *non-enjoyment* – drogowskazy dla europejskich rozwiązań prawnych w zakresie TDM?”, 137-138.

⁶³ W 2023 r. wytoczyły go: Sary Anderson, Kelly McKernan i Karla Ortiz przeciwko *Stability AI*, *Midjourney* i *DeviantArt*. *Stability AI*, *Midjourney* oraz *DeviantArt*, które to podmioty korzystają z programów wykorzystujących sztuczną inteligencję do generowania obrazów.

⁶⁴ Matthew Lindberg, „Applying Current Copyright Law to Artificial Intelligence Image Generators in the Context of *Anderson v. Stability AI, Ltd.*” *Mitchell Hamline School of Law*, nr 15 (2024).

⁶⁵ *Ibidem*, 48.

sprawie prawdopodobnym jest, że sąd uzna charakter przekształceń poczynionych przez AI za transformacyjne, ponieważ cel użytku jest inny niż cel obrazu. Celem obrazów tworzonych przez powodów jest doświadczanie, podziwianie i kontakt ze sztuką, natomiast celem wytworów sztucznej inteligencji jest dalszy rozwój AI. W trakcie procesu uczenia maszynowego dzieło przetwarzane jest w ten sposób, że nie jest ono czytelne dla człowieka, nakłada na nie tzw. „szumy”, by później otrzymać nowy wytwór.

Proces ten nasuwa pytanie o moment naruszenia prawa autorskich do utworu, czy będzie on miał miejsce na etapie wprowadzania dzieła do systemu AI (dane wejściowe), czy też na etapie wygenerowania wytworu na podstawie tych danych (dane wyjściowe). Lindberg stwierdza, że do takiego naruszenia dojdzie raczej w momencie wprowadzania danych – dzieje się to bowiem bez żadnej licencji lub zezwolenia^[66]. Zdaniem Lindberga istnieje jednak wysokie prawdopodobieństwo, że sąd uzna to za użytek transformacyjny związany z wykorzystaniem dzieł, co zaznaczono powyżej.

Zupełnie odmiennie kształtuje się sytuacja w przypadku trzech pozostałych przesłanek. Druga z nich dotyczy charakteru działania chronionego prawem autorskim. Jeśli dzieło, w tym wypadku wytwór, będzie zbyt zbliżony do swego pierwowzoru, wówczas dojdzie do naruszenia. W omawianej sprawie zarówno w kontekście danych wejściowych, jak i wyjściowych przesłanka ta może przemówić na korzyść powodów^[67]. Trzeci czynnik dotyczy ilości i istotności wykorzystanych danych – przesłanka ta odnosi się jedynie do danych wejściowych. Również ten warunek będzie przemawiać na korzyść twórców. Wykorzystane bowiem zostały całe zasoby ich dzieł, a ponadto wykorzystane zostało każde dzieło w całości. Ostatnia przesłanka dotyczy wpływu wytworu na rynek. Lindberg analizuje ją pod względem danych wejściowych. Jak zaznacza, doszło do szkody na rzecz twórców na gruncie danych wejściowych stanowiących utwory powodów, a które zostały pozyskane w celu szkolenia AI bez żadnej opłaty. Mamy więc do czynienia ze wspólnym rynkiem, ponieważ dane wejściowe i sporne obrazy stanowią ten sam przedmiot. Zdaniem Lindberga ta przesłanka także będzie przemawiać na korzyść powodów.

Kluczowym w odpowiedzi na pytanie, kto wygra spór jest to, co sąd uzna za istotniejsze – transformacyjność dzieł wytwarzanych przy wykorzystaniu technologii TDM, czy też trzy pozostałe przesłanki: istotność danych, wspólny rynek oraz charakter wytworu.

⁶⁶ Ibidem, 48-49.

⁶⁷ Ibidem, 51.

4 | Ochrona praw twórców w przypadku naruszeń z wykorzystaniem technologii TDM – nowe pole eksploatacji, ochrona dóbr osobistych?

