Securing multimedia-based personal data: towards a methodology for automated anonymization risk assessment seeking GDPR compliance

Mikel Aramburu[1], David Redó[1], and Jorge García-Castaño[1]

[1]Vicomtech Foundation, Mikeletegi 57, 20009 Donostia-San Sebastián (Spain)

ABSTRACT

Anonymizing personal data in multimedia content (image, audio and text) has become crucial for secure data-sharing while adhering to the rigorous data compliance requirements of the European Union (EU) General Data Protection Regulation (GDPR). Given the substantial volume of data involved, manual verification of anonymization accuracy is not feasible due to the high potential for human error and the impracticality of scaling such efforts. Consequently, automated or semi-automated processes are indispensable. However, it is important to note that these methodologies cannot guarantee absolute anonymization, potentially leading to inadvertent disclosure of personal information and associated legal and privacy implications. Therefore, when dealing with extensive multimedia datasets, it is strongly advised to conduct a comprehensive anonymization risk assessment. In response to this challenge, we introduce a novel methodology with an innovative design to quantitatively evaluate the effectiveness and reliability of the anonymization techniques by generating metrics to calculate risk indicators to conduct a comprehensive anonymization risk assessment. This methodology is built based on de-identification techniques to protect personal data while preserving data integrity. Our approach leverages a novel algorithmic framework that helps humans inspect the anonymized dataset, ensuring higher data privacy and security. The methodology detects non-anonymized personal data within an extensive dataset automatically. This is achieved by extracting characteristics related to personal data during the anonymization process and correlating attributes from the surrounding data using sophisticated AI-driven analysis. Afterwards, a rule-based algorithm is applied to the extracted characteristics from both processes to identify and qualitatively assess the anonymization risk. We demonstrate the applicability and effectiveness of our methodology through a focused application on license plates and face anonymization, utilizing a dataset of non-annotated vehicles and human images. By offering a scalable solution to evaluate anonymization risk while data-sharing, our methodology represents a pivotal step towards achieving GDPR compliance and processing practices, facilitating safer data-sharing environments across industries.

Keywords: Anonymization risk assessment, anonymization quality, automatic anonymization, de-identification, personal data, privacy preservation, GDPR, data-sharing

1. INTRODUCTION

In the digital age, multimedia content sharing—encompassing images, audio, and text—has become a crucial component of various industries, including healthcare, law enforcement, and social media. However, this content often contains personal data, making it a significant concern for privacy and security. The European Union's GDPR,[1] implemented in 2018, mandates stringent measures to protect personal data during processing and sharing, emphasizing the necessity of anonymization. Anonymization ensures that personal data cannot be attributed to specific individuals without additional information, thereby safeguarding privacy rights during data-sharing activities. Data-sharing is essential for innovation and collaboration across sectors such as healthcare, finance, and public safety, where the exchange of multimedia content can drive advancements in technology and

Further author information: (Send correspondence to M.A.)
M.A.: E-mail: maramburu@vicomtech.org, Telephone: +[34] 943 30 92 30
D.R.: E-mail: dredo@vicomtech.org, Telephone: +[34] 943 30 92 30
J.G.C.: E-mail: jgarciac@vicomtech.org, Telephone: +[34] 943 30 92 30

Artificial Intelligence for Security and Defence Applications II, edited by Henri Bouma, Radhakrishna Prabhu, Yitzhak Yitzhaky, Hugo J. Kuijf, Proc. of SPIE Vol. 13206, 132060C · © 2024 SPIE
0277-786X · doi: 10.1117/12.3031687

service delivery. However, it must be done in a manner that respects individuals' privacy rights. The challenge lies in balancing the need for data utility with the requirement to protect personal data from unauthorized access or misuse. While anonymization techniques offer a solution, the complexity and diversity of multimedia content make it difficult to ensure complete anonymization, particularly when large datasets are involved. This difficulty is exacerbated by the need to comply with GDPR, which places strict obligations on organizations to protect the privacy of individuals whose data they handle.

Despite advancements in anonymization techniques, ensuring personal data privacy during multimedia data-sharing remains a significant challenge. Current anonymization methods, though sophisticated, are not foolproof and often require manual inspection to verify their effectiveness. This process is time-consuming and prone to human error, which can result in personal data being inadvertently exposed during sharing. Such breaches not only infringe on individuals' privacy rights but also carry severe legal and financial repercussions under GDPR. The volume of multimedia data being shared across various platforms and industries further complicates this issue, highlighting the need for scalable, automated solutions that can assess the quality of anonymization. Without reliable and efficient methods for verifying that data has been properly anonymized, organizations risk violating GDPR and infringing on personal privacy rights when sharing data.

In response to these challenges, this paper introduces a novel methodology aimed at automating the anonymization risk of multimedia content, with a specific focus on enabling secure and compliant data-sharing. By introducing a quantitative metric for assessing the data-sharing risk related to anonymization, this methodology aims to provide organizations with the tools to share data securely, ensuring that no personal rights are infringed upon. This approach facilitates GDPR compliance and supports the broader goal of creating safer data-sharing environments across various industries. The primary objectives of this research are: (i) to develop a framework that ensures personal data is effectively anonymized while preserving the integrity and utility of multimedia content; (ii) to leverage advanced AI-driven analysis to detect and evaluate non-anonymized data within extensive datasets, thereby reducing the need for human inspection and minimizing the risk of privacy breaches. This paper contributes to the anonymization risk assessment seeking GDPR compliance with the following:

- A methodology to automatically detect non-anonymized personal data within a dataset.
- Quantitative evaluation metrics that reflect the anonymization success for further data-sharing risk indicators.
- Implementation and demonstration of the methodology with vehicles containing license plates and people's face-based anonymized datasets.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive review of related work, including an overview of existing anonymization techniques and the challenges associated with their application to multimedia content. Section 3 introduces the proposed methodology, detailing the algorithmic framework and the metrics developed for assessing the possible data-sharing risk. In Section 4, we present a case study focusing on license plate and facial anonymization, demonstrating our approach's applicability and effectiveness in seeking GDPR compliance and data-sharing. Section 5 discusses the results and highlights the advantages and limitations of our proposed solution. Finally, Section 6 concludes the paper with a summary of our findings and suggestions for future research.

