



dr Zbigniew Okoń

Radca prawny, partner w kancelarii Lubasz i Wspólnicy sp.k. (ORCID: <https://orcid.org/0000-0002-0999-780X>).

Memoryzacja utworów przez modele AI w świetle unijnego prawa autorskiego

Słowa kluczowe: *memoryzacja, sztuczna inteligencja, prawo autorskie, eksploracja tekstu i danych (TDM), dyrektywa 2001/29/WE*

Gwałtowny rozwój sztucznej inteligencji doprowadził do bezprecedensowej liczby sporów sądowych pomiędzy podmiotami praw autorskich a firmami technologicznymi. W Stanach Zjednoczonych pierwsze rozstrzygnięcia sądowe potwierdziły jednak, że trenowanie modeli sztucznej inteligencji na materiałach pozyskanych z Internetu bez zgody uprawnionych mieści się w granicach *fair use*¹. W Europie uzasadnieniem dla takiej praktyki może być z kolei wyjątek dotyczący eksploracji tekstu i danych, wynikający z art. 3 i 4 dyrektywy 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 i 2001/29/WE², co – mimo sceptycyzmu części doktryny – znajduje potwierdzenie w orzecznictwie³, przepisach rozporządzenia Parlamentu Europejskiego i Rady (UE) 2024/1689 z 13.06.2024 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) 2020/1828⁴ czy dokumentach Komisji⁵. Istnienie podstawy prawnej dla wkraczających w zakres praw autorskich czynności związanych z uczeniem maszynowym nie rozwiązuje jednak wszystkich problemów. Pojawia się bowiem pytanie, jak należy oceniać przypadki, gdy model „zapamięta” (zmemoryzuje) dane zawarte w korpusie danych treningowych i pojawiać się one będą w wynikach jego działania.

Problemu tego dotyczy niedawny wyrok Sądu Krajowego w Monachium w sprawie GEMA przeciwko OpenAI, w którym „zapamiętanie” (memoryzacja) fragmentów utworów przez model oraz zwracanie ich w odpowiedzi na polecenia (prompty)

użytkownika zostały uznane za naruszenie praw autorskich⁶. Ponieważ memoryzacja jest zjawiskiem, któremu – jak sugerują badania z zakresu informatyki i sztucznej inteligencji – nie da się całkowicie zapobiec, kwalifikacja taka istotnie zmienia sytuację prawną dostawców modeli AI, otwierając drogę do roszczeń nakierowanych na same modele. Wydaje się mieć też niekorzystny wpływ na rozwój europejskiej infrastruktury technicznej w zakresie sztucznej inteligencji oraz zwiększać ryzyko udostępniania i użycia modeli otwartych (*open source*). W artykule przedstawiam rozumowanie sądu, analizuję adekwatność leżących u jego podstaw ocen oraz staram się wskazać ich konsekwencje. Wskazuję też na potrzebę znalezienia rozwiązań prawnych, które umożliwiłyby lepsze wyważenie interesów dostawców modeli i podmiotów uprawnionych.

1 Zob. Andrea Bartz et al. przeciwko Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. 23.06.2025 r.), Richard Kadrey et al. przeciwko Meta Platforms, Inc., No. 3:23-cv-03417 (N.D. Cal. 25.06.2025 r.).

2 Dz.Urz. UE L 130, s. 92 – dalej dyrektywa 2019/790.

3 Zob. zwłaszcza wyrok OLG Hamburg z 10.12.2025 r., 5 U 104/24, Robert Kneschke przeciwko LAION e.V.

4 Dz.Urz. UE L 2024/1689 – dalej AIA, akt w sprawie sztucznej inteligencji. Chodzi tu o przepis art. 53 ust. 1 lit. c oraz motywy 105 i 106 AIA. Regulacja ta wyraźnie wskazuje na możliwość zastosowania wyjątku eksploracji tekstu i danych (*text and data mining*, TDM) w trenowaniu modeli ogólnego przeznaczenia.

5 Zob. komunikat Komisji do Parlamentu Europejskiego, Rady Europejskiej, Rady Europejskiego Komitetu Ekonomiczno-Społecznego i Komitetu Regionów z 25.04.2018 r., Sztuczna inteligencja dla Europy, COM(2018) 237 final, w którym TDM jest wymieniany w kontekście uczenia maszynowego.

6 LG München I, wyrok z 11.11.2025 r., 42 O 14139/24, GEMA przeciwko OpenAI.

1. Sprawa GEMA przeciwko OpenAI

Wyrok monachijskiego sądu zasługuje na uwagę nie tylko dlatego, że należy do pierwszych europejskich rozstrzygnięć dotyczących sztucznej inteligencji w kontekście prawa autorskiego. Po pierwsze, powód – niemiecka organizacja zbiorowego zarządzania GEMA – z powodzeniem udowodnił, że modele GPT faktycznie „zapamiętały” fragmenty kilku utworów objętych repertuarem GEMA i potrafią je odtworzyć. Po drugie, sprawa rozstrzygana była wprawdzie na gruncie niemieckiego prawa autorskiego, ale w zakresie objętym harmonizacją, co wymagało od sądu zastosowania konstrukcji wspólnych dla europejskiego systemu ochrony: szerokiego zakresu prawa zwielokrotniania z art. 2 dyrektywy 2001/29/WE Parlamentu Europejskiego i Rady z 22.05.2001 r. w sprawie harmonizacji niektórych aspektów praw autorskich i pokrewnych w społeczeństwie informacyjnym⁷, zamkniętego katalogu wyjątków i ograniczeń – w tym wyjątku dotyczącego eksploracji tekstu i danych z art. 4 dyrektywy 2019/790 – oraz koncepcji naruszenia praw wyłącznych i wynikających z tego roszczeń ukształtowanych przez dyrektywę 2004/48/WE Parlamentu Europejskiego i Rady z 29.04.2004 r. w sprawie egzekwowania praw własności intelektualnej⁸. Wnioski płynące z tego rozstrzygnięcia mogą być więc użyteczne zarówno przy wykładni prawa unijnego, jak i prawa państw członkowskich, w tym prawa polskiego.

Stan faktyczny leżący u podstaw wyroku LG München I nie był skomplikowany. GEMA, za pomocą relatywnie prostych promptów – takich jak nakazanie modelowi, żeby wcielił się w rolę eksperta od niemieckich piosenek, a następnie przytoczył fragmenty tekstu wskazanej piosenki – uzyskała od modeli gpt-4 i gpt-4o udostępnianych przez OpenAI fragmenty tekstów dziewięciu popularnych niemieckich utworów z repertuaru GEMA. Fragmenty te miały długość od kilkunastu do kilkudziesięciu słów. Sąd stwierdził, że w ten sposób doszło do naruszenia w dwóch obszarach. „Zapamiętanie” spornych utworów w modelach gpt-4 i gpt-4o zostało uznane za zwielokrotnienie. Zdaniem sądu możliwość uzyskania fragmentów utworów „na wyjściu” oznacza bowiem, że są one przechowywane w modelu, a więc model jest ich kopią. Natomiast obecność fragmentów utworów w wynikach działania modelu sąd zakwalifikował zarówno jako zwielokrotnienie, jak i publiczne udostępnienie tych utworów.

