

УДК 34:004+342.9.

**ДУБНЯК М.В.**, кандидат юридичних наук, завідувач наукової лабораторії правового забезпечення цифрової трансформації, Наукового центру цифрової трансформації і права ДНУ ІБП НАПрН України.  
ORCID: <https://orcid.org/0000-0001-7281-6568>.

## ВІДКРИТІ ДАНІ ТА ПОЯСНЮВАЛЬНИЙ ШТУЧНИЙ ІНТЕЛЕКТ: ПРАВОВІ ПЕРСПЕКТИВИ DOI...

**Анотація.** У статті комплексно проаналізовано правові проблеми відкритих даних. Досліджено поняття “генеровані громадянами дані” та визначено їх роль, місце та значення для правового регулювання у системі відкритих даних, описано ознаки відокремленості цієї категорії даних від схожих за принципом формування даних “громадської науки” та даних “громадської участі”. Осмислено процес формування наборів навчальних даних для розвитку технологій штучного інтелекту з використанням відкритих даних. Досліджено поняття та особливості “пояснювального штучного інтелекту”. Встановлено, що при розробці “алгоритмів надання пояснень” мають враховуватись людські процеси мислення та когнітивні упередження особи, яка формує, сприймає та оцінює рішення пояснювального штучного інтелекту. Досліджено питання використання відкритих даних у інших сферах, зокрема для аналізу показників Цілей Сталого Розвитку (ЦСР). Виявлено, що частина даних необхідних для порівняння індикаторів ЦСР, публікуються одними і тими ж Розпорядниками на різних платформах. Це призводить до дублювання частини даних, необхідності їх перетворення або компіляції для опублікування на Порталі відкритих даних та Відкритій платформі ЦСР в Україні.

**Ключові слова:** відкриті дані, пояснювальний штучний інтелект, баланс прав, публічна інформація, прийняття рішень, набори даних, захист даних, правові режими даних, звільнення даних, Цілі Сталого Розвитку.

**Summary.** The article comprehensively analyzes the legal issues of open data. The concept of “citizen-generated data” is explored, and its role, place and importance for legal regulation in the open data system are determined. The paper also describes the features that distinguish this category of data from similar “citizen science” and “public participation” data. The process of forming training datasets for the development of artificial intelligence technologies using open data is comprehended. The concept and features of “explanatory artificial intelligence” are investigated. It is established that the development of explanation algorithms should take into account the human thought processes and cognitive biases of the person who forms, perceives and evaluates the decisions of explanatory artificial intelligence. The use of open data in other areas, in particular for analyzing the indicators of the Sustainable Development Goals, is investigated. It was found that some of the data necessary for comparing SDG indicators are published by the same data managers on different platforms. This leads to duplication of some data, the need to transform or compile data for publication on the Open Data Portal and the Open SDG Platform in Ukraine.

**Keywords:** open data, explanatory artificial intelligence, balance of rights, public information, decision-making, datasets, data protection, data legal regimes, data liberation, Sustainable Development Goals.

**Постановка проблеми.** Технології штучного інтелекту (далі – ШІ) активно впроваджуються у всіх сферах суспільного життя, формуючи масштабні соціальні трансформації. Поява нових рішень з використанням технологій ШІ відбувається

швидше, ніж формування релевантного правового регулювання. Використання різних даних для навчання генеративних мовних моделей (наприклад, таких як чат GPT від OpenAI, Gemini від Google, Bing від Microsoft та інші), машинного навчання (розпізнавання облич, образів, усної мови), спричиняє правові та етичні проблеми. Зокрема, для формування наборів даних на етапі розробки і розгортання моделей ШІ, відбувається збір даних з усіх можливих джерел в мережі Інтернет. Це персональні дані, інформація з профілів користувачів соціальних мереж, об'єктів інтелектуальної власності, даних наукових репозитаріїв, веб-сайти різного тематичного спрямування. Зазначені дані охороняються за різними правовими режимами. Наприклад, для обробки персональних даних необхідна інформована згода суб'єкта даних. Через комбінацію у наборах навчальних даних можливе виконання критерію “конкретно ідентифікована особа”. Це означає, що через використання різних наборів даних, їх комбіноване поєднання в наборі навчальних даних для ШІ, розробники отримують набір даних, які дозволяють конкретно ідентифікувати особу, і такі набори даних підпадають під правову охорону персональних даних. Для використання в наборах навчальних даних зображень, текстів, музики, відео розробникам необхідно отримати дозволи та ліцензії від осіб, які володіють правами інтелектуальної власності. Оскільки в наборах навчальних даних використовуються сотні, а може і десятки тисяч окремих об'єктів, які охороняються за законодавством про інтелектуальну власність, укладання десятків або сотень ліцензійних угод на кожен такий об'єкт значно ускладнить розробку технології. У такому висококонкурентному середовищі, як розвиток технологій ШІ, швидке і законне отримання даних для формування навчальних наборів даних має виняткове значення.

Однак, на етапі формування навчальних наборів даних, ігнорування правових норм розробниками ШІ призводить до безпрецедентного порушення законодавства. Це негативно впливає на результати застосування технологій ШІ. Використання неповних, недостовірних даних має як наслідок формування неточних відповідей технологіями ШІ, що ускладнює застосовування розроблених програм за призначенням.

**Результати аналізу наукових публікацій.** Загальні правові проблеми застосування технологій ШІ для вирішення проблем прийняття рішень досліджувалось в декількох монографіях [1; 2]. Проблеми використання окремих категорій даних для навчання технологій ШІ, розглядалися на прикладі он-лайн репозитаріїв [3], медичних даних [4]. Застосування систем ШІ для вирішення практичних проблем та підтримки прийняття рішень досліджувалися на прикладі сфери публічних закупівель [5]. Проблематикою вивчення генерованих громадянами даних займалися: О. Corcho, J. Blanco, С. Morote, Simperl E. [6], А. Meijer, S. Potjer [7], А. Suman [8], М. Ponti, М. Craglia [9]. У цих наукових працях були запропоновані категорії генерованих громадянами даних, а також наведені їх характеристики для класифікації. Визначені масштаби їх збору: (від волонтерських ініціатив до урядової політики використання генерованих громадянами даних) [6, с. 10]. Але при цьому, не визначено змістовних ознак генерованих громадянами даних для правового регулювання, не сформовано правове визначення дефініції для закріплення в нормативних актах.

