

УДК 347.771

DOI <https://doi.org/10.24144/2307-3322.2025.89.1.64>

МАШИННЕ НАВЧАННЯ: ВИКЛИКИ АВТОРСЬКОМУ ПРАВУ В ЦИФРОВУ ЕПОХУ ЧЕРЕЗ ПРИЗМУ ПРАВА ЄС ТА УКРАЇНИ

Шалугіна О.Р.,
*студентка 2-го курсу бакалаврської програми «Право»
Українського Католицького Університету*

Шалугіна О.Р. Машинне навчання: виклики авторському праву в цифрову епоху через призму права ЄС та України.

У відповідь на стрімке зростання попиту на технології штучного інтелекту (далі – ШІ) ЄС сформував власне правове поле, що включає Директиву 2019/790/ЄС щодо авторського права від 17 квітня 2019 року та новий Регламент ЄС 2024/1689 (Акт про ШІ) від 13 червня 2024 року. Це окреме правове регулювання, присвячене виключно питанням ШІ.

Однак, дискусійною темою досі лишається можливість навчання ШІ на базі чужих результатів інтелектуальної і творчої діяльності, які є об'єктами авторського права. Попри відсутність прямого регулювання саме машинного навчання, ст. ст. 3 і 4 Директиви 2019/790/ЄС встановлюють винятки для так званого глибинного аналізу тексту та даних (text and data mining, далі – TDM).

Суперечливою темою у науковій літературі все ще залишається питання можливості прирівнювання TDM та машинного навчання ШІ, адже останній швидко розвивається і запроваджує все більш витончені механізми. У даній статті на основі технічного аналізу було визначено, що машинне навчання суттєво відрізняється від TDM, як за своєю метою, так і за обсягом оброблених даних та якістю цієї обробки. Було встановлено, що Акт про ШІ не додав визначеності у поставлене питання, не врегулювавши машинне навчання по суті.

У статті наголошується також на дискусійності запропонованого у Директиві 2019/790/ЄС регулювання, згідно із яким вона поширює свою дію і на процес машинного навчання. Окрему увагу у статті зосереджено на положеннях Директиви 2019/790/ЄС щодо законного доступу до контенту та механізму резервування прав.

В рамках наукового дослідження проаналізовано рішення гамбурзького суду у справі *LAION v Robert Kneschke* від 27 вересня 2024 року, в якому TDM було прирівняно для машинного навчання, через призму Директиви 2019/790/ЄС.

У роботі окремо розглянуто й українське законодавство, яке вже почали робити таким, що відповідає регулюванню ЄС. Як було визначено, цей процес потребує окремої обережності у транспозиції, з урахуванням відповідності міжнародним зобов'язанням України та потенційних загроз.

Ключові слова: машинне навчання, глибинний аналіз тексту та даних, штучний інтелект, авторське право.

Shaluhina O.R. Machine learning: challenges to copyright in the digital age through the prism of EU and Ukrainian law.

In response to the rapid growth in demand for artificial intelligence (AI) technologies, the EU has created its legal framework, which includes Directive 2019/790/EU on the copyright of 17 April 2019 and the new EU Regulation 2024/1689 (AI Act) of 13 June 2024. This is a separate legal framework dedicated exclusively to AI issues.

However, the possibility of training AI based on someone else's intellectual and creative work, which is the subject of copyright, is still a topic of debate. Despite the lack of direct regulation of machine learning, Articles 3 and 4 of Directive 2019/790/EU establish exceptions for the so-called deep analysis of text and data (text and data mining, hereinafter – TDM).

The possibility of equating TDM and machine learning AI remains a controversial subject in the scientific literature, as the latter is rapidly developing and introducing increasingly sophisticated

mechanisms. In this article, based on a technical analysis, it was determined that machine learning is significantly different from TDM, both in its purpose and in the quantity of processed data and the quality of this processing. It was found that the AI Act did not add certainty to the issue raised, without regulating machine learning in its essence.

The article also emphasizes the debatability of the regulation proposed in Directive 2019/790/EU, according to which it extends its scope to the process of machine learning. The article focuses on the provisions of Directive 2019/790/EU on legal access to content and the rights reservation mechanism.