Sąd odrzucił argumentację OpenAI zmierzającą do przypisania użytkownikom modelu odpowiedzialności za naruszenia w generowanych przez model wynikach. Podkreślił, że spółki OpenAI kontrolują architekturę modeli, proces treningu i wynikającą z niego memoryzację danych, a zastosowane przez GEMA proste prompty nie doprowadziły do utraty tej kontroli przez OpenAI. W konsekwencji sąd wykluczył możliwość zastosowania dodatkowych kryteriów odpowiedzialności właściwych dla podmiotów jedynie pośrednio przyczyniających się do naruszenia, takich jak aktywna rola dostawcy czy rozmyślny charakter jego działania. Sąd nie przychylił się również do argumentacji pozwanych, że ingerencja w prawo zwielokrotniania i częściowo prawo publicznego udostępniania może znajdować podstawę w wyjątkach i ograniczeniach praw autorskich – przede wszystkim wyjątku dotyczącym eksploracji tekstu i danych oraz

przewidzianego w § 57 niemieckiej ustawy o prawie autorskim wyjątku dotyczącego „nieistotnego dodatku” (niem. *unwesentliches Beiwerk*), będącego odpowiednikiem przewidzianego w art. 5 ust. 3 lit. i dyrektywy 2001/29/WE wyjątku dotyczącego „niezamierzonego włączenia” utworu do innego nośnika. Odnośnie do TDM sąd uznał, że objęcie memoryzacji jego zakresem wymagałoby rozszerzającej interpretacji, na co nie pozwalają zasady wykładni ograniczeń i wyjątków w prawie unijnym oraz test trójstopniowy uregulowany w art. 5 ust. 5 dyrektywy 2001/29/WE. Zastosowanie przepisu zezwalającego na użycie cudzego utworu jako „nieistotnego dodatku” ma z kolei uniemożliwiać wynikający z niemieckiej ustawy wymóg włączenia takiego utworu do innego utworu – przymiotu tego nie ma korpus danych treningowych OpenAI.

Ostatecznie sąd nakazał zaniechanie naruszeń polegających na zwielokrotnianiu tekstów piosenek w wynikach zwracanych przez te modele oraz oddzielnie w modelach językowych pozwanych. Odrzucił przy tym argumentację OpenAI, zgodnie z którą nakaz zaprzestania zwielokrotniania tekstów w modelu jest nieproporcjonalny, ponieważ jego wykonanie wymagałoby ponownego przetrenowania modeli. Po pierwsze, uznał, że niemiecki ustawodawca nie zaimplementował w ustawie o prawie autorskim i prawach pokrewnych⁹ przepisu art. 12 dyrektywy 2004/48/WE, pozwalającego w przypadku naruszeń, do których doszło nieumyślnie i „bez zaniebdania”, na zastosowanie środków alternatywnych, jeżeli środki naprawcze lub zakazy sądowe są nieproporcjonalne. Po drugie, wskazał, że doszło tu co najmniej do zaniebdania ze strony pozwanych, skoro problem memoryzacji niektórych utworów przez modele OpenAI był im znany co najmniej od publikacji w 2021 r. wyników badań naukowych dotyczących tego zagadnienia, w których uczestniczyli również pracownicy pozwanych.

2. Kwalifikacja memoryzacji jako zwielokrotniania w rozumieniu prawa autorskiego

Kluczową tezę wyroku jest uznanie za zwielokrotnienie memoryzacji danych treningowych przez model. Wyrok LG München I opiera się tu na następującym rozumowaniu: skoro w niektórych przypadkach modele potrafią – po odpowiednim poleceniu użytkownika – wygenerować identyczne lub jedynie nieznacznie zmodyfikowane fragmenty utworów, to sam model musi stanowić kopię tych utworów. Sąd odwołał się tu do znaczenia, jakie pojęciu zwielokrotniania nadaje się w doktrynie i orzecznictwie niemieckim, uznając je za zgodne z prawem unijnym¹⁰. Zakłada ono, że do zwielokrotnienia dochodzi, gdy utwór zostaje zapisany w sposób umożliwiający jego percepcję – bezpośrednią lub pośrednią (tj. za pomocą urządzeń technicznych, jak w przypadku plików cyfrowych). W konsekwencji uznał, że utwory zapisane w wagach modelu są możliwe do pośredniego odtworzenia: użytkownik może bowiem „odtworzyć” zmemoryzowane utwory za pomocą

7 Dz.Urz. WE L 167, s. 10, ze zm. – dalej dyrektywa 2001/29/WE.

8 Dz.Urz. WE L 157, s. 45 – dalej dyrektywa 2004/48/WE.

9 Gesetz über Urheberrecht und verwandte Schutzrechte z 9.09.1965 r. (BGBl. I S. 1273 ze zm.) – dalej UrhG.

10 Z powołaniem na wyrok TS z 29.07.2019 r., C-476/17, Pelham GmbH i in. przeciwko Ralfowi Hütterowi i Florianowi Schneiderowi-Eslebenowi, EU:C:2019:624, pkt 52 i n. oraz opinię Rzecznika Generalnego M. Szpunara z 12.12.2018 r., Pelham GmbH i in. przeciwko Ralfowi Hütterowi i Florianowi Schneiderowi-Eslebenowi, EU:C:2018:1002, pkt 46.

interfejsu czatu, co ma być sytuacją porównywalną do odtwarzania muzyki z odtwarzacza MP3.

Takie podejście przeważa w literaturze przedmiotu¹¹, i to nie tylko niemieckiej¹². Za nieistotne uznaje się przy tym, że zapis utworu ma postać zbioru wag sieci neuronowej, których analiza nie pozwala ani na odtworzenie zapamiętanych przez model treści, ani nawet na ustalenie, jakie dokładnie treści zostały zapamiętane. Stanowiska odmienne są w mniejszości. Taki mniejszościowy pogląd na gruncie prawa unijnego prezentują M. Leistner i L. Antoine. Autorzy argumentują, że w świetle orzecznictwa TS zharmonizowane pojęcie utworu wymaga, by przedmiot ochrony był „możliwy do zidentyfikowania z wystarczającą precyzją i obiektywnością”¹³. Warunek ten nie jest ich zdaniem spełniony w odniesieniu do potencjalnej reprodukcji danych treningowych w wyszkolonym modelu AI, skoro nawet ekspert nie jest w stanie wykryć i zidentyfikować konkretnego utworu w ramach parametrów modelu. Utwory wykorzystane do szkolenia mogą „ponownie pojawić się” jedynie w wyniku konkretnego polecenia użytkownika, lecz wynik ten jest wyrazem efektu probabilistycznego, którego nie można przewidzieć z wystarczającą pewnością. Skłania to M. Leistnera i L. Antoine do usytuowania memoryzacji poza zakresem prawa zwielokrotniania z art. 2 dyrektywy 2001/29/WE. Ich zdaniem takie ujęcie nie pozbawia uprawnionych możliwości dochodzenia roszczeń z tytułu naruszenia praw autorskich, mogą oni bowiem podnosić roszczenia w odniesieniu do naruszeń w wynikach generowanych przez model, racjonalizuje natomiast zakres ochrony prawnoautorskiej¹⁴.