2. BACKGROUND AND RELATED WORK

Having established the importance of secure data-sharing under GDPR, the following section reviews existing anonymization techniques and evaluates their effectiveness in various multimedia contexts.

The European Union (EU) has established comprehensive legal frameworks designed to achieve two crucial objectives: safeguarding individuals' privacy and data protection rights, while simultaneously promoting the sharing and processing of data to fuel innovation. Among these frameworks, the GDPR, the Law Enforcement Directive (LED), and the forthcoming Artificial Intelligence Act (AI Act) play key roles in regulating personal data processing within the EU.

The GDPR is particularly significant, applying to all entities, whether public or private, that handle personal data within the EU or related to EU residents. It sets out fundamental data protection principles that must be adhered to in all data processing activities (Article 5). Additionally, the GDPR distinguishes between general personal data and more "sensitive" data, which could lead to discrimination and therefore requires additional safeguards (Article 9(1)). However, it is vital to stress that all types of personal data—whether sensitive or not—pose substantial risks to individuals' rights and legitimate interests depending on how the data is processed. Ensuring privacy preservation when sharing data is paramount to minimizing these risks. The GDPR also places stringent obligations on data controllers and processors to secure the confidentiality and integrity of personal data throughout its lifecycle. Non-compliance with these rigorous requirements can result in severe penalties, including significant administrative fines.

In response to the rapid advancement of AI technologies, the European Commission introduced the Artificial Intelligence Act in 2021, which is expected to be adopted in 2024 and come into full effect by 2026. The AI Act seeks to regulate the development, deployment, and use of AI systems across various sectors, ensuring their transparency, accountability, and alignment with ethical standards. It categorizes AI systems based on their risk levels: unacceptable risk (prohibited AI practices), high risk, limited risk, and minimal risk. High-risk AI systems, in particular, must meet specific requirements, including robust data governance, risk management, detailed technical documentation, transparency, and human oversight.

The legal landscape described above presents a complex challenge for compliance, especially for developers with limited resources. Compliance requires navigating intricate requirements related to consent management, data protection principles, risk assessments, and accountability measures. Developers must prioritize understanding and implementing these regulations effectively across diverse technological contexts, which demands significant time, expertise, and resources. Failure to comply can lead to serious legal repercussions, underscoring the necessity of prioritizing privacy preservation and regulatory compliance throughout the development process, despite its complexity and resource demands.

Given these challenges, the automated anonymization risk assessment methodology proposed in this work offers a comprehensive solution that can significantly assist all stakeholders involved in the research, development, and deployment of AI-based technologies, helping to ensure that privacy is preserved and compliance is maintained.

Different privacy preservation techniques can be applied to cover or eliminate personal data. The main techniques fall under two subgroups concerning data-sharing issues and GDPR:

Pseudonymisation: Article 4(5) of the GDPR defines 'pseudonymisation' as "the processing of personal data [so] that the personal data can no longer be attributed to a specific data subject without the use of additional information".

Anonymization: Anonymization, in turn, is transforming personal data so that it can no longer be attributed to a specific individual, either directly or indirectly. Recital 26 of the GDPR defines anonymous information as "information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."

The pivotal distinction lies in the reversibility of the pseudonymization process, allowing for the potential re-identification of individuals. In contrast, anonymization is inherently irreversible, precluding the re-identification of individuals from their data. On the one hand, pseudonymized data falls under the application of regulations like the GDPR or other similar EU regulations, given its reversible nature. Consequently, additional documentation and efforts are required to render such data shareable. On the other hand, anonymized data is exempt from applying the above-mentioned legal frameworks, making it a preferred choice for data-sharing purposes. From a data-sharing perspective, optimal privacy preservation techniques involve automatically anonymizing all data, ensuring its integrity while mitigating the application of legal regulations and methods.

To ensure compliance with GDPR's data-sharing requirements, anonymization is a critical measure that must be implemented. Different multimedia data types (such as tabular, image, audio, and text) necessitate specific anonymization techniques to achieve GDPR compliance. For example, **tabular** data can be anonymized through various methods, as discussed in studies by Murthy *et al.* [2], Ferreira *et al.* [3], and Majeed *et al.* [4]. These techniques often focus on anonymizing healthcare data,[5] including masking, generalization, and data

suppression, among other methods. When considering **image**-based data, privacy preservation techniques are focused mainly on face and license plate anonymization. Techniques include synthetic generation using Generative AI, which prevents the disclosure of original information,[6] and GAN-based methods that preserve attributes during face anonymization.[7] Other methods involve face de-identification through caricature generation[8] and mask-guided facial image manipulation.[9] Additionally, real-time face blurring is employed as an alternative to generating synthetic faces.[10] For license plate anonymization obscuring or blurring images and videos to protect sensitive information is a common technique.[11] And services offered by companies like Brighter AI[12] for anonymizing vehicle license plates and faces are quite popular. Comprehensive reviews of visual privacy preservation techniques further detail these methods, as explained by Ravi *et al.* [13]. In the domain of **audio** data, anonymization techniques primarily target speaker identification, employing methods such as neural audio codec language models[14] and prosody cloning techniques.[15] For **text**-based data, anonymization is achieved through various methods, including word embeddings[16] and more recent approaches leveraging Large Language Models (LLMs).[17] Each of these anonymization techniques is crucial for safeguarding personal data and ensuring adherence to GDPR standards across different multimedia data types.