4.1. Nowe pole eksploatacji

W związku z omawianą tematyką nasuwa się pytanie, czy dotychczasowe regulacje polskie sprzed nowelizacji związanej z dyrektywą DSM wystarczająco chroniły twórców. W ramach rozwoju nowych technologii od kilku lat formułowany jest postulat modyfikacji art. 50 pr. aut. poprzez stworzenie nowego pola eksploatacji. Możliwość jego wprowadzenia jest związana z koniecznością spełnienia określonych kryteriów sformułowanych przez doktrynę.

Elżbieta Traple odrębne pole eksploatacji rozumie poprzez zróżnicowanie sposobu zwielokrotniania i rozpowszechniania utworu^[68]. Dalej wskazuje, że taki sposób zwielokrotniania powinien być odrębny pod względem technicznym oraz ekonomicznym. Natomiast Tomasz Targosz formułuje następujące kryteria świadczące o odrębności pola eksploatacji: ekonomiczność oraz dostrzegalność obrotu^[69]. Spełnienie wyżej wymienionych przesłanek przez technologię TDM sygnalizuje między innymi Karina Kunc-Urbańczyk^[70]. Zwolenniczką modyfikacji art. 50 pr. aut. jest także Iga Bałos, która stwierdza: „wykształcenie się nowego pola eksploatacji, jakim jest TDM, umożliwia korzystanie z utworu w sposób i w celu, który dotąd nie był możliwy. Można to przyrównać do nowego produktu lub usługi, które nie były wcześniej dostępne na rynku”^[71].

Wyodrębnienie nowego pola eksploatacji rodzi jednak pewne problemy, gdyż „istnieje niebezpieczeństwo, że za nowe pole eksploatacji zostanie uznane pole eksploatacji już istniejące albo objęte umową, co może doprowadzić nawet do podwójnego wynagrodzenia twórcy. Dotyczy to zwłaszcza

⁶⁸ Elżbieta Traple, „Komentarz do art. 50”, [w:] *Komentarz do ustawy o prawie autorskim i prawach pokrewnych*, red. Ryszard Markiewicz, Janusz Barta (Warszawa: Wolters Kluwer, 2011), teza 3, Lex.

⁶⁹ Traple, „Komentarz do art. 50”, teza 4.

⁷⁰ Karina Kunc-Urbańczyk, „Uczenie maszynowe jako pole eksploatacji utworu”, [w:] *Prawo sztucznej inteligencji i nowych technologii*, red. Bogdan Fischer, Adam Pązik, Marek Świerczyński (Warszawa: Wolters Kluwer, 2021), 227-237.

⁷¹ Iga Bałos, „Trenowanie wybranych modeli sztucznej inteligencji a uprawnień twórców” *Studia Prawnicze. Rozprawy i Materiały*, nr 2 (2019): 22.

takich sytuacji, gdy pojawia się pewna modyfikacja techniczna znanych już sposobów korzystania i trzeba podjąć decyzję, czy postanowienia wcześniej zawartych umów objęły swoim zakresem i tę modyfikację, czy też konieczne będzie zawarcie nowej umowy o korzystanie”^[72]. Dodatkowe problemy ze sformułowaniem nowego pola eksploatacji rodzi kwestia rozmytych kryteriów dotyczących stworzenia nowego pola – stanowiska doktryny w tym zakresie bywają bowiem rozbieżne.

Mimo powyższego zasadnym zdaje się wyodrębnienie nowego pola eksploatacji na potrzeby uczenia maszynowego systemów sztucznej inteligencji. Jak podkreśla Kunc-Urbańczyk, należy odrzucić koncepcję, zgodnie z którą uczenie maszynowe wpisuje się w inne pole eksploatacji utworu – zwielokrotnienia metodą cyfrową. Jak uzasadnia, uczenie maszynowe „nie skupia się na treści utworu, zamiast tego stara się wyekstrahować jego «istotę»”^[73].

Na potrzebę wyodrębnienia nowego pola eksploatacji zwracają również uwagę Grzegorz Tylec, Sebastian Kwiecień i Łukasz Sadowski w swoim artykule^[74], zauważają jednak różnice pomiędzy TDM a sztuczną inteligencją jako taką. Zdaniem autorów trudno jest jednoznacznie stwierdzić, czy działania związane z TDM wyczerpują pojęcie aktywności sztucznej inteligencji, czy też stanowią tylko część AI. Zatem stworzenie pola eksploatacji obejmującego TDM mogłoby spowodować pominięcie niektórych rodzajów działalności sztucznej inteligencji^[75], co doprowadziłoby do luki prawnej. Rozwiązaniem tego problemu mogłoby być wprowadzenie postulatu Bałos, zdaniem której nowe pole powinno zostać określone jako: „pole eksploatacji wytwarzania kopii cyfrowych utworu”^[76]. Tak określone pole objęłoby swoim zasięgiem zarówno działalność AI, jak i działalność związaną z TDM.

