Compliance of de-identified images with GDPR

De-identification of facial images in corporate databases for the enhancement of privacy while preserving visual similarity.

August 2019

Introduction

The GDPR[1] (General Data Protection Regulation), that came into effect in May 2018, introduced strict requirements for protection of Personally Identifying Information (PII) and the privacy rights of individuals. The EU regulations have set a new global standard for privacy rights and in doing so, changed the way organizations worldwide store and process personal data.

GDPR brings the importance of preserving the privacy of personal information to the forefront, yet the importance of facial images within this context is often overlooked. In this paper, we introduce a unique solution that helps organizations protect facial images and strengthen their compliance with the GDPR. D-ID™ is the first to fill the gap between the regulatory requirements and the facial image protection solutions currently available, while ensuring data utility through preserving the visual similarity.

Our Face is our Identity

Our face is the most fundamental and highly visible element of our identity. People recognize us when they see our face or a photo of our face. In recent years we have seen an exponential increase in the use, storage and dissemination of facial images in both private and public domains i.e in social networks, smart-city deployments, autonomous vehicle, digital media, government applications, and nearly every organization's databases.

The Rise of Face Recognition

Significant developments in facial recognition technologies and applications continue to emerge in the market. With the advancements in AI, face recognition algorithms have become more accurate than humans. More and more organizations are using people's faces as identifiers, when crossing border controls, withdrawing cash or accessing smartphones. Governments worldwide already apply face recognition for different use cases. In addition, financial services like Mastercard and retail chains such as Tesco and Walmart have incorporated facial recognition solutions.

Concerns & Risks

It's common knowledge that organizations and governments store our personal data. Most notable is the case of Facebook and Cambridge Analytica. Due to a loophole in the privacy protection, data of 50 million Facebook users was exposed to the consulting firm which helped Trump's election campaign. China collects personal data on payment credibility, social behavior and purchasing habits to rank its citizens. As part of this data, facial images are critical and when processed by facial recognition systems serious privacy concerns might be raised. – yet other stories appear in the press on an almost daily basis. More recently, Google agreed to pay $13 million in Street View privacy case and earlier in 2019 is became know that the Immigration and Customs Enforcement agency of the United States (dubbed ICE) used facial recognition to mine State driver's license databases without obtaining the drivers' consent. Moreover, state-owned biometric databases are subject to hacks and leaks as became evident when India's database was leaked and unauthorized access to personal

[1] https://gdpr-info.eu/

information of India's citizens was enabled in 2018. Since the beginning of 2019, we've learned that U.S. Customs and Border Protection database was breached and photos of travelers were obtained by unknown hackers. Only weeks later, it was reported that over 90 million records were leaked by Chinese public security department.

Facial images are particularly prone to misuse. Anyone with access to the images can gather personal information and hack into accounts. Stored facial images pose the risk of unauthorized tracking and identity theft. With the growing use of facial images, the size of facial (biometric) databases grows and consequently, so does the risk of hacking and violation of privacy rights. This troublesome reality signals the end of our basic public anonymity as we know it. This is one of the many reasons why facial images are now considered sensitive personal information.

Face Images as Sensitive Data under the GDPR

Recent regulatory initiatives, specifically the GDPR, have recognized these concerns and risks, and further highlight the importance of protecting personal data. Sensitive data is a special category of personal data that is subject to additional protections. According to the GDPR, biometric data constitutes a 'sensitive' category of personal data. The GDPR particularly states that the processing of biometric data for identification purposes is prohibited. Processing is only justified if explicit consent of the relevant person exists, specific legal obligations apply or if the processing is required for reasons of public interest. The GDPR defines biometric data as: "personal data resulting from specific technical processing relating to the physical, physiological or behavioral characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as **facial images**" Indeed, facial images are categorized as sensitive data under the GDPR and need to be protected.

