

Iga Bałos

dr, Krakowska Akademia im. Andrzeja Frycza Modrzewskiego

ORCID: 0000-0001-8606-9719

ibalos@afm.edu.pl

Trenowanie wybranych modeli sztucznej inteligencji a uprawnienia twórców

Wprowadzenie

Coraz mniej obszarów życia pozostaje bez wsparcia rozwiązań wykorzystujących sztuczną inteligencję (SI). Jej zastosowanie przestaje zaskakiwać w kontekście usprawniania czynności powtarzalnych, zwłaszcza tych wykonywanych za pośrednictwem aplikacji w smartphonie. Zainteresowanie i kontrowersje budzą natomiast sytuacje, w których SI realizuje zadania jak dotąd swoiste jedynie dla człowieka.

W ostatnich latach opisano co najmniej kilka wytworów generowanych przez SI, które naśladują styl danego autora. Osoby zaangażowane w projekt *The Next Rembrandt*¹, zakończony wydrukowaniem obrazu ludzaco przypominającego portrety słynnego malarza, twierdziły, iż z łatwością dałyby się przekonać, że patrzą na oryginał. Idea przyświecająca projektowi to wykorzystanie tekstu i danych poza kontekstem biznesowym, w sposób „dotykający ludzką duszę”². Podsumowując uzyskany efekt, wyrażono opinię, że Rembrandt byłby zadowolony, mogąc podziwiać swój obraz, ale nie swojego autorstwa. Na razie żadna ze znanych form SI nie jest w stanie pomóc w zweryfikowaniu powyższej tezy.

Można jednak założyć, że wśród niektórych z obecnie tworzących autorów, zwłaszcza pisarzy, podobne inicjatywy, bazujące na ich utworach, nie znajdują aprobaty. Obejmuje to również życzenie, by takie praktyki nie

¹ Opis projektu, wraz z wizualizacją, dostępny na stronie: <https://www.nextrembrandt.com/> [dostęp: 31.10.2019].

² Cel wyznaczony przez reprezentanta jednego ze sponsorów projektu, Rona Augustusa (Microsoft).

następowały nawet po śmierci twórcy. Uzasadnieniem może być unikalna więź z poszczególnymi utworami, jak i z całym dorobkiem. Nie bez znaczenia pozostaje także motywacja natury finansowej. Celem artykułu jest analiza czy obecnie obowiązujące przepisy dają podstawę do efektywnego przeciwstawienia się trenowaniu SI twórczością jednego autora w celu uzyskania wytworu naśladowującego styl danej osoby.

Uważam, że komentowanie wyżej opisanej postawy, zrekonstruowanej na podstawie prywatnych rozmów z twórcami, stwierdzeniem, iż styl nie podlega ochronie autorsko-prawnej, wynika z powierzchownej analizy i niezrozumienia złożoności problemu. Po pierwsze, skala „inspiracji”, jaka ma miejsce w przypadku AI, wykracza często poza możliwości człowieka, także nieprzeciętnego. Po drugie, należy zwrócić uwagę nie tylko na sam efekt, ale także na zakres korzystania z utworu na poszczególnych etapach. Po trzecie, nie bez wpływu na konsekwencje prawne pozostanie także sposób i okoliczności, w jakich będzie czyniony użytek z opisanego powyżej wytworu³.

Przykłady trenowania wybranych modeli SI

Powszechnie dostępne materiały opisujące realizację projektu *The Next Rembrandt* nie dostarczają dokładnych informacji o rodzaju wykorzystanej SI. Z materiału filmowego oraz zwięzłego opisu wynika, iż dokonano analizy 346 prac Rembrandta, badając każde dzieło piksel po pikselu. W celu otrzymania danych, na których można było prowadzić powyższe czynności, skorzystano z takich materiałów jak skany 3D o wysokiej rozdzielczości oraz pliki cyfrowe, które zostały przeskalowane za pomocą algorytmów uczenia głębokiego. Pozwoliło to uzyskać możliwie najwyższą rozdzielczość i jakość obrazu. Utworzona w ten sposób, bardzo rozbudowana baza danych została wykorzystana do stworzenia „kolejnego dzieła” Rembrandta⁴. Zastosowane narzędzia zostały określone w literaturze jako procesy z pogranicza eksploracji danych (*data mining*) i uczenia maszynowego⁵. Ustalono, że portret powinien przedstawiać spoglądającego w prawo białego mężczyznę pomiędzy 30 a 40 rokiem życia, z zarostem, w czarnym ubraniu z białym kołnierzem. Kolejny etap to selekcja cech charakterystycznych dla stylu i techniki ma-

³ Roszczenia na gruncie ustawy o zwalczaniu nieuczciwej konkurencji nie zostały omówione w ramach niniejszego artykułu. W większości przypadków przysługiwałyby innym podmiotom, niż twórcy, np. wydawcom.

⁴ Opis stanowi tłumaczenie tekstu zamieszczonego na stronie: <https://www.nextrembrandt.com/> [dostęp: 31.10.2019].

