Privacy-Preserving Machine Learning using Deidentification Techniques for HR and Payroll Data

Shanmugaraja Krishnasamy Venugopal
Independent Researcher
Carleton University , Ottawa ON, Canada

***Abstract*— The use of machine learning in Human Resources (HR) and payroll systems represents significant capability gains for automation, decision-making, and operational efficiencies. However, this transition process raises significant concerns related to the protection of sensitive employee data. Privacy-preserving computer scientific methods, like de-identification (removing or altering identifiers such that the individual cannot be identified), are becoming an increasingly important way to preserve privacy while maintaining the richness of datasets for use in analytical research.**

The purpose of this review article is to provide a comprehensive report on various privacy-preserving techniques, especially de-identification techniques common to HR and payroll data. The review will include real-world examples of privacy-preserving methods, including CV de-identification, clustering of quasi-identifiers (QI), narrative-level de-identification, and federated learning (model training without the data leaving the local source, but sharing model updates). The article looks at implications for privacy and utility trade-offs, assesses potential reidentification risks, and summarizes innovations in secured data, such as format-preserving transformations (anonymizing and preserving values such as IDs/dates) and voice anonymization. The review article considers issues identified in the current literature, as well as policy implications and future directions for utilizing secure and compliant machine learning frameworks in human resource settings.

***Index Terms*— Privacy-preserving machine learning, Deidentification techniques, HR data privacy, Payroll anonymization.**

1. Introduction

The rapid development of data-driven technologies in Human Resource (HR) and payroll systems has fostered advanced analytical capabilities, increased decision-making practices, and automated decision-making. Machine learning (ML) is a central element of this transformation and can be leveraged for predictive modeling of employee attrition, payroll fraud, performance analytics, etc. However, given the expansion of ML in HR and payroll systems, there are concerns regarding the privacy and management of employees' personally identifiable information (PII) and other confidential data. Thus, in order to demonstrate legal compliance and encourage the trust of employees—and to prevent misuse of employee data—it is necessary to create more privacy-preserving mechanisms, most notably de-identification mechanisms, specifically for use in HR and payroll systems.

This literature review will examine the development of privacy-preserving machine learning approaches, but will specifically focus on de-identification processes for Human Resource and payroll data. After the literature is reviewed, this paper will consider differing frameworks and approaches for the trade-off of utility for privacy, as well as practical obstacles, considerations for implementation, and possible next steps such as federated learning or voice anonymization.

2. Deidentification in HR Data Processing

In HR and payroll, the sensitivity of data necessitates robust de-identification processes. De-identification may be thought of as either removing or transforming PII to the point that individuals cannot be readily identified from the data. This is important not just from an ethical viewpoint, but also from a regulatory perspective regarding regulations such as the General Data Protection Regulation (GDPR). A proper de-identification process needs to minimize lost information while still allowing the possibility of ML algorithms to perform valid analytical processing.

The deidentification of resumes is a common practice used for recruiting. Before automated applications process candidate resumes, candidates should be stripped of identifiers such as the candidate's name, addresses, gender, and photo, in order to mitigate the potential for bias, either explicit or implicit. There is a relatively recent set of literature that has proposed a systematic method of deidentifying that does not greatly impact the information value of the resumes. [1] For instance, it is inevitable in human resource (HR) contexts that fairness and transparency are always entangled with the intent to maintain privacy, especially since hiring decisions have lasting organizational consequences.

More evidence indicates that CV de-identification can both safeguard privacy and mitigate biases present in algorithmic decision-making during a recruitment process [2]. For example, the inappropriate use of demographic characteristics could amount to incidental bias in the selection of candidates. In this way, de-identification techniques are beneficial for legal compliance to

equality legislation, and, perhaps more importantly, they provide an important method for enhancing fairness and diversity in hiring processes.