Przyrównywanie memoryzacji do typowych form reprodukcji utworów rzeczywiście nie uwzględnia technicznej specyfiki tego zjawiska. Model językowy podczas treningu uczy się wzorców statystycznych z danych treningowych. Wzorce te mogą mieć charakter całkowicie abstrakcyjny, obejmując reguły języka naturalnego, mogą dotyczyć faktów, takich jak to, że Paryż jest stolicą Francji, ale mogą również pozwalać na odtworzenie niektórych danych treningowych lub ich fragmentów. Oznacza to, że nie odpowiadają w pełni rzeczywistości ani twierdzenia, wedle których modele stanowią „bazy danych” składające się ze „skompresowanych” utworów wchodzących w skład korpusu treningowego¹⁵, ani argumentacja, zgodnie z którą modele zawierają jedynie abstrakcyjne wzorce obojętne

z punktu widzenia prawa¹⁶. W rzeczywistości wytrenowany model zawiera wzorce wyekstrahowane z danych treningowych, z tym że czasem te wzorce są na tyle szczegółowe, że pozwalają na dokładne odtworzenie fragmentów danych treningowych. Memoryzacja nie jest więc kategorią zero-jedynkową, ale pewnym ekstremum na kontinuum sposobów uczenia się modelu¹⁷, które występuje tylko czasami.

Według A. Coopera i J. Grimmelmanna oraz cytowanych w ich pracy badań K. Lee, D. Ippolito i innych dokładne oszacowanie częstotliwości memoryzacji jest trudne. W zależności od modelu i implementacji wskaźnik memoryzacji może wahać się między 0,1% a 100% treści wchodzących w skład korpusu danych treningowych. Ocenę utrudnia to, że samo generowanie wyników przez model jest do pewnego stopnia probabilistyczne. W ten sposób system generatywnej sztucznej inteligencji może np. w 10% przypadków dla danego polecenia powielić dane szkoleniowe jako wynik, a w pozostałych 90% przypadków generować inne wyniki¹⁸. W nowszym badaniu A. Coopera i współpracowników, obejmującym 50 książek i 17 otwartych modeli językowych, przeprowadzonym w kontrolowanym przez autorów środowisku, a więc bez użycia mechanizmów filtracji wyników, stwierdzono, że typowo modele nie memoryzują większości książek – ani w całości, ani w części. Przy przyjętym progu prawdopodobieństwa ekstrakcji na poziomie 10% większość książek wykazuje pokrycie poniżej 2%. Istnieją jednak znaczące wyjątki. Z modelu Llama 3.1 70B możliwe jest wyekstrahowanie całej książki *Harry Potter i kamień filozoficzny*: 43% tekstu można odtworzyć z prawdopodobieństwem przekraczającym 50%, a ponad 90% – z prawdopodobieństwem co najmniej 1%. Podobnie wysoką memoryzację wykazują: *Rok 1984*, *Wielki Gatsby* oraz *Hobbit*. Z drugiej strony książka *Sandman Slim* Richarda Kadreya – powoda we wspomnianej sprawie Kadrey przeciwko Meta – wykazuje pokrycie poniżej 0,4% dla wszystkich testowanych modeli i progów, a wzorec ten jest typowy dla większości badanych pozycji¹⁹. Memoryzacja jest więc zjawiskiem, które występuje w relatywnie niskim, ale mierzalnym stopniu. Konkretnie wartości zależą od pojemności modelu, duplikacji danych treningowych w zbiorze czy zastosowanych algorytmów uczenia maszynowego.

Obecnie technicznie nie jest możliwe całkowite wyeliminowanie memoryzacji. Dostawca modelu ma wprawdzie wpływ na skalę tego zjawiska – udowodniono na przykład, że przykłady wielokrotnie powtórzone w zbiorze treningowym są zapamiętywane częściej²⁰ – nie jest to jednak proces pozostający pod jego pełną kontrolą. Memoryzacja może bowiem wystąpić nawet wtedy, gdy dany przykład pojawia się w zbiorze treningowym tylko raz. Co więcej, wedle aktualnego stanu wiedzy pewien poziom memoryzacji

11 Tak m.in. A. Sesing-Wagenpfeil, *Trainierte KI-Modelle als Vervielfältigungsstücke im Sinne des Urheberrechts*, „Zeitschrift für geistiges Eigentum” 2024/2, s. 266; T.W. Dornis, *Urheberrecht und Training generativer KI-Modelle*, Baden-Baden 2024, s. 72; P.J. Pesch, R. Böhme, *Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern*, „GRUR. Gewerblicher Rechtsschutz und Urheberrecht” 2023/14, s. 1005; M. von Welsler, *Generative KI und Open Source Software*, „GRUR. Gewerblicher Rechtsschutz und Urheberrecht” 2024/19, s. 1411.

12 Tak. A.F. Cooper, J. Grimmelmann, *The Files are in Computer: on Copyright, Memorization, and Generative AI*, arXiv 2024, <http://arxiv.org/abs/2404.12590> (dostęp: 28.02.2025 r.), s. 20; P. Mezei, *Memorization and Generative AI – A Persistent Issue with Copyright Consequences?*, Social Science Research Network 2025, <https://papers.ssrn.com/abstract=5404367> (dostęp: 1.12.2025 r.), s. 5.

13 Z powołaniem na wyrok TS z 13.11.2018 r., C-310/17, Levola Hengelo BV przeciwko Smilde Foods BV, EU:C:2018:899, pkt 40.

14 M. Leistner, L. Antoine, *TDM and AI Training in the European Union – From ‘LAION’ to Possible Ways Ahead?*, „GRUR International” 2025/11, s. 1032–1033. Podobnie na gruncie prawa amerykańskiego O. Bracha, *The Work of Copyright in the Age of Machine Production*, „Harvard Journal of Law & Technology” 2024/1, s. 8, 24–25.

