



ПРАВОВІ ПІДСТАВИ ДЛЯ ВИКОРИСТАННЯ ОБ'ЄКТІВ АВТОРСЬКОГО ПРАВА В МАШИННОМУ НАВЧАННІ

Антон Тарасюк

аспірант НДІ інтелектуальної власності НАПрН України

ORCID: 0009-0003-5015-8797

У статті розглянуто правові підстави для використання об'єктів авторського права в машинному навчанні. Запропоновано визначення машинного навчання через призму авторського права. У цьому контексті проаналізовано можливість використання ліцензійного договору та визначено його особливості. Проаналізовано актуальну поточну судову практику США щодо застосування доктрини добросовісного використання (fair use) як правової підстави для використання об'єктів авторського права в машинному навчанні та виокремлено індикатори, що можуть свідчити про можливість/неможливість застосування такої доктрини на конкретних прикладах.

Ключові слова: добросовісне використання, машинне навчання, штучний інтелект, авторське право

Постановка проблеми. Машинне навчання є однією з найбільш трендових технологій сучасності. Розробники штучного інтелекту (ШІ), зокрема систем генеративного ШІ, використовують найрізноманітніші об'єкти — тексти, фотографії, музичні твори тощо — як матеріал для «навчання» нейронних мереж, що є основною відповідних ШІ-продуктів. Серед таких об'єктів можуть бути й об'єкти авторського права — комп'ютерний код, оригінальні фотографії, тексти книг тощо.

Питання правомірності формування дата-сетів (баз даних) з об'єктів авторського права та використання їх для навчання нейронних мереж стало наріжним каменем самої індустрії ШІ, що ілюструється поточними багатомільярдними судовими спорами стосовно правових підстав використання об'єктів авторського права для навчання нейронних мереж, зокрема, у США.

Актуальним є визначення суті процесу машинного навчання з точки зору авторського права та аналіз правових підстав, на основі яких компанії-власники нейронних мереж можуть використовувати об'єкти авторського права, зокрема аналіз ліцензійного договору та доктрини добросовісного використання.

Літературний огляд. Питання захисту авторських прав у машинному навчанні та в ШІ-проектах досліджували, зокрема, такі науковці: К. Зеров, О. Дорошенко, Л. Тарасенко, А. Штефан, О. Баранов, О. Костенко. Проте, урахувавши динамічність розвитку ШІ-технологій та їх вплив на правовідносини, питання, порушені в цій статті, залишаються дискусійними та потребують детального аналізу.

Метою дослідження є аналіз правових підстав для використання об'єктів авторського права в машинному навчанні, зокрема ліцензійного договору та доктрини добро-

совісного використання, і вироблення рекомендацій щодо умов застосування таких підстав з метою ефективного захисту авторського права, поряд зі стимулюванням розвитку інновацій.

Виклад основного матеріалу. Використання методів машинного навчання для створення рішень на основі технологій штучного інтелекту набуває все більшого розвитку.

Нейронні мережі, які лежать в основі відповідних мобільних або вебдодатків, являють собою масиви даних та комп'ютерний код, що дає змогу реалізувати обробку певної інформації через вказаний масив/базу даних. Така база даних може мати різний вигляд та зазвичай складається з шарів, у рамках яких вказуються ваги зв'язків нейронів, умови їх активації та зсуви. Значення параметрів такої нейронної мережі, зокрема значення вагів та зсувів, змінюються та коректуються шляхом та в результаті машинного навчання. Саме ж машинне навчання означає використання певних об'єктів для коректування вказаних параметрів з метою формування стану таких параметрів таким чином, щоб мати змогу отримувати конкретні вихідні дані в результаті обробки певних типів інформації [1].

Наприклад, у результаті використання методів машинного навчання за допомогою навченої/натренованої нейронної мережі можна визначити, якою саме є написана людиною цифра або чи є зображення на картинці зображенням kota тощо.

Відповідно, щоб отримати відповідь, чи вказане зображення є зображенням kota, у процесі машинного навчання необхідно використати зображення котів та позначити, на яких із зображень дійсно відображений кіт, а на яких — ні. У результаті «навчання» нейронної мережі значення її вищевказаних параметрів (ваги нейронів, зсуви) буде змінено і після запиту щодо наявності зображення kota на конкретній фотографії/малюнку користувач може отримати відповідь, чи дійсно там є кіт. Правильність визначення залежить від багатьох факторів: якості коду, ефективності методів машинного навчання, що застосовувалися, та, відповідно, якості самого датасету, який використовувався.

Іншими словами, датасет зображень стає інструментом для створення конкретного активу — нейронної мережі з визначеними параметрами, яка може використовуватися з комерційною і некомерційною метою.

Ситуація є аналогічною для генеративних систем ШІ, зокрема великих мовних моделей (LLM), коли в результаті використання нейронної мережі можна отримати текст, а матеріалами для навчання можуть бути вірші, книги, комп'ютерний код та інші текстові файли, які становлять єдиний датасет і можуть охоронятися авторським правом.