**Метою статті** є визначення правового режиму “відкритих даних” та категорії “генеровані громадянами дані”, з метою законного формування навчальних наборів даних для пояснювального ШІ, а також пошук переваг і особливостей застосування пояснювального ШІ для держави і суспільства.

Завдання роботи полягає у визначенні стану правового регулювання відкритих даних, на прикладі законодавства України; з'ясуванні ролі “генерованих громадянами

даних” (англ. – citizen-generated data, далі – CGD) в системі відкритих даних; встановленні особливостей, переваг і недоліків ХАІ у порівнянні з використанням інших технологій ШІ для держави і суспільства; з’ясуванні ролі відкритих даних для використання в окремих сферах, наприклад, для розробки пояснювального ШІ або для аналізу індикаторів Цілей сталого розвитку.

### **Виклад основного матеріалу.**

#### **1. *Праве регулювання відкритих даних в Україні.***

З 2015 року і до цього часу триває розвиток правового забезпечення відкритих даних в Україні. В основних нормативних актах визначені вимоги до формату, структури, періодичності оновлення даних, порядку оприлюднення, переліків наборів даних, які підлягають оприлюдненню у формі відкритих даних та механізм ведення Порталу відкритих даних [10 – 12].

Запровадження політики відкритих даних в Україні не супроводжувалося затвердженням єдиної концепції чи стратегії. Не були закріплені в програмних документах план дій, критерії та очікувані результати реалізації політики у сфері відкритих даних, тому оцінити стан її виконання неможливо [13]. З 2015 року до Положення про набори даних, які підлягають оприлюдненню у формі відкритих даних, зміни вносились понад 10 разів, в частині збільшення наборів даних та кількості Розпорядників. Так, в першій редакції Положення обов’язковому оприлюдненню підлягали 327 наборів від 35 Розпорядників [11 (в ред. від 21.10.15 р.)], станом на 2021 рік це вже 887 наборів від 68 Розпорядників [11 (в ред. від 03.03.21 р.)].

Згідно затвердженого Плану дій для реалізації принципів Міжнародної хартії відкритих даних усі заходи для формування організаційної та інституційної спроможності були заплановані на 2018 – 2020 роки [14]. У 2020 році підготовлено аналітичний звіт щодо стану виконання законодавства про відкриті дані за період з 2015 по 2019 роки. Загальними проблемами правового регулювання оприлюднення відкритих даних можна назвати наступні:

1. Надмірна кількість Розпорядників (понад 85 тисяч суб’єктів, за даними Державної служби статистики, станом на 2020 рік).

2. Оприлюднення даних низької якості (не розроблені вимоги виконання у відповідних форматах, немає стилізованих форм даних, і це все унеможлиблює їх автоматизовану обробку). Причинами цього є:

2.1. кадрова проблема (відсутність належної технічної компетенції відповідальних осіб, відсутність ІТ-спеціалістів у Розпорядників);

2.2. операційна проблема (створення нормативно-розпорядчих документів у формі таблиць, і їх подальше автоматичне перетворення у формати відкритих даних, що займає великий обсяг часу. Також це призводить до втрати частини даних через різну внутрішню структуру файлів та неправильне відображення атрибутів файлу; оприлюднення даних у форматах, непридатних до машинної обробки, наприклад геопросторові динамічні дані у форматі картинок, або файли у форматі “pdf.” – у вигляді сканованих картинок, без текстового змісту);

2.3 організаційна проблема (несвоєчасне оновлення необхідних наборів даних, невірне визначення Розпорядниками періодів оновлення наборів даних).

Відкриті дані низької якості мають такі властивості:

а) неповні дані (наприклад, дані транспортних засобів не містять їх ідентифікаторів, що унеможлиблює їх використання на он-лайн картах відстеження громадського транспорту);

б) дані, які неможливо порівняти в однакових наборах від різних Розпорядників (різна структура файлів, різні одиниці вимірювання, тощо);

в) несинхронне оновлення в однакових наборах даних від різних Розпорядників (наприклад, Розпорядники обирають різні періоди оновлення для однакових наборів даних) [15, с. 45].

3. Порушення принципу “відкритості за замовчуванням” усієї публічної інформації. Така ситуація, очевидно, складається через наявність нормативного акту, який визначає конкретні набори відкритих даних [11], а Розпорядники помилково сприймають це правове регулювання як таке, що встановлює пріоритет для публікації визначених наборів даних, замість публікації усієї публічної інформації у формі відкритих даних.

Таким чином, система відкритих даних має низку не вирішених до цього часу проблем, які потрібно усунути шляхом прийняття Стратегії відкритих даних із визначеними показниками, а також сформуванню в ній напрямки застосування таких даних для розробки технологій ШІ.

Формування Стратегії щодо використання відкритих даних для усунення наявних проблем правового регулювання може включати такі заходи:

- встановлення вимог до збору і створення даних одразу у стандартах, які допустимі для автоматизованої обробки (машиночитаний формат);
- встановлення пріоритетних наборів даних за чіткими вимогами (структура, атрибути, формат даних) та періодами їх створення та оприлюднення;
- впровадження стандартів до пріоритетних наборів даних;
- розробка та використання єдиних програмних засобів для всіх Розпорядників.

## **2. Роль та значення “генерованих громадянами даних”.**

Особливістю відкритих даних є те, що вони створюються та оприлюднюються Розпорядниками публічної інформації. Однак, у сучасному глобалізованому світі величезний обсяг даних генерується користувачами під час використання різних смарт-речей, мобільних додатків до них та інших програм. Це відбувається у процесі повсякденної життєдіяльності мільярдів користувачів. Частина таких даних збирається компаніями-власниками технічних пристроїв, та використовується ними для зміни і удосконалення власних бізнес-моделей [16, с. 65]. Як правило, зібрані у такий спосіб дані захищаються приватними компаніями як конфіденційна або бізнес-інформація, і отримати такі структуровані набори даних у вільному доступі неможливо. За умови набуття такими приватними компаніями статусу Розпорядника (наприклад, вони займають домінуюче становище на ринку, або наділені спеціальними чи виключними правами, або є природними монополіями) вони зобов'язані публікувати дані про умови постачання товарів, послуг та цін на них [10]. Такі дані часто є легкодоступними на веб-сайтах самих компаній, адже становлять частину їх маркетингової стратегії. Дані про товари, послуги та ціни на них є менш цінними для використання в наборах навчальних даних, якщо їх порівнювати з даними, які отримують компанії від громадян – користувачів смарт-речей, мобільних додатків та інших програм.