3. Governmental Adoption and Technical Challenges

Given increased societal concern about data privacy, government organizations are starting to deploy privacy-preserving technology. In the public sector specifically, payroll and HR datasets consist of a considerable amount of citizen personal data, so protecting that data is important to the nation as a whole. While the actual importance of safeguarding is clear, there are many technical and organizational challenges in the operationalization of the technologies, including interoperability between technologies, high computation costs, and no established framework for deidentification [3].

Governments are rapidly becoming responsible for the stewardship of significant administrative datasets, trying to balance usability and privacy concerns based not only on expectations created by institutions in private enterprise but also on a larger set of transparency obligations that come into play due to the greater number of stakeholders that exist when compared to a private enterprise. Accordingly, we expect privacy-preserving solutions in this vein to be scalable and interpretable. If there are limitations to either scalability or interpretability, this may quickly develop into misuse in the deployment of ML algorithms, and if one has to rely on the ability to use large quantities of high-quality data, this becomes more problematic.

Quasi-identifiers, such as age, date of birth, and occupational title, can also contribute to the re-identification of individuals when datasets are combined in order to resolve technical difficulties of data linking. This is especially alarming when considered in relation to payroll systems having the ability to aggregate data with quasi-identifiers across departments. Knowing when and how to utilize quasi-identifiers alongside robust anonymization measures will be helpful in mitigating the re-identification risk of data linking [4].

4. Language-Adaptable Anonymization Techniques

The deidentification of narrative records, especially in natural language format, is another expanding domain of interest for privacy-preserving ML. Free-text performance evaluations, employee complaints, and recruitment interviews are commonly collected in many HR systems. These narrative records often include identifiable information according to the context of the records, and during the redaction process, the information cannot be simply safeguarded by deidentification.

An anonymization framework with two phases applicable to a variety of language models is an alternative. The first phase is entity detection using NLP models, and the second is context-sensitive transformation or masking of the terms created through phase one [5]. These flexible anonymization models can be used in HR reports, audit trails of payroll records, and training evaluations with privacy for individual subjects and the richness of the original narrative data.

This matter is particularly important related to managing a multilingual workforce and sites around the world in affiliated locations. A model that can appropriately handle variation in language and use of language would bolster de-identification methods and usability around the world. Therefore, adaptable language processing is more than just a technical requirement, it is strategically beneficial in creating a global HR analytic platform.

5. Cluster-Based Anonymization of Quasi-Identifiers

Quasi-identifiers are data fields that by themselves may not uniquely distinguish an individual. However, together these fields can allow for the triangulation of individuals again. This presents a risk for payroll datasets that may contain relatively common features, such as job title, department assignment, or salary ranges, where a small number of individuals could triangulate back to a specific instance of themselves contained in the dataset and be identified. A different, more complex process than k-anonymization that can be used to help mitigate the risks described above would be to use a cluster-based k-anonymization approach.

This process groups similar records together such that the individual records within a group would not be distinguishable from one another. The anonymization process, therefore, generalizes or masks some of the quasi-identifiers amongst those records in that group. The table will then have a low risk of re-identification potential while having statistical utility for machine learning [6].

In terms of payroll systems, this approach is particularly useful when looking at pay trends, overtime trends, or fraud, since it pools our analytical models in line with the actual trends we are measuring without revealing individual pay or rankings. Cluster-based anonymization within a preprocessing pipeline can support HR departments responsibly scaling their ML capabilities.

6. Use of Federated Learning in Privacy-Sensitive Systems

Machine learning has relied on the premise that data existed in a centralized repository, which raises privacy issues, as has the drive to evolve from federated learning to decentralized learning, or both methods, to increase privacy. In federated learning, training occurs on end devices, and only an update—either reflecting the loss or an update to the aggregated model—is communicated at the end of training, so raw data is not transported from the storage configuration.