The article also examines the Hamburg court's decision in *LAION v Robert Kneschke* of September 27, 2024, in which TDM was equated with machine learning and considered through the prism of Directive 2019/790/EU.

The paper also separately examines Ukrainian legislation, which has already begun to be brought into line with EU regulations. As was determined, this process requires special caution in transposition, taking into account compliance with Ukraine's international obligations and potential threats.

Key words: machine learning, text and data mining, artificial intelligence, copyright.

Постановка проблеми. У 2022 році американська компанія OpenAI, Inc. створила ChatGPT, генеративний чат-бот зі штучним інтелектом, що ознаменувало початок нової ери розвитку технологій. Протягом наступних трьох років з'явилися й інші моделі: Copilot від Microsoft, Gemini від Google, китайський DeepSeek тощо.

Наукова новизна генеративних моделей ШІ полягає у їхній здатності до самостійного розвитку власних навичок, якості генерованих відповідей, що відбувається за допомогою технології машинного навчання. Спрощуючи, під час цього процесу моделі обробляють велику кількість даних, доступних у глобальній мережі, виділяють закономірності, формують власну базу та застосовують накопичені знання до власних відповідей на команди користувачів. На перший погляд, ця фантастична здатність комп'ютерних програм дозволяє спрощувати роботу людини у безлічі сфер, прискорювати науковий прогрес тощо.

Проте, застосування машинного навчання ШІ породило також і багато суперечностей. Однією з них є зріст справедливого невдоволення авторів, на чийх роботах проводив своє навчання ШІ. Розв'язання цієї проблеми пропонується знайти саме в рамках права інтелектуальної власності, яке може розмежувати права авторів з правами і обмеженнями ШІ.

Метою дослідження є з'ясування можливості та законності використання системами штучного інтелекту в процесі їх самонавчання (вдосконалення) чужих об'єктів інтелектуальної власності (здебільшого, авторського права).

Стан опрацювання проблематики. У науковій царині тема навчання ШІ та TDM є дуже актуальною та дискусійною. Втім, зауважуємо, що ця тема є недокінця розкритою в науковій літературі. Питання прирівняння TDM та ШІ досліджували Тім Дорніс, професор права у Стенфордському університеті, Гаррі Сюрден з Університету Колорадо. Саме з технічної точки зору це питання розглядали Кароліна Рістовська та Цвета Мартиновська Банде. Ларісса Берш окремо досліджувала відповідність Директиви 2019/790/ЄС Бернській Конвенції про охорону літературних і художніх творів 1971 року.

Виклад основного матеріалу. У відповідь на розвиток ШІ в рамках ЄС було оперативно розроблено нормативно-правову базу. У контексті цього дослідження ключовими є Директива 2019/790/ЄС про авторське право та суміжні права в єдиному цифровому ринку, що була прийнята 17 квітня 2019 року і деталізувала інші директиви у сфері авторського права (йдеться про Директиву 96/9/ЄС та Директиву 2001/29/ЄС), а також Акт про ШІ, Регламент ЄС № 2024/1689 від 13 червня 2024 року, що встановлює гармонізовані правила щодо штучного інтелекту.

Жоден з цих документів прямо не регулює машинне навчання ШІ, проте, висловлюється ідея, що цей процес підпадає під дію ст. 3, 4 Директиви 2019/790/ЄС, які стосуються регулювання виключення до використання TDM. Зокрема, з метою наукових досліджень науково-дослідні організації та установи зі збереження культурної спадщини можуть відтворювати та витягувати дані, до яких мають законний доступ, для проведення TDM [1]. А для інших цілей відтворення та витягування даних для TDM є допустимим лише за відсутності чіткого обмеження, здійснення суб'єктом прав інтелектуальної власності у належний спосіб. На відміну від TDM для наукових досліджень, дані при їх використанні у цьому випадку мають зберігатися лише протягом періоду здійснення аналізу [1].

