

ШМАТКОВ Д. І.,

кандидат педагогічних наук, доцент,
провідний науковий співробітник
науково-дослідної лабораторії з проблем
інформаційних технологій та протидії
злочинності у кіберпросторі
(Навчально-науковий інститут
№ 4 Харківського національного
університету внутрішніх справ)
Стаття поширюється на умовах
ліцензії CC BY 4.0

УДК 347.78:004.8(477)

DOI <https://doi.org/10.32842/2078-3736/2025.4.1.17>

ПРАВОВЕ РЕГУЛЮВАННЯ TDM В ЕПОХУ ШТУЧНОГО ІНТЕЛЕКТУ

У статті досліджуються виклики, що постають перед власниками українських вебсайтів у контексті використання контенту для цілей пошуку тексту і даних (TDM) і навчання штучного інтелекту. З огляду на стрімкий розвиток технологій машинного навчання, виникає нагальна потреба у юридичних механізмах, які дозволили б авторам і правовласникам контролювати цифровий контент і захищати його від несанкціонованого використання, зокрема автоматизованими системами.

Проаналізовано положення Директиви (ЄС) 2019/790 щодо охоплення TDM, а також імплементацію деяких її норм в українське законодавство. Встановлено, що чинна редакція Закону України «Про авторське право і суміжні права» у частині винятків для TDM є значно вужчою за європейську модель, обмежуючи вільне використання лише науковими публікаціями та дослідницькими цілями. У статті також аналізується поняття «належним чином зарезервованого права» та можливість застосування односторонніх заяв, таких як «All rights reserved», як машиночитуваного механізму заборони TDM. Наводяться приклади європейської судової практики, де національні суди по-різному тлумачать достатність таких застережень.

Обґрунтовано, що Україні доцільно розширити винятки для TDM, адаптувати термінологію до європейських підходів і впровадити технічні стандарти щодо машиночитуваного резервування прав. Розміщення чітких декларацій на вебсайтах може стати ефективним способом передачі намірів правовласника користувачам та автоматизованим системам. У статті також підкреслюється необхідність законодавчого балансу між охороною прав та забезпеченням свободи аналізу даних, що є ключовим чинником інноваційного розвитку.

Виявлено потребу у подальших емпіричних дослідженнях використання односторонніх заяв про авторські права на вебсайтах та порівняльного аналізу судової практики.

Ключові слова: *пошук тексту і даних, III, авторське право, вебсайт, вільне використання.*

Shmatkov D. I. Legal Regulation of Text and Data Mining in the Age of Artificial Intelligence

This article examines the challenges faced by Ukrainian website owners in the context of the use of content for the purposes of text and data mining (TDM)



and artificial intelligence training. Given the rapid development of machine learning technologies, there is an urgent need for legal mechanisms that would allow authors and rightsholders to control digital content and protect it from unauthorized use, particularly by automated systems.

The article analyzes the provisions of Directive (EU) 2019/790 on TDM and its partial implementation in Ukrainian legislation. It is established that the current version of the Law of Ukraine “On Copyright and Related Rights,” in the part concerning exceptions for TDM, is significantly narrower than the European model, limiting free use to scientific publications and research purposes only. The article also explores the concept of “properly reserved rights” and the potential use of unilateral declarations such as “All rights reserved” as a machine-readable mechanism to prohibit TDM. Examples of European case law are provided, demonstrating how national courts interpret the sufficiency of such reservations differently.

It is argued that Ukraine should expand TDM exceptions, align its terminology with European approaches, and introduce technical standards for machine-readable rights reservations. Posting clear declarations on websites may become an effective means of communicating a rightsholder’s to users and to automated systems. The article also emphasizes the need to maintain a legal balance between copyright protection and the freedom of scientific data analysis, which is a key driver of innovation.

The study identifies the need for further empirical research on the use of unilateral copyright notices on websites and a comparative analysis of relevant case law.

Key words: *text and data mining, AI, copyright, website, fair use.*

Вступ. З появою генеративного штучного інтелекту та стрімким розвитком технологій машинного навчання особливої актуальності набули питання використання вмісту вебсайтів для цілей пошуку тексту і даних (TDM). Ми бачимо актуалізацію проблеми правової охорони цифрового контенту. У цьому контексті зростає потреба в юридичних механізмах захисту від небажаного використання вмісту вебсайтів, зокрема для навчання штучного інтелекту (ШІ).