New investigations into the feasibility of federated architectures for medical data uses indicated, similarly to its evident uses in HR data uses, a privacy risk. This research, while also strictly aimed at the identification of rare medical conditions, has clear implications for payroll and HR data uses [7]. The technical architecture and the protocols and plans can be seen and can thus be used for HR data processing pipelines based on security and privacy principles that are strict and nominally specific.

Federative networks are also a method for continual learning which makes federative networks appropriate for the dynamic process of staffing which exhibits changes in workforce characteristics over time. Federative networks also provide a mechanism for a model to be updated in collaboration among teams or organizations while still complying with data privacy agreement(s) or legal obligations [8].

7. Privacy Utility Trade-offs and Scoring Systems

Balancing data utility and privacy protections is a significant challenge to implementing privacy-preserving mechanisms. Users may anonymize too much and not realize that they may lose some of the underlying quality and interpretability of the models being used in the analyses that follow. Under-anonymizing could create greater privacy risk with discipline-associated identifiable data. Using tabulated data methods, the concept of a scoring framework and/or method of determination of privacy has been proposed in deidentified tabulated datasets [9].

These systems are capable of providing privacy scores depending on re-identification risk and anonymization success. This can be a useful resource for auditing HR datasets after processing, to detect identifiable patterns that may persist. Applying these scores to HR data pipelines will increase responsibility and provide decision support for data scientists and compliance teams.

The following figure demonstrates a theoretical trade-off from a privacy-utility perspective (evaluating privacy risk against analytic effectiveness) for anonymized HR data simulated from experimental contexts, applying anonymization interventions across different portions.

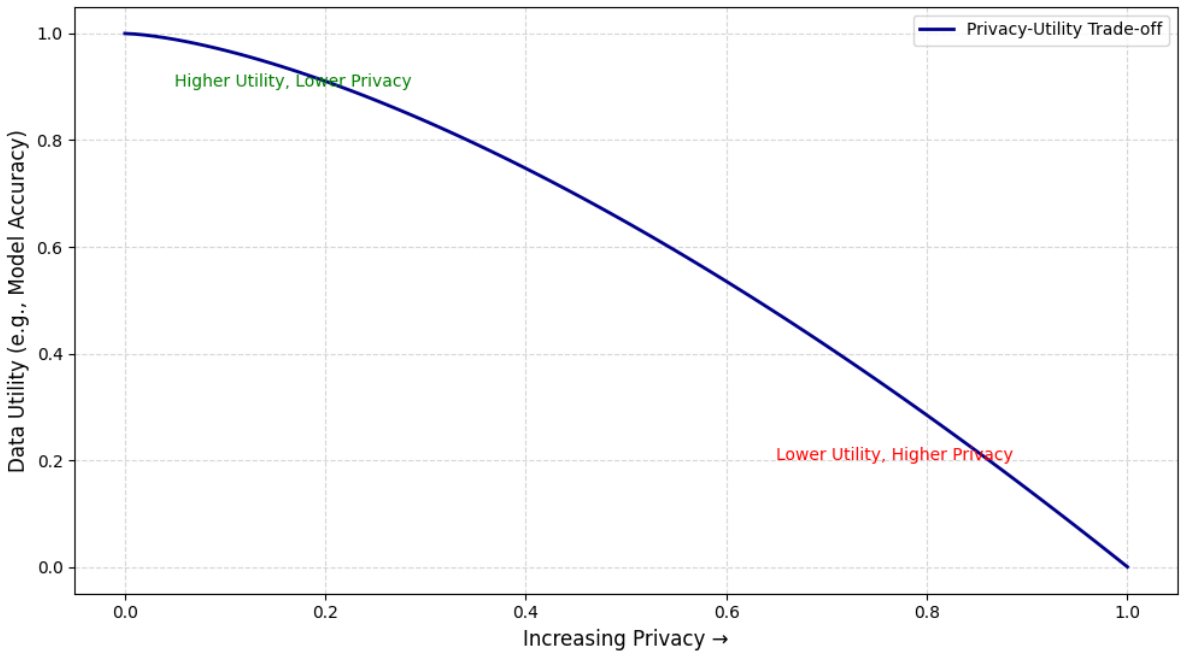

Figure 1: Privacy vs Utility Trade-Off in HR Data Anonymization

Adapted from [10]

As demonstrated, greater privacy tends to decrease utility. Achieving an ideal compromise will depend on the case, context-specific regulations, and data sensitivity.