However, the effectiveness of these anonymization techniques is not absolute, and they may fail to anonymize personal data. Efforts have been made to evaluate the quality of anonymization, including initiatives that measure the degree of pseudo-anonymization of a dataset from both legal and technical perspectives.[18] In the context of image data, particularly facial images, the focus of anonymization risk assessment is primarily on maintaining the anonymity of individuals and preventing re-identification, while still preserving certain facial characteristics.[19] However, these assessments often do not rigorously verify whether all personal data within the dataset has been fully anonymized. To address this gap, the present work seeks to develop an automated system for anonymization risk assessment. This system will ensure the accurate identification of personal data across the entire dataset and ensure its anonymization.

3. AUTOMATED ANONYMIZATION RISK ASSESSMENT

The methodology for automated anonymization risk assessment is grounded in the extraction of metadata from both the anonymization process and the correlated attributes present in the surrounding data. This metadata is then subjected to a rule-based analysis to determine whether the personal data has been effectively anonymized.

The initial stage involves extracting metadata from the anonymization process and the associated attributes, as illustrated in Fig. 1. These extractions are conducted independently for the anonymization process and the inspection of the surrounding data. The first component focuses on anonymizing the input data while simultaneously capturing metadata, such as the location of personal data within multimedia content. The second component, which involves inspecting the surrounding data, aims to identify correlated attributes that, while related to personal data, do not themselves constitute personal data. The metadata collected from both components is consolidated into a single file and prepared for analysis by the subsequent rule-based algorithm.

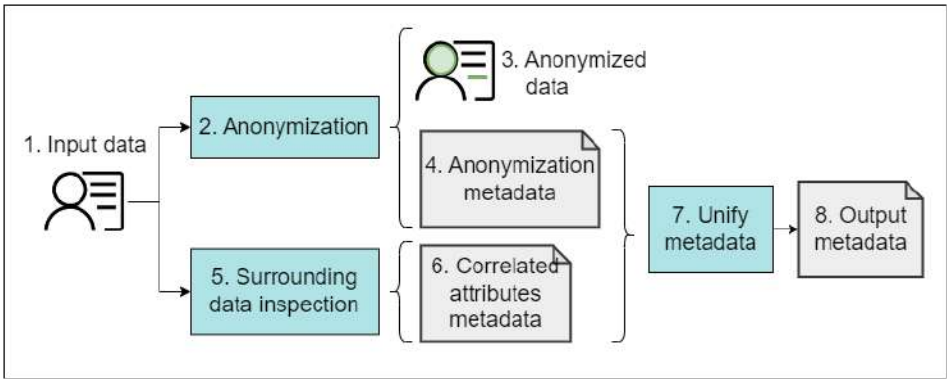

Figure 1. Metadata extraction from the anonymization and correlated attributes from the surrounding data.

The points are outlined as follows:

1. **Input Data**: This refers to the dataset or collection of data containing personal information that needs to be anonymized, along with the extraction of correlated metadata.
2. **Anonymization**: This step involves the anonymization of personal data within the input dataset, accompanied by the extraction of relevant metadata during the process.
3. **Anonymized Data**: The resulting data after the anonymization process, which no longer contains any personal identifiers.
4. **Anonymization Metadata**: Metadata that is extracted from the personal data present within the input dataset during the anonymization process.
5. **Surrounding Data Inspection**: A process to extract metadata that is related to, but not part of, the personal data within the input dataset.
6. **Correlated Attributes Metadata**: Metadata that characterizes or represents attributes correlated with the personal data.
7. **Unify Metadata**: A step to merge the metadata extracted from both the anonymization process and the correlated attributes.
8. **Output Metadata**: The final metadata file produced as the outcome of the entire process.

The purpose of the rule-based algorithm is to integrate and evaluate the metadata obtained from both the anonymization process and the surrounding data inspection to determine whether individuals' privacy has been adequately preserved. This algorithm is designed to detect non-anonymized personal data by comparing the results of the surrounding data inspection with those of the anonymization process. This step is crucial for identifying potential data leaks within the entire dataset and is dependent on both the nature of the data and the characteristics of the surrounding data.

The final step involves calculating an anonymization risk measure. This quantitative measure is derived by comparing the successfully anonymized and verified personal data against the total amount of personal data identified through the surrounding data inspection, see Eq. 1.

$$\mathrm{AQ} = \frac{\mathrm{PA} + \mathrm{SANP}}{\mathrm{ASA}}, \tag{1}$$

Where,

- AQ = Anonymization Quality
- PA = Properly Anonymized personal data
- SANP = Surrounding Attribute data with No Personal data
- ASA = All Surrounding Attribute data

Employing the anonymization quality the anonymization risk assessment measure can be inferred as depicted in Eq. 2. The result is expressed as a percentage, indicating the overall risk of the dataset containing non-anonymized data.

$$\mathrm{ARAM} = 1 - AQ, \tag{2}$$

Where,

- ARAM = Anonymization Risk Assessment Measure
- AQ = Anonymization Quality

4. METHODOLOGY IMPLEMENTATION AND CASE STUDY

The methodology has been implemented with two types of personal data in image multimedia datasets:

- **License plates on vehicles dataset.**

  An image-based vehicle dataset has been selected to assess the anonymization risk. The dataset contains different vehicle types in images that have mainly European-like license plates in various environments. The types of vehicles include primarily four-wheeled vehicles: cars, vans, trucks and buses, where the license plate of some vehicles is usually recognisable. The environments of the images are focused on CCTV-like images that can vary from vehicles on a highway to parked vehicles on a city street. The images were scraped from the internet, by selecting those images with open licenses to use and modify.

- **People's faces dataset.**

  The people-based dataset contains images of people of different ethnicities in various environments. The environments of the images are mobile camera-like images that can resemble social network images. I.e., the images usually show the whole body, or at least part of it, and the face is usually recognisable. The images were scraped from the internet, by selecting those images with open licenses to use and modify.

4.1 Application to vehicles dataset

To follow the methodology in the vehicles dataset, an anonymization process and a surrounding data inspection have been applied to the dataset. Fig. 2 resumes the processes to get the metadata from the input images and save them into files.

