PRIVATE AI

PRIVACY-ENHANCING TECHNOLOGIES AND THE GDPR

The Role of Private AI in GDPR Compliance

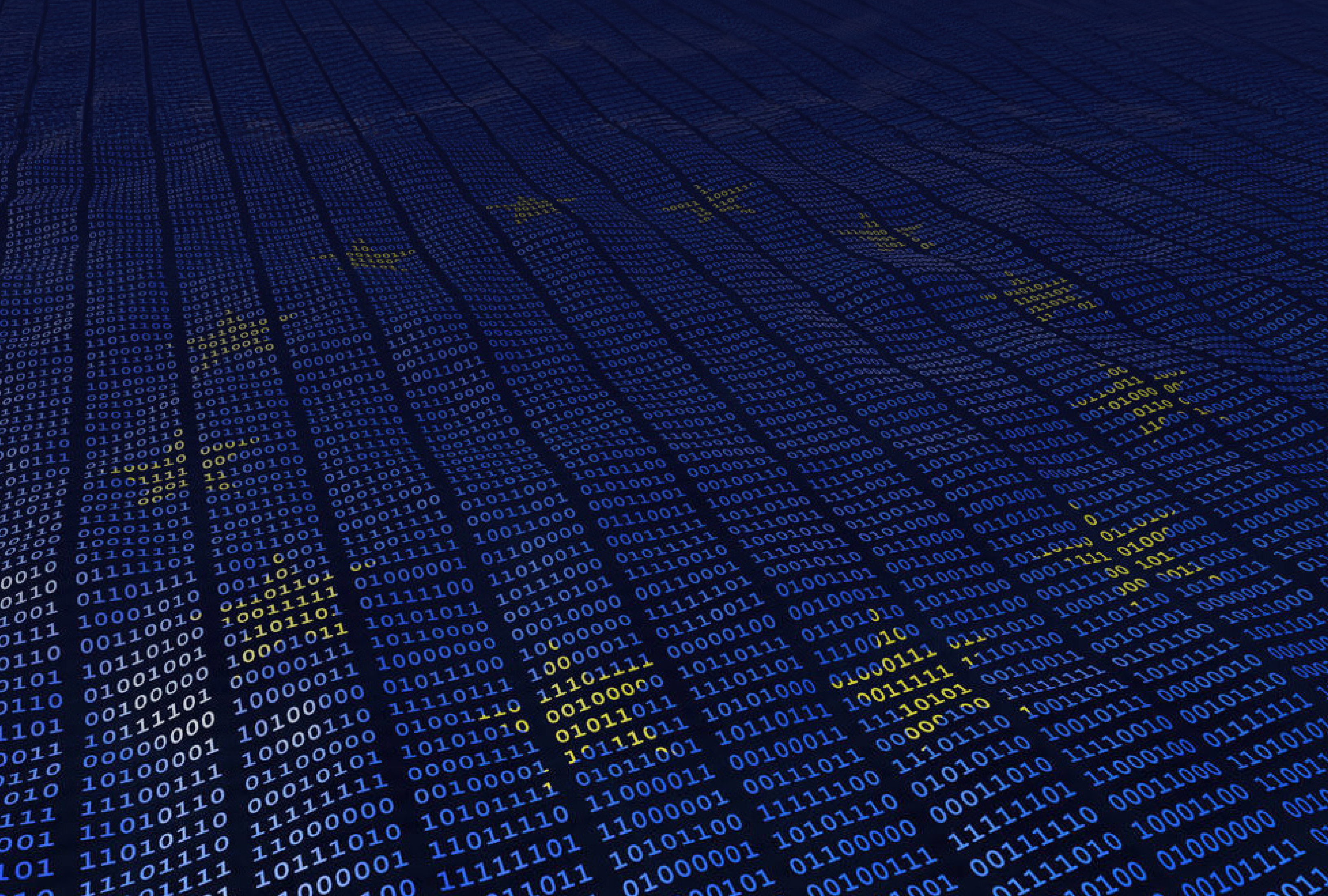

TABLE OF CONTENTS

INTRODUCTION

Successful businesses have traditionally competed on price and customer experience, both of which have been driving forces in technological advancements over the centuries. However, privacy and security are now emerging as essential characteristics of successful companies. In fact, this survey shows that 77% of businesses will not adopt commercial generative AI due to privacy concerns, and understandably so.

