IS IT POSSIBLE TO BREACH PRIVACY WHEN THERE IS NO HUMAN IN THE LOOP?

I. Introduction

The world is moving towards a digital era, where the human role is progressively being replaced by technology. Artificial intelligence (AI), being a disruptive technology, is particularly relevant when there is no human in the loop. However, privacy implications of technology are not consentaneous. The aim of this report is to demonstrate that, currently, it is still possible to breach privacy when there is no human in the loop.

To this end, section II highlights the role of technology and clarifies the main concepts related to data and privacy. Section III discusses privacy implications when there is no human in the loop and presents a case study in the current context, while section IV explores privacy implications in the future. Finally, section V concludes.

II. Context and framing

We are experiencing a technological revolution, where the application of information is used to generate knowledge and innovation [1], with a clear impact on the day-to-day life. Information and communication technologies have been increasingly used to carry out tasks and processes, especially those related to data, normally done by humans. Now automation is going beyond that. So, it is not surprising that, according to Gartner, *hyperautomation* is one of the top strategic technology trends over the next five to 10 years [2]. AI, as a disruptive technology, is of particular relevance in this regard. Considering that an algorithm is a set of instructions, AI may be seen as a set of algorithms, being able to learn and evolve based on experience, i.e. from patterns and features in the data it uses as input, in a way comparable to human intelligence. In a general manner, today's AI systems correspond to *narrow AI*, as they already have some characteristics of human intelligence [3, 4].

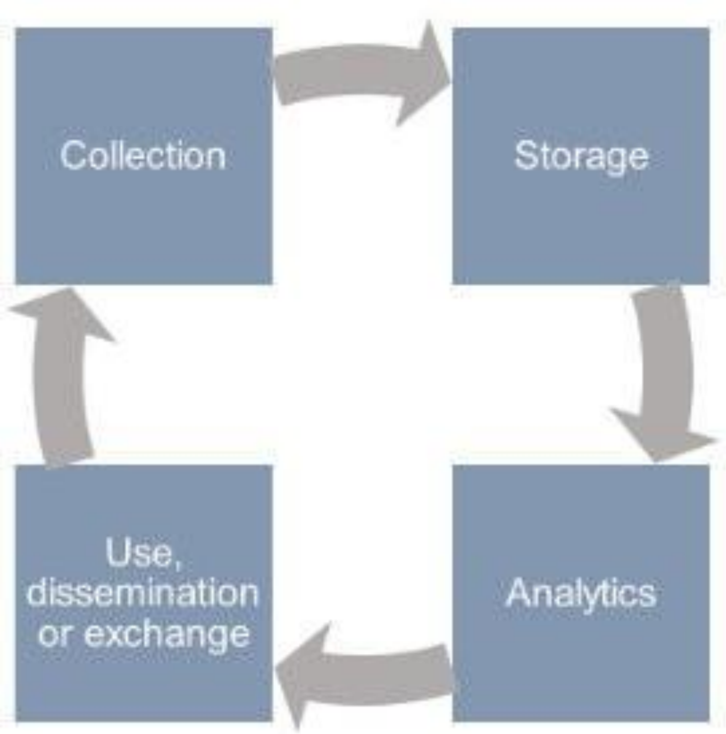
In the one hand, the advances in technology make it possible to increasingly collect, process and generate large amounts of data, much of them in real-time, from multiple sources, at great speed. In the other hand, technologies, such as AI, rely strongly on data. That is why data is a fundamental element in this technological revolution.

The public is becoming more aware of the value of their personal data and starting to raise concerns regarding digital governance, ethics and privacy in this data-driven world. Privacy, especially *data privacy*, has become a key topic nowadays. Privacy, as a fundamental human right established in the Universal Declaration of Human Rights, refers to the respect for private life. *Data privacy* regards to individuals' personal data, meaning that individuals have control over their personal data, as they decide who can access, use and disclose their personal data, although it is not always clear what personal data entails. Even though often used interchangeably, *data privacy* and *data protection* are not the same thing. *Data protection* refers to the management of personal information, protecting it from unauthorized access, use and disclosure. Therefore, both are needed to ensure privacy. It is recognized, in general, the need to define data privacy and protection rules to protect individuals and their data, but some have been arguing that strict rules can compromise innovation.

Nevertheless, data protection legislation is being adopted throughout the world as an attempt to protect personal data, for instance the General Data Protection Regulation in the European Union in 2018, the Data Protection Act 2018 in the United Kingdom, the Personal Data Protection Act in India in 2018 and the Consumer Privacy Act in California which came into effect in 2020. In a general manner, these legal frameworks are based on a "notice and consent" philosophy [5]: providing notice and requiring a consent of individuals to collect and disclose data, to guarantee that data is not used without individuals' knowledge nor for purposes that were not explicitly stated.