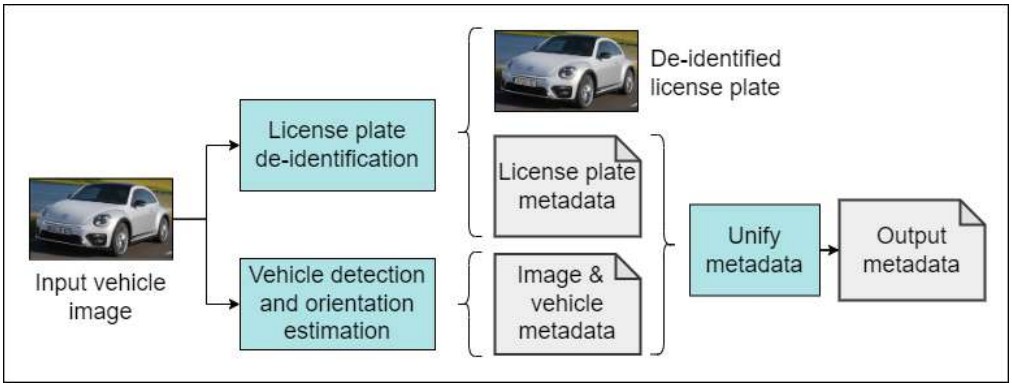

Figure 2. Metadata extraction from license plate anonymization and vehicle detection and orientation estimation.

The first process is license plate anonymization by de-identification. Fig. 3 depicts the general process that composes the anonymization by de-identification of license plates, which is divided into four main actions: license plate detection, synthetic license plate generation, style transfer, and in-painting. The process mirrors the steps depicted in the work of Baloukas *et al.* [20].

License plate(s) detection. This block is in charge of detecting all the license plates in the input data. To perform such a task, a regression network that estimates the four corner points of the polygon that wraps the license plate is employed. Conventional object detectors only provide bounding box information, which would remove additional information around the original license plate when the in-painting step is executed.

Synthetic license plate generation. Concurrently, a non-existent license plate number is created and rendered on a 2D template for EU-like license plates. The design has a blue strip on the left side containing two elements: the European flag symbol and a country code. Since any anonymization procedure must not reveal personal data, the license plate number is randomly generated with a seven-element sequence of numbers, characters, and spaces.

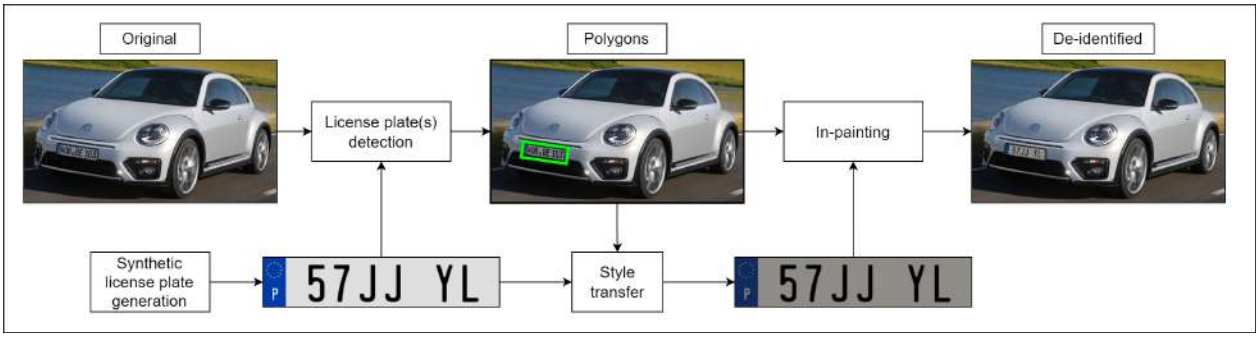

Figure 3. Applied de-identification technique to anonymize license plates and obtain metadata.

Style transfer. To mirror the statistical properties of the original data, this module is in charge of transferring the style of the detected license plate into the synthetic one using Generative Adversarial Networks (GANs) techniques.

In-painting. The last step involves replacing the original data with the synthetic one. This is performed by first projecting homographically the four corners of the customized license plate into the detected corners from the first module and then rendering the projected image on top of the original image.

To get the correlated attributes metadata, a vehicle detection and orientation estimation process is deployed. In this case, the method presented by Kumar *et al.* [21] has been applied to obtain the position and orientation of different vehicles. The metadata extracted from this process and the metadata extracted from the anonymization process serve as input for estimating the quality of the anonymization process and help in the identification of non-anonymized personal data.

First, the metadata of the overall image is extracted, which is the size of the image, i.e., width and height values. The License Plate De-identification process, precisely from the four corner points regression, provides the position, defined by the four points, and each license plate's detection score. The Vehicle Detection and Orientation Estimation extracts the position, defined by a BBOX, the detection score, and the orientation of each detected vehicle. The number of detected license plates and vehicles can be inferred from the metadata. Tab. 1 lists and defines each metadata value extracted from each image.

Once all of the metadata is gathered the next step is to apply the rule-based algorithm that will decide if the vehicles present any personal data (license plates in this case) and if they are correctly anonymized.

The anonymization quality for the license plates is determined by the sum of properly anonymized vehicles and vehicles that do not have recognisable license plates against the total detected number of vehicles. The rule-based methodology follows this logic: For each detected vehicle, there should be a license plate. The detected license plates should be inside the vehicles. However, this is not always true, as there are multiple cases where a vehicle could be detected, but the license plate is not.

For instance, a license plate in an image could be **occluded** by other vehicles or objects, such as trees, signals, etc. The variety of these cases is vast and can not be underestimated. The overlapping of two or more vehicles can be deduced using the vehicles' position. This way, if two vehicles appear in the same area, it can be considered that at least one of the license plates is not recognisable because it is hidden behind.

Another case occurs when the vehicle is **side-oriented** to the camera view. As the vehicles include the license plate in the front and the back, if the vehicle's orientation is from the side, looking for the camera view, the license plate will not be recognisable. The vehicle detection and orientation estimation module will provide information on these cases.