15 P. Hense, *Overfitting*, „MMR. Multimedia und Recht” 2024/6, s. 449.

16 A.F. Cooper, J. Grimmelmann, *The Files...*, s. 24–25.

17 A.F. Cooper, K. Lee, J. Grimmelmann, D. Ippolito i in., *Report of the 1st Workshop on Generative AI and Law*, Social Science Research Network 2023, <https://papers.ssrn.com/abstract=4634513> (dostęp: 30.11.2025 r.), s. 4–5.

18 A.F. Cooper, J. Grimmelmann, *The Files...*, s. 22, 49 z powołaniem na pracę K. Lee, D. Ippolito i in., *Deduplicating Training Data Makes Language Models Better*, arXiv 2022, <http://arxiv.org/abs/2107.06499> (dostęp: 4.12.2025 r.), s. 8.

19 A.F. Cooper, A. Gokaslan i in., *Extracting memorized pieces of (copyrighted) books from open-weight language models*, arXiv 2025, <http://arxiv.org/abs/2505.12546> (dostęp: 2.12.2025 r.), s. 7–8.

20 A.F. Cooper, J. Grimmelmann, *The Files...*, s. 31.

wydaje się niezbędny dla prawidłowego funkcjonowania modelu²¹. Dostawcy stosują oczywiście różnorodne środki zaradcze: deduplikację danych treningowych, optymalizację procesów uczenia maszynowego oraz mechanizmy filtracji wejścia (promptów użytkowników) i wyjścia (wyników zwracanych przez model), które ograniczają możliwość uzyskania przez użytkowników zmemoryzowanych treści²². Mechanizmy te nie eliminują jednak samej memoryzacji, tj. zapisu określonej informacji w wagach modelu, a jedynie ograniczają możliwość wywołania tych treści przez użytkownika.

Wreszcie pewne znaczenie mają działania użytkowników. Gdy model otrzymuje prompt, przewiduje na podstawie wyuczonych wzorców, które z możliwych uzupełnień jest najbardziej prawdopodobne. Jeśli model zwraca dane treningowe w odpowiedzi na dany prompt, czyni to dlatego, że nauczył się, że tekst z przykładowego treningowego stanowi najbardziej prawdopodobne uzupełnienie dla tego konkretnego promptu²³. Na wynik działania modelu ma więc wpływ również treść promptu, który musi „pobudzić” model do zwrócenia określonych treści np. przez użycie dłuższych promptów, czy wielokrotne powtarzanie tych samych lub bardzo zbliżonych prób²⁴. Już tylko z rachunku prawdopodobieństwa wynika, że podjęcie setek lub tysięcy prób, tak jak GEMA w omawianej sprawie, czyni uzyskanie w wynikach przynajmniej kilku fragmentów utworów bardziej prawdopodobnym.

Przyrównywanie memoryzacji w modelu do pliku MP3, a interfejsu chata do odtwarzacza muzyki wydaje się być więc nadmiernym uproszczeniem. W konstatacji tej kryje się zasadniczy zarzut pod adresem wyroku LG München. Istota problemu oceny memoryzacji w świetle prawa autorskiego nie wydaje się bowiem leżeć w tym, czy zakresem prawa zwielokrotniania objęte jest wykonanie kopii możliwej do „pośredniej” percepcji, a w tym, czy zwielokrotnianiem w znaczeniu normatywnym jest wykonywanie kopii nieprzeznaczonych do percepcji, w ramach procesu niesłużącego do zapisu utworów w celu ich późniejszego odtworzenia i zależnego w znacznym stopniu od efektów probabilistycznych. Twierdząca odpowiedź na tak postawione pytanie prowadzi do wniosku, że skoro zjawisku temu dostawca modelu nie może w pełni zapobiec, a dotyka ono w istocie każdego modelu, to każdy model należałoby uznać za kopię jakichś utworów. Otwartym pytaniem pozostaje jedynie – których.

3. Możliwość zastosowania ograniczeń i wyjątków praw autorskich

Przyjęcie założenia, że memoryzacja stanowi zwielokrotnianie utworów, przeszuwa ciężar oceny tego zjawiska z pojęcia zwielokrotniania na zastosowania ograniczeń i wyjątków praw wyłącznych (dozwolonego użytku). Pełna analiza przekracza ramy opracowania. Warto jednak zasygnalizować główne problemy prawne.

Podstawę dla czasowych czynności zwielokrotniania, do których dochodzi w pełni zautomatyzowanych procesach eksploracji tekstu i danych – w szczególności dla kopii tymczasowych powstających w pamięci operacyjnej podczas przetwarzania danych, może stanowić art. 5 ust. 1 dyrektywy 2001/29/WE²⁵,

który ustanawia obligatoryjny wyjątek od prawa zwielokrotniania dla czasowych czynności reprodukcji stanowiących integralną część procesu technologicznego²⁶. Takie podejście jest zgodne z intencją prawodawcy unijnego, który w motywie 18 dyrektywy 2019/790 wyraźnie zaznaczył, że wyjątek TDM „nie narusza istniejącego wyjątku dotyczącego czasowych czynności zwielokrotniania przewidzianego w art. 5 ust. 1 dyrektywy 2001/29/WE”²⁷. Artykuł 5 ust. 1 dyrektywy 2001/29/WE nie może jednak uzasadniać samej memoryzacji już choćby z tego powodu, że dotyczy wyłącznie zwielokrotnień czasowych, podczas gdy memoryzacja stanowi przypadek reprodukcji trwałej – parametry modelu, w których zapisane są informacje pozwalające na odtworzenie utworów, pozostają utrwalone przez cały okres eksploatacji modelu.

Bardziej złożona jest kwestia wyjątków dotyczących eksploracji tekstu i danych (art. 3 i 4 dyrektywy 2019/790). Możliwość ich zastosowania do trenowania modeli AI jest kwestionowana przez niektórych przedstawicieli nauki prawa²⁸. Niezależnie od przyjętego stanowiska doktryna jest jednak zgodna, że wyjątek TDM nie obejmuje późniejszych aktów publicznego udostępniania wyników eksploracji ani udostępniania zbiorów danych zawierających kopie chronionych treści²⁹. Wykraczałoby to poza cel regulacji, która – nieco upraszczając – umożliwiać ma jedynie ekstrakcję niechronionych informacji zawartych w utworach³⁰. Na przeszkodzie próbom rozszerzającej interpretacji art. 3 i 4 dyrektywy 2019/790, obok testu trójstopniowego z art. 5 ust. 5 dyrektywy 2001/29/WE, stoi zaś konstrukcja ograniczeń i wyjątków w prawie unijnym, gdzie są one traktowane jako wyjątek od zasady, zgodnie z którą oryginalne dzieła autorów zasługują na ochronę przy zachowaniu wysokiego poziomu tej ochrony³¹.

