

ЗАХИСТ ЯКОСТІ ДАНИХ ОФІЦІЙНОЇ СТАТИСТИКИ В ЕРУ ШТУЧНОГО ІНТЕЛЕКТУ



OECD > Blogs, videos & podcasts >

Protecting data quality in the AI era of official statistics

Generative AI is reshaping how people access official statistical data. How can we ensure that data quality remains intact and that users can trust the information they receive? This blog explores this emerging challenge and discusses solutions to help safeguard quality throughout the information chain.



Romina BOARINI

Director, OECD Centre on Well-being, Inclusion, Sustainability and Equal Opportunity (WISE)



Cameroon TIATI A BISCENE

OECD Centre on Well-being, Inclusion, Sustainability and Equal Opportunity (WISE)

У блозі ОЕСР опублікований матеріал «Захист якості даних офіційної статистики в еру штучного інтелекту».

Генеративний штучний інтелект змінює те, як люди отримують доступ до офіційних статистичних даних. Чи можна забезпечити збереження якості даних та довіру користувачів до отриманої інформації? Все більше людей, зокрема – журналістів, політиків, – отримують доступ до офіційної статистики через чат-ботів та ШІ- помічників, а не через офіційні веб-сайти національних статистичних інститутів чи урядів. Тепер інформація проходить через низку цифрових посередників, які її отримують, інтерпретують та перерозподіляють, що неодмінно буде призводити до втрат і викривлень. Можуть зникнути дані щодо контрольного періоду, примітки, методологічні пояснення, які дозволяють користувачеві судити, чи відповідає показник його конкретному питанню. ШІ може суттєво покращити доступ до офіційної статистики, полегшивши пошук даних та їх сприйняття. Замість безпосередньої взаємодії з таблицями чи інформаційними панелями, користувачі все частіше отримують згенеровані

ШІ дані з простими і зрозумілими коментарями. Це робить статистику більш доступною для користувачів-неекспертів, але така зручність має суттєвий недолік: базова інформація вже була відібрана, інтерпретована та узагальнена системою ШІ у такий спосіб, про який користувач майже ніколи не дізнається. Наприклад, виноска до редакції може не пережити шлях ШІ, як і відомості про джерело. Користувачі отримують велику кількість інформації, але не можуть розуміти контекст і базову структуру даних, що необхідно для оцінки якості інформації та її подальшого ефективного використання, адже користувач повинен знати про походження, належність, валідність та репрезентативність даних. Щойно штучний інтелект стає посередником у взаємозв'язку між даними та користувачами, питання якості виходить за рамки оцінювання показника. Тут важливо пам'ятати про відстежуваність, машиночитабельність та репрезентативність. Системи ШІ не мають вбудованої інформації про те, чи було джерело переглянуто, виправлено або вилучено. У дослідженні 2025 року ChatGPT було запропоновано оцінити 217 відкликаних дослідницьких статей і жодного разу чатбот не вказав користувачу на те, що йдеться саме про відкликані матеріали. Надійність відповіді ШІ залежить від структури джерела: точність пошуку досягає 91% для добре структурованого тексту і знижується до 86% для менш структурованого. Системи ШІ загального призначення навчаються на будь-яких даних, яких найбільше у відкритому доступі, але їхня кількість не є свідченням репрезентативності. Офіційна статистика є суспільним благом у строгому економічному сенсі: вона неконкурентна (використання однією особою не применшує використання іншою) і вона не підлягає виключенню (доступ до неї не може бути обмежений), тому вона традиційно надавалася державою, а не ринковими гравцями. ШІ починає змінювати ситуацію: контент залишається публічним, дані залишаються вільними для доступу, але змістовний доступ мігрує за преміум-асистентами, платними інтерфейсами та висококонцентрованою інфраструктурою (50% захищених інтернет-серверів світу розміщена в одній країні). Отже, суспільне благо може стати таким, що підлягає виключенню без будь-якого свідомого рішення про це з боку урядів і громад. Крім того, системи штучного інтелекту не є нейтральними провідниками інформації, оскільки кожна модель містить ціннісні судження щодо акцентів та заборон. Статистичні управління мають унікальні можливості для забезпечення достатньої структурованості, документованості й доступності офіційної статистики, щоб вона слугувала надійним орієнтиром для систем штучного інтелекту. Необхідним кроком є встановлення стандартів для неофіційних даних, включно з даними, створеними громадянами, які все частіше використовуються при ухваленні державних рішень.

Детальніше:

<https://www.oecd.org/en/blogs/2026/07/protecting-data-quality-in-the-ai-era-of-official-statistics.html>

Фото: скріншот

#НРАТ_Усі_новини #НРАТ_Науковцям_новини #НРАТ_Освітянам_новини
#НРАТ_ШтучнийІнтелект #НРАТ_OECD

2026-09-25

**Інформація з офіційного вебпорталу Національного репозитарію
академічних текстів**