Водночас, велика кількість дослідників не погоджуються з тим, що машинне навчання охоплюється виключенням TDM. Ми підтримуємо другу позицію, спираючись на технічний аспект операції ШІ та її розручу відмінність від ідеї TDM. Для цілей даної статті важливо провести розрізнення двох процесів.

Так, відповідно до ст. 2(2) Директиви 2019/790/ЄС, «TDM – це будь-який автоматизований аналітичний метод, призначений для аналізу тексту та даних у цифровій формі з метою отримання інформації, яка охоплює, серед іншого, моделі, тенденції та кореляції» [1]. Внаслідок розвитку технологій, машинне навчання вже вийшло за рамки традиційного TDM.

TDM запрограмований розрізняти закономірності та кореляції у великих базах даних [2,12]. Серед основних його функцій виділяють: ідентифікацію закономірностей, встановлення зв'язків, виведення висновків та ідентифікацію аномалій. TDM технології почали розвиватися у 1990-х роках і базувалися на статистичних методах, що дозволяла компаніям-користувачам виділяти важливі закономірності серед широкого поля інформації. Це допомагає приймати більш виважені рішення, маючи систематизовані дані [3,131]. TDM використовується у багатьох інших сферах: кліматології, бізнес-аналітиці, біомедицині тощо. Наприклад, TDM може проаналізувати великий масив даних щодо поширеності певного діагнозу та побачити кореляцію з часто повторюваним симптомом, що виникає на тлі цієї хвороби.

Натомість, машинне навчання, за визначенням його творця, американського математика Артура Семюеля, – це сфера застосування штучного інтелекту, яка характеризується наданням системам здатності автоматично навчатися та вдосконалюватися на основі даних або досвіду, без попереднього програмування [3, с. 131].

Метою машинного навчання є побудова прогнозувальних моделей, що здатні узагальнювати дані, продукувати рішення без втручання людини та розвивати свої навички в часі самонавчання. TDM послуговується строгими статистичними алгоритмами. У сфері навчання ШІ використовує опорні векторні машини, штучні нейронні мережі, що дозволяють застосувати вже т. зв. глибинне навчання (deep learning).

Глибинне навчання, підвид машинного навчання, дозволяє ШІ аналізувати складну неструктуровану інформацію, що часто стає проблемою для TDM [3, с. 132]. Цією інформацією також можуть виступати не тільки текст, а й зображення, музичні твори тощо. Це дозволяє моделям адаптуватися, генерувати інсайти у реальному часі та відповідати на все більш витончені запити користувачів.

Впливаючи з цих характеристик, можна зробити висновок, що відмінністю між даними технологіями є в, першу чергу, обсяг даних, що ними опрацьовується. Статичний, TDM зупиняється на результаті, отриманому внаслідок опрацювання заданого масиву даних. Орієнтований на постійний розвиток, адаптивний ШІ в ході машинного навчання буде поглинати все більшу кількість інформації для виведення більшої кількості параметрів [3, с. 132].

Іншим аспектом розрізнення є те, що «висмоктуючи» якнайбільше із доступних даних, ця технологія не здатна відрізняти їх ідейний та формальний склад. Базисом міжнародного права інтелектуальної власності є чітке розділення між цими поняттями. Ідея твору називається його семантикою, а зовнішній вираз – синтаксисом. Захисту підлягає саме останній, оскільки оригінальність твору, власне і полягає в його вираженні [2, 11]. У той час, коли TDM використовує лише ті дані, які необхідні для виконання його функцій, переважно для встановлення патернів, ШІ під час свого навчання поглинає і синтаксис, і семантику.

Іншими словами, з одного і того ж статистичного дослідження TDM візьме лише статистику, а генеративний ШІ «пропрацює» ще й спосіб викладення, стиль написання, висновки, зроблені людьми тощо. Власне, зовнішній вираз твору і є основною ціллю ШІ, оскільки це дозволяє генерувати правдоподібну відповідь на ускладнені запити користувачів [2,12]. Це дозволяє ШІ з кожним запитом все краще імітувати відповіді людини. Наразі програмування таких моделей, які були б здатні відрізняти синтаксис і семантику видається мало можливим.

