



Warszawa, 28 lutego 2024

Pan
Andrzej Wyrobiec
Podsekretarz Stanu
Ministerstwo Kultury i Dziedzictwa Narodowego

Szanowny Panie Ministrze,

Przedstawiamy niniejsze stanowisko w odpowiedzi na opublikowany 15 lutego 2024 r. projekt ustawy o zmianie ustawy o prawie autorskim i prawach pokrewnych oraz niektórych innych ustaw.

Nasza organizacja jest europejskim think tankiem zajmującym się cyfrowymi dobrami wspólnymi. W ramach naszej działalności uczestniczymy w debatach politycznych i pracach legislacyjnych dotyczących praw autorskich i technologii cyfrowych, zarówno na poziomie europejskim jak i w państwach członkowskich. Wraz z naszą partnerską organizacją COMMUNIA, Europejskim Stowarzyszeniem na rzecz Domeny Publicznej, aktywnie uczestniczyliśmy w debacie legislacyjnej na temat Dyrektywy 2019/790 w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym, oraz jej wdrażania w państwach członkowskich UE.

W niniejszym stanowisku chcemy skupić się na artykułach związanych z wdrożeniem wyjątków dotyczących eksploracji tekstu i danych (TDM).

W kwestii uzasadnienia wyłączenia tworzenia generatywnych modeli sztucznej inteligencji z zakresu wyjątku TDM

Uzasadnienie do projektu ustawy zawiera następujący zapis dotyczący decyzji o wyłączeniu wykorzystania utworów do tworzenia generatywnych modeli sztucznej inteligencji z zakresu wyjątków dotyczących eksploracji tekstu i danych:

Wdrożenie dyrektywy obecnie, w roku 2024, nakazuje odnieść się w tym miejscu do problematyki sztucznej inteligencji i kwestii, czy eksploracja tekstów i danych w rozumieniu

dyrektywy obejmuje również możliwość zwielokrotniania utworów na potrzeby uczenia maszynowego. Bez wątplenia w momencie przyjmowania dyrektywy w 2019 r. Możliwości sztucznej inteligencji nie były tak rozpoznawalne jak obecnie, gdy przy pomocy tej technologii zaczynają powstawać „utwory” o wartości artystycznej i komercyjnej porównywalnej z prawdziwymi utworami, tj. stworzonymi przez człowieka. Wydaje się zatem, że słuszne jest założenie, że ten rodzaj dozwolonego użytku nie był pomyślany na potrzeby sztucznej inteligencji. Zostaje zatem wprowadzone wyraźne doprecyzowanie, że zwielokrotnianie utworów w ramach eksploracji tekstów i danych nie może służyć tworzeniu generatywnych modeli sztucznej inteligencji.

Zgadza się, że pojawienie się w ciągu ostatnich dwóch lat generatywnych modeli sztucznej inteligencji jako produktów komercyjnych tworzy potrzebę wyjaśnienia jak unijne przepisy dotyczące praw autorskich regulują wykorzystywanie utworów chronionych prawem autorskim do tworzenia generatywnych modeli sztucznej inteligencji. Dostępne dowody wskazują jednak, że stan faktyczny jest inny niż założenia zawarte w cytowanym powyżej uzasadnieniu. Chociaż istnieje niewiele publicznie dostępnych dokumentów na temat tego, co ustawodawcy mieli na myśli, gdy uzgadniano strukturę wyjątków dotyczących eksploracji tekstu i danych, to z tego, co jest dostępne, jasno wynika, że rozwój sztucznej inteligencji był wprost uwzględniony. Zarówno [oświadczenie Parlamentu Europejskiego](#), jak i [wyjaśnienie Komisji Europejskiej dotyczące dyrektywy](#), opublikowane po przyjęciu dyrektywy w marcu 2019 r., wyraźnie stwierdzają, że wyjątek w art. 4 został wprowadzony "w celu przyczynienia się do rozwoju analizy danych i sztucznej inteligencji".

W marcu 2023 r. (kiedy wpływ generatywnej sztucznej inteligencji był już szerzej rozumiany), Komisja Europejska doprecyzowała te kwestie. W odpowiedzi na [pytanie parlamentarne](#),¹ które sugerowało, że "dyrektywa [CDSM] nie odnosi się do tej konkretnej kwestii", [Komisarz Breton wskazał](#), że wyjątki TDM w rzeczywistości "zapewniają równowagę między ochroną podmiotów praw autorskich, w tym artystów, a ułatwianiem TDM, w tym przez twórców sztucznej inteligencji".²