Podejście takie przyjął również LG München w omawianej sprawie, uznając, że wyjątek dotyczący eksploracji tekstów i danych (§ 44b UrhG, art. 4 dyrektywy 2019/790) może stanowić podstawę zwielokrotniania utworów na potrzeby trenowania modeli AI, z tym jednak zastrzeżeniem, że nie obejmuje

26 Szczegółowo na temat tego wyjątku w kontekście TDM zob. E. Rosati, *Copyright in the digital single market: article-by-article commentary to the provisions of directive 2019/790*, New York 2021, s. 29 i n.

27 Zob. także motyw 9 dyrektywy 2019/790, który potwierdza, że czynności TDM mogą w niektórych przypadkach kwalifikować się do wyjątku z art. 5 ust. 1 dyrektywy 2001/29/WE.

28 Zastosowanie wyjątków TDM do trenowania AI kwestionuje T.W. Dornis, *Urheberrecht...*, s. 88 i n. Odmienne m.in. J.P. Quintais, *Generative AI, copyright and the AI Act*, „Computer Law & Security Review” 2025/56, s. 2; M. Leistner, L. Antoine, *TDM and AI Training...*, s. 5–8; T. Margoni, M. Kretschmer, *A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology*, „GRUR International” 2022/8, s. 685 i n.

29 Zob. M. Veale, J.P. Quintais, *The Obligations of Providers of General-Purpose AI Models*, Social Science Research Network 2025, <https://papers.ssrn.com/abstract=5744602> (dostęp: 18.11.2025 r.), s. 16 („TDM exceptions do not cover subsequent acts of communication to the public or making available of the results”); podobnie J.P. Quintais, *Generative AI, copyright and the AI Act*, s. 6.

30 Zob. C. Geiger, G. Frosio, O. Bulayenko, *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Legal Aspects*, „Social Science Research Network” 2018, <https://www.ssrn.com/abstract=3160586> (dostęp: 18.11.2025 r.), s. 2.

31 Zob. wyroki TS: z 29.07.2019 r., C-469/17, Funke Medien NRW GmbH przeciwko Bundesrepublik Deutschland, EU:C:2019:623, pkt 69 – dalej wyrok C-469/17, Funke Medien; z 29.11.2017 r., C-265/16, VCAST Limited przeciwko RTI SpA, EU:C:2017:913, pkt 32; z 12.10.2016 r., C-166/15, Aleksandrs Ranks i Jurijs Vasiļevičs przeciwko Finanšu un ekonomisko noziegumu izmeklēšanas prokuratūra i Microsoft Corp., EU:C:2016:762, pkt 42, oraz z 16.07.2009 r., C-5/08, Infopaq International A/S przeciwko Danske Dagblades Forening, EU:C:2009:465, pkt 57.

21 A.F. Cooper, J. Grimmelmann, *The Files...*, s. 49 („memorization is a Goldilocks phenomenon; models are most useful when they memorize just the right amount, neither too little nor too much”).

22 A.F. Cooper, J. Grimmelmann, *The Files...*, s. 54–55.

23 A.F. Cooper, A. Gokaslan i in., *Extracting memorized...*, s. 2–3.

24 A.F. Cooper, A. Gokaslan i in., *Extracting memorized...*, s. 9–10.

25 M. Stieper, M. Denga, *Die Reichweite des EU-Urheberrechts nach der KI-VO*, „GRUR. Gewerblicher Rechtsschutz und Urheberrecht” 2024/20, s. 1475.

on tych przypadków trenowania modelu, które prowadzą do memoryzacji utworu. O ile z takim stanowiskiem można się w ogólnych zarysach zgodzić, o tyle wątpliwości budzi jego uzasadnienie. Sąd poddał bowiem cykl życia modelu analizie, wyróżniając trzy fazy: tworzenie materiału treningowego (objęte wyjątkiem TDM), trening modelu (w którym następuje memoryzacja, wykraczająca zdaniem sądu poza zakres wyjątku) oraz użytkowanie modelu. To rozumowanie nie uwzględnia, że proces uczenia maszynowego następuje stopniowo, przy czym w początkowych fazach model może przede wszystkim memoryzować dane, które dopiero wraz z przetworzeniem kolejnych przykładów zastępowane są zgeneralizowanymi wzorcami (tzw. *grokking*)³². Skoro w pewnej fazie trenowania model przede wszystkim memoryzuje dane, interpretacja przyjęta przez sąd prowadziłaby do niemożliwości zastosowania wyjątku TDM do procesu uczenia maszynowego w ogóle. Przyjmując założenie, że celem prawodawcy unijnego było objęcie tym wyjątkiem procesu uczenia maszynowego – co potwierdza brzmienie aktu w sprawie sztucznej inteligencji czy wspomniany komunikat Komisji – taka wykładnia uniemożliwiałaby zastosowanie wyjątku zgodnie z jego funkcją. Naruszałoby to więc wynikający z orzecznictwa TS wymóg interpretacji wyjątków w sposób zapewniający ich *effet utile*, zgodnie z którym wykładnia warunków wyjątku musi „chronić jego skuteczność i cel”³³.

Kontrowersyjną kwestią jest również zastosowanie przewidzianego w art. 5 ust. 3 lit. i dyrektywy 2001/29/WE wyjątku zezwalającego na „niezamierzone włączenie” utworu do „innego nośnika”. LG München odrzucił możliwość zastosowania jego odpowiednika w UrhG (§ 57) zezwalającego na zwielokrotnianie i rozpowszechnianie cudzego utworu, jeżeli stanowi on „nieistotny dodatek” do „właściwego przedmiotu”. Powołując się na orzecznictwo BGH uznał, że znajduje on zastosowanie tylko, gdy istnieje jakiś główny przedmiot wykorzystania sam w sobie będący utworem³⁴. Cech tych nie spełnia korpus treningowy Common Crawl wykorzystywany przez pozwane spółki w ramach praktyki *web scraping*. W odniesieniu do korpusu treningowego ocena ta wydaje się słuszna, choć z innych powodów niż wskazane przez sąd. Włączenie utworów do korpusu jest przecież celowe i systematyczne, skoro *web scraping* ma na celu właśnie gromadzenie możliwie dużych ilości danych. Nic jednak nie stoi na przeszkodzie, aby za „inny nośnik” uznać wytrenowany model AI. Przy takim założeniu na gruncie prawa unijnego można bronić stanowiska, wedle którego memoryzacja znajduje podstawę w omawianym wyjątku, skoro prawo

to nie wymaga, aby ów „inny nośnik” był utworem³⁵. Kwestia ta ma jednak znaczenie tylko dla ewentualnej oceny zgodności dostawcy modelu z wymogami określonymi w art. 53 ust. 1 lit. c AIA (zgodność z prawem autorskim Unii). Ze względu na fakultatywny charakter katalogu z art. 5 ust. 2 i 3 dyrektywy 2001/29/WE proponowana wykładnia nie rozstrzyga o zakresie zastosowania analogicznego wyjątku w prawie krajowym³⁶.