В літературі висловлюється позиція, що процес машинного навчання може бути прирівняно до процесу навчання людини, адже, як мозок людини за все своє життя опрацьовує величезний обсяг інформації й може потім генерувати ідеї, так і оперує ШІ. ШІ нових поколінь використовують штучні нейронні мережі, що імітують структуру людського мозку у вимірі коду. За таких умов, машинне навчання, як і процес генерування відповідей на промпт (запит) становив би лише процес вільного обороту ідей у суспільстві, а ідеї, як відомо, не становлять предмету авторського права.

На противагу варто зазначити, що машинне навчання все ще є запрограмованим процесом, який відбувається на ґрунті того, що модель здатна розпізнати серед масиву даних та витягнути з них. Натомість мозок людини містить у собі високі когнітивні властивості, здатний зчитувати не тільки фактичний контекст, але й емоційний. Гаррі Сюрден, професор права в Університеті Колорадо у Боулдері, проводить розрізнення на прикладі розпізнавання спам-повідомлень. За допомогою машинного навчання ШІ може виокремити закономірність, що найчастіше у таких повідомленнях використовують слово «безкоштовний» або встановити повторюваність місцеперебування адресата. Людина натомість також може звернутись до прикметника «безкоштовний», проте буде аналізувати його через призму соціального контексту, використовуючи абстрактне мислення та інші навички задля виокремлення змісту повідомлення [4, 4-5].

Прикладом правозастосовної практики, в якій прирівняно машинне навчання з TDM є рішення Гамбурзького окружного суду у справі *LAION v Robert Kneschke* від 27 вересня 2024 року. Німецький фотограф Роберт Кнешке подав позов проти некомерційної організації LAOIN за використання однієї з його фотографій для створення бази даних, призначеної для навчання штучного інтелекту. Суд відхилив позов, спираючись на секцію 60d Закону Німеччини про авторське право та суміжні права, що відтворила положення ст. 3 Директиви 2019/790/ЄС [5]. Попри критику даного рішення, що в першу чергу була пов'язана з неповною оцінкою судом діяльності LAOIN, це перше прецедентне рішення, в якому до машинного навчання було застосоване виключення TDM.

Однак, технічне розмежування TDM та машинного навчання є не єдиним проблемним питанням, що постають у контексті поєднання винятків TDM та застосування ШІ.

Ст. 4(3) Директиви 2019/790 встановлює, що для загальних цілей можуть бути використані лише ті роботи, до яких потенційний користувач має законний доступ [1].

У *LAION v Robert Kneschke* Гамбурзький окружний суд помилково визначив, що таким способом може бути публікація фотографії в Інтернеті. Той факт, що твір було опубліковано для загальної публіки хоч і означає, що він доступний публічно, проте не означає, що доступ до нього і його використання є правомірним [6,2]. Загалом, існує мало розуміння, що є тим самим законним доступом у розумінні ст. 4(3), що створює виклики, як для добросовісних розробників ШІ, так і для пересічних авторів.

Ст. 4(3) також передбачає, що правовласники можуть заперечити проти використання їхніх робіт «у належний спосіб, такий як за допомогою засобів машинного зчитування у випадку контенту, доведеного до загального відома онлайн» [1]. Цей належний спосіб, відповідно до параграфу 18 Директиви, може полягати у резервуванні за допомогою метаданих або умов вебсайту чи сервісу. Таке резервування буде поширюватись виключно на застосування TDM і не впливатиме на інші випадки резервування. Іншими застосовними засобами можуть вбачатися договірні угоди або односторонні декларації [1].

При цьому, може виникнути ситуація, коли власники прав можуть резервувати права на свою роботу неоднаково на різних ресурсах. Автори можуть завантажувати різні копії, а також це може відбуватися внаслідок різних ліцензійних політик окремих ресурсів, що займаються розповсюдженням текстів, зображень, музичних творів тощо. За таких умов, важко передбачити, яким чином поводитиметься ШІ під час опрацювання даних, що зарезервовані у різний спосіб і це може поставити авторів у скрутне становище.