Procedowane ostatnio [rozporządzenie w sprawie sztucznej inteligencji](#)³ zawiera przepis, który wskazuje, że twórcy generatywnych systemów sztucznej inteligencji muszą "wdrożyć politykę przestrzegania unijnego prawa autorskiego, w szczególności w celu identyfikacji i poszanowania, w tym za pomocą najnowocześniejszych technologii, zastrzeżeń praw wyrażonych zgodnie z art. 4 ust. 3 dyrektywy [CDSM]". Ponadto rozporządzenie to zawiera również motyw (60i), który wyjaśnia związek między szkoleniem systemów generatywnej

¹ https://www.europarl.europa.eu/doceo/document/E-9-2023-000479_EN.html

² https://www.europarl.europa.eu/doceo/document/E-9-2023-000479-ASW_EN.html

³

https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/CJ40/AG/2024/02-13/1296003EN.pdf

sztucznej inteligencji a wyjątkami określonymi w art. 3 i 4 dyrektywy o prawie autorskim. Motyw ten wyjaśnia, że prawodawcy unijni uważają, że wykorzystanie utworów do szkolenia generatywnych modeli AI wchodzi w zakres tych wyjątków.

Zgodność proponowanego wdrożenia z dyrektywą CDSM

Wyłączenie zwielokrotnień dokonywanych w kontekście szkolenia generatywnych modeli AI z zakresu wyjątków TDM zaproponowanych w art. 262 i art. 263 skutkowałoby wadliwym wdrożeniem dyrektywy do polskiego prawa.

Termin eksploracja tekstów i danych został zdefiniowany w art. 2 ust. 2 dyrektywy jako "zautomatyzowaną technikę analityczną służącą do analizowania tekstów i danych w postaci cyfrowej w celu wygenerowania informacji, obejmujących między innymi wzorce, tendencje i korelacje". Należy go uznać za autonomiczne pojęcie prawa UE, które nie może być modyfikowane przez państwa członkowskie w przepisach wdrażających dyrektywę. Takie działanie podważałoby sam cel dyrektywy, którym jest zapewnienie harmonizacji wyjątków i ograniczeń we wszystkich państwach członkowskich.

W tym kontekście należy podkreślić, że w przypadku przyjęcia obecnych propozycji implementacyjnych, Polska byłaby jedynym państwem członkowskim ze zmodyfikowanym zakresem wyjątków dotyczących eksploracji tekstów i danych. Istnieje powszechna zgoda co do tego, że w oparciu o definicję zawartą w dyrektywie, pojęcie eksploracji tekstów i danych obejmuje reprodukcje dokonywane w celu trenowania generatywnych modeli sztucznej inteligencji.

Aby uniknąć uznania polskiej implementacji za niezgodną z Dyrektywą, z tekstu projektowanej ustawy implementującej należy usunąć następujące fragmenty: "z wyłączeniem tworzenia generatywnych modeli sztucznej inteligencji" (art. 26(2)) oraz "z wyłączeniem tworzenia generatywnych modeli sztucznej inteligencji" (art. 26(3)).

Zawężony wyjątek TDM, jeśli zostanie przyjęty w polskim prawie, będzie stanowił przeszkodę dla rozwoju generatywnych rozwiązań sztucznej inteligencji w Polsce. Polska "[Polityka dla rozwoju sztucznej inteligencji w Polsce od roku 2020](https://www.gov.pl/web/ai/polityka-dla-rozwoju-sztucznej-inteligencji-w-polsce-od-roku-2020)"⁴ definiuje szereg celów, które zostaną utrudnione przez wąski wyjątek TDM. Są to m.in. zwiększenie podaży polskich rozwiązań AI, wzmocnienie polskich badań nad AI, usunięcie przeszkód prawnych dla rozwoju AI oraz stworzenie rozwiązań odpowiadających na specyficzne wyzwania stojące przed Polską (np. tworzenie polskich modeli językowych). Ostatni punkt jest szczególnie wart odnotowania, ponieważ wymienione wyzwania dotyczą również kwestii związanych z polską kulturą i dziedzictwem cyfrowym - takich jak tworzenie generatywnej sztucznej inteligencji w języku polskim oraz badanie jej potencjalnej wartości i wpływu kulturowego czy społecznego. W 2023 r. konsorcjum polskich instytucji badawczych rozpoczęło prace

⁴ <https://www.gov.pl/web/ai/polityka-dla-rozwoju-sztucznej-inteligencji-w-polsce-od-roku-2020>

nad PLLUM, niekomercyjnym modelem językowym o otwartym kodzie źródłowym. Dalsze prace nad tym modelem staną się niemożliwe, jeśli Polska przyjmie wyjątek TDM w proponowanej, wąskiej formie.

W tym kontekście szczególnie kontrowersyjna jest propozycja, zawarta w art. 26(2), aby poza zakresem wyjątku TDM znalazły się również niekomercyjne badania prowadzone przez instytucje badawcze i instytucje dziedzictwa kulturowego. Badania nad generatywną sztuczną inteligencją są kluczowym obszarem badań z zakresu szeroko pojętej informatyki, które polski rząd powinien wspierać.