De-Identification and Data Utility

De-identification is a method for privacy protection and describes the process of removing personal identifiers of certain data. The de-identification of facial images is becoming a critical need in order to protect personal information and prevent fraudulent use. Currently available de-identification techniques provide a range of tools – from relatively weak techniques that can modestly reduce privacy risks, to strong techniques that can effectively eliminate most or all privacy risk. But a common trade-off is that, the stronger the applied de-identification, the greater the loss of data utility and value.

When considering facial image de-identification, a crucial factor that comes into play in many scenarios is the stringent requirement regarding the preservation of the visual naturalness. Current anonymization methods include blurring, pixelation, face swapping, deterioration, quality reduction and K-SAME. However, these techniques dramatically affect the visual similarity of the facial image compared to the original. Ideally, the de-identified facial image remains highly similar to the original one in the eyes of the human viewer, in order to enable seamless use of the image for various purposes. To preserve the visual naturalness and ensure data utility, a large set of facial attributes needs to be maintained. However, that increases the vulnerability to unauthorized reuse of the images. The goal lies in creating win-win solutions, compared to current approaches where unnecessary trade-offs are made, enabling both privacy and data utility.

D-ID's Innovative Solution

D-ID™ is a ground-breaking technology that protects facial images from facial recognition technology. The product generates new faces and makes them a look-alike to the original face to prevent the identification by automated face recognition systems, while preserving the identification by humans. Moreover, it is designed in a way that makes it difficult to be overcome by AI.

D-ID™ uses only **generation methods** which were specifically tailored to be difficult for the human eye to distinguish between the original and the altered image. The visual naturalness is preserved, and seamless use of the image is enabled. While the modified facial image is almost identical to the original, it is resistant to facial recognition systems, effectively protecting it from misuse.

The generated faces are simply faces of persons that do not exist. The newly generated faces look similar to human eyes and fail facial recognition. During generation – a new facial image is generated from scratch and the attributes are simply copied into the generating model without any analysis. General attributes include, for example, age range, gender, ethnicity and very general look. Then the generated face is modified by an Artificial Intelligence Deep Learning (AI/DL) process to resemble the original image in the eyes of a human - all the time keeping it below the identification threshold of facial recognition engines.

Since the generated faces do not belong to any person, their classification as PII or Sensitive PII is greatly diminished.

Specific GDPR references

The process described above is in line with **GDPR Recital 51**[2]: "The processing of photographs should not systematically be considered to be processing of special categories of personal data as they are covered by the definition of biometric data only when processed through a specific technical means allowing the unique identification or authentication of a natural person. Such personal data should not be processed". D-ID assures protection and in reality, it does **not** process the original image, just the generated image is brought along to high similarity score.

GDPR Article 1[3] states: "This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data".... Which means that a person **has the right** to provide a **D-ID protected image** any time his or her facial image is required, without a need for face recognition operation. In this way the **purpose** of the person is better met under the principles of Privacy by Design.

As mentioned above, D-ID does not process the original image, but rather the generated image. Furthermore, the original image is not stored anywhere and is in fact deleted as soon as the protected image is generated. Thus D-ID fully meets GDPR Article 4[4] which states: "... (2) **'processing'** means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction." .

D-ID protects sensitive biometric data by generating a new images that does not belong to any existing person, but merely closely resembles the original to the human eye. The newly

[2] https://gdpr-info.eu/recitals/no-51/

[3] https://gdpr-info.eu/art-1-gdpr/

[4] https://gdpr-info.eu/art-4-gdpr/

generated (D-ID) image **is not biometric or sensitive data** as defined by GDPR Article 4[5] : "...(14) 'biometric data' means personal data resulting from specific technical processing relating to the physical, physiological or behavioral characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data".

D-ID protection helps organizations to better fit the Privacy by Design (PdB) principles which are in core of GDPR in Clauses 52-54, 78, 91 and Articles 25 and more.

The idea is that images displayed to humans would be our protected images for identification by a human when needed. Yet, in case of a leak, the privacy risk of computerized risk is greatly reduced thus better meeting the PdB principles.