Customers and regulators require assurance that data will be stored and managed safely, particularly in the wake of numerous high-profile scandals across different industries that have shaken consumer trust. As customers increasingly take control over their data due to increasing awareness of their value, and backed up by shifts in regulatory requirements, there is a growing sense that companies will lose the ability to leverage data to create value for themselves, customers, and society as a whole. It is thus time to look at the ability of technologies to enhance data privacy and provide a competitive edge by increasing the trustworthiness of organizations' data handling processes.

As the global gold standard of data privacy, any organization would be well advised to measure its privacy posture against the GDPR even if the regulation does not directly apply to them. Many privacy laws are also drawing inspiration from the GDPR and follow Europe's lead when developing their own standards, like Brazil's LGPD and California's CPRA. Thus, when providing guidance on what and how to use technology to enhance data privacy, we first set out the GDPR's requirements focusing on those that can most easily be met when technology does part of the job, namely identification, anonymization, and pseudonymization of personal data.

IDENTIFICATION OF PERSONAL DATA

One important and at the same time demanding prerequisite for data protection is knowing what data you hold and where. In this section, we set out for which GDPR provisions this knowledge is key, and how technology can help identify the personal data under your control.

The Law

Art. 4(1) of the GDPR defines personal data as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

There is no provision in the GDPR that specifically requires an organization to have at all times current knowledge of what personal data is under its control. However, the following GDPR requirements are very difficult to meet, and non-compliance is penalized, if the organization is not clear on what personal data is collected, exists in its systems, or is held on its behalf by data processors.

1. Transparency towards data subjects regarding the processing of their personal data, including the extent to and the purpose for which they are processed and whether any automated decision-making, incl. profiling, is undertaken on the basis of that personal information;
2. Limiting the collection of personal data to what is necessary to achieve the purposes

for which they are processed;

3. Strictly limiting the retention period of the personal data to a minimum;
4. Information obligations regarding the period for which the data is processed;
5. Determining the legal basis for the processing of the personal data, and if it is consent, the extent to which it covers the processing of the personal data;
6. Access to and rectification or erasure of personal data and the exercise of the right to object to the processing of information as well as the right to data portability;
7. Honoring the withdrawal of consent to certain or all processing of a data subject's personal data;
8. Breach notification of affected individuals and breach reporting to the supervisory authority that includes approximate numbers of affected individuals;
9. Risk assessments with regard to the scope of processing;
10. Pseudonymizing or anonymizing personal data; and
11. Conducting Privacy Impact Assessments.

Under the GDPR, the data subject shall remain in control of their own data, and legal as well as practical certainty should be enhanced (Recital 7). For data subjects to have certainty of what happens with their data, the organization holding it must first of all know what data it is holding. Only then can the many information obligations be met and appropriate protections be put in place.

The Tech

One possible solution to identifying personal data contained in datasets are Regular Expressions, also known as regexes. Regexes are a sequence of characters that form a search pattern. They are commonly used in computer science and programming to find and manipulate patterns of text, such as searching for specific words or characters in a document or ensuring a form entry matches the correct format.

When it comes to detecting personal data in a data set, regexes can be used to search for patterns of text that match specific formats or types of data. For example, a regex pattern might be used to search for email addresses, phone numbers, or social security numbers in a data set.

However, regexes have some limitations when it comes to detecting personal data. Some of the limitations include:

- Variations in formatting: Personal data can be represented in many different formats, such as with or without dashes or spaces, in different orders, or with variations in spelling or capitalization. For example, writing a credit card number like this: 11112222-3333 4444. These variations can make it difficult to create regex patterns that capture all possible variations of the data.
- Context-specific information: Personal data can be used in different contexts, and the meaning of the data can change based on the context. For example, a number that looks like a social security number might actually be a phone number or an account number, depending on the context in which it is used.
- Identifying partial matches: Personal data might be represented partially in a data set, or might be included in a larger block of text. Consider, for example, an ASR transcript that reads like this: "My credit card number is 1223. Got that? Yes, please read the next 4 digits. Ok, 7659." Regexes might

not be able to identify these partial matches or identify the relevant portions of text that contain personal data.