Також цікавим видається той факт, що Директива 2019/790 хоч і лишає певне право контролю для правовласнику вигляді резервування прав, але не передбачає жодних заходів матеріальної компенсації (винагороди) для авторів, чиї роботи були використані. Відтак, розробники ШІ мають можливість отримувати великі прибутки й генерувати значні клієнтські бази шляхом покращення функціоналу своїх моделей, але, водночас, автори, що вклали фізичні й ментальні ресурси у створення творів, не мають від такого використання жодної винагороди. У науковій літературі обговорюється можливість запровадження механізмів такої компенсації, хоча наразі таке право не передбачено [7, с. 18].

У 2024 році в ЄС було прийнято Акт про ШІ. Метою цього регулювання виступає забезпечення прозорості на ринку ШІ. Також, він класифікує моделі за різними ступенями ризику для прав громадян ШІ. Загалом, Акт є новачком у цій сфері, зважаючи на врахування технічних аспектів функціонування ШІ. Не оминає він і сферу авторського права. У ст. 53(3) Акт про ШІ ЄС покладає на держав-учасниць обов'язок привести національне законодавство у відповідність до Директиви 2019/790/ЄС та наголошує конкретно на вже згаданій ст. 4(3).

З тексту випливає, що держави мають звернутися до регулювання авторського права та суміжних прав, коли це стосується саме методів TDM [8]. Проте, глобально навчання ШІ може відбуватись і за допомогою традиційного TDM. Формулювання пар. 105 Акту виступає тому підтвердженням. Там зазначається, що ШІ розвивається внаслідок доступу до широкого кола інформації, куди входять і твори, захищені авторським правом, та визнається, що цей процес може спричинити виклики для авторів даних творів. Водночас, через буквально тлумачення цієї норми, для цього можуть використовуватись саме технології TDM, а не машинного навчання, яке, як було з'ясовано раніше, вже стало значно ширшим [8]. Тому, не можна стверджувати, що Акт поставив крапку у розмежуванні двох процесів і врегулював безпосередньо машинне навчання.

Євроінтеграційне спрямування нашої держави відкриває необхідність окремо дослідити українське законодавство на предмет положень, відповідних до винятків TDM або таких, що регулюють машинне навчання.

Закон України «Про авторське право і суміжні права» від 1 грудня 2022 року у п. 14 ч. 2 ст. 22 встановлює, що виготовлення копій творів, що зроблені із законного джерела з метою пошуку тексту і даних, включених до наукових публікацій або пов'язаних з ними, для дослідницьких цілей може бути здійснено безоплатно і без дозволу суб'єкта авторського права за умови що ім'я автора і джерело запозичення будуть зазначені. Це положення може застосовуватись лише в тому випадку, якщо використання творів не було прямо заборонено суб'єктами авторського права відповідним чином [9]. Тобто, в одному положенні поєднуються одразу обидві статті Директиви. Як зазначається у дослідженні ВОІВ, Україна є однією з небагатьох країн, що перенесли винятки TDM у своє законодавство [10, с. 24].

Проте, важливо зазначити, що Україна має бути обережною у процесі транспонування винятків TDM, оскільки вони можуть йти в розріз із її зобов'язаннями за Бернською конвенцією про охорону літературних і художніх творів 1971 року. Стаття 9(2) Конвенції запроваджує три кумулятивні умови, за яких відтворення творів може здійснюватись: (1) у певних особливих випадках, (2) за умови, що таке відтворення не завдає шкоди нормальному використанню твору і (3) не порушує будь-яким необґрунтованим способом законні інтереси автора [11].

Окремі науковці погоджуються з проходженням двох останніх критеріїв, однак більше питань викликає саме перший [12, с. 358]. Особливі випадки передбачають, що вони є одночасно «чітко визначеними» та «вузькими за обсягом та охопленням» [12, с. 355]. З тексту ст. 4(3) випливає, що вона охоплює всіх розробників ШІ (100%), які прагнуть відтворити твір, доки його автор не висловить своє заперечення. Відповідно, на кожен твір авторського права може претендувати безліч моделей ШІ, а застереження статті може вплинути лише на кількість творів, доступ до яких може бути отримано. Тому, не можна стверджувати, що можливість резервування якимось чином сприяє звуженню цих особливих випадків.