De-identification and standardization

Privacy standardization started under ISO Sub Committee 27's (SC27) Work Group 5 (WG5) which standardizes Privacy aspects. WG5 published ISO/IEC 20889[6] which standardized de-identification terms. Before D-ID introduced its technology, there was no practical implementation of machine de-identification. Another interesting standard is ISO/IEC 27771[7] (formly 27552) which maps **Privacy Terms** into the well known **Security** standard for server security ISO/IEC 27001[8] and ISO/IEC 27002[9] and thus allowing Privacy Assessment to be conducted within a well known security framework.

It is important to recognize that GDPR is practically developed by the European National members of SC27/WC5.

D-ID's Director of Product and Privacy, Mickey Cohen, is the Head of the Israeli Delegation to ISO SC27 Cyber and Privacy and to ISO SC37[10] Biometrics and works hand in hand with the world leaders in Privacy, cyber Security and Biometrics. Thus ensuring D-ID is as close as possible to the intimate work on Privacy, Cyber Security and Face Recognition.

Endorsement of D-ID's technology

D-ID received an official letter from **CNIL** of France that it plans to publish D-ID's technology on its technology website for those wishing to enhance their compliance with Privacy by Design and GDPR requirements for protection of private data.

D-ID is in discussion with the Canadian Privacy authority for similar publication of its technology.

NIST of the USA is supporting the technology and invited a lecture at IFPC 2018[11] about de-identification (our lecture is 42).

---

5 https://gdpr-info.eu/art-4-gdpr/

6 ISO/IEC 20889 PRIVACY ENHANCING DATA DE-IDENTIFICATION TERMINOLOGY AND CLASSIFICATION OF TECHNIQUES

7 ISO/IEC 27771 SECURITY TECHNIQUES -- EXTENSION TO ISO/IEC 27001 AND ISO/IEC 27002 FOR PRIVACY INFORMATION MANAGEMENT -- REQUIREMENTS AND GUIDELINES

8 ISO/IEC 27001 INFORMATION TECHNOLOGY -- SECURITY TECHNIQUES -- INFORMATION SECURITY MANAGEMENT SYSTEMS -- REQUIREMENTS

9 ISO/IEC 27002 INFORMATION TECHNOLOGY -- SECURITY TECHNIQUES -- CODE OF PRACTICE FOR INFORMATION SECURITY CONTROLS

10 SC37 Biometrics Website

11 NIST IFPC 2018

Germany's leading Biometric and Privacy conveners of standardization committees strongly support D-ID and invited D-ID to lecture in Darmstadt Biometrics Week https://www.eab.org/events/program/179.

Advisory Board Members

Prof. Adi Shamir of the Weizmann Institute, a renowned world leader in cryptography, cyber and security and a co-inventor of RSA Encryption Algorithm is on D-ID's advisory board guiding the company in the required security planning.

Dr. Ann Cavoukian, the 17yrs former Privacy Commissioner of Canada and the inventor of Privacy by Design (PdB) concepts (click here to download) is on D-ID's advisory board guiding the company in the required privacy planning. As stated above, GDPR fully endorses the PbD concepts. D-ID employs its inventor !

Conclusion

Global organizations that store and process facial images, collected in corporate databases, shared on social media or used for identification and authentication purposes, need to implement solutions that ensure the protection of privacy, as well as meet data protection requirements to comply with the GDPR. By **replacing** the biometric data with a **newly generated facial image** D-ID converts the image from *sensitive* to *non-sensitive. At the same time,*D-ID protects facial images from facial recognition technologies and their risks while preserving the visual similarity of the image. Thus, D-ID facilitates better compliance with GDPR.

D-ID's privacy enhancing solution reduces the risk of fines and lawsuits and relaxes notification requirements in case of a data breach. D-ID helps companies to better implement e 'Privacy by Design' concepts of GDPR[12], which promote privacy and data protection by integrating appropriate privacy principles into the development process from its initial phase. D-ID's advisory board members are the world leaders in security and privacy protection. Furthermore, D-ID enables companies to apply the 'Right to be Forgotten' concept and balance between privacy and data utility. At the moment, D-ID™ **is the sole available solution**.

[12] GDPR and Privacy by Design (PbD)