- False positives and negatives: Regexes may incorrectly identify non-personal data as personal data, or might fail to identify certain types of personal data. For example, a regex pattern might mistakenly identify a random string of numbers as a social security number, like "the part number for the oil filter your car needs is 324-45-3237." A regex might also miss personal information that uses an unconventional format. For example, an unconventional format is often used when putting email addresses on websites to reduce the amount of spam that is received, like "my email is george dot lloyd at me dot com."

A significantly more accurate method relies instead on machine learning models that are context aware. Private AI detects more than 50 different types of personal data across 52 languages. The ML models achieve 99.5%+ accuracy, with structured, semi-structured, as well as unstructured data. Private AI's linguists train these modules on locale-specific formats for numerical personal data and optimize them for the particularities of each language, as different languages pose different challenges to ML models.

The output is a report that tells the user exactly where each relevant data point is located in their data and what type of personal data it is. This information can then be used to disclose to data subjects what information is held by the organization, to check whether the data limitation principle is properly implemented, i.e., whether less personal data needs to be collected or more disposed of, to quickly respond to access and rectification request, to assess the exposure in case of and respond to data breaches by adhering to the information obligation, and much more. Getting the personal data identification piece right is essential for the majority of the compliance requirements under the GDPR.

ANONYMIZATION

Anonymized data fall outside of the scope of the GDPR because they do not meet the definition of personal information. Aside from not controlling any personal data at all, anonymizing data is the safest strategy to employ to protect data. Anonymizing data has the great benefit that an organization can do as it pleases with this data, e.g., share it with anyone, even across EU borders, use it for any purposes that come to mind, and retain it as long as desired. As good as that sounds, achieving anonymization is not a small feat.

The Law

While there is no definition of anonymized data in the GDPR, Recital 26 provides relevant guidance that can be summarized as follows:

- Anonymization requires that an individual cannot be identified by means of the anonymized data;
- To determine whether a natural person is identifiable, account should be taken of:
    - All the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly.
    - To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of

> time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

But what exactly are the requirements that must be met for data to be considered anonymized?

Article 29 Working Party's Opinion on Anonymization Techniques (WP 216) from 2014 provides helpful guidance that remains relevant today. For one, parts of it were included almost word-for-word in Recital 26. And second, the Working Party is the predecessor of the European Data Protection Board (EDPB) which confirmed the persistent relevance of WP 216.

The anonymization guidance of WP 216 requires that "the outcome of anonymisation as a technique applied to personal data should be, in the current state of technology, as permanent as erasure, i.e. making it impossible to process personal data."

This strict requirement cannot be met if the original data set is not deleted or highly aggregated by the data controller, as the WP specifies that "when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data. Only if the data controller would aggregate the data to a level where the individual events are no longer identifiable, the resulting dataset can be qualified as anonymous."

Depending on the task, the utility of the data may be reduced following this standard. A case study performed on medical images showed that 85% of the image needs to be covered with noise in order to reliably produce anonymized data under this strict standard, rendering the images useless for medical research purposes. On the other hand, for many use cases, the particular individuals to whom the data pertains is irrelevant, such as where a concentration of data points is of interest, as in traffic analysis or the origin and spread of a disease. Data utility would not be impacted by its anonymization.

WP 216 has been criticized for being too strict a standard, that no organization would anonymize data as the deletion of the original data set is not in their interest. It has also been pointed out that the Court of Justice of the European Union decided in 2016 in *Breyer* that an IP address, which does not directly identify an individual, would be considered personal information only in the hands of someone who has the legal means to access the information of the internet service provider which could, in combination with the IP address, identify an individual. An argument could thus be made that, in some circumstances, information that remains identifiable can still fall outside of the scope of the definition of 'personal information,' namely so long as the combination of data sets that would lead to the identification is not reasonably likely, so that the risk of identification is "insignificant."