⁵ M. Maggiore, *Artificial Intelligence, Computer Generated Works and Copyright*, [w:] *Non-Conventional Copyright: Do New and Atypical Works Deserve Protection?*, red. E. Bonadio, N. Lucchi, Edward Elgar Publishing, Chetlham–Northampton 2018, s. 383.

larza. Technologia 3D umożliwiła odwzorowanie struktury dzieła typowej dla Rembrandta. Ustalono liczbę warstw farby, stworzono mapę wysokości, uwzględniono typowe dla malarza ruchy pędzla.

Powyższy projekt został zrealizowany przy znaczącej pomocy sponsorów i partnerów biznesowych. Dzięki wsparciu finansowemu otrzymany rezultat jest imponujący. Sztuczna inteligencja jest także wykorzystywana przez pasjonatów tematu, którzy dzielą się, przede wszystkim w celach edukacyjnych, uzyskanymi efektami. Za pośrednictwem blogów i innych stron internetowych udostępniane są opisy przykładowego zastosowania sieci neuronowych do przewidywania tekstu. Wystarczająca ilość danych treningowych może być pozyskana z tekstu jednej książki⁶ lub około stu piosenek jednego artysty⁷. Oznacza to konieczność skopiowania i wprowadzenia bazy danych do programu. W przypadku książek wskazane jest wyeliminowanie zbędnego tekstu, co wpłynie na ograniczenie ilości potrzebnej pamięci. Przykładowo: rezygnuje się z pustych linii, informacji o wydawnictwie, wielokrotnych spacji, sprowadza się cały tekst do małych liter. Podczas treningu tekst będzie poddawany różnym operacjom, takich jak podziały, klasyfikacje, budowanie ciągów znaków.

Z kolei aby nauczyć sieci neuronowe wytwarzania chorałów w stylu J.S. Bacha, potrzebne było 352 utworów kompozytora. Bazę wzbogacono o transkrypcje dla różnych głosów, w ramach ustalonych wcześniej zakresów⁸.

Eksploracja tekstu i danych a prawa autorskie

Twórcy przedstawionych narzędzi nie skupiają się, co zrozumiałe, na aspektach prawno-autorskich czynności, jakim jest poddawany utwór. Sprowadza się go do danych i materiału wejściowego. Uważam, że takie podejście, mimo iż uzasadnione kontekstem, może stanowić skrót myślowy, mający konsekwencje przy analizowaniu zjawiska na gruncie prawa autorskiego. Nie ulega wątpliwości, iż informacje i dane nie są objęte monopolem prawno-autorskim. Nie jest to jednak równoznaczne ze stwierdzeniem, że żadna z czynno-

⁶ Ł. Prokulski, *LSTM – przewidywanie tekstu (sieci neuronowe, część 4)*, Łukasz Prokulski. Dane i Analizy, 5.03.2018, <https://blog.prokulski.science/> [dostęp: 31.10.2019]. Autor, prezentując możliwości sieci LSTM (ang. *Long Short Term Memory*), podtypu sieci rekurencyjnych, wykorzystuje powieść Remigiusza Mroza *Kasacja*. Ćwiczenie nie prowadzi do utworzenia wytworu, przedstawia jedynie, w jaki sposób można wykorzystać do tego celu sieć LSTM.

⁷ M. Ma'amari, *AI Generates Taylor Swift's Song Lyrics*, Towards Data Science, 28.09.2018, <https://towardsdatascience.com/ai-generates-taylor-swifts-song-lyrics-6fd92a03ef7e> [dostęp: 31.10.2019].

⁸ J-P Briot, G. Hadjeres, F-D. Pachet, *Deep Learning Techniques for Music Generation – A Survey*, ArXiv 2019, hal-01660772v3, s. 169, <https://hal.sorbonne-universite.fr/hal-01660772v3/document> [dostęp: 31.10.2019].

ści, którym jest poddawany utwór w celu uzyskania informacji i danych, nie wkracza w monopol uprawnionego. Jak wskazuje się w literaturze przedmiotu, „[...] [n]iezależnie od przyjętego modelu teoretycznego faz analizy danych co najmniej część takich działań jest nierozzerwalnie powiązana z reprodukcją co najmniej kilku fragmentów utworów. Co więcej, czynności takie mogą być powtarzane w kolejnych etapach *data mining*”⁹. Przychyłam się także do stanowiska, iż w przypadku eksploracji tekstów i danych (ang. *text and data mining*, dalej jako TDM), nie znajduje usprawiedliwienia wyjątek na rzecz tymczasowego przejściowego lub incydentalnego charakteru dokonywanego zwielokrotnienia. Przemawiają za tym poszczególne etapy maszynowej analizy, z uwzględnieniem konieczności usuwania niektórych fragmentów danych¹⁰.

Modele SI, wykorzystywane do generowania wytworów w stylu danego twórcy, bazują na eksploracji tekstów i danych. Ta metoda analizy komputerowej jest wykorzystywana obecnie w wielu obszarach nauki i przemysłu. Może być stosowana na wszystkich poziomach organizacji, w szczególności w przedsiębiorstwach, w celu planowania zasobów czy zarządzania klientami. Eksploracja danych służy także budowaniu systemów wspomaganie decyzji, systemów informacji geograficznej lub systemów poprawy jakości i konkurencyjności¹¹.