Отже, як можна визначити машинне навчання на таких наборах даних через призму авторського права?

Одне з технічних визначень машинного навчання наведено С. Б. Могильним: «Процес, під час перебігу якого система опрацьовує велику кількість прикладів, виявляє закономірності і використовує їх, щоб прогнозувати вихідні характеристики для нових вхідних даних» [2, с. 5].

З точки зору авторського права вказаний процес можна визначити так: використання об'єктів авторського права з метою зміни характеристик та можливостей системи (бази даних, що виражена у вагах нейронів, зсувах тощо та програмного коду) задля покращення можливості роботи системи з об'єктами такого типу, зокрема, з метою розпізнавання чи генерування подібних об'єктів.

Теоретико-правове підґрунтя спорів щодо законності використання об'єктів авторського права в машинному навчанні. Спори виникають, коли автор твору або інший суб'єкт майнових прав вважає, що його авторські права порушено шляхом незаконного використання об'єктів для машинного навчання. Власники ж нейромереж/особи, що використовують вказані об'єкти для машинного навчання,

часто позиціонують таке використання як законне через призму концепту добросовісного використання (*fair use*).

Необхідним убачається аналіз суті вказаних правовідносин, умов охорони та захисту прав суб'єктів таких відносин щодо об'єктів та в контексті змісту правовідносин загалом.

Як зазначає Н. М. Павлюк, «цивільні правовідносини слід сприймати як врегульовані нормами цивільного права майнові та особисті немайнові відносини, засновані на вільному волевиявленні, юридично рівних учасників, що є носіями суб'єктивних цивільних прав» [3, с. 3].

Убачається за можливе припустити, що правовідносини стосовно «навчання» нейронної мережі з використанням об'єктів авторського права складатимуться з таких елементів: (а) Суб'єкт 1 (фізична або юридична особа, яка здійснює дії щодо навчання/організації навчання нейронної мережі); (б) Суб'єкт 2 (особа, якій належать авторські права на об'єкт, що використовується для навчання нейронної мережі); (в) Об'єкт 1 (конкретна штучна нейронна мережа, що виражена у вигляді комп'ютерного коду та масиву даних та стосовно якої відбувається процес навчання); (г) Об'єкт 2 (конкретний об'єкт авторського права, що використовується для навчання нейронної мережі); (д) Зміст правовідносин: дії Суб'єкта 1, які виражаються у використанні Об'єкта 2 для зміни/покращення Об'єкта 1 з дозволу Суб'єкта 2 або на підставі закону.

Ключовим правовим елементом і наріжним каменем цих правовідносин є правова підстава, на якій Суб'єкт 1 використовує об'єкти авторського права. Такою підставою може бути договір або право користувача, засноване на доктрині добросовісного використання (*fair use*) або *fair dealing*. Науковий та практичний інтерес обумовлює застосування ліцензійного договору, урахуваючи інноваційність правовідносин, зокрема в контексті їх змісту, а також об'єктів (один об'єкт може включатися у численні набори даних), способу їх використання та умов подальшого використання відповідної нейронної мережі.

Ліцензійний договір на використання об'єкта авторського права з метою навчання нейронної мережі. Отже, Суб'єкт 2 (суб'єкт авторського права) може надати Суб'єкту 1 (організатору навчання) дозвіл на використання об'єкта/низки об'єктів для вчинення конкретних дій — навчання конкретної нейронної мережі. Такі правовідносини можуть бути врегульовані ліцензійним договором, у якому, серед іншого, доцільно детально регламентувати такі умови: список об'єктів авторського права, що можуть бути використані для навчання; визначена нейронна мережа/нейронні мережі, для навчання яких можуть використовуватись об'єкти; оплатність/безоплатність використання; строк використання; суть навчання — конкретний перелік способів, яким може бути використаний об'єкт; етичні вимоги до можливостей використання нейронної мережі (наприклад заборона генерування об'єктів з певними ознаками, заборона використання нейронної мережі для надання медичних порад тощо); види технічних продуктів, у яких може використовуватися нейронна мережа (програмне забезпечення для перекладу чи створення зображень тощо); правові вимоги щодо можливості відтворення конкретних частин об'єкта в результаті операцій натренованої нейронної мережі (точне відтворення фрагмента твору на запит користувача відповідного продукту тощо); практичні засоби/способи контролю автора за умовами використання його творів та захисту ділової репутації в контексті атрибуції (згадування імені автора у разі видачі неправдивих матеріалів тощо).

Такий ліцензійний договір може мати багато застережень/обмежень, залежно від домовленостей сторін. Можливо, розвиток індустрії буде сприяти укладанню угод з організаціями колективного управління правами, які отримуватимуть право укладати такі договори від імені чи в інтересах багатьох авторів, і, відповідно, власники нейронних мереж будуть мати можливість підтверджувати «безпечність» своїх продуктів у контексті авторського права шляхом інформування про наявність у них відповідних прав, отриманих на об'єкти, що використовувалися в процесі машинного навчання.