⁷² Aleksandra Flegel, „Powstanie nowych pól eksploatacji utworów i ich wpływ na treść umów prawnoautorskich w perspektywie prawnoporównawczej” *Folia Iuridica Universitatis Wratislaviensis*, nr 2 (2021): 46.

⁷³ Kunc-Urbańczyk, „Uczenie maszynowe jako pole eksploatacji utworu”, 229.

⁷⁴ Tylec, Kwiecień, Sarowski, Sadowski, „Is It Possible to License Works Used in the Learning Process of Artificial Intelligence Algorithms?”, 9.

⁷⁵ Ibidem, 10.

⁷⁶ Bałos, „Trenowanie wybranych modeli sztucznej inteligencji a uprawnienia twórców”, 19.

5.2. Ochrona praw twórcy w ramach art. 23 k.c. oraz praw osobistych twórcy

W przypadku, w którym do naruszenia praw twórcy dojdzie bez uprzednio zawartej umowy, nasuwa się pytanie o zabezpieczenie praw autora w inny sposób. Jedną z metod ochrony mogłoby być dochodzenie roszczeń z art. 16 pr. aut. Omawiany przepis stanowi bowiem o prawach osobistych twórcy, m.in. prawie do nadzoru nad sposobem korzystania z utworu oraz nienaruszalności treści i formy utworu oraz jego rzetelnego wykorzystania^[77].

W trakcie uczenia maszynowego sztuczna inteligencja niejako „rozbija” utwór, nakłada na niego „szumy”, by potem wyekstrapolować najistotniejsze części dzieła wykorzystanych do uczenia. Czy takie naruszenie można jednak traktować jako naruszenie integralności utworu w rozumieniu ustawy? Utrwalonym w doktrynie i orzecznictwie poglądem jest, że udowodnienie wkroczenia w prawa z art. 16 pr. aut. wymaga wykazania naruszenia więzi twórcy z utworem^[78]. Zmiany w utworze dokonane w ramach uczenia maszynowego zdają się być na tyle znaczne, by do naruszenia takiej więzi doszło, zasadnym byłoby więc powoływanie się na analizowaną regulację. Wartym zauważenia jest jednak stanowisko Bałos, zgodnie z którym w procesie uczenia maszynowego nie dochodzi do zerwania więzi twórcy z utworem ze względu na fakt, że proces ten nie jest widoczny dla twórcy. Innymi słowy, więź twórcy z utworem zbudowana na bazie emocjonalnego stosunku twórcy do dzieła nie może być naruszona, jeśli twórca nie jest w stanie takiego naruszenia zauważyć. Co więcej, więź twórcy może być naruszona jedynie z konkretnym utworem, nie zaś z całym zbiorem utworów, dlatego dochodzenie roszczeń na podstawie art. 16 pr. aut. jest w tym aspekcie w pewien sposób ograniczone^[79].

⁷⁷ Jak podkreśla Sąd Najwyższy w wyroku z dnia 28 lutego 2023 r., sygn. akt: II CSKP 245/22, LEX nr 3509939: „twórczość artystyczna stanowi aktywność służącą wyrażaniu określonych idei i przeżyć oraz dawaniu świadectwa co do określonych postaw i preferowania pewnych wartości. Z tego punktu widzenia uzasadniona jest ochrona prawnoautorska przed wykorzystywaniem utworów w sposób naruszający ich integralność oraz podważający więź, która łączy z nimi ich twórców”.

⁷⁸ Wyrok Sądu Apelacyjnego w Łodzi z dnia 4 lutego 2016 r., sygn. akt: I ACa 1107/15, LEX nr 2665807.

⁷⁹ Iga Bałos, „Trenowanie wybranych modeli sztucznej inteligencji a uprawnień twórców”, 22.