Arguably, this may mean that erasure of the original data set is not strictly required but that it only needs to be ensured that the risk of linking back to the original data set is insignificant. It is important to note that this would still be a standard that can hardly be met within an organization who wishes to keep the original data set but "anonymize" it, for example for use in its testing environments. It would seem that in this scenario the GDPR would continue to apply to the data set, but may not apply if the same data set were to be disclosed to another entity, provided the risk of linking it back to the original data set retained by the organization is insignificant.

The latest development, notably in the EU General Court in Case T-557/20, SRB v EDPS from April 2023, points towards a slight tendency to move away from the strict interpretation of anonymization entertained in WP 216. The court points out that with regard to the requirement of the definition of personal data that it is 'related' to an individual, we must consider whether the content, purpose, or effect is linked to a person, and that it does not suffice to look at the disclosed data point out of context. Hence, the court found that a determination of comments containing views and opinions as personal data on this basis alone was a presumption that could not hold up in court. Furthermore, applying the Dreyer decision to this case, the court held that it is required to determine whether data constitutes personal data in the hands of the recipient, and whether re-identification is possible by the recipient, not by the original data controller. It remains to be seen whether data disclosed within an organization where the recipient department has no access to the original data can be considered anonymized.

In summary, anonymization is included in the GDPR under Recital 26. The exact definition is vague, however, with WP 216 providing guidance on its interpretation. WP 216 has been criticized as very strict, and recently we have seen some signs of more lenient interpretations, such as the SRB v EDPS case.

The Tech

GDPR-compliant anonymization is a very high bar. There are a number of techniques that can get you partially there, but they vary with regard to the complexity of their administration and their success. According to WP 216, robust anonymization is achieved along three vectors, namely the prevention

- of singling out an individual,
- of linking at least two records related to an individual back to them,
- and of inferring information concerning an individual.

Two approaches for anonymizing structured data that can help achieve anonymization, according to the WP 216 guidance, are randomization and generalization. Randomization refers to the technique of altering the individual values in the dataset, resulting in slightly inaccurate data points that retain overall distribution of the attributes throughout the dataset, and thus the usefulness for some data analytics and statistical purposes. Three techniques considered that fall under randomization are 1) addition of noise, 2) permutation, and 3) differential privacy.

The addition of noise to a dataset alone will not render it anonymous. In particular, singling out and linking records to an individual will still be possible, although the accuracy of the values attributed to an individual is less reliable. A danger with this technique is that a false value can expose an individual to greater risk than an accurate one, if it is disclosed and taken for accurate.

Permutation describes a randomization technique where attributes that pertain to certain individuals are switched with attributes pertaining to other individuals in the dataset. This can be useful when it is important to retain the exact attribute distribution within a dataset. By breaking the correlation between values and data subjects, if done correctly and attributes with strong logical correlations are shuffled around together, can help reduce reliable linkability of records as well as singling out. Inference attacks are rendered much harder, as the attacker must assume that the inference is based on a flawed hypothesis.

Differential privacy, a concept first introduced in 2006, is a mathematical framework that determines how much noise must be added to the output of a query performed on a dataset to ensure the privacy of individuals whose data is contained in the dataset. This is achieved by ensuring that the output of the query does not differ significantly depending on a particular individual's data's presence in or absence from the dataset. Differential privacy is a different approach compared to adding noise just by virtue of the fact that the dataset to which noise is added is not released, but rather retained by the data controller and simply queried by a third party. Strictly speaking, under the WP 216's understanding of anonymization, the output would nevertheless not be anonymized data, as the original dataset is left intact.

Generalization, on the other hand, describes the reduction of the level of detail by enlarging numerical intervals, e.g., providing an age range rather than the age of an individual, or the combination of several categories of data into one. Generalization techniques can be further divided into 1) Aggregation and K-anonymity, and 2) L-diversity/T-closeness.

K-anonymity, introduced in 1998, is a technique that aims to ensure that each individual's attribute value is shared with at least k other individuals in the dataset. This can be achieved by aggregation, or grouping, e.g., when the granularity of an attribute value is lowered from a date or other numerical value to an interval. This technique is strong in preventing singling out and linkability, but less so inference attacks.