Паралельно зі зростанням актуальності TDM та навчання ШІ, у 2023 році в Україні набули чинності норми, що частково імплементують положення Директиви (ЄС) 2019/790 про авторське право на Єдиному цифровому ринку (далі – Директива). У результаті, власники вебсайтів отримали юридичні інструменти для обмеження пошуку тексту і даних у об’єктах авторського права, що їм належать.

Таким чином, сучасна цифрова реальність формує нові виклики для українських власників вебсайтів, що полягає у захисті вмісту від автоматизованого збору для TDM та навчання ШІ.

Постає потреба у аналізі чинного законодавства на предмет відповідності сучасним викликам та практикам Цифрової Епохи.

У контексті захисту контенту від автоматизованого збору для TDM та навчання ШІ Регламент (ЄС) 2024/1689 Європейського Парламенту та Ради від 13 червня 2024 року про штучний інтелект в контексті авторського права не робить нового внеску, тому учені схильні коментувати статті 3 та 4 Директиви. Ці статті формують умови до вільного використання вмісту вебсайтів із застосуванням TDM. Останні дослідження свідчать про наступне [1–7]:

- TDM є критично важливим інструментом для машинного навчання, цифрових гуманітарних наук і соціальних досліджень.
- TDM викликає очевидний конфлікт із авторським правом, але застарілі закони по всьому світу створюють юридичні бар’єри, які стримують розвиток технології.
- Бернська Конвенція та міжнародні договори не вимагають запровадження винятків для TDM, залишаючи це на розсуд юрисдикцій, тим не менш, у країнах з більш «відкритими» винятками технічний прогрес є стрімкішим.



- Ідеальна модель правового регулювання TDM повинна дозволяти як наукові, так і комерційні цілі, а також охоплювати всіх користувачів, а не тільки дослідників.
- Через різні підходи в національних законах дослідники часто не впевнені у легальності TDM.
- Навіть якщо в законі передбачено виняток для TDM (наприклад, для наукової діяльності), існує ряд механізмів обмежити таке використання, наприклад, через ліцензії з авторами.
- Обов'язкові винятки для TDM на базі положень Директиви повинні впроваджуватись у національні закони, Директива є першим кроком, але далеко не останнім.

При цьому автоматичні системи можуть не розрізняти допустимість використання або ж просто ігнорувати правові застереження [8] через недосконалість або відсутність стандартів контролю над автоматичним доступом [7], тому все ще існує потреба у пошуку балансу врахування інтересів зацікавлених сторін.

В Україні ці проблеми поки не знаходять такого активного обговорення.

Постановка завдання. Метою статті є визначення шляхів вдосконалення національного регулювання з урахуванням викликів, пов'язаних із використанням вмісту вебсайтів для цілей TDM та навчання штучного інтелекту, а також розробка поєднання технічних та юридичних механізмів захисту цифрового вмісту, які могли б бути імплементовані в українське правове поле.

Результати дослідження. Пунктом 14 частини другої статті 22 Закону України «Про авторське право і суміжні права» у редакції від 15.11.2024 (далі – Закон) визначено, що «без дозволу суб'єктів авторського права і безоплатно, але із зазначенням імені автора і джерела запозичення, допускається 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. Це положення застосовується за умови, що використання творів не було прямо заборонено суб'єктами авторського права відповідним чином, зокрема за допомогою засобів, придатних для зчитування комп'ютером з цифрового контенту, що є доступним у мережі Інтернет».

Порівняємо формулювання із Закону зі згаданими положеннями статей 3 та 4 Директиви.

По-перше, український Закон чітко обмежує вільне використання творів науковими та пов'язаними з ними публікаціями і дослідницькими цілями, якщо таке використання не заборонено правовласниками. У статті 3 Директиви мова йде про вільне використання будь-яких творів для наукових цілей, у статті 4 – будь-яких творів для будь-якої мети у межах TDM, якщо таке використання не обмежено правовласниками.