Ще одним напрямом розвитку таких правовідносин можуть бути ліцензійні договори, укладені компаніями-власниками онлайн-маркетплейсів / соціальних мереж / глобальних онлайн-форумів.

Оскільки нейронну мережу, зокрема велику мовну модель (LLM), можна навчати на текстах з Інтернету, всі написані користувачами повідомлення можуть розглядатись як матеріал для такого навчання, тому право на використання таких текстів може зацікавити компанії, що здійснюють розробки у сфері штучного інтелекту.

У 2024 році ліцензійний договір на використання контенту для навчання нейронної мережі було укладено між американською компанією Reddit та власником продукту з використанням технологій штучного інтелекту [4]. Вартість такої угоди — близько 60 млн дол. щорічно. Оскільки оплата відбувається на періодичній основі, можна припустити, що об'єктом ліцензування будуть і старий датасет (архів), і поточні датасети контенту, що будуть створюватись у реальному часі користувачами платформи.

Правовою засадою цього ліцензійного договору виступає користувацька угода компанії Reddit Inc., у якій також зазначено, що якщо в разі використання сервісів створюється/завантажується контент, користувач підтверджує, зокрема, таке: надання все-світньої, безоплатної, безстрокової, безвідкличної, невиключної ліцензії з можливістю передачі та субліцензування на використання, копіювання, зміни, адаптації, підготовки похідних робіт, розповсюдження, зберігання, виконання та відображення контенту і будь-якого імені, імені користувача, голосу або образу, наданого у зв'язку з контентом, у всіх медіаформатах і каналах, відомих зараз або розроблених пізніше в будь-якій точці світу. Ця ліцензія включає в себе право робити контент доступним для трансляції, розповсюдження або публікації іншими компаніями, організаціями чи особами, які є партнерами Reddit; Reddit може видалити метадані, пов'язані з контентом, і користувач безповоротно відмовляється від будь-яких претензій і тверджень щодо моральних (особистих немайнових) прав або авторства щодо контенту [5].

У цілому можна дійти висновку, що наявність такого широкого тексту ліцензії, що інкорпорований у публічні правила користування сайтом, включає в себе можливість власника такого сайту надавати відповідний контент, який, зокрема, містить об'єкти авторського права, у ліцензію для машинного навчання. У певному сенсі це новий вид генерування доходів для корпорацій, що мають у своєму розпорядженні такий актив. При цьому, ураховуючи, що вказані матеріали створюються/завантажуються користувачами, завжди існують ризики потрапляння до датасету для навчання нейронної мережі об'єктів авторського права, завантажувати які користувачі не мали права.

Отже, укладання ліцензійного договору на використання об'єктів для навчання нейронної мережі може мати місце як з власниками відповідних майнових прав на такі об'єкти, наприклад фотографіями/авторами книг, так і з корпораціями, які мають у розпорядженні створені користувачами тексти та інші об'єкти авторських прав і використовують механізм дозволів, закріплений в угоді користувача, щоб мати юридичні підстави для передачі всього наявного контенту для навчання нейромереж, які належать третім особам.

Важливо зазначити, що наявність права на використання об'єктів авторського права шляхом укладання ліцензійного договору з авторами означатиме для компанії-власників відповідних нейронних мереж, що імплементовані в технологічні рішення (мобільні та веб-додатки тощо), можливість вести мову про певну правову безпеку в контексті авторського права, наприклад генеративних систем ШІ, для кінцевих споживачів-компаній.

Добросовісне використання (fair use) та fair dealing об'єктів авторського права в машинному навчанні

Можливість застосування підходів *fair use* та *fair dealing* до машинного навчання є дискусійним питанням.

У своїй роботі Н. Чередник аналізує застосування та особливості цих підходів у США, Великій Британії та Україні. Зокрема акцентується на гнучкості доктрини *fair use* у США через відсутність необхідності внесення змін до законодавства у відповідь на виклики часу та нових технологій, порівняно з підходом визначення вичерпного переліку умов застосування концепту *fair dealing*, заснованого на триступеневому тесті Бернської конвенції [6, с. 7].

Стрімкий розвиток технологій штучного інтелекту і нейронних мереж та використання об'єктів авторського права в машинному навчанні стали тими викликами, що поставили під питання адекватність застосування вищевказаних підходів у контексті справедливого врегулювання відповідних правовідносин.

Говорячи про концепцію Fair Dealing, варто звернути увагу на ст. 27 Закону України «Про авторське право і суміжні права», яка передбачає дії з вільного використання твору у формі тимчасового відтворення, які мають проміжний або випадковий (додатковий) характер і є невід'ємною складовою частиною технологічного процесу та можуть бути здійснені за визначених умов [7].

