ЕТИКА РОБОТИ З ДАНИМИ В НАУКОВИХ ДОСЛІДЖЕННЯХ

Чуканова Світлана Олександрівна, кандидат педагогічних наук, Наукова бібліотека Національного університету «Києво-Могилянська академія», ORCID: http://orcid.org/0000-0002-5717-5050, e-mail: chukanovaso@ukma.edu.ua

Abstract. In the contemporary digital landscape, data serves as a fundamental asset for scientific inquiry, yet its handling, particularly of personal data, necessitates rigorous adherence to ethical standards and legal frameworks. This thesis explores the ethics of data management, emphasizing anonymization as a critical mechanism for safeguarding privacy while maintaining data's analytical value. Anonymization involves transforming personal data to prevent individual identification, extending beyond removing direct identifiers to mitigating re-identification risks through comprehensive strategies. Key challenges include achieving a balance between privacy protection and data utility, adapting to evolving technologies that may compromise anonymized datasets, and considering contextual factors that influence identifiability. Identifiability is conceptualized as a dynamic spectrum, ranging from direct identification via names or IDs to indirect via combinations of attributes like age, location, or profession. Three core indicators—singling out (isolating individual records), linkability (connecting data across sources), and inference (deducing personal details)—are examined to assess risks. Drawing from legal principles such as GDPR's «motivated intruder» test, the analysis evaluates practical threats from motivated actors using publicly available resources. The dynamic nature of risks, driven by technological advancements, new data releases, and societal shifts, underscores the need for ongoing risk monitoring and reassessment, tailored to data sensitivity levels (e.g., frequent reviews for high-risk categories like medical data). Countermeasures include techniques such as masking, generalization, and suppression. This work provides a broad framework for ethical data practices, offering researchers guidelines to ensure compliance, minimize harms, and enhance the integrity of scientific research in data-driven fields.

Вступ. У сучасному світі дані є невід'ємною частиною наукових досліджень, але їх обробка, особливо коли йдеться про персональні дані, вимагає суворого дотримання етичних норм та законодавства, наприклад GDPR (General Data Protection Regulation) (Загальний регламент про захист даних). Актуальність теми зумовлена швидким зростанням обсягу даних, ризиками порушення приватності та потребою в етичних підходах, що забезпечують баланс між науковою користю та захистом прав осіб, що беруть участь у дослідженнях. Метою дослідження є узагальнення принципів етики роботи з даними з акцентом на анонімізацію як інструмент запобігання ідентифікації.

Матеріали і методи. Дослідження ґрунтується на аналізі нормативно-правових актів ЄС щодо захисту даних (GDPR, рекомендації колишньої Article 29 Data Protection Working Party та поточні настанови EDPB[1]) (1, 2), науковій літературі з етики обробки даних, а також ілюстративних прикладах із медичної, освітньої та інших сфер (3,4,5). Застосовувалися такі методи: систематичний огляд літератури, моделювання сценаріїв ризиків реідентифікації (зокрема тест «мотивованого зловмисника»), оцінка ключових індикаторів ідентифікованості (виділення, зв'язуваність, виведення висновків).

Результати дослідження. Дослідження демонструє, що етика роботи з даними охоплює широкий спектр аспектів, де анонімізація виступає ключовим елементом для забезпечення конфіденційності та відповідності правовим нормам (2). Ідентифікованість даних трактується як динамічний спектр, який залежить від прямих ідентифікаторів (наприклад ім'я, номер ID) та непрямих (демографічні характеристики, місце проживання,

---

[1] EDPB (European Data Protection Board) — це Європейська рада з охорони даних, незалежний орган ЄС, створений відповідно до Загального регламенту з охорони даних (GDPR). Вона складається з представників національних органів з охорони даних країн ЄС та координує застосування GDPR на рівні ЄС. Настанови EDPB (guidelines) — це офіційні рекомендаційні документи, які надають роз'яснення щодо ключових положень GDPR, допомагаючи організаціям, контролерам даних та обробникам дотримуватися норм. Вони не є юридично обов'язковими, але мають значний вплив на практику та судову інтерпретацію.

професія), що можуть комбінуватися для ідентифікації. Три основні індикатори ризику детально проаналізовані:

- виділення (ізоляція окремих записів про особу, наприклад унікальної особи за віком і стажем);
- зв'язуваність (можливість інтеграції даних із різних джерел, таких як медичні записи та реєстри нерухомості);
- виведення висновків (статистичні передбачення, наприклад діагноз на основі покупок в аптеці) (4).

Ризики значно посилюються через технологічні зміни (розвиток машинного навчання, обчислювальної потужності), зростання доступності публічних даних (соціальні мережі, витоки баз) та еволюцію соціальних контекстів, що вимагає регулярного моніторингу ризиків — щомісяця для чутливих категорій (медичні, фінансові дані), щороку для загальних наборів. Рекомендації включають упровадження методів маскування, генералізації та придушення рідкісних комбінацій, використання договорів про конфіденційність для обмеження доступу, а також адаптацію стратегій анонімізації залежно від контексту (публічний доступ на противагу внутрішньому використанню) (3,5). Ці заходи забезпечують етичну обробку даних, зберігаючи їхню наукову цінність без значних втрат для дослідницьких цілей.

Висновки. Етика роботи з даними є основою відповідальної науки, де анонімізація мінімізує ризики ідентифікації, забезпечуючи правовий захист та збереження корисності. Теоретичне значення: узагальнення спектру ідентифікованості та ризиків для ширшого застосування. Практичне значення: загальні рекомендації для дослідників щодо етичних практик, моніторингу та технік захисту, що сприяють розвитку даних як етичного ресурсу.

Використана література:

1. General Data Protection Regulation (GDPR) [Інтернет]. General Data Protection Regulation (GDPR) – Legal Text; [цитовано 30 верес. 2025]. Доступно на: https://gdpr-info.eu/.

2. European Commission [Інтернет]. Article 29 Data Protection Working Party; [цитовано 30 верес. 2025]. Доступно на: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

3. Elliot M., Mackey E., O'Hara K. The Anonymisation Decision-Making Framework: European Practitioners' Guide [Інтернет]. Manchester: UKAN; 2020. 117 с. Доступно на: https://eprints.soton.ac.uk/445373/1/adf_2nd_edition_1.pdf

4. Narayanan A., Shmatikov V. Robust de-anonymization of large sparse datasets. In: 2008 IEEE Symposium on Security and Privacy (SP 2008); 2008 May 18-21; Oakland, C.A. Piscataway (NJ): IEEE; 2008. p. 111-25.

5. Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Rev. 2010; 57:1701-77. (U of Colorado Law Legal Studies Research Paper No. 9-12). Available from: https://ssrn.com/abstract=1450006