4. Konsekwencje w zakresie środków ochrony prawnej

Uznanie modelu za kopię utworów wykorzystanych w procesie uczenia maszynowego prowadzi do pewnych trudności w sferze stosowania środków ochrony praw wyłącznych, szczególnie roszczenia o zaniechanie i roszczeń prewencyjno-restytucyjnych, w tym zniszczenia bezprawnie wytworzonych kopii. Wynikają one ze specyfiki technologii generatywnej AI. W przypadku tradycyjnych kopii cyfrowych, np. plików MP3 zapisanych na karcie pamięci odtwarzacza, aby przywrócić stan zgody z prawem, wystarczy usunąć nielegalne pliki, co pozostaje bez wpływu na pracę urządzenia oraz pozostałe zapisane w nim informacje. W przypadku sieci neuronowej nie istnieje analogiczna możliwość „wymazania” konkretnych wag w celu usunięcia zmemoryzowanego utworu. Pomijając samą niemożność ustalenia, w których wagach zapisane są dane utwory, sieciowa natura modelu sprawia, że próba bezpośredniej ingerencji w pojedyncze wagi prowadziłaby do jego uszkodzenia³⁷. Powstaje zatem pytanie, w jaki sposób naruszciciel ma wykonać sądowy nakaz zaniechania dalszych naruszeń albo usunięcia ich skutków.

Przyjmując konsekwentnie rozumowanie, że model jest kopią zmemoryzowanych utworów, należałoby stwierdzić, że do wykonania nakazu sądu nie wystarczy zastosowanie na wyjściu z modelu filtrów blokujących określone treści. Filtr nie usuwa utworu z modelu ani nie prowadzi do zaniechania jego zwielokrotniania razem z modelem, a jedynie usuwa możliwość „wywołania” utworu przez użytkownika. Co prawda w przypadku tzw. modeli zamkniętych, tj. znajdujących się pod pełną kontrolą dostawcy, przestaje być możliwe udowodnienie, że utwory są nadal „w modelu”, ale literalnie nakaz zaniechania naruszeń taki, jak sformułowany przez monachijski sąd, nie zostanie wykonany. Niewątpliwie zaniechanie naruszeń i przywrócenie stanu zgodnego z prawem mogłoby nastąpić przez wycofanie całego modelu i ponowne przeprowadzenie procesu uczenia maszynowego. To rozwiązanie wydaje się jednak nadmiernie dolegliwe dla pozwanego, szczególnie jeżeli uwzględnić rangę naruszeń w omawianej sprawie i okoliczność, że koszt trenowania modeli serii gpt-4 szacuje się na dziesiątki milionów dolarów³⁸.

Z prawnego punktu widzenia najlepszym rozwiązaniem byłoby „odtrenowanie” sieci neuronowej (*machine unlearning*). Ale tu na przeszkodzie stają z kolei ograniczenia techniczne, gdyż dostępne obecnie metody są czasochłonne, zawodne i prowadzą

32 A. Power, Y. Burda, H. Edwards, I. Babuschkin, V. Misra, *Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets*, arXiv 2022, <http://arxiv.org/abs/2201.02177> (dostęp: 6.11.2025 r.). Autorzy wykazali, że sieci neuronowe mogą osiągać generalizację dopiero znacznie po osiągnięciu pełnego dopasowania do danych treningowych (*overfitting*), przez długi okres pozostając w fazie memoryzacji.

33 Zob. zwłaszcza wyrok TS z 29.07.2019 r., C-516/17, Spiegel Online GmbH przeciwko Volker Beck, EU:C:2019:625, pkt 36, oraz wyrok C-469/17, Funke Medien, pkt 51. Przyjęta wykładnia wyjątku nie może więc eksplorację tekstu i danych procesem niemożliwym do przeprowadzenia. Por. co do samego pojęcia eksploracji tekstu i danych E. Rosati, *Copyright...*, s. 37 i n.

34 Chodzi tu o wyrok BGH z 17.11.2014, I ZR 177/13 – Möbelkatalog. BGH faktycznie wskazał na konieczność wąskiej interpretacji nośnika, jednak nie rozstrzygał, czy włączenie ma nastąpić do przedmiotu spełniającego przesłanki ochrony prawnoautorskiej. Zob. szerzej F.L. Stang, *Enge Auslegung für Qualifizierung eines Werkes als „unwesentliches Beiwerk” – Möbelkatalog*, „GRUR. Gewerblicher Rechtsschutz und Urheberrecht 2015/7, s. 671.

35 Tak S. von Lewinski [w:] *European Copyright Law: A Commentary*, red. M. Walter, S. von Lewinski, Oxford 2010, s. 1053.

36 Zob. krytyczne uwagi Ch. Geiger, F. Schönherr [w:] I. Stamatoudi, P. Torremans, *EU Copyright Law*, Cheltenham–Northampton 2021, s. 319.

37 Zwracają na to uwagę R. Konertz, R. Schönhof, *Rechtsfolgen der Urheberrechtsverletzung bei generativer Künstlicher Intelligenz: Über die Möglichkeit des „Ver-gessens” in Neuronalen Netzen*, „Wettbewerb in Recht und Praxis” 2024/5, s. 539.

38 Szerzej B. Cottier, R. Rahman, L. Fattorini, N. Maslej, T. Besiroglu, D. Owen, *The rising costs of training frontier AI models*, arXiv 2025, <http://arxiv.org/abs/2405.21015> (dostęp: 14.11.2025 r.).

do obniżenia jakości modelu³⁹. Co więcej – jak trafnie zwraca się uwagę – skoro w procesie trenowania nie rejestruje się wpływu pojedynczych danych treningowych na wagi modelu, to nie jest technicznie możliwe ustalenie, czy „odtrenowanie” rzeczywiście doprowadziło do usunięcia z modelu tych danych⁴⁰.

Trudności w zastosowaniu środków ochrony prawnej oraz ich dolegliwość i niedostosowanie do specyfiki technologii są jednak prostą konsekwencją uznania memoryzacji za zwielokrotnianie oraz pozostawienia jej poza zakresem ograniczeń i wyjątków praw autorskich. Jak się wydaje, bardziej efektywne byłoby przeniesienie ciężaru oceny z samego modelu na generowane przez model treści. Zdaniem P. Mezei tendencją taką można zaobserwować w orzecznictwie amerykańskim⁴¹. W sprawach Bartz przeciwko Anthropic oraz Kadrey przeciwko Meta sądy uznały bowiem za wystarczające zastosowanie przez pozwanych mechanizmów filtracji wyników (*post-training mitigations*), traktując je jako element oceny transformatywnego charakteru użycia w ramach doktryny *fair use*⁴².