The **size of the detected vehicle compared to the image** is a critical factor because smaller vehicles may have license plates that are not detectable by the algorithm, thereby reducing the risk of privacy breaches. When a vehicle is detected with the front or back orientation but no license plate related to this vehicle and no occlusion is present, the reason can be related to the size. If the size of the vehicle is small, which can be

Table 1. Metadata extracted from the license plate anonymization and vehicle detection processes.

Source	Metadata	Definition
Input Image	Width	Original image's width.
	Height	Original image's height.
License Plate De-identification	Four corner points polygon	Position and size of the detected license plate, which is the four-point coordinates clockwise ordered as: [Top left,Top right, Bottom right, Bottom left].
	Detection score	Detection score of the detected license plate.
Vehicle Detection and Orientation Estimation	BBOX	Position and size of the detected vehicle, which is the bounding box coordinates ordered as [Top left, Bottom right].
	Detection score	Detection score of the detected vehicle.
	Vehicle orientation	Detected vehicles' orientation as front, back or side.

compared with the total image size, the license plate might also be too small to be detected. In this case, the size makes the license plate's characters unrecognisable. Thus, the license plate is considered anonymous as it does not reveal personal information.

The **detection scores** from both the license plates and vehicles can elucidate why a license plate is not considered. The quality of the image can affect the detection score of the vehicles. If a detection score is not high enough, it could serve as a threshold to consider a vehicle not clearly visible and, therefore, the license plate unrecognisable.

Based on the aforementioned points a rule-based algorithm has been designed to check this logic for each image. Alg. 1 resumes the logic steps.

The rule-based algorithm indicates the sum of the properly anonymized vehicles and those considered side-oriented, too small, overlapping or not considered by the detection score. The total number of vehicles in the image is extracted from the metadata given by the vehicle detection process. To get the final anonymization risk measure Eq. 3 and 4 are applied.

$$\mathrm{VAQ} = \frac{\mathrm{PAV} + \mathrm{VNRLP}}{\mathrm{TDV}}, \tag{3}$$

$$\mathrm{VARAM} = 1 - \mathrm{VAQ}, \tag{4}$$

Where,

- VARAM = Vehicles' Anonymization Risk Assessment Measure
- VAQ = Vehicle Anonymization Quality
- PAV = Properly Anonymized Vehicles
- VNRLP = Vehicles with No Recognisable License Plates
- TDV = Total number of Detected Vehicles

Algorithm 1 Anonymization Verification for License Plates

```
for each detected vehicle do
    for each detected license plate do
        if license plate is not assigned to any vehicle then
            check if the license plate is inside the vehicle centre
            if license plate belongs to the vehicle then
                assigns the license plate to the vehicle
                vehicle is anonymized
                break                          ▷ No need to check other plates if anonymized
            end if
        end if
    end for
    if vehicle is not anonymized then
        if vehicle orientation is from a side or vehicle relative size to the image is small or vehicle overlaps with other vehicle or vehicle detection score is too low then
            license plate is not recognisable
        end if
    end if
end for
```

4.2 Application to people dataset

As in the previous vehicle dataset use case, an anonymization process and a surrounding data inspection have been applied to the dataset to follow the methodology. Fig. 4 resumes the steps to get the metadata from the input images and save them into files.

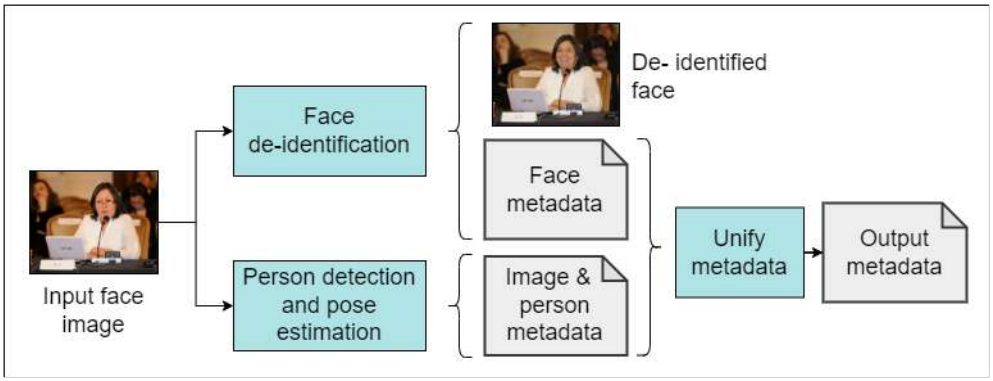

Figure 4. Metadata extraction from face anonymization and person detection and pose estimation.

The face anonymization de-identification is based on two steps: detecting the faces in the image and applying generative AI to create a synthetic face in place of the original. Fig. 5 depicts these steps.

Face detection. This block detects all faces in the input data, the output of which is a bounding box for each face. To detect the face, DeepFace[22] with YoloV8 was used.

Synthetic face generation and anonymization. The rise of generative AI has enabled direct anonymization, compressing several steps into one. For this reason, this block is responsible for generating a synthetic face, applying the original style, and painting the new one on top of the original one. It is obtained using inpainting abilities from diffusion models,[23] which only needs the detected face as an input mask.