Automatycznej analizie mogą zostać poddane zarówno wytwory niechronione, jak i podlegające ochronie prawno-autorskiej. Dostrzegając korzyści ze stosowania TDM oraz ich wpływ na rozwój gospodarki cyfrowej oraz nauki, należy brać pod uwagę ewentualność naruszenia prawa autorskiego. Okoliczność ta została uwzględniona w treści dyrektywy z dnia 17.4.2019 r. w sprawie prawa autorskiego i praw pokrewnych na jednolitym rynku cyfrowym (dalej jako dyrektywa JRC). W tekście preambuły odnotowano, iż

[...] [w] niektórych przypadkach eksploracja tekstów i danych może obejmować czynności chronione prawem autorskim lub prawem do bazy danych *sui generis*, w szczególności zwielokrotnianie utworów lub innych przedmiotów objętych ochroną lub pobieranie treści z bazy danych, co ma miejsce na przykład w przypadku normalizacji danych podczas eksploracji tekstów i danych. Jeżeli nie znajduje zastosowanie wyjątek lub ograniczenie, konieczne jest uzyskanie zezwolenia na dokonanie takich czynności od podmiotów uprawnionych¹².

⁹ K. Gienas, *Data mining jako problem prawa autorskiego*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2017, z. 1, s. 26.

¹⁰ *Ibidem*, s. 33.

¹¹ M. Mirończuk, *Przegląd i klasyfikacja zastosowań, metod oraz technik eksploracji danych*, „Studia i Materiały Informatyki Stosowanej” 2010, t. 2, nr 2, s. 41.

¹² 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 z 17.5.2019 r., L 30/92.

Dyrektywa zobowiązuje państwa członkowskie do wprowadzenia¹³ wyjątku od monopolu eksploatacyjnego przysługującego podmiotom prawa autorskiego, praw pokrewnych i prawa producenta bazy danych w odniesieniu do zwielokrotnień utworów i pobrań danych pochodzących z chronionej bazy, 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 dostęp¹⁴.

Poza powyższym rozwiązaniem, przewidzianym dla celów badań naukowych, dyrektywa nakazuje wprowadzić kolejny wyjątek lub ograniczenie odnoszące się generalnie do eksploracji tekstów i danych, bez wskazania kontekstu. W takim przypadku wyjątek lub ograniczenie mają zastosowanie pod warunkiem, że korzystanie z utworów i innych przedmiotów objętych ochroną nie zostało wyraźnie zastrzeżone w odpowiedni sposób przez podmioty uprawnione.

W kontekście uprawnień twórców, którzy nie są entuzjastami trenowania SI w celu naśladowania ich stylu, istotne jest ustalenie, jak do kwestii TDM odnoszą się obecnie obowiązujące przepisy ustawy o prawie autorskim i prawach pokrewnych¹⁵ oraz jak na sytuację autorów, w nakreślonym zakresie, wpływają postanowienia dyrektywy JRC.

Przychyłam się do stanowiska, iż maszynową analizę tekstów i danych należy kwalifikować jako nowe, nienazwane pole eksploatacji wytwarzania kopii cyfrowych utworu¹⁶. Nie przekonują mnie argumenty, z powodów o których wspominam już we wprowadzeniu do niniejszego artykułu, iż „proces analizy zmierzającej do absorpcji niewielkiego zasobu danych nie ingeruje w dobro chronione”¹⁷. Za jego wyodrębnieniem przemawiają przesłanki techniczne i ekonomiczne. Za trafne uważam spostrzeżenie, iż ustalając, czy nowy sposób korzystania stanowi nowe pole eksploatacji, należy

¹³ Zgodnie z art. 29 ust. 1 dyrektywy JRC państwa członkowskie wprowadzą w życie przepisy ustawowe, wykonawcze i administracyjne, niezbędne do wykonania niniejszej dyrektywy do dnia 7 czerwca 2021 r. Niezwłocznie przekazują one Komisji tekst tych przepisów.

¹⁴ Art. 3 dyrektywy JRC.

¹⁵ Ustawa o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (tekst jedn. z 6.06.2019 r., Dz.U. 2019, poz. 1231); dalej w tekście jako: PrAut.

¹⁶ K. Gliściński, *Komentarz do art. 50, [w:] Ustawa o prawie autorskim i prawach pokrewnych. Komentarz*, red. A. Michalak, C.H. Beck, Warszawa 2019, s. 390.

¹⁷ Stanowisko odnotowane, lecz zdaje się – niepodzielane przez Wojciecha Machałę, który przytaczając je, zwraca jednocześnie uwagę, iż pod względem technicznym TDM wkracza w co najmniej jedno pole eksploatacji utworu; W. Machała, *ACTA 2 czy Nihil novi? Pierwsze refleksje na temat dyrektywy Parlamentu Europejskiego i Rady o prawie autorskim na jednolitym rynku cyfrowym*, „Monitor Prawniczy” 2019, nr 18.

zbadać, czy generuje on korzyści, w tym także finansowe, niezależnie od tych, które pochodzą z innych pól eksploatacji¹⁸. Postulat, by przy ustalaniu istnienia nowego pola eksploatacji brać pod uwagę wytworzenie możliwości udostępnienia utworu nowej publiczności, nie stanowi przeszkody dla powyższej kwalifikacji TDM. Należy zauważyć, iż wspomniana przesłanka została ukształtowana w orzecznictwie w określonym stanie faktycznym. Bywa także rozpatrywana jako kontrargument w kontekście rozważań nad tym, czy należy uznać za nowe pole eksploatacji działania mające na celu poprawę jakości utworu lub jego odbioru¹⁹.