8. Format-Preserving Transformations

Another innovation that offers value is format-preserving deidentification that allows for the use of data without destroying shape-space. Unlike classic anonymous forms that obliterate the shape-space of a dataset, format-preserving transformations of a dataset preserve the dataset's structural integrity. A payroll system, for instance, may need to have the formats of employee IDs and dates maintained for usability.

A recent article on format preserving strategies reports a performance measure as a function of privacy and utility across different deidentification strategies [11]. This paper finds that even if some loss of utility is unavoidable, format preserving strategies are able to maximize privacy and utility with data collected for administrative reasons like payroll audits, where accuracy of the data is important and compatibility with existing systems is an issue.

[Table 1: Comparison of Deidentification Techniques in HR and Payroll Data]

Technique	Use Case	Strengths	Weaknesses	Privacy Risk Level
CV Deidentification	Recruitment screening	Reduces bias, easy to implement	Can reduce context	Low
Cluster-Based Anonymization	Payroll analysis	Retains patterns, scalable	Computationally intensive	Medium
Two-Stage Narrative Anonymization	Performance reviews	Adaptable to languages, accurate	Requires NLP expertise	Low
Format-Preserving Transformation	Payroll system compatibility	Retains schema, system-friendly	Limited against sophisticated attacks	Medium
Federated Learning	Workforce analytics	No raw data sharing, collaborative	Complex setup, limited support	Very Low

Compiled using findings from [1] to [11]

9. Voice Anonymization in HR Applications

In addition to written text and numerical values, HR systems increasingly leverage multimedia data formats—especially voice recording data. Most of the spoken data are likely tied to an interview, grievance, employee feedback, or internal communications. As machine learning models are built to analyze and interpret speech, a new privacy dimension emerges with the ability to potentially identify individuals by the sound of their voice.

To obscure and/or alter voice data without leaving identifying characteristics but retain audio accuracy for downstream tasks (such as sentiment analysis and/or keyword spotting), various voice anonymization methodologies are evolving. A possible area is adapting generative adversarial networks (generative adversarial networks (GANs)) to be able to modify acoustical features and maintain content accuracy. This area is emerging in medical and surveillance, as well as in HR, especially for larger firms operating globally to manage and/or investigate multilingual voice data in context from anywhere in the world [12].

When voice anonymization is incorporated with privacy-preserving machine learning (ML) frameworks, the human speech of individuals contains biometric identifiers, possibly to the detriment of the individual using the service. Additionally, the voice anonymizing methodologies operate under the minimum data and storage principles by only capturing the minimally sufficient data needed to identify the individual and excluding extraneous, unnecessary data. In the ongoing trajectories in our societies, as organizations begin utilizing AI- and machine-driven HR voice data analysis automation, these new preserving voice anonymization methods will become an unbiased and legitimized legal rationale based in ethics.

10. Data Re-identification Risks and Defensive Strategies

Although de-identification is an important layer of privacy protection, it alone cannot ensure total safety. Adversaries can potentially use the remaining identities of the data or use other data sources to identify someone from de-identified HR and payroll data. In response to this vulnerability, scoring models have been developed to estimate the probability of identification through measuring the rarity of elicited attributes, the strength of the association, and the risk of re-individualization based on disclosure risk and prior risk.