Дослідник К. Зеров аналізує ст. 27 у контексті машинного навчання, зокрема через призму судової практики ЄС, і акцентує на тому, що положення цієї статті можуть застосовуватися за умови дотримання вимоги ст. 22 вказаного закону в контексті захисту прав та законних інтересів автора твору, а тому вільне використання творів для навчання нейронної мережі (тренування системи ШІ), результатом операцій якої буде генерування творів, що будуть конкурувати з творами, використаними для навчання, потребуватиме згоди відповідного автора. Учений зазначає, що «посилання авторів на винятки не повинні порушувати «триступеневий тест», а саме — не обмежувати безпідставно законні інтереси відповідних суб'єктів авторського права, зокрема шляхом конкуренції на ринку AI-outputs з оригінальними творами» [8, с. 5, 9].

З цією тезою можна погодитись. Якщо в результаті роботи готового продукту (веб-додатка), в основі якого буде нейронна мережа, зокрема система генеративного ШІ, можна буде, наприклад, отримати повний текст певного розділу книги; отримати практично таку саму книгу (з однаковими сюжетними лініями, однак, наприклад, іншими іменами героїв); згенерувати комп'ютерний код, що є практично ідентичним коду, використаному для навчання нейронної мережі та розробленому конкретним програмістом, права відповідного автора можуть бути порушені.

Порушення можуть, зокрема, виражатися в тому, що доступ до твору або його частини буде надано необмеженому колу користувачів цього додатка, що створить конкуренцію з нормальним використанням твору.

До того ж така натренована нейронна мережа, вбудована у відповідний додаток, доступний для третіх осіб, є лише засобом, що створює можливості для його користувачів. Саме користувач дає інструкції (робить запит), отримує результат і вирішує, як його використовувати. Чи сприятиме сама наявність та доступність такого засобу збільшенню конкурентної продукції для авторів? Безумовно, так. Однак звичайний ксерокс також є технічним засобом, за допомогою якого можна зробити несанкціоновані копії творів і розповсюджувати їх, завдаючи збитків авторам.

Ба більше, машинне навчання застосовується не лише з метою створення систем генеративного ШІ. У технологіях комп'ютерного зору (computer vision) у процесі навчання нейромережі виявляти перешкоди на дорозі, зокрема тварин (котів), використовувалися відповідні зображення. Навряд чи функціонування програми для безпілотного керування автомобілем, яка, серед іншого, потенційно здатна розрізнити котів на дорозі, створить яку-небудь конкуренцію для фотографів — авторів фотографій котів, які використовувалися для машинного навчання. Однак чи заслуговують такі фотографії на винагороду за використання їхніх творів, що стали одним з ключових чинників у розробленні комерційно успішного продукту? Цілком можливо.

Багато корпорацій, що інвестують у розвиток та розроблення продуктів на основі штучного інтелекту, як і компанії, об'єкти яких використовуються для машинного навчання, зареєстровані та діють у США. Відповідно, у разі спорів щодо можливості вільного використання об'єктів авторського права для машинного навчання в цій юрисдикції ключовим є практичне застосування доктрини *fair use*.

Положення цієї доктрини наведені у секції 107 Закону США «Про авторське право» шляхом виділення таких елементів для визначення, чи мало місце добросовісне використання: (1) мета та характер використання, включно з тим, чи є таке використання комерційним або здійснюється з некомерційною освітньою метою; (2) природа твору, охоронюваного авторським правом; (3) кількість і суттєвість використаної частини стосовно охоронюваного авторським правом твору в цілому; (4) вплив використання на потенційний ринок або вартість охоронюваного авторським правом твору [9].

Одним з ключових судових рішень для подальшого розвитку доктрини *fair use* можна вважати рішення Верховного суду США в справі *Google LLC v. Oracle Am., Inc.* (2021) [10]. Суть спору полягала у визначенні, чи було використання відповідачем комп'ютерних програм допустимим у контексті *fair use*. У рішенні, ухваленому на користь відповідача, зокрема, було зазначено про факт запозичення, його малий обсяг, і, головне, про мету використання об'єктів — створення можливостей для будь-яких розробників створювати нові продукти, і про вплив на розвиток розвитку комп'ютерних програм у цілому, і що таке використання було трансформаційним. [10].

Повертаючись до застосування доктрини *fair use* саме в машинному навчанні, варто звернути увагу на працю Марка А. Лемлі та Браяна Кейсі «Добросовісне навчання». Учені вказують на зосередження доктрини добросовісного використання на трансформативному використанні творів, зазначаючи, що «в процесі машинного навчання не відбувається трансформативного використання, адже нейронні мережі не трансформують бази даних, на яких вони навчаються; вони використовують всю базу даних, причому з комерційною метою, і суди можуть вважати, що таку діяльність слід заборонити» [11]. Водночас автори вказують, що «іноді люди (або машини) копіюють певний об'єкт (його вираження), але вони зацікавлені лише в тому, щоб вивчити ідеї, які передаються цим вираженням, і у більшості випадків саме це відбувається з навчальними даними» [11].