Цінність “генерованих громадянами даних” полягає у можливостях різноманітного використання, наприклад, в наборах навчальних даних для розвитку ШІ, або для різноманітних комерційних цілей. Існує велика інституційна структура у формі Порталів відкритих даних для публікації наборів відкритих даних Розпорядниками. Важливо дослідити, чи є на цих Порталах можливості для публікації даних не тільки Розпорядниками, а і самими громадянами. Така соціальна модель опублікування даних

дозволить усунути із процесу збору генерованих громадянами даних приватні компанії, і зробити ці дані відкритими і публічними для всіх зацікавлених осіб.

Серед дослідників немає єдиної дефініції для поняття “генеровані громадянами дані” (англ. – Citizen-generated data, далі – CGD). Зокрема, є наступні погляди на цю категорію:

- Дані, свідомо створені людьми, є відкритими та доступними у суспільному надбанні [7].
- Ресурси, які створюють громадяни, дозволяють ставити під сумнів вироблені експертні знання. За допомогою оцінки цих даних можна сформувані реальні альтернативні погляди на одну і ту саму проблему. У підсумку ці дані можна розглядати як форму соціального протесту [8].
- Відкриті дані необхідні для державного управління, оскільки їх використання сприятиме узгодженню спільних дій зацікавлених осіб для вирішення суспільних проблем [9].
- Дані, які окремі люди чи спільнота громадян створюють для безпосереднього моніторингу, або стимулювання змін у питаннях, які їх стосуються [17]. Тобто як один з інструментів для участі в управлінні державними справами.
- Дані, які забезпечують пряме представлення їх інтересів і є альтернативним набором даних, у порівнянні з даними які збирають уряди та міжнародні організації [6, с. 8].

Отже, поняття CGD сильно залежить від контексту дослідження і описується як через інформаційні категорії: “дані”, “збір”, “обробка”, так і категорії інтелектуальної власності: “суспільне надбання”. Це поняття також як самостійний інструмент прямої демократії у формі участі громадян в управлінні державними справами: “вирішення питань, які їх стосуються”, “альтернативні погляди”.

Зокрема, у національних Порталах Чехії, Франції, Польщі, Іспанії та Об’єднаного королівства вже існують модулі для CGD. Однак активність громадян у публікаціях є низькою, менше 1 % у порівнянні з даними, які оприлюднюють державні органи. Найчастіше публікують геопросторові дані, та дані з розділів “питання і відповіді” з інших веб-порталів органів державної влади, також “опитування” та “статистика”. Такі набори даних корисні для врахування у процесі прийняття рішень органами влади різних рівнів, але розробники контент-проектів на базі відкритих даних ці дані не використовують [6, с. 15-17].

У якості експерименту перевірено декілька національних порталів, зокрема, Чехії (data.gov.cz), Франції (data.gouv.fr), Великобританії (data.gov.uk) за пошуковим запитом “згенеровані громадянами дані”. Виявлено такі набори даних: середовище проживання, заяви про оренду квартир, дослідження трафіку громадського транспорту, опитування підприємців, обстеження торговельної мережі (декілька окремих міст), перелік запитань і відповідей Президента щодо декількох напрямків політики, дані температурних рекордів британського шельфу, дослідження морської біологічної асоціації, дані про пошкоджені дерева.

За такими категоріями розміщених даних не дивно, що поняття CGD змішується із “даними про участь громадян в управлінні державними або місцевими справами”, або “громадською наукою” (краудсорсингова або волонтерська наука) Для відмежування цих понять від CGD наведемо декілька прикладів.

“Громадська (краудсорсингова, волонтерська) наука” характеризується широким залученням зацікавлених осіб для надання конкретно визначених даних, які будуть

використані в наукових дослідженнях [19]. Тобто кваліфікуючими ознаками цих даних є такі:

- мета їх збору — для наукової діяльності;
- визначена тематика даних (наприклад, про стан здоров'я).

Така співпраця може бути в різних формах, наприклад розміщення будь-якою особою, на спеціально визначених інформаційних ресурсах конкретно визначених даних. Наприклад, за таким принципом було створено найбільшу базу симптомів коронавірусу Covid 19 – Zoe [20].

Для правового регулювання CGD суттєвими характеристиками є такі:

- дані генеруються окремими громадянами, а не Розпорядниками даних;
- дані генеруються громадянами у процесі їх повсякденної життєдіяльності через використання різних смарт-речей, мобільних застосунків, інших програм;
- у громадян є можливість самостійно завантажувати необхідні набори даних в спеціально створені модулі на Порталах відкритих даних;
- набори даних завантажуються у машиночитаних форматах.

Для удосконалення правового регулювання більш цікавий погляд на дослідження CGD у контексті юридичних зобов'язань, порядку розміщення даних, залучення громадськості до публікації даних за власною ініціативою. Саме ця категорія є найбільш цікавою для нашого аналізу, оскільки забезпечує побудову спільнот та екосистем даних, де різні зацікавлені сторони залучені до всього життєвого циклу обороту інформації в кожній екосистемі даних. Це дозволяє у публічному відкритому доступі розмістити дані, доступні для повторного використання без участі корпорацій, які розробляють смарт-речі, через використання яких йде масовий збір статистичних даних у власних цілях окремих виробників.

Отже, *генеровані громадянами дані* – це дані, які генерують особи у процесі використання технологій та програм, та можуть бути самостійно розміщені такими особами як відкриті дані при використанні спеціальних модулів Порталу відкритих даних.