One way to assess the identity disclosure risk is by examining how much an anonymized record resembles known values from publicly available data repositories. This is particularly salient within the context of payroll data, where high-ranking jobs, unusual benefit plans, and unusual salary ranges can serve as identifiers and thus make individuals easier to re-identify [10]. Individuals trying to adversarially re-identify individuals may also use machine learning to exploit sensitive characteristics from partially anonymized datasets.

All of the defensive strategies discussed (e.g., k-anonymity, l-diversity, t-closeness) are intended to reduce the chance of re-identification — they are not meant to eliminate the chance of re-identification altogether or orthogonally — by creating records that are indistinguishable from at least ‘k’ other records, or by providing plausible deniability that sensitive attributes are sufficiently dispersed across anonymised records within a group. These models can fail: none of them are perfect and, thus, cannot operate in isolation; however, when we combine models with more complex anonymisation modes created for specific domains, we create a layer of defence against the challenge with this approach.

In addition to direct identifiers, structural characteristics – which are not direct identifiers, but can be used together and correlated to serve as quasi-identifiers – should be a part of effective anonymization of payroll data. For example, sometimes we can uniquely identify someone by job grade, work location, and years of employment. It is this type of implicit identifier that needs to be addressed and should be addressed through dynamic measurement of risk based on preprocessing the data, as opposed to just relying on a set of static rules to redact identifiers.

11. Measuring and Optimizing the Privacy-Utility Trade-Off

A fundamental dilemma in privacy-preserving machine learning is the trade-off between data privacy and utility. Data that is processed to become extremely anonymized can be made statistically useless, and in turn, can lessen the ability of machine-learning models to learn patterns with meaning. Alternatively, if the data is not adequately anonymized, then privacy can be compromised. Hence, measurable estimators are required to balance privacy and utility.

An example of how to analyze the trade-off claimed is through privacy scoring systems, which offer an objective means to assess the level of aggregation, are informed by the potential utility of what might be retained for, or in some use-cases, in service of machine learning tasks such as: information loss, classification accuracy, and outlier analysis, which have all been documented means for evaluating confounding anonymization techniques against one another on the same dataset [10].

Furthermore, format-preserving anonymization approaches offer a greater trade-off with data where there is other usability afforded by a maintained structural syntax (the data can still be used by legacy HR systems). This may bring value in the event of data pathways that are incredibly brittle or connected in an enterprise software stack. Therefore, these methods provide a seemingly compatible format while still functioning and providing privacy with minimal structural change [11].

In real-world applications, like fraud deterrence or maximizing benefits to employees in payroll systems, there can be instances where the loss of data fidelity attributed to anonymization could offset the ethical/legal benefits of the privacy. Thus, privacy-utility tradeoffs are more about prioritizing the strategy than they are about some outcomes for technical performance.

12. Technological Feasibility and Operational Integration

Shifting from the idea of de-identifying to practical application within organizational HR systems can be complex and difficult. This can involve anything from using old, pre-existing datasets; operationalizing data from one older system for availability to another; and ensuring it is real-time data. Further, there is a call for multi-modal FHP because HR data often contains mixed modalities of information (e.g., numeric payroll data compared to performance reviews and transcripts of comments and other forms of scripting aloud).

Scalable frameworks, such as federated learning, can help address some of these operational limitations. Because the data never leaves the storage system, federated systems avoid data clean-up, transfer costs, and exposure risks. A tradeoff to federated systems is the increased complexity of the system and the need to create custom APIs, as data needs to be exchanged with other organizations, requiring coordination [8].

Furthermore, when using Privacy Enhancing Technologies (PETs), organizations must prove that the solutions are appropriate for their culture and values in terms of regulatory obligations. Many organizations do not have the technical sophistication, nor awareness, to be in a position to make assessments of re-identification risk or to build a framework, or an endless intervention, to anonymize data. Employee education, engagement, and support across departments and leadership are extremely important to build a culture around privacy and data protection within their HR departments.