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 rynku²⁰. Dodatkowo, liczba podmiotów potencjalnie zainteresowanych maszynową analizą oraz zróżnicowanie sytuacji gospodarczych, w których może znaleźć ona zastosowanie, przemawiają za uznaniem TDM za nowe pole eksploatacji.

Biorąc pod uwagę stosunkowo niedługi czas, w którym dochodzi do generowania wytworów, posiadających niektóre cechy utworów wedle definicji legalnej, jest za wcześnie, by postulować wyodrębnienie jeszcze bardziej szczegółowego pola eksploatacji niż eksploracja tekstów i danych, mianowicie eksploracji tekstów i danych pozyskiwanych z przedmiotów ochrony autorstwa konkretnej osoby. Nie można tego jednak w przyszłości wykluczyć.

W obecnie obowiązującym stanie prawnym i faktycznym twórca może wpływać na legalność korzystania z własnej twórczości w celu trenowania algorytmu poprzez nieprzenoszenie i nieudzielanie licencji na korzystanie z utworu na takim polu eksploatacji, jak eksploracja tekstów i danych. W celu uniknięcia wątpliwości, dokonując rozporządzeń lub czynności prawnych o innym charakterze, odnoszących się generalnie do wywarzania kopii cyfrowych (np. w celu umożliwienia przeprowadzenia prac redakcyjnych lub korzystania z utworu w inny, objęty treścią umowy sposób), warto sprecyzować, iż nie dotyczy to zwielokrotnień w ramach TDM.

Powstaje pytanie, od kiedy wytwarzanie kopii cyfrowych w związku z TDM można uznać za znane pole eksploatacji – jest to istotne w kontekście interpretacji umów już zawartych. Waga zagadnienia byłaby mniejsza,

¹⁸ K. Gliściński, *Wyodrębnianie się nowych pól eksploatacji i ich wpływ na obrót prawami do utworów*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2010, nr 3, s.51.

¹⁹ Zob. Wyr. ETS z 7.12.2006 r. w sprawie *Sociedad General de Autores y Editores de España (SGAE) przeciwko Rafael Hoteles SA*, sprawa C-306/05.

²⁰ T. Targosz, *Komentarz do art. 50, [w:] Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Wolters Kluwer, Warszawa 2014, s. 721.

gdyby sądy dokonywały restrykcyjnej wykładni art. 41 ust. 2 PrAut²¹. Wedle przepisu umowa o przeniesienie autorskich praw majątkowych lub umowa o korzystanie z utworu obejmuje pola eksploatacji wyraźnie w niej wymienione. Wykładnia językowa prowadzi do wniosku, iż niewymienienie w umowie pola eksploatacji, jakim jest wytwarzanie kopii cyfrowych utworu w związku z TDM, oznacza utrzymanie monopolu eksploatacyjnego przy twórcy²². Uwzględniając jednak liberalną wykładnię umów dokonywaną przez polskie sądy, potwierdzenia powyższego skutku nie można być pewnym.

Według art. 41. ust. 4 PrAut umowa może dotyczyć tylko takich pól eksploatacji, które są znane w momencie jej zawarcia. Ma to znaczenie dla oceny legalności działań podejmowanych na podstawie umów podlegających obecnie wykonaniu. Istnieje wątpliwość, od którego momentu eksplorację tekstów i danych należy uznać za znane pole eksploatacji. Za początkowy okres wyodrębnienia eksploracji danych spośród innych dziedzin nauk komputerowych uznaje się lata '80. XX w. Znaczący wzrost zainteresowania metody odnotowano w połowie lat 90.²³ Miernikiem popularności są organizowane konferencje tematyczne oraz założenie w 1997 r. czasopisma „Data Mining and Knowledge Discovery”. Sformułowanie „data mining” wyparło opisowe określenie „pozyskiwanie informacji z dużych zbiorów danych”. „Data Science Journal” – pierwsze czasopismo, poświęcone omawianiu systemów danych, ich publikacji w Internecie oraz możliwymi zastosowaniami, a także odnoszące się do kwestii prawnych – założono w 2002 r.²⁴ Kolejne lata przynoszą coraz większy zachwyty możliwościami stwarzanymi przez TDM. W kontekście interpretacji postanowień przedmiotowego zakresu umowy przenoszącej autorskie prawa majątkowe (lub licencji) powyższe ramy czasowe mają charakter jedynie orientacyjny. W literaturze przedmiotu wskazuje się, że „[p]ojawienie się nowych możliwości nie wpływa samo w sobie na wykładnię umowy”²⁵. Istotna jest także świadomość obu stron, w momencie

²¹ Opowiadają się za nią m.in.: J. Barta, R. Markiewicz, *Prawo autorskie*, Wolters Kluwer, Warszawa 2016, s. 338; T. Targosz, *Nakaz wyraźnego wymienienia pól*, [w:] T. Targosz, K. Włodarska-Dziurzyńska, *Umowy przenoszące autorskie prawa majątkowe*, Wolters Kluwer, Warszawa 2010, s. 26 i n.