Ідея того, що використання під час машинного навчання об'єктів авторського права відбувається саме з метою отримання ідей/концептів, що закладені в цих типах об'єктів, а не відтворення самих об'єктів як таких, є надзвичайно цікавою. Людина, прочитавши певний твір, може частково або повністю запам'ятати його і використати певні ідеї/концепти в ході створення своїх власних творів, то чому ж не буде добросовісним використання таких творів для зміни вагів зв'язків нейронів та інших параметрів у нейронній мережі? Відповідь на це питання формуватимуть судова практика та подальші наукові дослідження.

Ураховуючи рівень розвитку ШІ-технологій у США та наявність відповідних поточних судових спорів, судова практика в цій державі багато в чому стане визначальною для встановлення конкретних індикаторів правомірності/неправомірності використання об'єктів авторського права для навчання нейронних мереж без згоди відповідного правовласника.

Актуальна судова практика США щодо авторського права в машинному навчанні. Предметом спорів щодо авторського права в машинному навчанні найчастіше стає питання законності використання об'єктів авторського права під час навчання відповідних нейронних мереж, що було здійснене певною компанією.

Ураховуючи динамічність та інноваційний характер правовідносин, у рамках яких виникають спори, у багатьох справах ще немає рішення, яке набуло б законної сили.

Справа № 3:22-cv-07074 щодо можливого порушення авторських прав веб-продуктами Copilot, Codex [12].

До Окружного суду штату Північна Дакота, Каліфорнія, було подано колективний позов до (серед інших) компаній Microsoft corporation, Github inc., OpenAI inc. Позивачі стверджують, що відповідачі порушили (серед іншого) умови відкритих ліцензій, на яких був завантажений комп'ютерний код на онлайн-ресурсе GitHub. Спосіб порушення — використання коду, з попереднім видаленням всієї інформації про авторів та ліцензійні умови, шляхом «згодовування» його системі ШІ (нейромережі) з метою навчання відповідної нейромережі та, відповідно, створення комерційного продукту. Суть продукту — онлайн ШІ-помічник для написання коду, у результаті використання якого, серед іншого, пропонується оригінальний код, що був згодований без атрибуції (зазначення імені автора) [13].

Відповідачі аргументували свою позицію тим, що в позові відсутні обґрунтування збитків і порушення є гіпотетичними, а також можливістю використовувати доктрину добросовісного використання. Серед аргументів варто також виділити запевнення про те, що оскільки код був публічним, то ні будь-якій людині, «ні будь-якому комп'ютеру не може бути заборонено читати та вчитися з такого публічно розміщеного коду» [14, с. 4].

Рішення в цій справі може мати значний вплив на подальшу практику застосування доктрини добросовісного використання в процесі навчання нейронних мереж-систем генеративного ШІ.

Справа № 1:23-cv-00135-UNA щодо можливого порушення авторських прав/прав на торгівлю марку між GETTY IMAGES INC. та STABILITY AI, INC [15].

У рамках позову, що розглядається в Окружному суді округу Делавер, GETTY IMAGES INC (позивач), серед іншого, зазначає, що компанія STABILITY AI, INC (відповідач) незаконно, зокрема, порушивши правила користування сервісами позивача та відповідні умови ліцензій, автоматизованим шляхом вивантажила понад 12 млн фотографій, їх деталізованих описів та метаданих і використала їх, зокрема, для навчання своєї нейронної мережі, яка використовується для генерування фотографічних творів. Позивач акцентує увагу на високій якості фотографічного датасету як такого, зокрема, через високу якість колекції фотографій та їх детальну каталогізацію та опис, що, на думку позивача, робить вказану колекцію надзвичайно цінною для навчання нейронних мереж. За твердженням позивача, «відповідач використовує вказану нейронну мережу у своїх комерційних продуктах, що створюють конкуренцію з сервісами позивача, а також розповсюджує вказану нейронну мережу на умовах вільної ліцензії, що завдає позивачу значної шкоди» [16].

Як можна побачити, у цьому позові акцент зроблений не лише на сам факт використання датасету для навчання нейронної мережі, а й на спосіб, у який ця навчена нейронна мережа використовується, зокрема — з конкурентною метою.

Справа № 1:23-cv-11195 щодо можливого порушення авторських прав на тексти журналістських статей у продукті CHAT GPT [17].

Справа розглядається в Окружному суді Південного округу Нью-Йорка. Позивачем виступає THE NEW YORK TIMES COMPANY, відповідачами — *inter alia* — MICROSOFT CORPORATION, OPENAI, INC.

Позивач, серед іншого, акцентує увагу, що відповідачі з порушенням авторських прав позивача використали мільйони охоронюваних авторським правом новинних статей, розслідувань та інших матеріалів для створення своїх ШІ-продуктів (шляхом навчання своїх нейронних мереж). Датасет зі статей позивача є надзвичайно цінним для навчання нейронних мереж через свою високу якість. На думку позивача, відповідачі створили продукт, метою якого є заміна сервісу позивача, адже результати, які видає натренована нейронна мережа відповідача, мімікрують (уподібнюються/стають такими, що бентежно схожі) під ті статті, які застосовувалися в процесі машинного навчання такої мережі, тобто таке використання статей не може вважатися добросовісним використанням, на чому наполягали відповідачі в рамках досудових переговорів.