L-diversity is an approach that builds on k-anonymity to safeguard against deterministic inference attacks. It works by ensuring that each attribute in every equivalence class has at least l distinct values, thereby constraining the occurrence of poor attribute variability. However, this technique cannot always prevent information leakage if attributes within a partition have limited variability or semantic meanings.

To address this limitation, **t-closeness** refines the L-diversity method by creating equivalent classes that resemble the initial attribute distribution in the table. It achieves this by requiring each class to have not only at least l different attribute values but also that each value is represented sufficiently to reflect the original distribution of each attribute. T-closeness is particularly useful when it's crucial to keep the data as close as possible to its original form.

A combination of the two techniques, randomization and generalization, is often used to enhance data protection while retaining data utility. In addition, data suppression (i.e., deletion of certain identifiers), masking, and tokenization are added for additional privacy protection. We discuss these in the next section under Pseudonymization. For now, it is important to note that the removal of personal identifiers alone does not meet the high standard of anonymization under the GDPR. The additional steps of determining the re-identification risk and then mitigating it by means of the techniques described in this section are necessary.

A difficulty that presents itself in addition to how data should be anonymized is the re-identification risk quantification. When have enough changes been made to the original data to consider it anonymized?

Sophisticated mathematical models may be required to quantify the risk to the data subjects and to guide their reduction. Open-source tools are available and this paper provides a thorough assessment of 13 solutions for structured tabular data anonymization and risk quantification developed by academic institutions, setting out their strengths and weaknesses. Note that only three of them have been employed outside of research tasks in real-world application, namely Q-Argus, sdcMicro, and ARX. Commercial tools are not in scope of this research as little is publicly known about their functionalities.

With regard to unstructured data (such as free text, images, and recordings), the 2022 article "The GDPR and unstructured data: is anonymization possible?" makes the case that under the WP 216 standard of anonymization, unstructured data cannot be anonymized as long as the original data set still exists because, in the two case studies that were undertaken, it is easy to tie the data set back to the original with any data remaining intact. However, some room for anonymization of unstructured data remains when the risk-based approach to anonymization is taken.

PSEUDONYMIZATION

Anonymization is not the only type of de-identification considered by the GDPR. In fact, there is an entire range of de-identified data, with anonymized data at the furthest end of the spectrum. Still subject to the GDPR but less stringently protected than identifiable data is pseudonymized data which is personal data that is not attributable to a specific individual without the use of additional information. This additional information must be kept separate and subjected to technical and organizational safeguards.

The Law

Pseudonymization generally refers to replacing an identifier by a different value. Pseudonymizing personal data allows its processing for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes.

Pseudonymized data is, under the GDPR's definition, reversible. Whether it has to otherwise meet the same standard as anonymized data in unclear. Note that the descriptions of anonymized and pseudonymized data in the Recitals differ in more regards than the reversibility. The former requires that information cannot be 'related' to an individual, whereas the latter speaks of data not being 'attributable' to an individual without additional information.

According to WP 216, pseudonymization is not a method of anonymization, as the ones set out above under generalization and randomization because it cannot adequately address the three identified risks, i.e., singling out, linkability, and inference attacks. Rather, pseudonymization of data is a security measure that can be used alone or in conjunction with anonymization techniques. WP 216 argues that pseudonymization only reduces the linkability of records to an individual. Identifiability is often still possible, albeit indirectly. And singling out of an individual's records is easily achieved, at least if the same new value is always used to replace a particular identifier, i.e., a name is always replaced by the same number. However, all of the anonymization techniques described above, which WP 216 also considers, have weaknesses with regards to one or more of the three risks, hence it is unclear what constitutes an anonymization technique vs. a security measure according to WP 216.

The Tech

Pseudonymization can be achieved in several different ways. As mentioned above, pseudonymizing data means replacing an identifier with a different value. The new value can be chosen independently of the original identifier or be derived from it. Different techniques with different degrees of reversibility and privacy can be employed.