The person position and pose detection are calculated using the libraries from Mmpose[24] together with the YOLOX[25] and YOLO-Pose.[26] The final pose estimation is done by and from the previous model's output to decide if the body is front, side or back-oriented. Similarly to the license plate use case, the metadata extracted

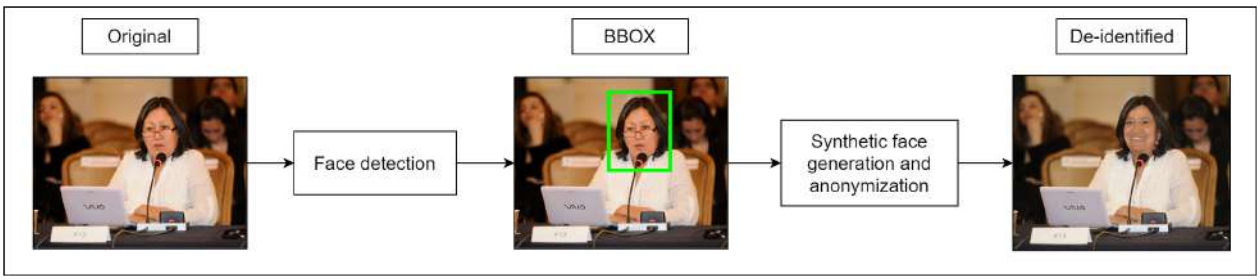

Figure 5. Applied de-identification technique to anonymize license plates and obtain metadata.

from this process and additional metadata extracted from the anonymization process will serve as input for the risk assessment methodology. Tab. 2 resumes the metadata extracted from each person-containing image.

Table 2. Metadata extracted from the face anonymization and body detection.

Source	Metadata	Definition
Input Image	Width	Original image's width.
	Height	Original image's height.
Face detection	BBOX	Position and size of the detected face, which is the bounding box coordinates ordered as [Top left, Bottom right].
	Detection score	Detection score of the detected face.
Person Detection and Pose Estimation	BBOX	Position and size of the detected person, which is the bounding box coordinates ordered as [Top left, Bottom right].
	Detection score	Detection score of the detected person.
	Person orientation	Detected person´s orientation as front, back or side.

As in the license plate case, the quality of the anonymization process is ruled by the sum of properly anonymized people and the people whose faces are not recognisable against the number of detected people. The rule-based methodology follows this logic: For each detected person there should be one face. The face should be on the upper part of the person's body. But, as explained in the previous vehicle case, this is not always true, as the faces can be occluded while the person is still recognisable.

The **occlusions** by other people or objects; **person orientations** whether it is from the front, side or back; the **size of the faces compared to the image**, and the **detection scores** also play a critical role in this case. Similar to the previous case, the decision if a person's identity is covered should be taken based on the metadata extracted from these processes.

Based on the aforementioned points a rule-based algorithm has been designed to check this logic for each image. The proper anonymization verification for the faces in Alg. 2 resumes the steps taken to apply the methodology.

The rule-based algorithm indicates the sum of the properly anonymized people and the ones considered side-oriented, too small concerning the image, overlapping or not considered by the detection score. The total number

Algorithm 2 Anonymization Verification for Faces

```
for each detected person do
    for each detected face do
        if face is not assigned to any person then
            check if the face is inside the person's centre
            if face belongs to the person then
                assigns the face to the person
                person is anonymized
                break                                  ▷ No need to check other faces if anonymized
            end if
        end if
    end for
    if person is not anonymized then
        if person's orientation is from the back or person's relative size to the image is small or person overlaps with another person or person detection score is too low then
            face is not recognisable
        end if
    end if
end for
```

of people in the image is extracted from the metadata given by the people detection process. To get the final anonymization risk measure the following Eq. 5 and 6 are applied.

$$\mathrm{PAQ} = \frac{\mathrm{PAP} + \mathrm{PNRF}}{\mathrm{TDP}}, \tag{5}$$

$$\mathrm{PARAM} = 1 - \mathrm{PAQ}, \tag{6}$$

Where,

- PARAM = People's Anonymization Risk Assessment Measure
- PAQ = People Anonymization Quality
- PAP = Properly Anonymized People
- PNRF = People with No Recognisable Face
- TDP = Total number of Detected People

5. RESULTS

The metrics and results presented in this section quantify the anonymization risk while data-sharing and related to GPDR compliance of the datasets mentioned above by applying the automated anonymization risk assessment from Sec. 4. So, the analysis is not focused on achieving high anonymization rates but on correctly identifying the personal data leaks from the anonymization processes and generating risk indicators.

5.1 Vehicles dataset

The automated anonymization risk assessment yielded a score of **13%** for the vehicles dataset. This result indicates that the anonymization process effectively covers the majority of recognisable license plates. However, the score is not flawless, suggesting that some vehicles with identifiable license plates were not successfully anonymized. Fig. 6 illustrates examples of such instances.

Several vehicles with recognisable license plates were not successfully anonymized in the images presented. The methodology effectively identifies these vehicles and flags them for user review to ensure compliance with

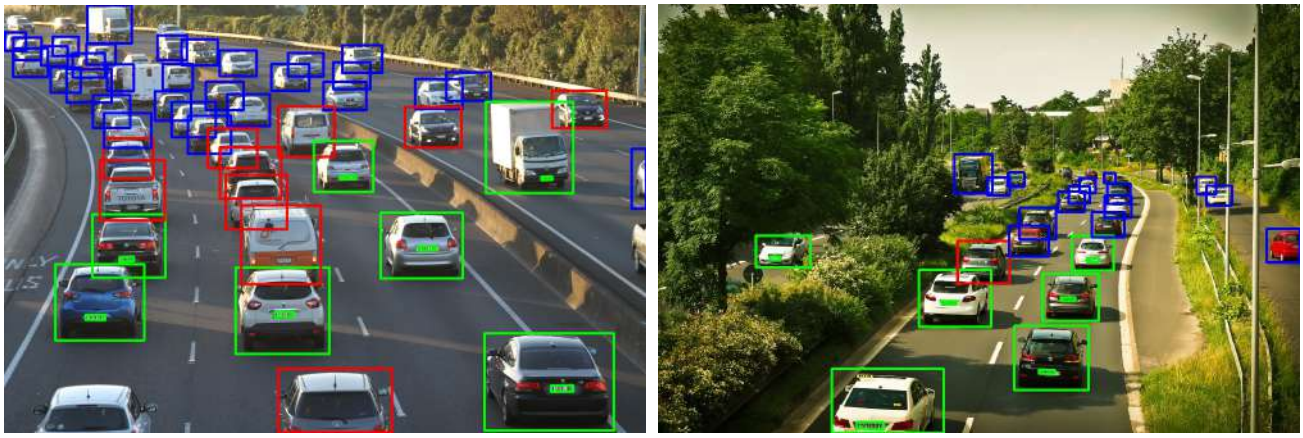

Figure 6. Examples of non-totally anonymized vehicle images. The methodology captures those vehicles whose license plates have been properly anonymized (in green), vehicles with no recognisable license plates (in blue) and vehicles with possibly recognisable license plates without anonymization (in red).