В Україні відсутнє правове регулювання такої категорії даних, що є недоліком національного законодавства. Враховуючи заяву Міністерства цифрової трансформації про формування візії та стратегії розвитку відкритих даних в Україні [21], така категорія повинна неодмінно з'явитись в майбутній Стратегії, законах і інших нормативно-правових актах про відкриті дані.

Звичайно, що для CGD будуть характерні всі описані проблеми оприлюднення відкритих даних. Щоб вони не повторилися в стратегічних документах, необхідно передбачити формування рівня правової культури у громадян:

- навіщо їм потрібно публікувати дані зі своїх смарт-речей;
- як знеособити дані і зберегти конфіденційність;
- як підвищити рівень технічної та цифрової грамотності;
- як забезпечити порівнюваність даних – через їх періодичну публікацію в стандартизованих оприлюднених форматах та інші питання, які мають бути комплексно враховані у майбутній Стратегії відкритих даних.

### **3. Особливості пояснювального штучного інтелекту у порівнянні з іншими технологіями ШІ.**

Системи підтримки прийняття рішень на основі ШІ (далі – Системи з ШІ) доступні у багатьох різноманітних сферах. Однак, ефективність застосування цих систем

обмежена через відсутність можливостей перевірки кінцевими користувачами надійності та достовірності наданих рекомендацій, висновків і прогнозів.

Розробка технологій пояснювального ШІ (ХАІ) вимагає прозорості, а також різних інтерпретацій наданих прогнозів за запитом користувача. Прозорість – це здатність зрозуміти, як модель ШІ дає рекомендації, а можливість інтерпретації – це здатність зрозуміти, чому саме ці рекомендації надає ШІ. Відсутність прозорості та різних інтерпретацій наданих результатів – прогнозів ШІ, призводить до відсутності довіри до його рішень. Моделі ШІ потребують великої кількості масивів різноманітних високоякісних даних, щоб відповідати критеріям прозорості, точності та інтерпретації. Ці масиви повинні бути надійними та правдивими, а самі Системи ШІ потрібно навчати на різноманітних, достовірних, якісних і навчальних наборах Великих даних [4].

Сучасні технології ШІ містять так звану “чорну скриньку” в якій і відбувається процес формування рішення. Для таких сфер як медицина, право, безпека і оборона критично важливо розуміти причини прийнятого рішення, а не лише погодитись із згенерованим результатом. Такою ж “чорною скринькою” є набори навчальних даних, тому невідомо, які саме дані проаналізував ШІ і чи дійсно вони враховують усю специфіку сфери.

Агентство передових оборонних дослідницьких проєктів США (Defense Advanced Research Projects Agency, USA) з 2018 року підтримує розробку пояснювального ШІ. Зокрема, очікується створення нових методів машинного навчання, які допоможуть:

1. створювати більш зрозумілі моделі ШІ, зберігаючи при цьому високий рівень ефективності навчання (точність передбачення); і

2. дозволять користувачам належним чином зрозуміти, довіряти та ефективно застосовувати системи підтримки прийняття рішень на основі ШІ [22].

Отже, пояснювальний штучний інтелект (інтерпретований ШІ) – метод машинного навчання, який надає пояснення стосовно прийнятого рішення.

Особливістю ХАІ є врахування контексту та середовища, в якому вони працюють [22], а не тільки наборів навчальних даних та комбінацій різних методів машинного навчання на етапі розробки технології.

Як бачимо з Таблиці 1, у ланцюжку роботи ШІ присутній “кінцевий користувач” – людина, яка формує запити для отримання рішення. Системи підтримки прийняття рішень на основі ШІ – це не тільки алгоритми та методи машинного навчання. Це ще і психо-когнітивні процеси, рівень розуміння технологій та рівень компетенцій конкретного користувача. А тому досягнення когнітивної науки для процесу розробки ХАІ є важливим, оскільки дозволяє врахувати особливості людського сприйняття пояснень, які генерують ХАІ. Це значно розширює набір методів та комбінацій аргументів для користувача, у процесі формування різних та зрозумілих йому пояснень.

У правових дослідженнях ШІ багато уваги приділено проблемі “прозорості”. Вона розглядається з трьох точок зору:

- прозорості процесу розробки технології;
- політики взаємодії компанії-власника технології з користувачами;
- готовності розробників систем з ШІ нести відповідальність.

• Втрата переваги на ринку високих технологій для компаній-розробників ШІ була б тоді, якби вони розкривали усі технологічні особливості роботи ШІ, при тому, що розробка таких систем вимагає значних матеріальних ресурсів. Враховуючи це, розробка ХАІ переводить дискусію щодо проблеми “прозорості” від проблеми “розкриття корпоративних секретів”, у дискусію щодо можливого “переконання” кінцевого користувача у логічності процесу міркування та вирішення певної, заданої

користувачем, проблеми. Він також може побачити і перевірити хід міркування та альтернативних суджень ХАІ.

Таблиця 1.  
Порівняння моделей сучасного ШІ та ХАІ

<b>Сучасний ШІ</b>					<b>Проблемні питання кінцевого користувача</b>	
1. Набори навчальн. даних	2. Методи машинного навчання	3. Конкретні функції ШІ	4. Запит кінцевого користувача до ШІ	5. Згенерований ШІ результат (прогнозна аналітика, зображення, текст, відео, нові зв'язки і залежності в заданому наборі даних тощо)	Як ти отримав такий результат? Чому саме це рішення? Чи є помилки в твоїх міркуваннях? Як я можу довіряти твоїм рішенням? Як мені виявити та усунути помилку?	
<b>Пояснювальний ШІ (ХАІ)</b>					<b>Нові можливості кінцевого користувача</b>	
1. Набори навчальних даних	2. Нові методи машин. навчання	3. Поясн. модель прийняття рішення	3.1. Поясн. інтерфейс	4. Запит кінцевого користувача до ШІ	5. Згенерований ШІ результат (прогнозна аналітика, зображення, текст, відео, нові зв'язки і залежності в заданому наборі даних тощо)	Я розумію як ШІ отримав такий результат. Я розумію, чому саме це рішення, а не альтернативне. Я розумію як саме ШІ отримав це рішення, можу довіряти. Я розумію де і як шукати помилку.

Прим. Сформовано автором з урахуванням [22].