Another operational framework that can be beneficial is hybrid systems, that allow data to have a legitimate purpose, limiting its access to only the processing module, and the processing module performing a real-time dynamic anonymization of its data when shared with the ML models. The separation of duties and roles, as well as access and audits, provides tremendous technical and procedural safeguards. These models work within a zero-trust architecture, and seem to be more and more commonly written into enterprise security policy these days.

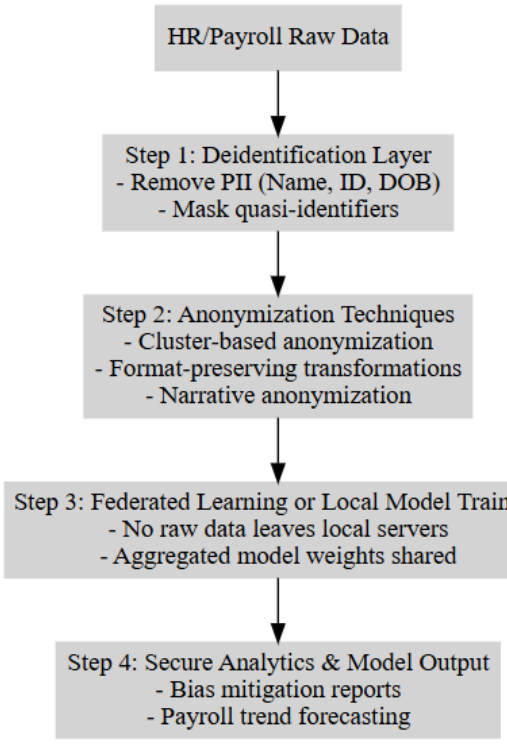

Figure 2: Overview of Privacy-Preserving ML Framework for HR and Payroll Data

Adapted from [1], [3], [6], [8]

The figure above illustrates a step-by-step process for a process of privacy-preserving machine learning applied to HR and payroll data across the entire lifecycle of data, from raw data deidentification to anonymization, federated model training, analytics, and finally, securely shared outputs.

13. Future Trends and Research Directions

The field of privacy-preserving ML for HR is relatively novel; we also come upon trends that are just forming and will likely evolve over the course of the next couple of years. One area of particularly interesting potential evolution, for example, is the intersection of deep learning models with differential privacy inserts, as a way to enhance privacy protections. Noising of some sort is then added to the inserts, to the data, or to the model outputs to mask any references to the individual and to also provide enhanced privacy protections. Although differential privacy has primarily existed within the realms of healthcare and finance, the addition of a differential privacy perspective would be an advantage to the HDR process and would be a unique approach to HR models that adds an appropriate amount of returns and opportunities.

A further area of promise is the generation of synthetic data to create plausible non-real datasets for ML training. Synthetic datasets attempt to preserve the statistical properties of the original data, yet do not contain any records of real people. While this does not bring it fully to the realm of deidentification, it has positive implications for privacy in that the researcher is able to train the model without ever having to see the original data.

Interest in multimodal anonymization is also on the rise. Given that most current HR systems preserve data records in text, tabular form, audio, and even images (example ID scans or profile photos from sites like LinkedIn), future deidentification pipelines may need to treat more than one data type at a time. Models that can anonymize across modalities will build a wider coverage to achieve a more comprehensive approach to reduce privacy leaks.

The policy landscape will adapt, as labor laws will generate new digital privacy rights, and future legislators may also impose economic and employment regulations on the use of bridge-gap anonymity or transparency in work-related decisions made using AI. Organizations that are making current investments in privacy-preserving infrastructures will be positioned to capture new opportunities in the future. They will be in a position to indicate they are compliant, or compliant better when employment regulations are developed. Organizations may also use privacy infrastructure, and the perception of privacy by individuals, as a means of attracting individuals who may put a premium on privacy.