²² Wniosek odnosi się oczywiście także do sytuacji innych podmiotów uprawnionych niż autor. Uwzględniając jednak tematykę artykułu, ograniczam się do omówienia okoliczności z perspektywy twórcy.

²³ F. Coenen, *Data Mining: Past, Present and Future*, „The Knowledge Engineering Review”, Vol. 26, Issue 1, s. 25.

²⁴ G. Press, *A Very Short History Of Data Science*, „Forbes” 28.5.2013, <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#73453e9355cf> [dostęp: 31.10.2019].

²⁵ P. Ślęzak, *Pola eksploatacji utworów audiowizualnych*, Branta, Bydgoszcz–Katowice 2006, s. 34.

zawierania umowy, istnienia nowej formy eksploatacji utworu. Powyższe wywody mogą okazać się przydatne w celu wykluczenia liberalnej interpretacji, w myśl której eksploracja tekstów i danych, mimo iż niewysłowiona w treści umowy, byłaby nią, wbrew interesom twórcy, objęta.

Odnosząc się do przepisów dyrektywy JRC, uważam, iż, co do zasady, powoływanie się na dozwolony użytek w przypadku trenowania algorytmu utworami autorstwa wyłącznie jednego twórcy znajdzie usprawiedliwienie w działalności organizacji badawczych i instytucji dziedzictwa kulturowego. Konieczną przesłanką jest jednak korzystanie z utworów przez powyższe podmioty wyłącznie w celach badań naukowych. Będą one zapewne interpretowane szeroko, umożliwiając także analizę wytworu uzyskanego w wyniku stosowania TDM, a nie samego przebiegu maszynowej analizy. Wątpliwości budzą procesy, które można przeprowadzić z zastosowaniem dowolnego materiału treningowego. Dopóki jednak – o ile w ogóle – eksploracja tekstów i danych jednego twórcy nie będzie stanowiła odrębnego pola eksploatacji, stosowanie dozwolonego użytku z art. 3 dyrektywy JRC nie jest uzależnione od przedmiotu prowadzonych badań.

Wyjątki i ograniczenia ustanowione w art. 4 dyrektywy JRC nie zawierają ograniczenia podmiotowego ani przedmiotowego. Trzeba mieć jednak na względzie cele, dla których owe wyjątki i ograniczenia są ustanawiane. Wedle preambuły mają one umożliwić

[...] osiągnięcie sprawiedliwej równowagi między prawami i interesami twórców i innych podmiotów uprawnionych, z jednej strony, a prawami i interesami użytkowników, z drugiej strony. Znajdują zastosowanie jedynie w niektórych szczególnych przypadkach, które nie kolidują z normalną eksploatacją utworów lub innych przedmiotów objętych ochroną ani nie powodują nadmiernego uszczerbku dla uzasadnionych interesów podmiotów uprawnionych²⁶.

Na tle tej ogólnej zasady należy rozpatrzyć cele, do osiągnięcia których ma się przyczynić stosowanie TDM, i zestawić je z prognozowanymi pozytywnymi skutkami ustanowienia ograniczeń i wyjątków. Zgodnie z treścią preambuły „eksploracja tekstów i danych umożliwia przetwarzanie dużych ilości informacji w celu uzyskania nowej wiedzy i odkrycia nowych tendencji”; i dalej: „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 technologii²⁷”.

²⁶ Pkt 6 preambuły do dyrektywy JRC.

²⁷ Pkt 18 preambuły do dyrektywy JRC.

W przypadku trenowania algorytmu wyłącznie w oparciu o teksty i dane pochodzące z twórczości jednego podmiotu, skorzystanie z dozwolonego użytku nie znajduje uzasadnienia, jeżeli celem nie jest odkrywanie nowych zastosowań czy usprawnień w obrębie TDM, a jedynie możliwość wykorzystania powstałego wytworu. Uważam, że odbywałoby się to z naruszeniem art. 35 PrAut, godząc w słuszne interesy twórcy. Biorąc pod uwagę cele ustanowienia wyjątku i ograniczenia w art. 4 dyrektywy JCR, opisany sposób i efekt trenowania algorytmu nie stanowi normalnego korzystania z utworu.

Trenowanie SI a autorskie prawa osobiste twórcy

Poczynione dotychczas uwagi odnoszą się do uprawnień wynikających z autorskich praw majątkowych. Uprawnienia te mogą zostać przez twórcę zbyte lub obciążone na rzecz osób trzecich. Niektórzy autorzy wyrażają obawę przed ewentualnymi przyszłymi praktykami wydawnictw, które na skutek zawartych umów stają się wyłącznie uprawnionymi z tytułu autorskich praw majątkowych. W takiej sytuacji powstaje pytanie o inną podstawę prawną umożliwiającą zakwestionowanie określonego sposobu wykorzystywania SI.