Зазвичай, людина приймає рішення через емоцію суб'єктивного задоволення, радість від того, що вона розуміє причинно-наслідкові зв'язки між явищами і контрфактичними поясненнями. Такі пояснення вказують на те, які альтернативні варіанти вирішення проблеми міг би надати ШІ, у разі формування альтернативного пошукового запиту. Різна комбінація методів машинного навчання та стратегій пояснень призводить до різних інтерпретацій користувачами отриманих від ШІ результатів.

Розглянемо приклад. Особі було відмовлено у наданні кредиту. Якщо користувач поставить ХАІ питання про те, чому саме йому відмовили в кредиті, він може очікувати наступні пояснення:

- встановлення причинно-наслідкових зв'язків: тобто пояснень того, як саме ХАІ отримав таке рішення;



- які вхідні дані користувача пов'язані з навчальними даними ХАІ: наприклад, це дані про те, що в кредитах було відмовлено іншим заявникам з подібним рівнем заробітної плати, професійним статусом, кредитною історією, як у користувача.

- яка мета рішення ХАІ: наприклад, що рішення такого характеру зменшують ризик дефолту через невиконання кредитів.

Кожен тип пояснення призведе до того, що користувач буде по-іншому розуміти різні аспекти системи ШІ, а це впливає на довіру до його рішень [23, с. 6539].

Наслідком для розробників ХАІ є те, що принаймні в деяких ситуаціях користувачі можуть віддати перевагу простим поясненням рішень, які були видані ХАІ. Багато пояснень в ХАІ здійснюються на основі зміни одного вхідного параметра, наприклад, “якби ви заробили на \$10 000 більше, ваша заявка на кредит була б схвалена”.

Інші особливості довіри до ХАІ пов'язані із відмінностями сприйняття дійсності, і як наслідок – різне причинно-наслідкове дискусювання отриманих фактів в думках самих користувачів. Це різні процеси мислення, за якого різним фактам надаються різні пріоритети для врахування і отримання висновку. При цьому інші факти пригнічуються (не враховуються, ігноруються у процесі обдумування). Тобто, такі користувачі можуть знехтувати запропонованим причинно-наслідковим поясненням від ХАІ, якщо вони знають інші альтернативні причини. Таким чином, потенційним наслідком прийняття рішень ХАІ є необхідність враховувати, що надійність і точність пояснень, які надаються користувачам, можуть бути відхилені людьми на основі власної інтерпретації відповідних причинно-наслідкових зв'язків.

Отже, щоб автоматизовані пояснення, надані ХАІ, були зрозумілими для користувачів, необхідно врахувати ключові відкриття когнітивної науки про людську пояснювальну здатність. Це має важливе значення для розробки ХАІ, щоб гарантувати, що пояснення, згенеровані Системами ШІ, можуть безпосередньо бути сприйняті користувачами [23, с. 6542].

Для формування якісного набору навчальних даних ці дані не повинні знаходитись в режимі правової охорони, бути відкритими та доступними для розробників ХАІ. Неповні, неточні, неякісні дані призводять до формування наборів даних з упередженнями, відтак якісного пояснення ХАІ з такими наборами навчальних даних не нададуть.

Запропонуємо гіпотезу: *відкриті дані за своїм правовим режимом можуть використовуватись в наборах навчальних даних для ХАІ без порушення правових режимів охорони даних.*

#### **4. Роль відкритих даних для розробки ХАІ і досягненні Цілей Сталого Розвитку.**

Аналіз проблем у сфері відкритих даних дозволяє обґрунтувати положення про те, що дослідження правових аспектів застосування технологій ШІ і відкритих даних можна розглядати з двох позицій.

Перша позиція: програмні рішення на базі ШІ можуть покращити роботу Порталу відкритих даних – вони дозволять аналізувати сформовані Розпорядниками набори даних, виявляти в них неточності та протиріччя, невідповідність форматів даних та змісту інформації, що значно удосконалив технічний супровід Порталу.

Друга позиція: використання якісних відкритих даних у наборах навчальних даних для ШІ дозволить розробити спеціалізовані програми для підтримки процесів прийняття рішень в органах державної влади усіх рівнів. Це включає як кількісну оцінку даних, так і оцінку перспектив їх використання у різних бізнес-моделях, в тому числі для навчання ХАІ.

В Україні вже існує позитивний приклад впровадження технологій ШІ до систем моніторингу публічних закупівель. Зокрема, модулі ШІ можуть бути використані для:

- обробки великого обсягу даних;
- формування прогнозу і плану закупівель для конкретного органу державної влади;
- виявлення порушень тендерних процедур;
- контролю вірогідності при неправильному заповненні тендерної документації;
- знаходження тендерів з несправедливим вибором переможця;
- виявлення несправедливої дискваліфікації учасників;
- пошуку змови між учасниками;
- обґрунтування неправомірної відмови в участі в торгах.

У сфері публічних закупівель алгоритми ШІ можуть:

- інформувати та консультувати учасників щодо останніх змін в законодавстві;
- здійснювати моніторинг заявок та аналітичні прогнози;
- автоматизувати трудомісткі, рутинні, стандартні, типові та повторювані операції;
- розробити стратегії управління закупівлями.

Наразі технології ШІ у сфері публічних закупівель реалізовані як поодинокі модулі підтримки окремих процесів (спеціалізовані чат-боти для вирішення окремих задач, формування прогнозів за визначеними ризик-індикаторами) [5, с. 132-133].

Даний приклад ілюструє, як можуть вирішуватись проблеми у певній сфері завдяки застосуванню технологій ШІ, який натренований на наборах необхідних даних.

Застосування технологій ШІ можна розділити на 3 покоління (етапи становлення):

1 етап – ШІ здатний автоматизувати рутинні процеси, які не змінюються з часом.

2 етап – ШІ здатний адаптуватися до змін автоматизованого процесу та вчитися на них.

3 етап – ШІ активно надає нові та інноваційні ідеї, маючи можливість аналізувати свої попередні дії та вчитися на них.

Тобто, ШІ на кожному етапі виконує принаймні одну з наступних функцій: моніторинг, виявлення, прогнозування, інтерпретація, взаємодія з фізичним середовищем, взаємодія з людьми та взаємодія з машинами.