5. Wnioski końcowe

W prawie amerykańskim położenie nacisku na treści generowane przez model jest możliwe ze względu na elastyczną konstrukcję doktryny *fair use*, pozwalającą na uwzględnienie całości okoliczności danego przypadku, w tym celu i charakteru użycia, istotności i ilości wykorzystanej treści czy wreszcie wpływu na rynek utworu. Europejskie wyjątki i ograniczenia praw autorskich elastyczności tej są pozbawione. Poszukiwanie właściwego wyważenia interesów podmiotów uprawnionych i dostawców modeli AI wymagałoby więc raczej ponownego spojrzenia na treść prawa zwielokrotniania. Rysują się tu dwa rozwiązania.

Po pierwsze, w ślad za M. Leistnerem i L. Antoine można twierdzić, że ze względu na rozproszony charakter zapisu w modelu wpływ efektów probabilistycznych oraz działań użytkownika, kopia relewantną z punktu widzenia treści prawa zwielokrotniania jest dopiero kopia generowana jako wynik działania modelu. W tym ujęciu zapis utworu w parametrach modelu, choć konieczny, aby doszło do takiej generacji, ze względu na swoją niepercepcyjność i nieokreśloność pozostawałby poza zakresem tego prawa wyłącznego. Po drugie – i to rozwiązanie zdaje się oferować więcej możliwości równoważenia interesów – można rozważać pewną „subiektywizację” pojęcia zwielokrotniania, w sposób analogiczny do tego, w jaki TS uczynił to w odniesieniu do prawa publicznego udostępniania utworów⁴³ oraz prawa *sui generis*

do baz danych⁴⁴, tj. przez sformułowanie dodatkowych kryteriów, których spełnienie jest konieczne, aby uznać, że doszło do ingerencji w dane prawo wyłączne. W analizowanym przypadku mogłoby to polegać na objęciu prawem zwielokrotniania tylko tych modeli, które intencjonalnie odtwarzają utwory wykorzystane do trenowania lub w przypadku których dostawca nie podjął wiarygodnych działań, żeby takiemu odtwarzaniu zapobiec.

Zaakceptowanie wykładni zgodnej z podejściem LG München, tj. opartej na założeniu, że memoryzacja stanowi zwielokrotnienie objęte treścią praw wyłącznych, do którego nie znajdują zastosowania wyjątki i ograniczenia praw autorskich, ma trudne do przyjęcia konsekwencje. Najbardziej oczywiste dotyczą relacji pomiędzy uprawnionymi a dostawcami modeli. Skoro dostawca nie może w pełni wyeliminować memoryzacji i zjawisko to dotyczy każdego modelu, to potencjalnie każdy model jest narażony na roszczenia uprawnionych. W takim ujęciu wykorzystanie treści na podstawie wyjątku dotyczącego eksploracji tekstu i danych wiązać się musi z niemożliwym do wyeliminowania ryzykiem, co zwiększa presję w kierunku wykorzystywania w procesie uczenia maszynowego wyłącznie treści licencjonowanych. Trudno nie zauważyć, że w ten sposób dochodzi do podważenia skuteczności wyjątków z art. 3 i 4 dyrektywy 2019/790.

Ponadto, skoro jedynym środkiem dowodowym pozwalającym na ustalenie, że doszło do memoryzacji, jest użycie odpowiednich promptów i analiza odpowiedzi zwracanych przez model, to uznanie memoryzacji za formę zwielokrotniania w uprzywilejowanej pozycji stawia dostawców modeli zamkniętych, którzy mogą skutecznie blokować jej wykrycie przez filtry wejścia i wyjścia. Możliwość takiej pozbawieni są dostawcy modeli otwartych, których wagi są publicznie dostępne. Dla Europy, nieposiadającej własnych znaczących modeli zamkniętych, oznacza to pogłębienie uzależnienia od amerykańskich dostawców i paradoksalne wzmocnienie pozycji konkurencyjnej rozwiązań zamkniętych kosztem otwartego ekosystemu, który mógłby stanowić podstawę cyfrowej suwerenności Europy.

Wreszcie problem memoryzacji uwidacznia niesatysfakcjonujący poziom harmonizacji wyjątków i ograniczeń prawa autorskiego z art. 5 dyrektywy 2001/29/WE. O ile na gruncie prawa unijnego uzasadnieniem memoryzacji może być wyjątek dotyczący „przypadkowego włączenia” utworu do innego nośnika, o tyle w przepisach państw członkowskich może być on nieobecny lub interpretowany bardziej restrykcyjnie. W rezultacie dostawca modelu, który spełnia wymogi unijnego prawa autorskiego – a tym samym zachowuje zgodność z aktem w sprawie sztucznej inteligencji – może zostać zmuszony do rezygnacji z infrastruktury zlokalizowanej w państwach o bardziej restrykcyjnych przepisach krajowych. Model trafi wówczas na rynek UE, ale z wykorzystaniem serwerów znajdujących się poza Unią. Zniechęca to dostawców do budowy europejskiej infrastruktury AI i pogłębia zależność od centrów przetwarzania danych zlokalizowanych poza UE.

39 J. Xu, Z. Wu, C. Wang, X. Jia, *Machine Unlearning: Solutions and Challenges*, „IEEE Transactions on Emerging Topics in Computational Intelligence” 2024/3.

40 A. Sesing-Wagenpfeil, *Trainierte KI-Modelle...*, s. 261–262; R. Konertz, R. Schönhof, *Rechtsfolgen der Urheberrechtsverletzung...*, s. 540.

41 P. Mezei, *Memorization and Generative AI...*, s. 3.

42 Zob. Andrea Bartz et al. przeciwko Anthropic PBC, No. 3:24-cv-05417 (N.D. Cal. 23.06.2025 r.), Richard Kadrey et al. przeciwko Meta Platforms, Inc., No. 3:23-cv-03417 (N.D. Cal. 25.06.2025 r.).

43 W odniesieniu do prawa publicznego komunikowania z art. 3 dyrektywy 2001/29/WE Trybunał rozwinął linię orzecniczą, zgodnie z którą uznanie, że doszło do ingerencji w to prawo jest uzależnione od spełnienia dodatkowych kryteriów, takich jak istnienie nowej publiczności, odmienność środka technicznego, zarobkowy charakter korzystania czy wiedza (lub jej brak) o bezprawności treści. Zob. wyrok TS z 13.02.2014 r., C-466/12, Nils Svensson i inni przeciwko Retriever Sverige AB, EU:C:2014:76; postanowienie TS z 21.10.2014 r., C-348/13, BestWater International GmbH przeciwko Michael Mebes i Stefan Potsch, EU:C:2014:2315;

wyrok TS z 8.09.2016 r., C-160/15, GS Media BV przeciwko Sanoma Media Netherlands BV i inni, EU:C:2016:644 oraz wyrok TS z 9.03.2021 r., C-392/19, VG Bild-Kunst przeciwko Stiftung Preußischer Kulturbesitz, EU:C:2021:181.