Odwołanie się do autorskich praw osobistych twórcy zdaje się być naturalną konsekwencją powodu, dla którego niektórzy autorzy, w szczególności pisarze, są niechętni omawianym praktykom. Za naganne uważają „zerowanie” na tekstach, które autorzy traktują często jako ekspresję swojej osobowości. Dodatkowo, poszczególne dzieła, ich tematyka, sposób prowadzenia narracji, budowania zdań, konkretne fragmenty, styl – stanowią tożsamość pisarza, czyli wizję własnej osoby. Uzasadnia to szczególną ochronę więzi twórcy z utworem.

Mimo iż odwołanie w tej sytuacji do autorskich praw osobistych zdaje się być oczywiste, w praktyce może okazać się nieefektywne. Moim zdaniem nie umożliwi twórcom skutecznego sprzeciwu wobec trenowania SI za pomocą wyłącznie ich twórczości.

W przypadku zarzutu naruszenia autorskich praw osobistych analizuje się wpływ określonego działania na więź twórcy z utworem. Ustawodawca wyliczył przykładowo, w jakich okolicznościach najczęściej dochodzi do jej naruszenia. Obecnie wykorzystywane algorytmy są trenowane za pomocą danych pochodzących ze zbioru utworów. Powstaje pytanie o to, czy art. 16 PrAut sankcjonuje ochronę więzi twórcy jedynie z konkretnym utworem, czy też także z całym dorobkiem lub jego częścią. Zasadna zdaje się być pierwsza z wymienionych opcji.

Nie można także wykluczyć, iż zostanie skonstruowany taki model SI, któremu w celach treningowych wystarczy jedna książka. Mogłoby się zda-

wać, że naruszenie prawa do autorstwa mogłoby nastąpić w przypadku przypisywania twórcy wytworu, który został jedynie wygenerowany na podstawie uprzednio napisanej książki. Trudno jednak określić, z jakim utworem więź zostałaaby naruszona. *Ratio legis* art. 16 PrAut nie sprowadza się do swobody decydowania o tym, z jakim wytworem niematerialnym twórca życzy mieć sobie więź. Problem należałoby raczej rozpatrywać na gruncie ogólnych dóbr osobistych twórcy, a nie autorskich praw osobistych. Nie jest także uzasadnione domaganie się twórcy, by oznaczyć jego autorstwem wytwór wygenerowany przez SI.

Prawo do integralności utworu należy rozważać nie w kontekście samego wytworu, ponieważ nie jest on tożsamy z tym dobrem niematerialnym, względem którego twórcy przysługuje prawo do decydowania o jego treści i formie. Zastanawiając się, czy dochodzi do naruszenia podczas trenowania algorytmu, uważam, że należy udzielić odpowiedzi przeczącej. Operacje, jakim jest poddawany tekst, zasadniczo nie są możliwe do zaobserwowania przez twórcę. Nie jest także prezentowany wytwór stanowiący oryginalny utwór ze zmianami nieznanymi akceptacji uprawnionego. Uznanie, iż podczas treningu dochodzi do naruszenia integralności utworu, oznaczałoby ochronę twórcy przez złymi emocjami, spowodowanymi świadomością, iż dzieło jest poddawane procesom, prowadzącym do powstania wytworu, którego wygenerowania autor sobie nie życzy. Tak ujęty interes autora, moim zdaniem, nie jest objęty zakresem art. 16 pkt 3 PrAut.

W opisanych sytuacjach nie aktualizuje się przesłanka, o której mowa w art. 16 pkt 4 PrAut. Zakładam, że twórca nie może być uznany za autora wytworu powstałego w omawiany sposób. W konsekwencji – nie przysługuje mu prawo do decydowania o jego pierwszym udostępnieniu publiczności. Na marginesie rozważań: za wątpliwą należy także uznać kwalifikację wspomnianego wytworu jako utworu w rozumieniu prawa autorskiego.

Odnosząc się do uprawnienia do sprawowania nadzoru nad korzystaniem z utworu, uważam, iż nie zostaje ono naruszone. Co do czynności, do których sprowadza się korzystanie, aktualne są uwagi poczynione powyżej w kontekście praw do integralności utworu. Dotychczasowe rozumienie treści uprawnienia, wynikające z orzecznictwa i prezentowane przez doktrynę, nie pozwala uznać, iż trenowanie algorytmu twórczością danego autora narusza jego prawo do nadzoru nad korzystaniem z utworu²⁸. Podkreślam także,

²⁸ Prawo do nadzoru nad sposobem korzystania z utworu jest postrzegane także w kontekście stosunków umownych i uprawnień twórców objętych art. 56, 58, 60 PrAut (m.in. prawo twórcy do odstąpienia od umowy ze względu na istotne interesy twórcze; w przypadku rozpowszechniania utworu w nieodpowiedniej formie albo ze zmianami, którym twórca mógłby się słusznie sprzeciwić; uprawnienie twórcy do nadzoru autorskiego przed rozpowszechnie-

że zastrzeżenia niektórych twórców budzi nie sam proces, czyli sposób korzystania, ale efekt, jaki można na jego skutek osiągnąć.