Висновки. Проведений аналіз засвідчив, що нормативне поле як ЄС, так і України лишає відкритим питання законності використання об'єктів авторського навчання під час машинного навчання ШІ. При цьому передбачені т. зв. винятки використання для TDM, але ні Директива (ЄС) 2019/790, ні Акт про ШІ не передбачають вичерпної відповіді на питання, чи охоплюється цим винятком і машинне навчання ШІ.

Водночас, проведений технічний аналіз двох процесів засвідчує, що TDM лише частково може охоплювати машинне навчання. TDM є більш обмеженим підходом, що спирається на статистичні алгоритми, які дозволяють виводити закономірності з конкретно заданих масивів інформації. Машинне навчання, своєю чергою, послуговується вищими технологіями (штучні нейромережі, глибинне навчання тощо). Це спричиняє і відмінність у ступені обробки об'єктів авторського права. ШІ використовує об'єкти авторського права не як джерело даних для проведення аналізу, а як матеріал для створення нових алгоритмів і шаблонів самої системи.

З огляду на це, актуальним є формування спеціального правового режиму у праві ЄС, так і законодавстві України, що б спрямовувався конкретно на тренування моделей ШІ. Цей режим має передбачати чіткі межі доступу до захищених творів, механізми надання згоди чи відмови авторами цих творів. Потенційно може бути включена також і грошова компенсація за їх використання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance.), OJ L 130, 17.5.2019. EUR-Lex. URL: <https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>.
2. Dornis, Tim W., The Training of Generative AI Is Not Text and Data Mining (October 19, 2024). European Intellectual Property Review (E.I.P.R.), forthcoming 2/2025. URL: <https://ssrn.com/abstract=4993782>.
3. Ristovska, K., and Martinovska Bande C., Machine Learning vs. Data Mining: Understanding the Differences and Intersections. *International Journal of Scientific and Research Publications*, vol. 15, no. 1, January 2025. ISSN 2250-3153. URL: <https://www.ijsrp.org/research-paper-0125/ijsrp-p15718.pdf>.
4. Surden H., Machine Learning and Law: An Overview (September 3, 2020). Forthcoming Chapter in «Research Handbook on Big Data Law». URL: <https://ssrn.com/abstract=3686055>.
5. Hamburg District Court, *LAION v Robert Kneschke*, 310 O.22723, 27 September 24. URL: <https://openjur.de/u/2495651.html>.
6. Rosati E., Is text and data mining synonymous with AI training?, *Journal of Intellectual Property Law & Practice*, Volume 19, Issue 12, December 2024, Pages 851–852, URL: <https://doi.org/10.1093/jiplp/jpae092>.
7. de la Durantaye, K., Control and Compensation. A Comparative Analysis of Copyright Exceptions for Training Generative AI (June 17, 2024). Forthcoming in 56 International Review of Intellectual Property and Competition Law (IIC) (2025), URL: <https://ssrn.com/abstract=5021889>.
8. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), OJ L, 2024/1689, 12.7.2024. EUR-Lex. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689>.
9. Про авторське право і суміжні права: Закон України від 1 грудня 2022 року. URL: <https://zakon.rada.gov.ua/laws/show/2811-20/conv#n200> (дата звернення 05.05.2024).
10. Xalabarder R., Standing Committee on Copyright and Related Rights, WIPO, SCCR/44/4, 12-13 (WIPO Oct. 17, 2023). URL: https://www.wipo.int/edocs/mdocs/copyright/en/sccr_44/sccr_44_4.
11. Бернська конвенція про охорону літературних і художніх творів (Паризький Акт від 24 липня 1971 року змінений 2 жовтня 1979 року) URL: http://zakon3.rada.gov.ua/laws/show/995_051.
12. Bersh, L. (2024). An exceptional formality under berne: evasion of copyright protection via the eu's text and data mining exception. *Harvard Journal of Law & Technology*, 38(1). URL: <https://jolt.law.harvard.edu/assets/articlePDFs/v38/6-Bersh.pdf>.