14. Conclusion

Utilizing deidentification technology in the context of privacy-enhancing machine learning represents a critical aspect of the ethical and compliant use of AI systems in people management and payroll scenarios. Options to use alternative methods to deidentify resumes to alleviate bias, cluster-based methods to deidentify payroll data, federated learning architectures, and voice

deidentification methods for soliciting survey feedback have all emerged as options now possible. The option to use privacy-preserving data technology is rapidly evolving.

The findings of the prior literature suggest that there is not one fix for privacy in ML HR applications. More specifically, privacy-preserving solutions will need to be contextualized and balance protections against privacy, utility, feasibility, and what is typically allowed by law. As the literature shows, organizations will continue to adopt ML into their workforce management processes, and protecting and de-identifying personal information will be critical for developing ethical HR analytics.

15. References

[1] Löbner, S., Serna, J., Tronnier, F., Tesfay, W., & Rannenberg, K. (2025, August). Mitigating Bias in Recruitment: A Practical Approach to CV De-identification Considering Privacy Sensitive Information. In *International Conference on Availability, Reliability and Security* (pp. 174-192). Cham: Springer Nature Switzerland.
[2] Rannenberg, K. Mitigating Bias in Recruitment: A Practical Approach to CV De-identification Considering Privacy Sensitive Information.
[3] Prabowo, S., Putrada, A. G., Oktaviani, I. D., Abdurohman, M., Janssen, M., Nuha, H. H., & Sutikno, S. (2025). Privacy-Preserving Tools and Technologies: Government Adoption and Challenges. *Ieee Access*.
[4] Borrero-Foncubierta, A., Rodriguez-Garcia, M., Muñoz, A., & Dodero, J. M. (2025). Protecting privacy in the age of big data: exploring data linking methods for quasi-identifier selection. *International Journal of Information Security*, *24*(1), 37.
[5] Jia, J., & Nishi, H. (2025). A flexible two-stage anonymization framework for narrative medical records adapting to various language models. *Computers in Biology and Medicine*, *195*, 110624.
[6] Biswas, S., Tripathi, K., Khare, N., Jain, P., Agrawal, P., & Shukla, S. (2025). Enhancing Privacy Through Cluster-Based Anonymization of Quasi Identifiers in Correlated Datasets. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, 1-22.
[7] Burns, G., Kauffman, C., Manion, M., Pai, R. A., Milla, C., O'Connor, M. G., ... & Bjornson-Pennell, H. (2025). Feasibility of Machine Learning Analysis for the Identification of Patients with Possible Primary Ciliary Dyskinesia. *medRxiv*, 2025-04.
[8] Santini, F., Wasserthal, J., Agosti, A., Deligianni, X., Keene, K. R., Kan, H. E., ... & Pichiecchio, A. (2025). Deep Anatomical Federated Network (Dafne): An Open Client-Server Framework for Continuous, Collaborative Improvement of Deep Learning–based Medical Image Segmentation. *Radiology: Artificial Intelligence*, *7*(3), e240097.
[9] Genge, B., & Haller, P. (2025, April). Balancing Privacy and Utility. In *Computer Security. ESORICS 2024 International Workshops: DPM, CBT, and CyberICPS, Bydgoszcz, Poland, September 16–20, 2024, Revised Selected Papers, Part I* (Vol. 15263, p. 124). Springer Nature.
[10] Folz, J., Vidanalage, M. D., Aufschläger, R., Almaini, A., Heigl, M., Fiala, D., & Schramm, M. (2025). Scoring System for Quantifying the Privacy in Re-Identification of Tabular Datasets. *IEEE Access*.
[11] Mesana, P., Vial, G., Jutras, P., Caporossi, G., Crowe, J., & Gambs, S. (2025). Measuring privacy/utility tradeoffs of format-preserving strategies for data release. *Journal of Business Analytics*, *8*(3), 147-169.
[12] Meyer, S., & Vu, N. T. (2025). Use Cases for Voice Anonymization. *arXiv preprint arXiv:2508.06356*.