Istnieje oczywista potrzeba wyważenia różnych interesów w kontekście rozwoju generatywnej sztucznej inteligencji: interesów podmiotów budujących systemy sztucznej inteligencji generatywnej, interesów twórców i innych właścicieli praw, których utwory są wykorzystywane w szkoleniu modeli AI, oraz interesu publicznego. Podejście zaproponowane w dyrektywie i potwierdzone w rozporządzeniu w sprawie sztucznej inteligencji zapewnia tę równowagę. Wyjątek dla eksploracji tekstów i danych wprowadzony w art. 4 Dyrektywy stanowi nowatorskie podejście oparte na zastrzeżeniach praw (opt-out). Pozwala to tym podmiotom praw autorskich, którzy chcą zastrzec swoje prawa, by to uczynili - i potencjalnie oferowali opcje licencjonowania swoich utworów - jednocześnie umożliwiając badania TDM i szkolenia w zakresie sztucznej inteligencji na szerokiej puli publicznie dostępnych utworów.

Zgodność z przepisami o prawie autorskim zawartymi w przygotowywanym rozporządzeniu w sprawie sztucznej inteligencji

Jak wskazano powyżej, przygotowywane rozporządzenie o sztucznej inteligencji, które będzie bezpośrednio stosowane w Polsce, zawiera szereg przepisów mających na celu zapewnienie zgodności z unijnymi ramami prawa autorskiego dostawców modeli sztucznej inteligencji ogólnego przeznaczenia (general purpose AI), w tym generatywnych modeli sztucznej inteligencji). Opierają się one na założeniu, że wykorzystywanie utworów chronionych prawem autorskim do tworzenia takich modeli mieści się w zakresie wyjątków określonych w art. 3 i 4 dyrektywy o prawie autorskim. Wdrożenie w Polsce dyrektywy w sposób, który jest bezpośrednio sprzeczny z założeniami leżącymi u podstaw odpowiednich części rozporządzenia o sztucznej inteligencji, spowoduje znaczną, dodatkową niepewność prawną.

Wyzwania związane z praktycznym zastosowaniem art. 4 ust. 3 dyrektywy CDSM

Wyjątki TDM w dyrektywie o prawie autorskim zostały opracowane w celu zapewnienia starannie opracowanej równowagi między interesami podmiotów praw autorskich a interesem ogółu społeczeństwa w zakresie korzystania z coraz większej ilości publicznie dostępnych informacji i kultury za pomocą metod analizy obliczeniowej i uczenia maszynowego (w tym między innymi obecnej fali technologii AI). Równowaga ta składa się z czterech elementów:

1. Oba wyjątki TDM mają zastosowanie wyłącznie do utworów legalnie dostępnych;
2. TDM do celów badań naukowych jest dozwolone bez ograniczeń, zapewniając uprzywilejowaną pozycję publicznym instytucjom badawczym;
3. Właściciele praw mogą zastrzec sobie prawo do zezwalania na eksplorację tekstów i danych w celach innych niż badania naukowe, co pozwala im kontrolować, czy i w jaki sposób ich utwory mogą być wykorzystywane;
4. Wszystkie inne utwory mogą być wykorzystywane do celów eksploracji tekstów i danych przez każdego i w dowolnym celu.

System ten zapewnia, że utwory i inne treści chronione prawem autorskim, które nie są aktywnie zarządzane przez posiadaczy praw, pozostają dostępne do użytku TDM.

Mechanizm opt-out został wyraźnie zaprojektowany w celu zapobieżenia sytuacji podobnej do dzieł osieroconych, w której informacje i dzieła kultury pozostają niedostępne do wykorzystania, ponieważ podmiotom praw autorskich przyznano wyłączne prawa, z których te jednak nie zamierzają korzystać. Mechanizm opt out adresuje w ten sposób jedno z kluczowych niedociągnięć istniejących ram prawa autorskiego i jest również obecny w innych częściach dyrektywy.⁵

Istnieje obecnie duża niepewność co do tego, jak zastrzeżenie praw zgodnie z art. 4 ust. 3 dyrektywy ma działać w praktyce. Szereg podmiotów tworzących generatywne modele sztucznej inteligencji zaczęło oferować możliwość wyłączenia treści z tworzonych baz danych szkoleniowych. Nie ma jednak obecnie ogólnie uznanych standardów ani protokołów dotyczących wyrażania zastrzeżeń w formie odczytywalnej maszynowo.