44 Zob. wyrok TS z 3.06.2021 r., C-762/19, CV-Online Latvia SIA przeciwko Melons SIA, EU:C:2021:434, pkt 47, w którym TS uznał, że pobieraniem i wtórnym wykorzystywaniem danych są tylko te czynności, które przynoszą szkodę inwestycji producenta bazy, co pozwala wyłączyć z zakresu tego prawa czynności niewpływające na zwrot z tej inwestycji.

Abstract

dr Zbigniew Okoń

Legal advisor, partner in the Lubasz i Wspólnicy sp.k. law office, Poland (ORCID: <https://orcid.org/0000-0002-0999-780X>).**Memorisation of Works by AI Models in the Light of EU Copyright Law****Keywords:** memorisation, artificial intelligence, copyright, text and data mining (TDM), Directive 2001/29/EC

The rapid development of artificial intelligence has resulted in an unprecedented number of court disputes between copyright holders and technology companies. However, first court rulings in the United States have confirmed that AI model training on data obtained from the Internet without the right holders' consent remains within fair use boundaries. In Europe, such a practice may be justified by the exception for text and data mining under Articles 3 and 4 of Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, which – despite scepticism of some researchers – is confirmed by judgments, provisions of Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 or Commission documents. However, the existence of a legal basis for copyright-encroaching machine learning activities does not address all issues, since the question arises as to how to assess a situation where a model “remembers” (memorises) data from the training data corpus and these data appear in the model's outputs.

The recent judgment of the national court in Munich, *GEMA v OpenAI*, which found that the memorisation of parts of lyrics by the model and reproducing them in response to user prompts constituted an infringement of copyrights, pertains to this issue. As memorisation is a process that – as suggested by IT and AI research – cannot be entirely prevented, this classification significantly changes the legal situation of AI model providers, opening the door to claims directed at the models themselves. It also seems to have a negative impact on the development of the European technical AI infrastructure and increase the risk of sharing and using open-source models. In the article, the author presents the court's reasoning, analyses the adequacy of the underlying opinions and attempts to identify their consequences. He also indicates a need for finding legal solutions that would better balance the interests of model providers and right holders.

Bibliografia/References

Bracha O., *The Work of Copyright in the Age of Machine Production*, „Harvard Journal of Law & Technology” 2024/1.
Cooper A.F., Gokaslan A. i in., *Extracting memorized pieces of (copyrighted) books from open-weight language models*, arXiv 2025, <http://arxiv.org/abs/2505.12546> (dostęp: 2.12.2025 r.).
Cooper A.F., Grimmelmann J., *The Files are in Computer: on Copyright, Memorization, and Generative AI*, arXiv 2024, <http://arxiv.org/abs/2404.12590> (dostęp: 28.02.2025 r.).
Cooper A.F., Lee K., Grimmelmann J., Ippolito D. i in., *Report of the 1st Workshop on Generative AI and Law*, Social

Science Research Network 2023, <https://papers.ssrn.com/abstract=4634513> (dostęp: 30.11.2025 r.).

Cottier B., Rahman R., Fattorini L., Maslej N., Besiroglu T., Owen D., *The rising costs of training frontier AI models*, arXiv 2025, <http://arxiv.org/abs/2405.21015> (dostęp: 14.11.2025 r.).
Dornis T.W., *Urheberrecht und Training generativer KI-Modelle*, Nomos Verlagsgesellschaft mbH & Co. KG 2024.

European Copyright Law: A Commentary, red. M. Walter, S. von Lewinski, Oxford 2010.

Geiger C., Frosio G., Bulayenko O., *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Legal Aspects*, „SSRN Electronic Journal” 2018.

Hense P., *Overfitting*, „MMR. Multimedia und Recht” 2024/6.
Konertz R., Schönhof R., *Rechtsfolgen der Urheberrechtsverletzung bei generativer Künstlicher Intelligenz: Über die Möglichkeit des „Vergessens“ in Neuronalen Netzen*, „Wettbewerb in Recht und Praxis” 2024/5.

Lee K., Ippolito D., Nystrom A., Zhang C., Eck D., Callison-Burch C., Carlini N., *Deduplicating Training Data Makes Language Models Better*, arXiv 2022, <http://arxiv.org/abs/2107.06499> (dostęp: 4.12.2025 r.).

Leistner M., Antoine L., *TDM and AI Training in the European Union – From ‘LAION’ to Possible Ways Ahead?*, „GRUR International” 2025/11.

Margoni T., Kretschmer M., *A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology*, „GRUR International” 2022/8.

Mezei P., *Memorization and Generative AI – A Persistent Issue with Copyright Consequences?*, Social Science Research Network 2025, <https://papers.ssrn.com/abstract=5404367> (dostęp: 1.12.2025 r.).

Pesch P.J., Böhme R., *Artpocalypse now? – Generative KI und die Vervielfältigung von Trainingsbildern*, „GRUR. Gewerblicher Rechtsschutz und Urheberrecht” 2023/14.

Power A., Burda Y., Edwards H., Babuschkin I., Misra V., *Grokking: Generalization Beyond Overfitting on Small Algorithmic Datasets*, arXiv 2022, <http://arxiv.org/abs/2201.02177> (dostęp: 6.11.2025 r.).

Quintais J.P., *Generative AI, copyright and the AI Act*, „Computer Law & Security Review” 2025/56.

Rosati E., *Copyright in the digital single market: article-by-article commentary to the provisions of directive 2019/790*, New York 2021.

Sesing-Wagenpfeil A., *Trainierte KI-Modelle als Vervielfältigungsstücke im Sinne des Urheberrechts*, „Zeitschrift für geistiges Eigentum” 2024/2.

Stamatoudi I., Torremans P., *EU Copyright Law*, Cheltenham–Northampton 2021.

Stang F.L., *Enge Auslegung für Qualifizierung eines Werkes als „unwesentliches Beiwerk” – Möbelkatalog*, „GRUR. Gewerblicher Rechtsschutz und Urheberrecht” 2015/7.

Stieper, M., Denga, M., *Die Reichweite des EU-Urheberrechts nach der KI-VO*, „GRUR. Gewerblicher Rechtsschutz und Urheberrecht” 2024/20.

Veale M., Quintais J.P., *The Obligations of Providers of General-Purpose AI Models*, Social Science Research Network 2025, <https://papers.ssrn.com/abstract=5744602> (dostęp: 18.11.2025 r.).

Welser M. von, *Generative KI und Open Source Software*, „GRUR. Gewerblicher Rechtsschutz und Urheberrecht” 2024/19.

Xu J., Wu Z., Wang C., Jia X., *Machine Unlearning: Solutions and Challenges*, „IEEE Transactions on Emerging Topics in Computational Intelligence” 2024/3.