Для прикладу, система ШІ Міністерства юстиції Великої Британії може сортувати, індексувати та узагальнювати документи, на “привілейовані” та “непривілейовані”. Однак, використання ШІ в державному секторі є радше винятком, ніж звичайною практикою [18].

Відкриті дані можуть сприяти швидкому розвитку технологій ШІ через доступ до великих масивів відкритих даних, а технології ШІ, навчені на відкритих даних, можуть виявити незвичайні закономірності.

Отже, поєднання відкритих даних, алгоритмів ШІ та баз даних бізнесу надають можливість виявити такі моделі (Таблиця 2 [24, с. 13]).

Важливо сформулювати застереження стосовно використання відкритих даних для розробки конкретних рішень на базі ШІ і їх застосування для підтримки процесів прийняття рішень. У масивах відкритих даних уже існують дані, при порівнянні яких можна виявити корупційні фактори та інші рішення, які приймались посадовими особами та привели до порушень ними чинного законодавства. Скопійовані приховані проблеми, які потрапили в навчальні набори даних для програм на базі ШІ, призведуть до масштабування соціальних проблем і надання в аналітичних звітах упереджених рішень з високим корупційним фактором.

Таблиця 2.  
Матриця бізнес-моделей ШІ та відкритих даних

<b>Комерційна модель</b> Відкритий алгоритм Закриті дані	<b>Відкритий алгоритм</b>	<b>Відкрита модель</b> Відкритий алгоритм Відкриті дані
<b>Закриті дані</b> (персональні дані, об'єкти інтелектуальної власності, різні види таємниць – прим. авт.)	<b>Спільна модель</b> Спільний алгоритм Спільні дані	<b>Відкриті дані</b>
<b>Закрита модель</b> Закритий алгоритм Закриті дані	<b>Закритий алгоритм</b>	<b>Комерційна модель алгоритму</b> Закритий алгоритм Відкриті дані

Наприклад: “керівник Укрзалізниці, декларує у 2017 році заробітну плату в розмірі 157 тисяч грн. Протягом 2018-го року зарплата суттєво збільшується. У березні 2018 року це становить 1 мільйон 205 тисяч гривень, у лютому 2019 – понад 2 мільйони. Якихось кардинальних позитивних змін у діяльності Укрзалізниці не відбулося. Так за що ж отримуються такі захмарні винагороди, за які успіхи вони нараховуються?” [25, с. 26].

Отже, якщо ШІ буде порівнювати дані з декларації службовців, у яких є постійний приріст заробітної плати, та дані про якісні показники діяльності очолюваної таким службовцем інституції, в якій об'єктивно не сталось жодних покращень. Цей показник можна об'єктивно доповнити даними CGD, наприклад, замість гнівного посту в соціальних мережах розвивати культуру відгуків про роботу громадського та залізничного транспорту на Порталі відкритих даних). Помітно, що такий масив даних вже містить суспільну проблему корупційного характеру, коли керівника “винагороджують за бездіяльність”. Хоча зрозуміло, що такі дані мали б бути використані для притягнення його до відповідальності.

Інший приклад: застосування відкритих даних відкриває можливості для аналізу показників Цілей сталого розвитку (Таблиця 3 [26; 11]).

У поєднанні з технологіями ШІ, натренованими на відкритих даних про:

- показники погоди;
- супутникові зображення місцевості;
- історичні тенденції;
- частоту та інтенсивність санітарних вирубок;
- даних про рекреаційну привабливість зони (лісовий туризм).

Також можна передбачити лісові пожежі, чого неможливо досягти шляхом порівнянь статистичних даних.

Виявлення взаємозв'язків у таких наборах даних сприятиме кращому аналізу індикаторів Цілей сталого розвитку, зокрема ЦСР 15 “захист та відновлення екосистем суші” (раціональне лісокористування, боротьба з опустелюванням, припинення і повернення назад (розвертання) процесу деградації земель та зупинка процесу втрати біорізноманіття).

Таблиця 3.  
Співвідношення даних національних індикаторів ЦСР  
та наборів відкритих даних

<i>Дані національних індикаторів Цілей сталого розвитку [26]</i>	<i>Набори відкритих даних [11]</i>
<b><i>Ціль 3. Міцне здоров'я і благополуччя. Розпорядник даних МОЗ</i></b>	
3.3.1. Кількість хворих з діагнозом ВІЛ, на 100 тис. населення. 3.3.2. Кількість хворих з діагнозом активного туберкульозу, на 100 тис. населення.	Медична статистична звітність (за регіонами, віковими групами, видами захворювань тощо – прим. авт.).
<b><i>Ціль 4. Якісна освіта. Розпорядник даних МОН</i></b>	
4.6.1. Частка чоловіків серед педагогічних працівників.	Знеособлені дані про педагогічних працівників закладів освіти (але з поділом за гендером, та віковими групами – прим. авт.).
<b><i>Ціль 7. Доступна та чиста енергія. Розпорядник даних Міненерго</i></b>	
7.1.1. Виробництво електроенергії, млрд. кВт/год.	Виробництво електричної енергії атомними електростанціями (за блоками). Прогнозний баланс електричної енергії об'єднаної енергетичної системи України. Показники роботи атомних електростанцій.

*Прим.* Сформовано автором на підставі порівняльно-правового дослідження [26; 11].

При дослідженні Таблиці 3 можемо встановити, що частина даних, які необхідні для індикаторів ЦСР, публікуються одними і тими самими Розпорядниками, як на Порталі відкритих даних (<https://data.gov.ua>), так і передаються Державній службі статистики для розміщення на Відкритій платформі ЦСР в Україні (<https://sdg.ukrstat.gov.ua/uk>). Але з іншого боку, не всі національні індикатори ЦСР співвідносяться з відкритими даними, і спеціально створюються для подання звітності в Держстат. Таким чином, є дублювання частини даних на різних платформах, є зміна даних, оскільки для ЦСР необхідна лише частина з відкритого набору даних, є компіляція декількох наборів відкритих даних для визначення одного індикатора ЦСР. Така проблема існує через прийняття правового регулювання відкритих даних та ЦСР у різний час, слабку інституційну систему координації роботи щодо можливих сфер використання даних, не приведення у відповідність переліків наборів відкритих даних та даних індикаторів ЦСР.