Trenowanie SI a dobra osobiste twórcy

O ile nie znajduję podstaw do uznania, iż trenowanie algorytmu twórczością jednego autora stanowi naruszenie autorskich praw osobistych, inaczej oceniam roszczenia w oparciu o art. 23 KC. Jednym z dóbr osobistych nazywanych przez ustawodawcę jest twórczość artystyczna. Uważam, że w braku definicji legalnej, a z uwzględnieniem aksjologicznych przesłanek ochrony, przez „twórczość artystyczną” należy rozumieć nie tylko konkretne jej efekty. Jak zauważa SA w Krakowie, ochronie podlega także „węzeł emocjonalny, łączący twórcę z jego dorobkiem”²⁹. W cytowanym wyroku pada także inne stwierdzenie, mogące uzasadniać sprzeciw autorów wobec opisywanych praktyk. Ochronę niemajątkową w oparciu o art. 23 KC usprawiedliwia osiągnięcie celu nieaprobowanego przez twórcę dla eksploatacji jego twórczości. Stwierdzenie zdaje się być aktualne w sytuacji, gdyby z wytworów SI miało korzystać wydawnictwo, wprowadzając je do sprzedaży, a także w odniesieniu do działalności osób fizycznych, specjalizujących się w maszynowym generowaniu literatury. Należałoby rozważyć, czy uprawnienia twórcy w powyższym kontekście nie powinny umożliwiać sprzeciwu wobec działań w oparciu o dozwolony użytek z art. 3 dyrektywy JRC. Konieczne byłoby zestawienie dwóch lub więcej wartości podlegających ochronie, co w oderwaniu od konkretnych okoliczności jest co najmniej trudne.

Czy warto formułować argumenty de lege ferenda?

Ponieważ znajduję zrozumienie dla ewentualnego sprzeciwu twórców wobec niektórych sposobów wykorzystywania SI, postulowałabym wprowadzenie efektywnych narzędzi umożliwiających respektowanie woli twórców. Nie jestem natomiast przekonana, czy wszystkie winny być wprowadzone do modelu ochrony autorsko-prawnej. W kwestii autorskich praw majątkowych takich wątpliwości nie mam. Postulowałabym jednoznaczne wskazanie, iż eksploracja testów i danych, a także inne rozwiązania, umożliwiające trenowanie SI z wykorzystaniem utworów, w szczególności autorstwa jednej osoby, stanowią odrębne pole eksploatacji. Uzasadniona jest jednak ochrona także tych twórców, którzy nie są uprawnieni z tytułu autorskich praw majątko-

niem utworu); zob. J Barta, Ryszard Markiewicz, *op. cit.*, s. 150 i 151. Powyższe uprawnienia nie dotyczą okoliczności, w których odbywa się trenowanie SI.

²⁹ Wyr. SA w Krakowie z 5.03.2004 r., I ACa 35/04, OSA 2004/10/33.

wych. Można rozważyć wprowadzenie do KC stypizowanego dobra osobistego. Sprzeciw twórców wobec omawianych praktyk powinien być wówczas motywowany względami innymi niż majątkowe, co może znaleźć odzwierciedlenie w katalogu roszczeń.

Najistotniejszy problem, poza wątpliwościami natury prawnej, to faktyczna możliwość ustalenia, iż w danym procesie korzystano z twórczości określonej osoby. Problem ten jest także aktualny przy rozważaniach na temat objęcia wytworów SI ochroną autorsko-prawną lub w ramach praw pokrewnych³⁰. Dodatkowo, przy założeniu, że wytwór uzyskany przez samouczącą się SI nie stanowi plagiatu, a jest jedynie inspirowany, trudno znaleźć obecnie podstawę do nakazania komukolwiek, by ujawnił stosowane praktyki.

Są to jednak zagadnienia wymagające kompleksowych, nie tylko prawnych rozważań. Być może także upływu czasu. Umożliwi to zaobserwowanie tendencji w modelach biznesowych i zachowaniach podmiotów, które byłyby zainteresowane trenowaniem SI w omawianym celu. Rozwiązanie powyższych problemów nie jest możliwe w ramach niniejszego artykułu.

Bibliografia

- Barta J., Markiewicz R., *Prawo autorskie*, Wolters Kluwer, Warszawa 2016.
- Briot J.-P., Hadjeres G., Pachet F.-D., *Deep Learning Techniques for Music Generation – A Survey*, ArXiv 2019, hal-01660772v3, s. 169, <https://hal.sorbonne-universite.fr/hal-01660772v3/document> [dostęp: 31.10.2019].
- 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 z 17.5.2019 r., L 30/92.
- Gienas K., *Data mining jako problem prawa autorskiego*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2017, z. 1.
- Gliściński K., *Komentarz do art. 50*, [w:] *Ustawa o prawie autorskim i prawach pokrewnych. Komentarz*, red. A. Michalak, C.H. Beck, Warszawa 2019.
- Gliściński K., *Wyodrębnianie się nowych pól eksploatacji i ich wpływ na obrót prawami do utworów*, „Zeszyty Naukowe Uniwersytetu Jagiellońskiego” 2010, nr 3.
- Ma'amari M., *AI Generates Taylor Swift's Song Lyrics*, Towards Data Science, 28.09.2018, <https://towardsdatascience.com/ai-generates-taylor-swifts-song-lyrics-6fd92a-03ef7e> [dostęp: 31.10.2019].