Data encryption is the process of converting plain, readable data into an unintelligible form known as ciphertext, which can only be decrypted and understood by authorized individuals who have the necessary key or password. Encryption uses complex algorithms and mathematical functions to scramble the data in a way that makes it unreadable to anyone who doesn't have the decryption key or password.

Data hashing is a process of taking a piece of input data (such as a file or message) of any size and applying a mathematical algorithm to it, which produces a fixed-size output, known as a hash value, or simply a hash. This hash is unique to the input data and cannot be reversed back to the original input. A salted hash function is a type of hashing technique where a random value, known as the "salt," is added to the data being hashed before running it through a one-way hash function. This makes it more difficult for an attacker systematically guessing the input value, as they now also have to guess not only the input value but also the "salt," significantly increasing the number of possible combinations they would need to try.

Tokenization involves replacing sensitive data with a non-sensitive equivalent called a "token." Tokenization can be achieved through an encryption mechanism or the assignment of a randomly generated number which is then indexed to allow for reidentification. The token thus still retains some information about the original data and can be used to perform certain functions, such as authorization or verification.

Synthetic data can be generated to replace personal data with entirely new, fabricated data that does not contain any information about the original data. This may render the data less useful for some types of data analysis, yet it preserves the privacy of the data subject and can still retain some utility, e.g., statistical similarity to the original dataset or the ability to train large language models on this data.

Tokenization and synthetic data generation are Private AI's specialty. Using the latest advancements in Machine Learning, Private AI achieves unparalleled levels of accuracy, even for unstructured data. Trained to detect and redact over 50 entity types of personal information, health information, and payment card information in 52 languages, the time-consuming work of redacting personal information with high accuracy becomes three lines of code. If compliance with the GDPR's pseudonymization provisions is the goal, there is no better tool available on the market at this time. To see the tech in action, try our web demo, or request an API key to try it yourself on your own data.

OTHER DE-IDENTIFIED DATA

There is also a third category of de-identified data that we will refer to as Article 11 data. Article 11(2) contemplates the situation where "the controller is able to demonstrate that it is not in a position to identify the data subject" to whom the personal data pertains. Presumably this means that there cannot be a known, systematic way of reliably re-identifying the data subjects, but the possibility remains. In these instances, the controller is released from several obligations under the GDPR, that is, the data subject has no right to access, rectify, erase, or restrict the processing of this data, and the right to portability of the data subject is also precluded.

Arguably, the individual can exercise these rights if they voluntarily provide additional data that would then allow the data controller to identify the data that pertains to the requesting individual. See also Recital 57 on this note, which clarifies that the data controller should not refuse the individual's additional data if it is provided.

From the wording of Article 11 it appears that it is the data controller's particular ability that needs to be considered when determining whether data qualifies as Article 11 data or not. If that is correct, a dataset may be Article 11 data in the hands of some but not of others, depending on the available resources to identify a data subject.

Meeting the de-identification requirement under Article 11(2) will likely require the same techniques as contemplated above under Pseudonymization, only the standard for determining the re-identifiability is a different one. We can therefore refer you again to Private AI's redaction solution for the most efficient and accurate solution to achieve de-identification of your data.

CONCLUSION

In conclusion, privacy-enhancing technologies provide a valuable toolset for organizations to achieve GDPR compliance by identifying, anonymizing, and pseudonymizing personal data. With the rise of big data and AI, privacy concerns have become a central issue for organizations that handle personal data.

By implementing privacy-enhancing technologies, organizations can reduce the risk of data breaches, protect individual privacy rights, and build trust with their customers. While different techniques have their strengths and weaknesses, combining them can provide a comprehensive approach to data privacy. However, it is important to note that compliance with GDPR is an ongoing process, and organizations should regularly review and update their privacy-enhancing measures to keep pace with changing technology and regulatory requirements.

ABOUT THE AUTHOR

Kathrin Gardhouse is the part-time Privacy Evangelist at Private AI. She also works at a Canadian Bank where she is responsible for the privacy program management. As a CIPP/C certified German- and Ontario-trained lawyer with credentials from TorontoMU in Cyber Security Policy she brings additional data privacy, legal, and compliance expertise to Private AI.