Виявлені проблеми правового регулювання “оприлюднення публічної інформації у формі відкритих даних” має значно більші негативні наслідки, які виходять за межі прогалін законодавства у сфері доступу до публічної інформації. Зокрема, через неякісні набори даних гальмується інноваційний розвиток, через проблеми дублювання, зміни, компіляції даних посилюється навантаження на Розпорядників даних, і як кінцевий фактор – поява помилок та їх масштабування.

І це приносить значну шкоду в кінцевих рішеннях різних проблем, аналізу сфери відкритих даних та даних для індикаторів Цілей сталого розвитку.

**Висновки.****1. Щодо правового регулювання відкритих даних в Україні.**

У 2015 року шляхом внесення статті 10-1 “Публічна інформація у формі відкритих даних” до Закону України “Про доступ до публічної інформації” та прийняття інших нормативно правових актів на рівні Кабінету Міністрів України. При цьому, окремих документів стратегічного рівня про формування та розбудову системи відкритих даних, прийнято не було.

Після моніторингу Порталу відкритих даних вдалося виявити проблеми:

- надмірної кількості Розпорядників;
- оприлюднення даних низької якості через існування кадрових, організаційних та операційних прогалин в законодавстві.

- порушення принципу “відкритості за замовчуванням” усієї публічної інформації

Ці проблеми існують через неоднакове розуміння вимог щодо формату, структури, періодичності оновлення, переліку наборів даних, які були сформовані в підзаконних нормативних актах.

Для усунення виявлених недоліків, та розвитку правового забезпечення відкритих даних, необхідне системне оновлення законодавства з урахуванням перспектив розвитку сучасних технологій.

**2. Щодо ролі та значення “генерованих громадянами даних”.**

Поряд із відкритими даними, які створюють органи державної влади та інші Розпорядники, існує цілий пласт даних, доступ до яких наразі мають приватні компанії.

Вони досягають цього через особливі способи збору таких даних (метадані зі смарт-речей та програм). Для цілей нашої роботи, *генеровані громадянами дані* – дані, які генерують особи у процесі використання технологій та програм, та які можуть бути самостійно розміщені такими особами як відкриті дані на спеціальних модулях Порталу відкритих даних.

Правове регулювання “генерованих громадянами даних” з однієї сторони, дозволить вирівняти баланс прав в частині формування рівних умов для розміщення, аналізу, повторного використання даних усіма зацікавленими сторонами.

А з іншої сторони, це породить ряд проблем, як для держави, так і для громадян:

- для держави – створення та підтримки спеціальних модулів на Порталі відкритих даних;

- для громадян – користувачів програм та технологій – вимагатиме від них підвищення рівня технічної та цифрової грамотності і формування культури надання даних.

Генеровані громадянами дані відрізняються від схожих категорій “дані волонтерської науки”, “дані про участь в управлінні” за метою, способом їх збору, та конкретною визначеністю тематики даних.

**3. Щодо особливостей пояснювального штучного інтелекту у порівнянні з іншими технологіями ШІ.**

Пояснювальний штучний інтелект (інтерпретований ШІ) – це метод машинного навчання, який надає пояснення стосовно прийнятого ним рішення, або згенерованої відповіді. Інші технології ШІ мають, так звану “чорну скриньку”, коли навіть розробники технології не можуть пояснити, після аналізу яких саме даних, ШІ пропонує проект рішення чи відповіді. За умов існування “чорної скриньки”, неможливо перевірити прозорість, достовірність та точність запропонованої ШІ відповіді. Розробка ХАІ трансформує дискусії проблеми “прозорості” у формі “розкриття корпоративних секретів”, до дискусій щодо можливого “переконання” кінцевого користувача у

зрозумілості наданих ХАІ пояснень. У перевірці процесу міркування ХАІ, вирішення певної, заданої користувачем, проблеми, з можливістю побачити і перевірити хід висновків, наданих ШІ, та альтернативних суджень до них.

ХАІ може надавати пояснення про:

- причинно-наслідкові зв'язки між запитом та усіма альтернативними варіантами рішень;
- те, як запит користувача пов'язаний з навчальними даними ХАІ, які обумовили формування саме такого рішення на користувацький запит;
- яку глобальну мету переслідував ХАІ при формуванні пояснення.

На ступінь зрозумілості і довіри до наданих ХАІ пояснень впливає особливості індивідуального когнітивного сприйняття користувача, що має бути враховано при розробці різних алгоритмів формування пояснень ХАІ.

#### **4. Щодо ролі відкритих даних для розробки ХАІ і досягненні Цілей Сталого Розвитку.**

“Відкриті дані” та “генеровані громадянами дані” за своїм правовим режимом можуть використовуватись в наборах навчальних даних для ХАІ без порушення правових режимів охорони даних. Якісні набори відкритих даних можуть використовуватись для тренування ШІ, який потенційно може бути застосований у адміністративних відносинах.

Необхідно системно переглянути правове забезпечення різних режимів даних, наприклад шляхом розробки Національної стратегії у сфері даних, яка буде формувати екосистему обороту якісних даних для: прийняття громадянами, політиками, державними службовцями рішень, на базі повної та якісної інформації; підвищення прозорості діяльності органів державної влади, розробки якісних наборів даних для навчання ШІ та використання даних в інших інноваційних проектах, у тому числі для досягнення Цілей Сталого Розвитку.

Такий перегляд доцільно зробити для усіх етапів інформаційного процесу: збір, аналіз, ідентифікація інформації, яка має бути оприлюднена, обробка, конвертація, знеособлення, опублікування наборів даних, використання наборів даних у комерційних рішеннях.

Це можливо зробити через побудову відповідної соціально-правової моделі, та дослідження адміністративних та інформаційних відносин у цій моделі.

Правове регулювання відкритих даних має важливе значення для аналізу індикаторів ЦСР. Зокрема, цими даними можна доповнити статистику аналізу показників ЦСР в Україні, вони доступні для моніторингу та відстеження прогресу в різних сферах розвитку суспільства, важливі для прийняття та коригування рішень, можуть бути вирішальними для оцінки ефективності досягнення ЦСР та виявити області недоліків правового регулювання та прогалин законодавства.