One cannot forget that there are different phases throughout data lifecycle. For simplification purposes, a diagram is proposed in figure 1, to highlight the main phases of data lifecycle. Normally, data, that is acquired, extracted or collected, is stored. This data is generally used for analytical purposes. Therefore, it is prepared, processed, analysed and usually used to make profiling or predictions. This process, typically, generates new derived data (e.g. inferences, profiling, predictions) which is used by the same entity, disseminated or exchanged (shared with or sold to third parties).

Figure 1: Data lifecycle



As technologies evolve, the information-gathering process become easier, faster and cheaper, data storage capabilities are higher at a lower cost, computing power is increasing and new methods of analysis are developed. Despite these advances, digital platforms are at an advantage, since they obtain individuals' data in exchange for delivering a service and they have the infrastructure to deal with data along its lifecycle [6]. This contributes to the concentration of data value, that sustains their business models [6].

Though privacy breaches and harms are frequently associated to accessing and disclosing personal information about a person, privacy can be threatened and breached along the different phases of the data value chain. For instance, different privacy breaches may be identified in each phase [7], as for example:

- **Collection**: scrutiny and surveillance, questioning or probing for information.
- **Storage**: loss or theft of data, misuse, unauthorized access to the information, disclosure of personal information, modification, destruction.
- **Analytics**: aggregation and/or linking information from different datasets that may reveal sensitive information about people, secondary use, improper access due to insecurity, exclusion.
- **Use, dissemination or exchange**: breach of confidentiality, disclosure of information, exposure of sensitive information derived from the data, increased accessibility, distortion of the information, appropriation of the information to serve other objectives, blackmail.

The human factor has been seen as a threat to privacy. With the advances of technology, the human involvement along the data value chain is progressively decreasing, while machines assume a central role. However, the privacy implications of using technology when dealing with data are not unanimous, as it will be further discussed.

III. Privacy implications

This section outlines the main arguments that sustain the opposing viewpoints nowadays: (a) incompatibility between technology and privacy and (b) compatibility between technology and privacy. A case study is also presented, in point (c), to reflect this duality.

a. Incompatibility between technology and privacy

Today’s algorithms and AI systems are not yet automating decisions, they are just automating tasks. Being programmed by humans and following human rules, algorithms and AI systems typically act based on how they are programmed [8], so they are no more objective than the people who develop them [9]. This means that the human being, intentionally or not intentionally, may interfere with privacy. In the one hand, inputs and algorithms can be manipulated in order to obtain certain decisions, serving certain interests. On the other hand, the algorithm, once it depends on humans’ decisions that are subjective, can be based on discriminatory behaviour. So, the algorithm itself will perpetuate or exacerbate the discrimination, leading to unfair and biased outcomes.

It is important to note that as technology advances, algorithms and systems are becoming more and more complex, requiring more and more data. That is why, some consider that several principles related to data privacy and protection, reflected in most data protection legislations, are somehow challenged by AI [3, 9, 10]:

- **Data minimisation**: personal data collected and used should be limited to only what is necessary and retained for the time necessary.
- **Purpose limitation**: the purpose for which personal data is collected should be clearly specified to the individual when the data is collected.
- **Fairness and non-discrimination**: personal data should be collected and used by fair means.
- **Transparency**: the way personal data is collected and used should be transparent to those from whom it is collected.

Being software-based or embedded in hardware devices, AI is a data-driven technology, collecting high volumes of data related to all forms of privacy, most of the times more than the necessary for the purpose defined, across multiple sources and devices. There is some uncertainty about future data use, which makes it difficult for organizations to set a specific purpose at the time data is collected [10]. Further, the initial purpose may change as machines learn from data or new purposes may arise from potential new uses of data. Thus, organizations have an incentive to over-collect data and keep it in the long term, not only to address the purpose defined, but also future purposes, as AI increases the scope and value of data use [3, 11].

This, not only is contrary to the idea of data minimisation and purpose limitation, but also represents data privacy and security risks. Keeping such amounts of data stored increases data vulnerability to hackers and scammers [11]. The causes of data breaches have also evolved with the advances of technology, nowadays they often result from targeted hacking and ransomware attacks [11]. As more systems and devices are networked, in particular due to the spreading of internet of things, they become more vulnerable to cyberattacks and new forms of attack may appear.