GDPR. The user can then manually anonymize these images or exclude them from further data-sharing or use. For example, in the image on the left in Fig. 6, some license plates on vehicles red bounding boxes remain identifiable, and the assessment correctly marks them as non-anonymized. However, the methodology also flags other vehicles in red bounding boxes as non-anonymized even though their license plates are not visible or recognisable. For instance, the two vehicles in the right part of the image. These instances represent edge cases where it is advisable to conduct a thorough review rather than dismiss them. The methodology is designed to err on the side of caution by flagging these edge cases as non-anonymized and recognisable to prevent potential data leakage, thereby facilitating a more comprehensive dataset review for users.

When proper anonymization is applied, the methodology successfully differentiates between visible and recognisable personal data. Fig. 7 provides two examples of this scenario. Although some vehicles do not have anonymized license plates in the image on the left, the methodology correctly determines that these plates are not recognisable due to factors such as the vehicles' size or orientation. Conversely, the image on the right contains vehicles oriented to the side, where the license plates are not visible, and the methodology accurately indicates that no personal data is present.

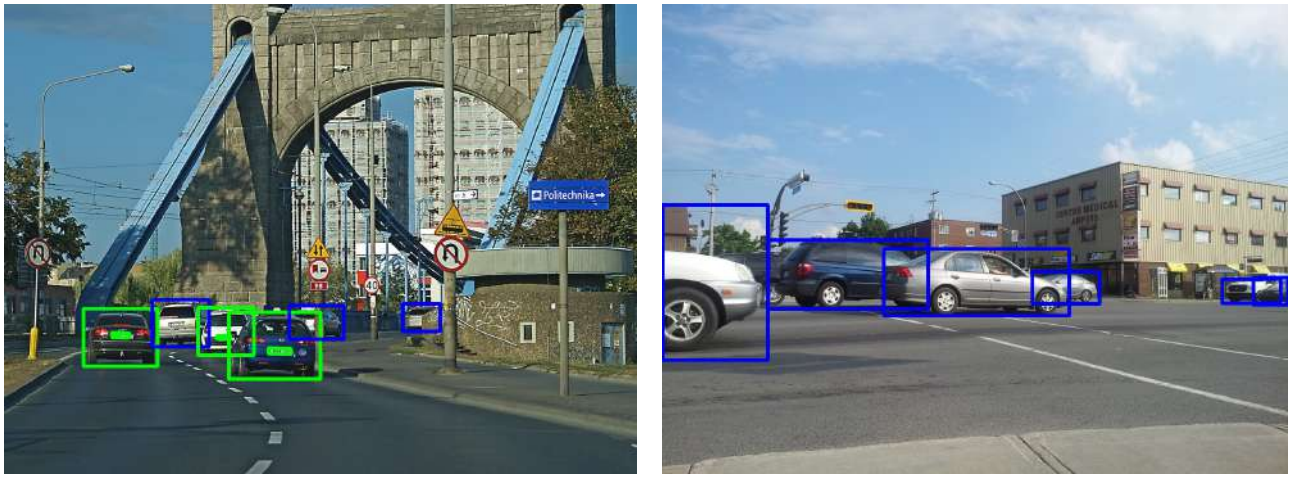

Figure 7. Examples of properly anonymized vehicle images. The methodology is capable of ensuring that there is no recognisable license plate.

5.2 People dataset

The automated anonymization risk assessment of the people dataset yielded a score of **4%**. This low score indicates that the anonymization process was highly effective, successfully anonymizing most of the recognisable faces. However, the fact that the score is not perfect (0%) suggests that some images may still contain identifiable personal data, specifically non-anonymized faces. Fig. 8 provides an example of such a case.

Figure 8. Example of a non-anonymized side-oriented recognisable face (left). And an image with properly anonymized faces (right). The methodology identifies the non-anonymized images.

In this instance, the methodology proves to be instrumental in identifying faces that have not been properly anonymized, ensuring that these faces, which remain visible and potentially recognisable, are flagged for further attention. This step is crucial in preventing the inadvertent exposure of personal data, which could otherwise lead to privacy breaches. The methodology's ability to detect such lapses highlights its value in enhancing the overall effectiveness of the anonymization process and ensuring compliance with GDPR standards.

6. DISCUSSION AND CONCLUSIONS

This paper presents a methodology for assessing the quality of anonymization in multimedia data, specifically targeting GDPR compliance for secure data-sharing. The approach focuses on automating the detection of non-anonymized personal data within large multimedia datasets, reducing the dependency on manual inspection and the associated risks of human error.

The core of the methodology lies in a rule-based algorithm that integrates metadata extracted from both the anonymization process and the surrounding data context. This integration allows for a comprehensive assessment of whether personal data has been successfully anonymized, ensuring that no sensitive information is inadvertently exposed. The quantitative evaluation metrics provide a reliable measure of anonymization risk, offering organizations a valuable tool for mitigating privacy risks in data-sharing.

The effectiveness of this methodology was demonstrated through case studies involving license plates and facial anonymization. The results showed that our approach could accurately identify both properly anonymized data and cases where anonymization was incomplete or ineffective. With anonymization risk measures scores of 13% for the vehicle dataset and 4% for the people dataset, the methodology proves to be robust, properly indicating the non-anonymized personal data within the dataset.