Powyższa analiza prowadzi do wniosku, że ochrona praw twórcy na bazie art. 16 pr. aut. mogłaby być nieskuteczna, choć nie brak autorów, którzy nie odmawiają możliwości dochodzenia roszczeń na tej podstawie prawnej^[80].

Drugą proponowaną metodą ochrony praw twórców jest art. 23 k.c., zawierający otwarty katalog dóbr osobistych podlegających ochronie, przy czym przepis wprost stanowi, że ochronie podlega twórczość artystyczna. Relację pomiędzy dobrami osobistymi z art. 23 k.c. a art. 16 pr. aut. wyjaśnia Jan Błęszyński, zdaniem którego art. 23 k.c. oraz 24 k.c. chronią swobodę pewnej działalności natomiast art. 16 pr. aut. chroni określony rezultat twórczości, jakim jest dzieło^[81].

Jest to więc możliwa forma ochrony twórcy, którego prawa zostały naruszone przez sztuczną inteligencję. Taką metodę zaaprobował m.in. Ryszard Markiewicz w odniesieniu do dzieł naukowych. Dochodzenie praw twórcy na tej podstawie prawnej, miałoby zdaniem Markiewicza, polegać na ochronie autorskiej fragmentów tekstu użytego przez ChatGPT, których autorem jest pokrzywdzony twórca. Jak wskazuje Bałos art. 23 k.c. jest o wiele bardziej elastyczny, ponieważ pozwala na łączną ochronę wszystkich naruszonych utworów, nie zaś konkretnego jak w przypadku art. 16 pr. aut.^[82] Należy zgodzić się, że koncepcja oparcia roszczeń o przepis art. 23 k.c. mogłaby znacznie ułatwić twórcom ochronę praw, które zostały naruszone z wykorzystaniem technologii TDM.

⁸⁰ Jak wskazuje R. Markiewicz: „nie jest także wykluczone sięganie do ochrony autorskich dóbr osobistych, gdy w wytworach ChatGPT eksploatowane są elementy twórcze utworu. Chodzi tu zwłaszcza o naruszenie prawa do autorstwa i prawa do integralności. W wyjątkowych przypadkach, przynajmniej teoretycznie, można powoływać się na tę ochronę także wtedy, gdy elementy twórcze utworu są wykorzystywane. Może to dotyczyć np. rozpowszechniania kontynuacji utworu stworzonej przez ChatGPT czy przyjęcia w takim utworze imienia własnego bohatera literackiego”. Ryszard Markiewicz, „ChatGPT i prawo autorskie Unii Europejskiej” *Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace z Prawa Własności Intelektualnej*, nr 2 (2023): 159.

⁸¹ Jan Błęszyński, *Prawo autorskie* (Warszawa: Polskie Wydawnictwo Naukowe, 1988), 117-118.

⁸² Bałos, „Trenowanie wybranych modeli sztucznej inteligencji a uprawnienia twórców”, 25.

5 | Podsumowanie

Pytanie o TDM jest tak naprawdę pytaniem o wyważenie dóbr twórcy oraz rozwoju gospodarczego. Dokonana analiza jasno wskazuje, że Unia Europejska postanowiła położyć znacznie większy nacisk na ochronę praw twórców. Art. 3 oraz 4 dyrektywy DSM wyznaczają niezwykle restrykcyjne wymogi wykorzystywania utworów w ramach technologii TDM, co potwierdza także analiza prawnoporównawcza. Niektóre z regulacji zawartych w dyrektywie budzą jednak znaczne kontrowersje. Niedostatecznie wyjaśniona zdaje się być relacja pomiędzy legalnością źródła a legalnością dostępu. Nie jest także jednoznaczne, czy podmioty wykorzystujące utwory pochodzące z nielegalnych źródeł będą obowiązane do wypłaty twórcom odpowiednich rekompensat finansowych. W związku z powyższym zasadnym zdaje się przyjęcie, by podmioty wykorzystujące utwory w ramach TDM dołożyły wystarczającej staranności w weryfikacji źródła pochodzenia danych.