По-друге, у Законі використане обмеження трансформовано у «заборонено», у Директиві – «reserved» (збережено, зарезервовано [9]) – ці терміни мають різні юридичні значення. Цікаво, що на офіційному порталі Верховної ради України це слово у перекладі Директиви представлено у третій спосіб – «обмежено» – такий переклад, очевидно, за семантикою та юридичним значенням є ближчим до «збережено» або «зарезервовано», ніж до «заборонено». Крім того, заборона у Законі застосовується до використання TDM відносно наукових публікацій для дослідницьких цілей, у Директиві можливість обмеження використання TDM правовласниками – до будь-яких матеріалів з будь-якою метою.

Ще Стаття 5 Директиви 2001/29/ЕС щодо гармонізації певних аспектів авторського права і суміжних прав в інформаційному суспільстві передбачала низку дозволів на відтворення в пресі, сповіщення публіці або надання доступу до опублікованих статей з актуальної економічної, політичної чи релігійної тематики, або до трансляцій та інших об'єктів подібного характеру, якщо таке використання не було прямо збережено правовласником. У преамбулі до Директиви 2001/29/ЕС було зазначено: «Технологічний розвиток полегшить розповсюдження творів, особливо в мережах, що змусить правовласників краще ідентифікувати твір або інший об'єкт, автора чи будь-якого іншого правовласника, а також надавати інформацію про умови використання твору або іншого об'єкта, щоб полегшити управління



пов'язаними з ним правами. Необхідно заохочувати правовласників, щоб під час розміщення власних творів або інших об'єктів у мережах вони використовували позначення, що, на додаток до наведеної вище інформації, вказують, зокрема, на надання ними дозволу».

Таким чином, у той час, як дослідники говорять про необхідність розширення свободи здійснення TDM за межі Директиви, а у деяких країнах такий підхід вже впроваджено [2], Закон, навпаки, суттєво обмежує свободи, визначені у Директиві. Чи може бути захист прав зацікавлених осіб пріоритетнішим за потреби суспільства? Світова наукова спільнота каже, що ні. Утилітарний підхід до інтелектуальної власності дає таку саму відповідь. У той час, як норму Директиви вже імплементовано в національне законодавство держав-членів ЄС [10], формулювання, використане у Законі, виглядає застарілим.

Розглянемо також питання резервування прав. З однієї сторони, як у Директиві, так і у Законі не лімітовано засоби для резервування, придатні для зчитування комп'ютером. Очевидно, що мова не йде про якийсь стандарт програмування, але навіть вигадувати щось нове, якщо такий стандарт існує і широко застосовується дизайнерами вебсайтів? Одностороння заява «All rights reserved» у нижній частині вебсайту виглядає адекватним задоволенням такої потреби.

Усім звичне застереження має давні витоки – воно слугувало підтвердженням збереження авторських прав відповідно до вимог Буенос-Айреської конвенції 1910 року. Варто нагадати, що ця конвенція була підписана лише Сполученими Штатами та певними країнами Латинської Америки. Однак із приєднанням більшості держав до Бернської конвенції у 1970-х роках (Україна приєдналася до неї у 1995 році) положення Буенос-Айреської конвенції втратили свою чинність та значення, при чому лише для згаданих панамериканських країн. Це було пов'язано з тим, що Бернська конвенція встановила принцип автоматичного виникнення авторських прав із моменту створення твору, без необхідності дотримання будь-яких формальностей.

До сьогодні заява «All rights reserved» скоріше відіграє соціальну [11; 12], UI/UX [13], репутаційну та/або маркетингову функції. Але її юридичне значення тепер може отримати нову силу, соціальна та інформаційна функція застереження «All rights reserved», цілком ймовірно, може пройти нормативне навантаження. Розробка техніко-юридичного стандарту для таких заяв виглядає компромісом між інтересами праволасників, розробників ШІ, потребами суспільства тощо.

Чи може фраза «All rights reserved» (або «всі права збережено/зарезервовано»), розміщена, наприклад, у нижній частині вебсайту, вважатися «належним чином вираженим застереженням прав», як це передбачено Директивою? Ця заява в основному відображається як звичайний текст HTML-сторінки, і тому є доступною для будь-яких автоматизованих систем. Сучасні TDM-інструменти здатні розпізнавати текст у форматах HTML, PDF, зображеннях із OCR-обробкою тощо. Тобто, формально вимогу машинозчитуваності виконано.