While the rule-based algorithm employed in this study has proven effective, future research could explore the potential of applying Machine Learning (ML) and Deep Neural Networks (DNNs) to this task. These advanced techniques could offer enhanced capabilities for identifying patterns and correlations within data that may not be easily captured by rule-based systems. By leveraging the power of ML and DNNs, the anonymization risk assessment could become even more accurate and adaptive, potentially identifying hidden details within the data that could pose privacy risks. Nevertheless, while machine learning and deep neural networks could offer enhanced

pattern recognition capabilities, they also require substantial computational resources and may introduce new challenges in model transparency and explainability.

Overall, this research contributes to the ongoing efforts to balance data utility with privacy protection, ensuring that the benefits of data-sharing can be realized without compromising individual privacy rights.

7. ACKNOWLEDGEMENTS

This paper is performed in the H2020 project LAGO ("Lessen Data Access and Governance Obstacles"). This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101073951.

REFERENCES

[1] "Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (general data protection regulation)." https://eur-lex.europa.eu/eli/reg/2016/679/oj (2016). Accessed: 2024-08-12.

[2] Murthy, S., Abu Bakar, A., Abdul Rahim, F., and Ramli, R., "A comparative study of data anonymization techniques," in [*2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*], 306–309 (2019).

[3] Marques, J. F. and Bernardino, J., "Analysis of data anonymization techniques," in [*International Conference on Knowledge Engineering and Ontology Development*], (2020).

[4] Majeed, A. and Lee, S., "Anonymization techniques for privacy preserving data publishing: A comprehensive survey," *IEEE Access* **9**, 8512–8545 (2021).

[5] Olatunji, I., Rauch, J., Katzensteiner, M., and Khosla, M., "A review of anonymization for healthcare data," *Big Data* (03 2022).

[6] Kamelski, T. and Olivos, F., "Ai-replicas as ethical practice: Introducing an alternative to traditional anonymization techniques in image-based research," (2024).

[7] Barattin, S., Tzelepis, C., Patras, I., and Sebe, N., "Attribute-preserving face dataset anonymization via latent code optimization," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 8001–8010 (June 2023).

[8] Laishram, L., Lee, J. T., and Jung, S. K., "Face de-identification using face caricature," *IEEE Access* **12**, 19344–19354 (2024).

[9] Khojasteh, M. H., Moradzadeh Farid, N., and Nickabadi, A., "Gmfim: A generative mask-guided facial image manipulation model for privacy preservation," *Computers & Graphics* **112**, 81–91 (2023).

[10] Piao, X., Piao, Z., Yoo, S. J., and Gu, Y. H., "Robust sensitive-information de-identification framework based on relative-position estimation of objects in closed-circuit television videos," *Alexandria Engineering Journal* **89**, 172–183 (2024).

[11] CR, B. N., S, D. M., Navya, D., S, B., Uttarakumari, M., and Manonmani, S., "License plate detection and privacy-aware masking in surveillance imagery," in [*2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*], 1–8 (2023).

[12] Brighter AI, "Brighter ai." https://brighter.ai/ (2024). Accessed: 2024-08-12.

[13] Siddharth Ravi, P. C.-P. and Florez-Revuelta, F., "A review on visual privacy preservation techniques for active and assisted living," *Multimedia Tools and Applications* **83**, 14715–14755 (2024).

[14] Panariello, M., Nespoli, F., Todisco, M., and Evans, N., "Speaker anonymization using neural audio codec language models," in [*ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 4725–4729 (2024).

[15] Meyer, S., Lux, F., Koch, J., Denisov, P., Tilli, P., and Vu, N. T., "Prosody is not identity: A speaker anonymization approach using prosody cloning," in [*ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 1–5 (2023).

[16] Hassan, F., Sánchez, D., and Domingo-Ferrer, J., "Utility-preserving privacy protection of textual documents via word embeddings," *IEEE Transactions on Knowledge and Data Engineering* **35**, 1058–1071 (2023).

[17] Pissarra, D., Curioso, I., Alveira, J., Pereira, D., Ribeiro, B., Souper, T., Gomes, V., Carreiro, A. V., and Rolla, V., "Unlocking the potential of large language models for clinical text anonymization: A comparative study," (2024).

[18] Kolain, M., Grafenauer, C., and Ebers, M., "Anonymity assessment – a universal tool for measuring anonymity of data sets under the gdpr with a special focus on smart robotics," *Rutgers University Computer & Technology Law Journal* **48**(2) (2022). Available at SSRN: https://ssrn.com/abstract=3971139.

[19] Ye, M., Shen, W., Zhang, J., Yang, Y., and Du, B., "Securereid: Privacy-preserving anonymization for person re-identification," *IEEE Transactions on Information Forensics and Security* **19**, 2840–2853 (2024).

[20] Baloukas, C., Papadopoulos, L., Demestichas, K., Weissenfeld, A., Schlarb, S., Aramburu, M., Redó, D., García, J., Gaines, S., Marquenie, T., Eren, E., and Erdogan Peter, I., "A risk assessment and legal compliance framework for supporting personal data sharing with privacy preservation for scientific research," in [*Proceedings of the 19th International Conference on Availability, Reliability and Security*], *ARES '24*, Association for Computing Machinery, New York, NY, USA (2024).

[21] Kumar, A., Kashiyama, T., Maeda, H., Omata, H., and Sekimoto, Y., "Real-time citywide reconstruction of traffic flow from moving cameras on lightweight edge devices," *ISPRS Journal of Photogrammetry and Remote Sensing* **192**, 115–129 (2022).

[22] Serengil, S. I. and Ozpinar, A., "Lightface: A hybrid deep face recognition framework," in [*2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*], 23–27, IEEE (2020).

[23] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B., "High-resolution image synthesis with latent diffusion models," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 10684–10695 (June 2022).

[24] Contributors, M., "Openmmlab pose estimation toolbox and benchmark." https://github.com/open-mmlab/mmpose (2020).

[25] Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J., "YOLOX: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430* (2021).

[26] Maji, D., Nagori, S., Mathew, M., and Poddar, D., "Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 2637–2646 (2022).