У зв'язку з «галюціонуванням» продуктів відповідача (видача ШІ-системою неістинного/викривленого контенту), кінцеві споживачі отримують неправдиву інформацію (статті, які начебто створені позивачем), що завдає останньому репутаційних збитків [18].

Як можна побачити, у цьому позові акцент також зроблено на факт конкурентного використання нейронної мережі, для навчання якої використовувалися статті, а також на репутаційні втрати від «невдалого» мімікрування (уподібнення) результатів роботи системи під статті позивача.

Відповідачі ж акцентують увагу на можливості використання доктрини fair use і на підтвердження своїх тез, зокрема, звертаються до тексту справи Google LLC v. Oracle Am., Inc. [10], зазначаючи, що «існує довга історія прецедентів, згідно з якими цілком законно використовувати вміст, охоронюваний авторським правом, як частину технологічного процесу, який (як тут) призводить до створення нових, різноманітних та інноваційних продуктів» [19].

Справа № 1:23-cv-08292 щодо можливого неправомірного використання повних текстів літературних творів (книжок) для навчання нейронної мережі [20].

Справа також розглядається в Округному суді Південного округу Нью-Йорка. Це позов групи авторів книг до компанії OPEN AI, INC та інших афілійованих компаній у контексті продукту Chat GPT.

Суть позиції позивачів є такою: нейронній мережі було «згодовано» (використано для навчання) повні тексти книжок авторів, які є письменниками і для яких дохід від книжок є ключовим джерелом доходу. Позивачі не надавали дозвіл на таке використання своїх творів і не отримали винагороду за їх використання. У результаті навчання нейронна мережа може генерувати книги, які схожі, мімікрують (уподібнюються), є переробкою оригінальних творів і такий інструмент доступний для необмеженої кількості людей. Вказані дії змінюють ринок для позивачів, адже Chat GPT фактично стає їх конкурентом, постійно використовуючи їхні ж роботи [21].

Згідно з позицією відповідачів машинне навчання мало на меті отримати інформацію з книг — частоту слів, синтаксичні моделі та тематичні маркери, що не є порушенням, оскільки не здійснюється копіювання тексту. У процесі машинного навчання не відбувається контакт охоронюваного виразу з людською аудиторією, так само це характерно і для процесу звичайного використання нетренованої нейронної мережі (моделі). Метою моделі не є копіювання виразу тексту, для цього існують набагато легші способи. Метою є створення нового матеріалу, якого раніше не було, на основі розуміння мови, аргументації та світу. Модель «вчиться» так само, як люди — на тому, що було раніше, що відповідає добросовісному використанню в контексті популяризації науки і прогресу [22].

На цьому етапі у вказаних справах ще немає рішень, які набули б законної сили, проте, аналізуючи їх у цілому, зокрема позиції сторін, варто зазначити таке. У всіх вказаних справах ідеться про системи генеративного ШІ, що дає можливість позивачам надавати суду реальні докази взаємодії з технічними продуктами відповідачів. Наприклад, (а) у справі № 3:22-cv-07074 позивач додає екземпляри коду, згенеровані продуктом Copilot; (б) у справі 1:23-cv-00135-UNA позивач додає зображення, згенеровані продуктом Stability AI, які надто схожі на оригінальні фотографії Getty images та навіть містять зображення, однозначно схоже на логотип (водяний знак), що мав місце на оригінальній фотографії; (в) у справі № 1:23-cv-11195 додаються зіставлення журналістських статей, які майже повністю відтворені в продукті відповідача при конкретному запиті; (г) у справі № 1:23-cv-08292 як доказ подаються відповіді Chat GPT на запити щодо генерування текстів книг, у рамках яких видаються зміст, суть, елементи книг авторів з деякими змінами.

У всіх справах позивачі акцентують увагу не лише на факті неправомірного використання об'єктів авторського права для навчання нейронних мереж, а й на способі, у який ці нейронні мережі використовуються, що призводить, зокрема, до створення кон-

курентних продуктів, тобто — заповідання реальних збитків. Теоретично доктрина добросовісного користування в контексті машинного навчання може використовуватись як правова підстава для суб'єкта, що здійснює машинне навчання, проте кожен випадок необхідно розглядати окремо, ураховуючи реальний вплив натренованої нейронної мережі, покладеної в основу відповідного технічного продукту, на права, законні інтереси та ділову репутацію авторів.

Ураховуючи вищевикладене, на думку автора цієї статті, ключовими індикаторами, що можуть свідчити про неможливість застосування доктрини добросовісного використання у процесі машинного навчання нейронної мережі, можуть бути: (а) використання об'єктів авторського права, що є основоположним активом певної компанії (упорядкована колекція фотографій тощо), коли можливим було б використання інших подібних об'єктів, які, наприклад, є суспільним надбанням; (б) реальний вплив на ринок унаслідок створення конкурентних кінцевих продуктів, що завдає авторам шкоди (збитків/упущеної вигоди); (в) часткове/повне відтворення оригінальних об'єктів авторського права для кінцевих користувачів продукту; (г) завдання репутаційної шкоди авторам унаслідок генерування неточних/спотворених об'єктів, що можуть асоціюватися з авторами/їхньою творчістю.