#### **Використана література**

1. Баранов, О.А. (2018). Інтернет речей: теоретико-методологічні основи правового регулювання: монографія. Харків: Право, 2018. Т. 1: Сфери застосування, ризики і бар'єри, проблеми правового регулювання. 2-ге вид. 342 с.
2. Баранов О.А. (2022). Трансформація: соціальна & цифрова & правова: монографія у 3-х т. Т. 1. Порятунк цивілізації: економіка результату. Одеса: Видавничий дім “Гельветика”, 272 с.
3. Lilkendey J. (2023). Conserve the open media ecosystem! Legal and ethical considerations when using online repositories for AI training in ecological research. Doi: 10.1111/ele.14311.

4. G. Alam, I. McChesney, P. Nicholl, J. Rafferty (2023). Improving Prospective Healthcare Outcomes by Leveraging Open Data and Explainable AI, 3rd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, 2023. Pp. 486-492. DOI: 10.1109/ICCIT58132.2023.10273878.
5. Карпенко О., Карпенко Ю., Герман Д. (2023). Штучний інтелект як інструмент зниження корупційних ризиків у сфері публічних закупівель. *Public administration aspects*.
6. Publications Office of the European Union, Corcho, O., Jiménez Blanco, J., Morote, C. et al., (2022) Data.europa.eu and citizen-generated data – Opportunities and challenges associated to the inclusion of citizen-generated data in data.europa.eu. Publications Office of the European Union. URL: <https://data.europa.eu/doi/10.2830/137589>
7. Meijer A., Potjer S. (2018). Citizen-generated open data: an explorative analysis of 25 cases. *Government Information Quarterly*, Vol. 35, № 4, Elsevier. Pp. 613-621. URL: <https://doi.org/10.1016/j.giq.2018.10.004>
8. Suman A. B., Schade S. Abe, Y. (2020). Exploring legitimization strategies for contested uses of citizen-generated data for policy. *Journal of Human Rights and the Environment*. Vol. 11, № 3, Edward Elgar Publishing. Pp. 74-102. URL: <https://www.elgaronline.com/view/journals/jhre/11-3/jhre.2020.03.04.xml>
9. Ponti M., Craglia M. (2020). Citizen-generated data for public policy, Joint Research Centre. URL: <https://ec.europa.eu/jrc/communities/en/community/citizensdata/document/citizen-generated-data-public-policy>
10. Про доступ до публічної інформації (2011): Закон України № 2939-VI, в ред. 08.10.23 р. URL: <https://zakon.rada.gov.ua/laws/show/2939-17#Text>
11. Положення про набори даних, які підлягають оприлюдненню у формі відкритих даних (2015): постанова Кабінету Міністрів України від 21.10.15 р. № 835. URL: <https://zakon.rada.gov.ua/laws/show/835-2015-%D0%BF#Text>
12. Деякі питання оприлюднення публічної інформації у формі відкритих даних (2016): Постанова Кабінету Міністрів України від 30.11.16 р. № 867. URL: <https://zakon.rada.gov.ua/laws/show/867-2016-%D0%BF#Text>
13. Пояснювальна записка до проекту Стратегії державної політики у сфері відкритих даних (2021). URL: <https://thedigital.gov.ua/storage/uploads/files>
14. План дій з реалізації принципів Міжнародної хартії відкритих даних (2018): Розпорядження Кабінету Міністрів України від 21.11.18 р. № 900-р. URL: <https://zakon.rada.gov.ua/laws/show/900-2018-%D1%80#Text>
15. Зелена книга “Політика відкритих даних” (2020). Електронне видання: під. ред Дорогань О., Лавриненко І, Кобець Р. Київ. 78 с.
16. Дубняк М.В. (2023) Економіка даних: правовий та етичний аспект. *Інформація і право*. № 3(46)/2023. С. 64-74. DOI: [https://doi.org/10.37750/2616-6798.2023.3\(46\).287](https://doi.org/10.37750/2616-6798.2023.3(46).287).
17. Sieber R., Johnson P. (2015). Civic open data at a crossroads: dominant models and current challenges. *Government Information Quarterly*, Vol. 32, № 3, Elsevier. Pp. 308-315. URL: <https://doi.org/10.1016/j.giq.2015.05.003>
18. Publications Office of the European Union, Corcho, Ó., Jiménez, J., Simperl, E., (2023) Data.europa.eu and Citizen-generated Data – Identification of other types of citizen-generated data, Publications Office of the European Union. URL: <https://data.europa.eu/doi/10.2830/95762>
19. Kullenberg C. Kasperowski D. (2016) What is citizen science? – a scientometric meta-analysis. *PLoS ONE*, Vol. 11, № 1, PLOS, San Francisco, California, United States. URL: <https://doi.org/10.1371/journal.pone.0147152>
20. Zoe Health Study. URL: <https://zoe.com/learn/category/covid>
21. Мінцифри презентувало візію розвитку відкритих даних в Україні (2024). – (Урядовий портал від 08.03.24 р.). URL: <https://www.kmu.gov.ua/news/mintsyfry-prezentuvala-viziiu-rozvytku-vidkrytykh-danykh-v-ukraini>
22. Defense Advanced Research Projects Agency. Explainable Artificial Intelligence (XAI). URL: <https://www.darpa.mil/program/explainable-artificial-intelligence>

---

23. Byrne, R.M. (2023). Good Explanations in Explainable Artificial Intelligence (XAI): Evidence from Human Explanatory Reasoning. International Joint Conference on Artificial Intelligence. P. 6536-6544.

24. J. Keller, L. Chauvet, J.Fawcett O. Thereaux and oth. (2018) The role of data in AI business models. Open Data Institute Report, 32 p.

25. Корж І.Ф. (2019). Право на відкриті дані – як право приватного характеру. *Інформація і право*. № 1(28)/2019. С. 19-28.

26. Питання збору даних для моніторингу реалізації Цілей Сталого Розвитку (2019): Розпорядження Кабінету Міністрів України від 21.08.19 р. № 686-р.

~~~~~ \* \* \* ~~~~~