Kolejną kontrowersją związaną z omawianą dyrektywą jest instrument *opt-out*. Samo zastosowanie tej instytucji jest poddawane w wątpliwość ze względu na swoje zbyt radykalne skutki, jakie rodzi dla branży IT. Nie do końca znane są także środki, za pomocą których autorzy mogliby dokonać skutecznych zastrzeżeń, tak by nie doszło do naruszeń ich praw. Na aprobatę zasługuje natomiast art. 53 ust. 1 lit. d rozporządzenia, który daje twórcom możliwość uzyskania informacji na temat tego, czy ich dzieło zostało wykorzystane w procesie uczenia maszynowego.

Przyjęte regulacje z uwagi na ich niedookreśloność nie rozwiązują wszystkich problemów, z którymi muszą zmierzyć się twórcy w obliczu coraz szybszego rozwoju AI. Tym samym praktyczny wymiar ochrony ich praw pozostaje niejasny. W tym aspekcie koniecznym zdaje się wprowadzenie nowego pola eksploatacji utworu do polskiego prawa, co umożliwiłoby skuteczne zabezpieczenie praw twórców. Niemniej jednak w przypadku dokonanych naruszeń w związku z technologią TDM nie jest wykluczone dochodzenie ochrony również na gruncie dóbr osobistych oraz autorskich praw osobistych.

Bibliografia

- Bagieńska-Masiota Aleksandra, „Dozwolony użytek w zakresie eksploatacji tekstów i danych w świetle Dyrektywy Parlamentu Europejskiego i Rady (UE) 2019/790” *Annales Universitatis Paedagogicae Cracoviensis | Studia De Cultura*, nr 1 (2022): 118-128. <https://doi.org/10.24917/20837275.14.1.8>.
- Bałos Iga, „Trenowanie wybranych modeli sztucznej inteligencji a uprawnienia twórców” *Studia Prawnicze. Rozprawy i Materiały*, nr 2 (2019): 15-28.
- Błęszyński Jan, *Prawo autorskie*, Warszawa: Polskie Wydawnictwo Naukowe, 1988.
- Fernández-Molina Juan Carlos, Fernando Esteban de la Rosa, „Copyright and Text and Data Mining: Is the Current Legislation Sufficient and Adequate?” *Portal. Libraries and the Academy*, nr 3 (2024): 653-672.
- Flegel Aleksandra, „Powstanie nowych pól eksploatacji utworów i ich wpływ na treść umów prawnoautorskich w perspektywie prawnoporównawczej” *Folia Iuridica Universitatis Wratislaviensis*, nr 2 (2021): 42-54. <https://doi.org/10.34616/141808>.
- Gisburg Jane. C., „Fair Use in The United States: Transformed, Deformed, Reformed?” *Singapore Journal of Legal Studies* (2020): 265-294.
- Griffiths Jonathan, Tatiana Eleni Synodinou, Raquel Xalabarder, „Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market” *European Copyright Society*, 3 maja 2022.
- Have I Been Trained? Frequently Asked Questions. <https://haveibeentrained.com>.
- Keller Paul, „A First Look at the Copyright Relevant Parts in the Final AI Act Compromise” *Kluwer Copyright Blog*, 11 grudzień, 2023. <https://copyright-blog.kluweriplaw.com/2023/12/11/a-first-look-at-the-copyright-relevant-parts-in-the-final-ai-act-compromise/>.
- Keller Paul, „Machine Readable or not? – Notes on the Hearing in LAION e.v. vs. Kneschke”. *Kluwer Copyright Blog*, 22 lipca 2024. <https://copyrightblog.kluweriplaw.com/2024/07/22/machine-readable-or-not-notes-on-the-hearing-in-laion-e-v-vs-kneschke/>.
- Kretschmer Martin, Thomas Margoni, Pinar Oruç, „Copyright Law, and the Lifecycle of Machine Learning Models” *IIC – International Review of Intellectual Property and Competition Law*, nr 1 (2024): 110-138. <https://doi.org/10.1007/s40319-023-01419-3>.
- Kunc-Urbańczyk Karina, „Uczenie maszynowe jako pole eksploatacji utworu”, [w:] *Prawo sztucznej inteligencji i nowych technologii*, red. Bogdan Fischer, Adam Pązik, Marek Świerczyński. 211-240. Warszawa: Wolters Kluwer, 2021.
- Kupiec Mateusz, „Amerykański fair use i japoński non-enjoyment – drogowskazy dla europejskich rozwiązań prawnych w zakresie TDM?” *Zeszyty Naukowe*