Ключовими індикаторами можливості застосування вказаної доктрини можуть бути: (а) отримання інформації з об'єктів щодо суті, структури, ідей, закладених у цих об'єктах; (б) трансформаційна та інноваційна мета — створення нового продукту, наприклад системи комп'ютерного зору, що здатна виявляти об'єкти на дорозі; (в) відсутність відтворення повного/часткового вираження об'єктів для кінцевого користувача/можливості створення переробок таких об'єктів; (г) відсутність конкуренції з авторами, реального впливу на ринок та реальних збитків для авторів.

Висновки. Машинне навчання з використанням об'єктів авторського права можна охарактеризувати як використання об'єктів авторського права (фотографій/текстів/тощо) з метою зміни характеристик та можливостей конкретної нейронної мережі задля покращення можливості роботи системи з об'єктами такого типу, зокрема, з метою розпізнавання, аналізу чи генерування подібних об'єктів.

Правовідносини щодо використання об'єктів авторського права для машинного навчання можуть бути врегульовані ліцензійним договором між власником відповідної нейронної мережі та суб'єктом авторських прав.

Окремо слід виділити ризики використання вже «натренованих» нейронних мереж, які, зокрема, можуть бути отримані на умовах відкритих ліцензій. Якщо для машинного навчання таких нейронних мереж об'єкти авторського права були використані неправомірно, подальше їх навчання та використання породжує правові ризики через неправомірну основу функціонування відповідного генеративного ШІ.

Окремою правовою підставою, на яку спираються власники нейронних мереж, що використовують об'єкти авторського права для машинного навчання, є fair dealing та доктрина добросовісного використання (fair use). Можливість застосування цих концептів слід розглядати окремо та комплексно в кожному випадку. Окремо — у контексті конкретного датасету (набору даних) об'єктів авторського права. Комплексно — у рамках подальшого застосування такої нейронної мережі в кінцевих продуктах.

Перелік використаних джерел

1. Шаранов О. Д., Матвійчук А. В. *Розвиток алгоритму зворотного поширення помилки в задачах оптимізації параметрів нейронних мереж*. Київ: ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана». URL: <https://ir.kneu.edu.ua/server/api/core/bitstreams/d31589ad-62ba-4fdb-a641-af6fbd3c6746/content> (дата звернення: 04.03.2024).

-
2. Могильний С. Б. *Машинне навчання з використанням мікрокомп'ютерів : навч.-метод. посіб. / за ред. О. В. Лісового та ін. Київ, 2019. 226 с. URL: <https://api.man.gov.ua/api/assets/man/54c0ee59-b490-4ff3-a346-90a89fd67e30/> (дата звернення: 04.03.2024).*
 3. Павлюк Н. М. *Правова природа та поняття цивільних правовідносин: загально-теоретичне дослідження. Юридичний електронний науковий журнал. 2022. № 3. С. 100–102.*
 4. *Reddit Signs AI Content Licensing Deal Ahead of IPO. 2024. URL: <https://www.bloomberg.com/news/articles/2024-02-16/masayoshi-son-seeks-to-build-a-100-billion-ai-chip-venture?srnd=technology-vc> (дата звернення: 04.03.2024).*
 5. *Угода користувача (Terms of service) компанії Reddit Inc. URL: <https://www.redditinc.com/policies/user-agreement-september-25-2023> (дата звернення: 04.03.2024).*
 6. Чередник Н. *Інститут вільного використання об'єктів авторського права в Україні та зарубіжних країнах — проблеми та перспективи розвитку. Порівняльно-правовий аналіз. Теорія і практика інтелектуальної власності. 2014. № 4. С. 81–88.*
 7. *Про авторське право і суміжні права : Закон України від 1 грудня 2022 року № 2811-IX. URL: <https://zakon.rada.gov.ua/laws/show/2811-20#Text> (дата звернення: 23.04.2024).*
 8. Зеров К. *Чи мріють системи генеративного штучного інтелекту про електричних овець? Поняття та умови охороноздатності об'єктів, згенерованих системами генеративного штучного інтелекту в Україні. Теорія і практика інтелектуальної власності. 2023. № 4–5. С. 90–100.*
 9. *Copyright Act of 1976. URL: <https://www.copyright.gov/title17/92chap1.html> (дата звернення: 23.04.2024).*
 10. *Google LLC v. Oracle Am., Inc. 141 S. Ct. 1183, 1201–09 (2021). URL: https://www.supremecourt.gov/opinions/20pdf/18-956_d18f.pdf (дата звернення: 29.04.2024).*
 11. *Lemley, Mark A. and Casey, Bryan, Fair Learning. 30.01.2022. URL: <https://ssrn.com/abstract=3528447> (дата звернення: 23.04.2024).*
 12. *Матеріали справи № 3:22-cv-07074. Окружний суд США штату Північна Дакота, Каліфорнія. URL: <https://www.courtlistener.com/docket/65669506/doe-1-v-github-inc/> (дата звернення: 23.04.2024).*
 13. *Текст позову у справі № 3:22-cv-07074. Окружний суд США штату Північна Дакота, Каліфорнія. URL: https://githubcopilotlitigation.com/pdf/07074/1-0-github_complaint.pdf (дата звернення: 23.04.2024).*
 14. *Матеріали справи № 1:23-CV-00135. Окружний суд США округу Делавер. URL: <https://www.courtlistener.com/docket/66788385/getty-images-us-inc-v-stability-ai-inc/> (дата звернення: 23.04.2024).*
 15. *Текст відповіді на Позов у справі № № 3:22-cv-07074. Окружний суд США штату Північна Дакота, Каліфорнія. URL: https://githubcopilotlitigation.com/pdf/07074/1-0-github_complaint.pdf (дата звернення: 23.04.2024).*
 16. *Текст позову у справі № 1:23-cv-00135. Окружний суд США округу Делавер. URL: https://cdn.vox-cdn.com/uploads/chorus_asset/file/24412807/getty_images_vs_stability_AI_delaware.pdf (дата звернення: 23.04.2024).*
 17. *Матеріали справи № 1:23-cv-11195. Окружний суд США Південного округу Нью-Йорк. URL: <https://www.courtlistener.com/docket/68117049/parties/the-new-york-times-company-v-microsoft-corporation/> (дата звернення: 04.03.2024).*
 18. *Текст позову у справі 1:23-cv-11195. Окружний суд США Південного округу Нью-Йорк. URL: <https://www.courtlistener.com/docket/68117049/1/the-new-york-times-company-v-microsoft-corporation/> (дата звернення: 23.04.2024).*
-