W tej kwestii istnieje pilna potrzeba, aby Unia Europejska i jej państwa członkowskie zapewniły znormalizowane ramy dla uczynienia takich klauzul opt-out bardziej dostępnymi, skutecznymi i skalowanymi zarówno dla posiadaczy praw, jak i firm tworzących modele sztucznej inteligencji. Taka interwencja, w połączeniu z obowiązkami wprowadzonymi w

⁵ Przykładowo art. 8 ust. 4 dyrektywy umożliwia posiadaczom praw wyłączenie się z mechanizmu zezwalającego instytucjom dziedzictwa kulturowego na korzystanie z utworów i innych treści objętych ochroną, będących poza obrotem handlowym.

rozporządzeniu o sztucznej inteligencji, zapewniłaby tym podmiotom praw autorskich, którzy chcą wyłączyć swoje utwory z tworzenia generatywnych modeli sztucznej inteligencji (lub licencjonować je w tym celu), taką samą ochronę prawną, jaką miałyby zapewnić proponowane przez polskiego ustawodawcę ogólne wyłączenie tworzenia generatywnej sztucznej inteligencji z zakresu wyjątków TDM. Jednocześnie takie rozwiązanie zapewnia dostęp na potrzeby badań naukowych, a także dostęp do tych utworów, w przypadku których ich twórcy i inne podmioty praw autorskich nie mają zastrzeżeń co do takich zastosowań.

Jak wzmocnić pozycję twórców i posiadaczy praw w ramach dyrektywy CDSM

System wprowadzony przez wyjątki TDM w dyrektywie pozwala twórcom wymagać wynagrodzenia za korzystanie z ich utworów w celu tworzenia generatywnych modeli sztucznej inteligencji, jeśli nie są to badania naukowe. Wymaga to od nich zastrzeżenie praw i zaoferowania licencji na korzystanie z utworów w tym celu. Biorąc pod uwagę skalę działań związanych z tworzeniem generatywnych modeli sztucznej inteligencji, takie licencje musiałyby być oferowane jako licencje zbiorowe przez organizacje reprezentujące szeroką grupę twórców (lub przez duże instytucjonalne podmioty praw, na przykład przez wydawców). Już teraz obserwujemy, że organizacje zbiorowego zarządzania w całej UE zastrzegają prawa w tym zakresie w imieniu swoich członków, tak aby móc następnie licencjonować treści. Takie formy dobrowolnego zbiorowego zarządzania, rozwijane na bazie mechanizmów opt-out, są odpowiednim sposobem adresowania kwestii interesów ekonomicznych podmiotów praw autorskich. Powinny one być wspierane poprzez tworzenie warunków dla przejrzystej i efektywnej dystrybucji przychodów z licencji.

Jednym z nowatorskich aspektów obecnej generacji generatywnych modeli sztucznej inteligencji jest fakt, że są one szkolone na bardzo dużej liczbie publicznie dostępnych utworów. Rodzi to pytania dotyczące zrównoważonego wykorzystywania publicznych zasobów informacyjnych, które są dziś wykorzystywane do tworzenia sztucznej inteligencji, i w przypadku których ich twórcy lub właściciele praw nie widzą potrzeby zastrzeżenia praw. Aby tę kwestię zaadresować, należy wprowadzić opłatę od przychodów dostawców generatywnych modeli sztucznej inteligencji szkolonych na publicznie dostępnych danych. Taki mechanizm finansowy gwarantowałby, że podmioty korzystające ze wspólnych i publicznie dostępnych zasobów wnoszą wkład w ich utrzymanie oraz w tworzenie w nowych zasobów. Przychody uzyskane z takiej opłaty powinny zasilać publiczny fundusz wspierający szeroki zakres inicjatyw wspierających rozwój wspólnej wiedzy i kultury. Beneficjentami takiego funduszu byłiby indywidualni artyści i twórcy, organizacje kulturalne, instytucje publiczne zarządzające cyfrowymi dobrami wspólnymi (na przykład biblioteki cyfrowe lub repozytoria czasopism o otwartym dostępie) lub inicjatywy obywatelskie (takie jak Wikipedia). Byłoby optymalnie, gdyby taki fundusz miał zasięg europejski, ale możliwy

jest również system funduszy krajowych. Dobrym punktem odniesienia dla tego podejścia jest Polski Instytut Sztuki Filmowej (PISF), który wspiera polską kinematografię dzięki opłatom pobieranym od podmiotów z szeroko rozumianego sektora filmowego. Innym przykładem funduszu publicznego mającego na celu wspieranie cyfrowych dóbr wspólnych jest niemiecki Sovereign Tech Fund. Należy zauważyć, że tego typu fundusz nie ma być formą wynagrodzenia dla indywidualnych twórców i podmiotów praw autorskich za wykorzystanie ich utworów do tworzenia generatywnej sztucznej inteligencji (cel ten zostałby osiągnięty poprzez opisany powyżej mechanizm opt-out). Zamiast tego ma on zapewnić wsparcie ekonomiczne dla dóbr kultury i wiedzy, rozumianych jako dobro wspólne.