«All rights reserved» прямо сигналізує, що автор не надає жодних дозволів, окрім прямо зазначених, отже TDM заборонено, якщо не передбачено інше. Розміщення застереження у нижній частині вебсайту може вважатися частиною умов користування, згідно з загальноновизнаною юридичною практикою, адже ця частина – традиційне місце для застережень і правових умов. Такий підхід дозволяє розглядати фразу як умову доступу до контенту.

На відміну від ліцензійних посилань, RDF-дескрипторів чи політик, «All rights reserved» є зрозумілим як для людини, так і для машини, без необхідності спеціального уніфікованого ідентифікатора ресурсу URI чи форматів. Це відповідає поняттю інтуїтивно зрозумілого сигналу для алгоритму або користувача.

Проте технічне значення «machine-readable» у Директиві може трактуватися вужче: як структурований, стандартизований, формалізований спосіб подання умов (наприклад, через <meta>, robots.txt, RDF, rel="terms" тощо). Простий текст не завжди дозволяє автоматизованим системам визначити, які саме дії заборонено, особливо якщо мова йде про часткові обмеження (наприклад, дозвіл на TDM лише у наукових цілях) – як було зазначено вище, про це повідомляють й інші дослідники.



Суди держав-членів ЄС поки не надають єдиного підходу у трактуванні поняття «належним чином збережених прав»:

- У Німеччині (Kneschke v. LAION) суд визнав достатнім просте текстове повідомлення на вебсайті, якщо воно є чітко вираженим.
- У Нідерландах (DPG v. HowardsHome) суд звернув увагу на важливість адресності повідомлення – якщо воно не супроводжується додатковими поясненнями, його сила зменшується.
- В Угорщині (справа 9.Pf.20.353/2024/6-II) апеляційний суд у справі щодо застосування згаданих статей Директиви не дав однозначного тлумачення, яке можна було б інтерпретувати до розглянутого питання.

Ми бачимо, що у межах Європейського Союзу практика судового застосування демонструє фрагментацію. Це ускладнює передбачуваність права і створює ризики як для дослідників, так і для правовласників.

Для України, що лише починає імплементувати ці європейські стандарти, важливо враховувати такі відмінності, аби не допустити перенесення правової невизначеності до національного поля. Попри те, що використання «All rights reserved» є стандартом дизайну вебсайтів і є доступним для машинного зчитування, відповідність вузькому розумінню «machine-readable» у контексті європейського законодавства все ще залишається дискусійною. Але такий підхід виглядає цілком логічним і доктринально обґрунтованим способом реалізації права.

Іншою проблемою є розбіжності у використанні згаданої односторонньої заяви на практиці. Саме по собі її застосування є стандартом, але формати часто відрізняються, як у світовому так і (особливо) в українському контексті. Подібні розбіжності можна дуже часто побачити, починаючи від вже обговореного написання фрази – «захищено», замість «reserved» / «зарезервовано» / «збережено», і закінчуючи невірним зазначенням правовласника, наприклад, у абсурдних форматах на кшталт «права належать вебсайту», «зарезервовано торговельною маркою» тощо [14].

Цікаво, що Законом визначено можливість, за якої «суб'єкт авторського права для повідомлення про свої майнові права може використовувати знак охорони авторського права, що складається з латинської літери «С», обведеної колом – ©, поряд з яким зазначаються ім'я (найменування) суб'єкта майнових прав на твір, рік першого опублікування твору», але у цій статті 9 немає інформації про функціональність додаткових фраз.

Отже, проблема є багатосторонньою як у європейському, так і в українському контекстах, але вітчизняна практика, на жаль, поки що демонструє обмежене реагування на згадані виклики. Особливу роль повинна відігравати держава у встановленні мінімальних технічних і правових стандартів доступу до цифрового контенту. За відсутності таких рамок рішення щодо формату оповіщення щодо використання TDM перекладається на приватних осіб, що може породжувати хаос.

