APPLICATION OF ARTIFICIAL INTELLIGENCE IN SOLVING THE PROBLEM OF DATA ANONYMITY

Kalinin, D. V., Severin, V. P.
National Technical University, «Kharkiv Polytechnic Institute», Kharkiv

Nowadays, the question related to ensuring the confidentiality of information during its storage, transmission over communication lines (e.g., on the Internet), and processing has gained significant importance. The problem of data protection is particularly relevant in the medical field. Adhering to medical ethics, which requires non-disclosure of patient information, is essential. However, when storing data on digital media, there is a risk of unauthorized access. Therefore, it is necessary to present data in a way that cannot be directly linked to specific individuals. Methods aimed at transforming information to prevent person re-identification fall under the umbrella term "anonymization methods".

Anonymization methods must consider the specific of the processed data: its size and homogeneity as well as the presence of outliers (anomalies) can be exploited by potential attackers to identify an individual and reveal associated sensitive information [1]. The proposed approach for anonymization focuses on enhancing classical models such as *k*-anonymity, *l*-diversity, and *t*-closeness [2]. While these methods have certain advantages, they also exhibit expected limitations, including insufficient data protection, inefficiency against specific attack types, and excessive information loss.

To achieve effective anonymization, it is proposed to apply transformations, partial replacements, and augmentation with synthetic data. Artificial intelligence methods can be successfully involved in generating such "synthetic" data [3]. The task of preserving the utility of medical data is considered as an applied problem for testing the developed information system. The specifics of such type of data is being considered when developing a generative neural network that will create the necessary artificial data.

References:

1. El Emam Kh., Arbuckle L. *Anonymizing Health Data*. O'Really Media Inc., 2013. 228 p.

2. Olatunji I. E, Rauch J., Katzensteiner M., Khosla M. A Review of Anonymization for Healthcare Data. *Big Data*. 2022. URL: https://arxiv.org/pdf/2104.06523. DOI: 10.1089/big.2021.0169.

3. El Emam Kh., Mosquera L., Hoptroff R. *Practical Synthetic Data Generation*. O`Really Media Inc., 2020. 166 p. URL: https://books.google.com.ua/books/about/Practical_Synthetic_Data_Generation.html?id=XWnnDwAAQBAJ&redir_esc=y