- Uniwersytetu Jagiellońskiego. *Prace z Prawa Własności Intelektualnej*, nr 2 (2021): 134-146.
- Kusuma Fauzan Eka, Rahmi Rahmi, „The Use of Shadow Libraries at Universitas Indonesia”. *First Monday*, nr 10 (2023): 1-24. <https://doi.org/10.5210/fm.v28i10.12947>.
- Lindberg Matthew, „Applying Current Copyright Law to Artificial Intelligence Image Generators in the Context of Anderson v. Stability AI, Ltd.” *Mitchell Hamline School of Law*, nr 1 (2024): 38-62.
- Manteghi Maryna, „Can Text and Data Mining Exceptions and Synthetic Data Training Mitigate Copyright-Related Concerns in Generative AI?” *Law, Innovation and Technology*, nr 2 (2024): 663-686. <https://doi.org/10.1080/17579961.2024.2392928>.
- Margoni Thomas, “Saving Research: Lawful Access to Unlawful Sources Under Art. 3 CDSM Directive?” *Kluwer Copyright Blog*, 22 grudzień, 2023. <https://copyright-blog.kluweriplaw.com/2023/12/22/saving-research-lawful-access-to-unlawful-sources-under-art-3-cdsm-directive/>.
- Markiewicz Ryszard, „ChatGPT a prawo Unii Europejskiej” *Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace z Prawa Własności Intelektualnej*, nr 2 (2023): 143-171.
- Markiewicz Ryszard, *Prawo autorskie na jednolitym rynku cyfrowym: Dyrektywa Parlamentu Europejskiego i Rady (UE) 2019/790*. Warszawa: Wolters Kluwer, 2021.
- Nowakowski Michał, *Sztuczna inteligencja. Praktyczny przewodnik dla sektora innowacji finansowych*. Warszawa: Wolters Kluwer, 2023.
- Traple Elżbieta, „Granice eksploracji tekstów i danych na potrzeby maszynowego uczenia się przez systemy sztucznej inteligencji”, [w:] *Prawo sztucznej inteligencji i nowych technologii*, red. Bogdan Fischer, Adam Pązik, Marek Świerczyński. 24-25. Warszawa: Wolters Kluwer, 2021.
- Traple Elżbieta, „Komentarz do art. 50”, [w:] *Komentarz do ustawy o prawie autorskim i prawach pokrewnych*, red. Ryszard Markiewicz, Janusz Barta, teza 3. Warszawa: Wolters Kluwer, 2011. Lex.
- Tylec Grzegorz, Sebastian Kwiecień, Łukasz Sadowski, „Is It Possible to License Works Used in the Learning Process of Artificial Intelligence Algorithms?”: 1-19.
- VentureBeat, „OpenAI will Release a Media Manager Tool in 2025, Allowing Creators to Block AI Training”. <https://venturebeat.com/ai/openai-will-release-a-media-manager-tool-in-2025-allowing-creators-to-block-ai-training/>.
- Vesala Juha, „Developing Artificial Intelligence-Based Content Creation: Are EU Copyright and Antitrust Law Fit for Purpose?” *IIC – International Review of Intellectual Property and Competition Law*, nr 3 (2023): 351-380. <https://doi.org/10.1007/s40319-023-01301-2>.

Whitepress, *Robots.txt* – do czego służy i jakich błędów unikać przy jego tworzeniu?.

<https://www.whitepress.com/pl/baza-wiedzy/642/robots-txt-do-czego-sluzyi-jakich-bledow-unikac-przy-jego-tworzeniu>.

Ziaja Gina Maria, „The Text and Data Mining Opt-Out in Article 4(3) CDSMD: Adequate Veto Right for Rightholders or a Suffocating Blanket for European Artificial Intelligence Innovations?” *Journal of Intellectual Property Law & Practice*, nr 5 (2024): 453-459. <https://doi.org/10.1093/jiplp/jpae025>.



This article is published under a Creative Commons Attribution 4.0 International license.
For guidelines on the permitted uses refer to
<https://creativecommons.org/licenses/by/4.0/legalcode>