Висновки. Таким чином, TDM є ключовим інструментом для машинного навчання і досліджень, однак через недостатність обов'язкових винятків і різноманітність національних підходів, TDM стикається з правовими бар'єрами, що уповільнюють технічний прогрес, тоді як ефективне правове регулювання має передбачати ці аспекти у більш широкому масштабі.

Українське законодавство має бути удосконалено з урахуванням цих викликів та з метою підтримки розвитку технологій. Удосконалення будуть ефективними у напрямках розширення сфер вільного використання TDM та адаптації термінології до європейської та світової практики, а також юридичної сили та сенсу заяв.

Використання односторонніх заяв про авторські права на вебсайтах мають перспективи стати засобами, «придатними для зчитування комп'ютером з цифрового контенту» у міжнародному масштабі, що значною мірою вирішить проблему розбіжностей у регулюванні питання на рівнях національних юрисдикцій, але цей аспект потребує подальших емпіричних досліджень та порівняльного аналізу судової практики.



Список використаних джерел:

1. Fernández Molina J. C., Esteban De La Rosa F. Copyright and Text and Data Mining: Is the Current Legislation Sufficient and Adequate?. *Libraries and the Academy*. 2024. 24, № 3. P. 653–672.
2. Ueno T. The flexible copyright exception for ‘non-enjoyment’ purposes – recent Amendment in Japan and its implication. *GRUR International*. 2021. 70, № 2. P. 145–152.
3. Quintais J. P. Generative AI, copyright and the AI Act. *Computer Law & Security Review*. 2025. 56. Article № 106107.
4. Fiil-Flynn S. M., Butler B., Carroll M., Cohen-Sasson O., Craig C., Guibault L., Contreras J. L. Legal reform to enhance global text and data mining research. *Science*. 2022. 378, № 6623. P. 951–953.
5. Ducato R., Strowel A. M. Ensuring text and data mining: remaining issues with the EU copyright exceptions and possible ways out. *European Intellectual Property Review*. 2021.
6. Шматков Д. І., Коломійцев С. О. Використання інформації з відкритих джерел та право інтелектуальної власності. *Право та інновації*. 2024. 2(46). С. 195–201. DOI 10.37772/2518-1718-2024-2(46)-29.
7. Carroll M. W. Copyright and the progress of science: why text and data mining is lawful. *UC Davis Law Review*. 2019. 53. P. 893–964.
8. Tóth A. K. Algorithmic copyright enforcement and AI: Issues and potential solutions through the lens of text and data mining. *Masaryk University Journal of Law and Technology*. 2019. 13, № 2. С. 361–388.
9. Бондарчук Ю. А., Бороліс І. І., Вишнеvsька М. О. Англо-український словник юридичних термінів : навч. посіб. Київ : КНУТД, 2021. 256 с.
10. Sganga C., Contardi M., Turan P., Signoretta C., Bucaria G., Mezei P., Harkai I. Copyright flexibilities: mapping and comparative assessment of EU and national sources [Електронний ресурс]. 2023. Режим доступу: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4325376.
11. Hook S. A. Protecting Content Online: The Interface of Copyright and Design for Websites, Apps, and GUIs. AIPLA Spring Meeting. 2016.
12. Portmann L. Crafting an audience: UX writing, user stylization, and the symbolic violence of little texts. *Discourse, Context & Media*. 2022. 48. Article № 100622. Режим доступу: <https://doi.org/10.1016/j.dcm.2022.100622>.
13. Gray C. M., Gairola R., Boucaud N., Hashmi M., Chivukula S. S., Menon A. R., Duane J. N. Legal Trouble?: UX Practitioners' Engagement with Law and Regulation. Companion Publication of the 2024 ACM Designing Interactive Systems Conference. 2024. P. 106–110.
14. Шматков Д. І., Пашнєв Д. В., Хлестков О. В., Коломійцев С. О. Юридичне значення односторонніх заяв про авторські права на вебсайтах. *Вісник Харківського національного університету внутрішніх справ*. 2024. 107(4). С. 231–243. doi: 10.32631/v.2024.4.21.

Дата першого надходження рукопису до видання: 19.07.2025

Дата прийнятого до друку рукопису після рецензування: 21.08.2025

Дата публікації: 26.09.2025