-
19. Текст відповіді на позов у справі № 1:23-cv-11195. Окружний суд США Південного округу Нью-Йорк. URL: <https://static01.nyt.com/newsgraphics/documenttools/82a013b9ba852548/9d4b1790-full.pdf> (дата звернення: 23.04.2024).
 20. Справа № 1:23-cv-08292. Окружний суд США Південного округу Нью-Йорк. URL: <https://www.courtlistener.com/docket/67810584/authors-guild-v-openai-inc/> (дата звернення: 23.04.2024).
 21. Текст позову у справі 1:23-cv-08292. Окружний суд США Південного округу Нью-Йорк. URL: <https://www.courtlistener.com/docket/67810584/1/authors-guild-v-openai-inc/> (дата звернення: 23.04.2024).
 22. Текст відповіді на позов у справі 1:23-cv-08292. Окружний суд США Південного округу Нью-Йорк. URL: <https://www.courtlistener.com/docket/67810584/75/authors-guild-v-openai-inc/> (дата звернення: 23.04.2024).

Anton Tarasiuk

Ph.D. student of the Intellectual Property Scientific Research Institute of the National Academy of Legal Sciences of Ukraine

Legal basis for using copyright objects in machine learning

In this article, I characterize the legal basis for using copyright objects in machine learning. In this regard, the definition of machine learning from the perspective of copyright law is provided. In the work, legal relations that occur in the context of machine learning using copyright objects are studied.

The two key legal grounds for the usage of copyright objects to train models/neural networks are defined: license agreement, fair dealing, and fair use doctrine. Specific terms of the license agreement on copyright objects for machine learning are defined.

In the analysis of the doctrine of fair use as a legal ground to use copyright objects without permission/license of the author, four actual and ongoing disputes in the USA regarding possible unlawful usage of copyright objects for machine learning without getting a license or providing payment to the authors are examined. As a result of this study, the author concludes that in order to determine whether it is possible to use the fair use doctrine as a legal basis, there must be a separate examination for each data used, and the analysis should be complex, including the way the new product based on the respective trained neural network will be used and the effect it may have on the market.

As a result, key indicators of the pros and cons of the possibilities of applying the fair use doctrine are defined. Pro indicators: (a) obtaining information from objects about the essence, structure, and ideas embedded in these objects; (b) a transformational and innovative goal; (c) lack of reproduction of the full/partial expression of objects for the end user/the possibility of creating modifications of such objects; (d) lack of competition with authors. Con indicators: (a) the use of extremely valuable/highest quality copies of copyright objects in the industry that are a fundamental asset of a particular company (an organized collection of photographs, etc.), while it was possible to get objects in the public domain for the stated goals; (b) a real impact on the market of authors due to the creation of competitive end products that cause real harm to them (real/hypothetical); (c) partial/full reproduction of original copyright objects for end users of the final product; (d) causing reputational damage to authors.

Keywords: machine learning, artificial intelligence, copyright, fair use

*Рекомендовано відділом авторського права і суміжних прав
НДІ інтелектуальної власності (протокол № 4 від 25.04.2024 р.)*

Подано / Submitted: 30.04.2024
Прийнято до публікації / Accepted: 01.05.2024