³⁰ Na zagadnienia te zwracał uwagę prof. Ryszard Markiewicz podczas wykładu inauguracyjnego rok akademicki 2018/2019; zob. R. Markiewicz, *Sztuczna inteligencja i własność intelektualna*, [w:] *Uniwersytet Jagielloński w Krakowie. Inauguracja roku akademickiego 2018/2019*, WUJ, Kraków 2019, s. 50. Tekst dostępny także online: https://www.uj.edu.pl/documents/10172/140821974/SI_prof_Markiewicz.pdf/ [dostęp: 31.10.2019].

- Machała W., *ACTA 2 czy Nihil novi? Pierwsze refleksje na temat dyrektywy Parlamentu Europejskiego i Rady o prawie autorskim na jednolitym rynku cyfrowym*, „Monitor Prawniczy” 2019, nr 18.
- Maggiore M., *Artificial Intelligence, Computer Generated Works and Copyright*, [w:] *Non-Conventional Copyright: Do New and Atypical Works Deserve Protection?*, red. E. Bonadio, N. Lucchi, Edward Elgar Publishing, Chettenham–Northampton 2018.
- Markiewicz R., *Sztuczna inteligencja i własność intelektualna*, [w:] *Uniwersytet Jagielloński w Krakowie. Inauguracja roku akademickiego 2018/2019*, WUJ, Kraków 2019.
- Mironczuk M., *Przegląd i klasyfikacja zastosowań, metod oraz technik eksploracji danych*, „Studia i Materiały Informatyki Stosowane” 2010, t. 2, nr 2.
- Press G., *A Very Short History Of Data Science*, „Forbes” 28.5.2013, <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/#73453e9355cf> [dostęp: 31.10.2019].
- Prokułski Ł., *LSTM – przewidywanie tekstu (sieci neuronowe, część 4)*, Łukasz Prokułski. Dane i Analizy, 5.03.2018, <https://blog.prokułski.science/> [dostęp: 31.10.2019].
- Ślęzak P., *Pola eksploatacji utworów audiowizualnych*, Branta, Bydgoszcz–Katowice 2006.
- Targosz T., *Komentarz do art. 50*, [w:] *Prawo autorskie i prawa pokrewne. Komentarz*, red. D. Flisak, Wolters Kluwer, Warszawa 2014.
- Ustawa o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (tekst jedn. z 6.06.2019 r., Dz.U. 2019, poz. 1231).

Abstract

Authors’ Rights in the Context of Selected Models for AI Training

Artificial Intelligence (AI) is no longer about automatization of routine or repetitive tasks. Its application extends far beyond plain computational power and efficiency. “AI-generated works” is one of the most popular research areas in this field. It raises the question about intelligent systems singularity and their true capability to perform human-like tasks. This article examines selected models for AI training which generate works mimicking the unique style of given artists provided their works serve as input data. Certain authors, perhaps especially writers, are reluctant to accept such forms of the exploitation of their works. Even though classification of AI-generated works is a challenge, significant legal issues arise at the earlier stage. Does text and data mining (TDM) infringe the copyright monopoly? Is using a collection of works of a single author as training data affects the lawfulness of a process? Are authors entitled, on the basis of their economic or moral rights, to oppose such practices?

Key words: AI generated works; copyright; AI; text and data mining

Streszczenie

Trenowanie wybranych modeli sztucznej inteligencji a uprawnienia twórców

Sztuczna inteligencja (SI) przestała być wykorzystywana jedynie w celu automatyzacji powtarzalnych czynności lub przeprowadzania zadań wymagających skrupulatności obliczeniowej. Szczególną ciekawość budzi funkcjonowanie SI prowadzące do uzyskania efektów, które dotąd były osiągalne jedynie dla człowieka. Wśród nich można wyróżnić m.in. wytwory posiadające cechy utworu. Artykuł odnosi się do szczególnego rodzaju trenowania SI, odbywającego się wyłącznie na podstawie utworów jednego twórcy, ze szczególnym uwzględnieniem procesów prowadzących do powstania wytworu „w stylu” danego twórcy. Niektórzy autorzy, zwłaszcza pisarze, nieprzychylnie odnoszą się do tego typu praktyk. Wątpliwości natury prawnej powstają już na etapie trenowania modelu SI. Czy eksploracja tekstów i danych (TDM) stanowi wkroczenie w monopol prawno-autorski? Czy trenowanie SI twórczością wyłącznie jednego twórcy wpływa na legalność procesu? Czy powołując się na autorskie prawa majątkowe lub osobiste, można zakazać wykorzystywania swojej twórczości do trenowania SI?

Słowa kluczowe: eksploracja tekstów i danych; prawo autorskie; SI; sztuczna twórczość