Once collected, data can be used as input for the purpose of sorting, classifying, scoring, evaluating and ranking individuals, which may result in discrimination. As AI learns from data used to train algorithms, the results produced by the algorithms may reflect data inaccuracies and biases, creating discrimination and even exclusion [5]. These risks may be amplified by the algorithm, exacerbating existing prejudices and inequalities in the society [3, 8]. Further, the inputs may not be representative of the population, generalising an outcome that does not correspond to reality. So, even when the algorithm is unbiased, managing fairness can be difficult.

There is also lack of transparency and interpretability regarding AI systems, as they sometimes do not act exactly the way they were originally programmed [10], being seen as "black boxes" [3, 8, 9, 10]. The result produced is not always easily understood.

This limitation concerning AI systems contributes to increase the lack of transparency regarding the use of individuals' data. In practice, users do not have control on their private data that is collected, processed and used by others. In a general manner, AI extends information asymmetry between individuals and organizations [11]. Therefore, individuals are unable to fully understand what data is collected and how it is used. So, the philosophy of "notice and consent" does not seem to fit in the context of AI.

Other risks and challenges, especially regarding data privacy, protection and security, may arise from AI-powered technology. As for example, the combination of different sets of information can enable to link data to individuals, meaning that non-personal information can become personal information [3]. In addition, AI can provide means of discovering information about individuals, namely by inferring or predicting sensitive information from non-sensitive forms of data [4]. Sometimes it is still possible to reverse engineer what the machine saw during the training [12].

These are some of the reasons that support the idea that AI and data privacy cannot be achieved simultaneously, by considering that AI increases the risk of data privacy breaches. Some of these concerns have also been used as justification for more policy, regulatory and legal solutions when dealing with AI. Even some big tech companies have already call for more regulation around AI, as a way to explore the potentials of AI and in the same time design trustworthy solutions. Although some consider that existing data protection frameworks already regulate how AI systems can process personal data [4], there is no specific legal framework currently governing AI.

b. Compatibility between technology and privacy

AI plays a significant role nowadays, as it has applications in all domains, allowing automation, optimization and customization. It is already transforming all industries and business areas. From the analytical point of view, AI brings many advantages, not only in terms of scale and speed, but also improving data analysis efficiency and accuracy, both descriptive, predictive and prescriptive. Therefore, less individuals and/or organizations are involved along the data value chain, decreasing the human errors and bias. In a general manner, AI capabilities accelerate business decisions, creating a competitive advantage for its organizations.

AI allows synergies between machines and humans, evidencing that technology and privacy are not incompatible. In fact, a number of algorithmic techniques have been already developed to preserve user privacy. In one hand there are *privacy enhancing technologies*, which aim to improve privacy in existing systems, in the other hand there are the *privacy preserving technologies* which are aimed at building in privacy by design in new systems [13]. Although there is a lot of references of *privacy preserving technologies* in literature [14, 15, 16, 17, 18, 19, 20], there is no standard terminology used.

Regarding *privacy preserving techniques*, the most common refers to *data anonymization*, *data distortion* or *cryptography*. In general, these techniques use some form of transformation on the original data, to "hide" sensitive information, before it is released for data analytics, although there are also techniques that modify the content of the result.

Some machine learning algorithms have been also developed in order to enhance or preserve privacy when dealing with data. The implementation of privacy by design in AI-based solutions has been largely mentioned in this regard [3, 9, 10, 13]. This approach requires to imbed data privacy into the design and architecture of AI systems, as part of the technology, anticipating privacy concerns. For example, *privacy preserving data mining algorithms*, which include *classification mining*, *association rule mining*, *clustering* and *Bayesian networks* [16], have been used to protect privacy. Moreover, *privacy preserving data publishing* are widely mentioned in the literature [14, 16, 17, 19, 21], in particular the *k-anonymity* model that takes advantage of anonymization techniques. The risk of privacy disclosure, inherent to this model, has led to the introduction of new privacy models, as *l-diversity*, *m-invariance* or *t-closeness*. Other works have provided new frameworks to enhance privacy. This is the case of *differential privacy*, that consists in adding a random noise to data before it is stored, so that data can be used for analysis purpose without revealing personal information while ensuring that the predictions are accurate [9, 11, 21]. *Federated learning* is based on a decentralised approach in regards the data used for training that is kept on the devices [21, 22]. Apart from these, *homomorphic encryption* has been gaining some relevance too, as it enables the processing of data whilst it is still encrypted [9, 21].

These are only some examples of multiple approaches that have been developed in order to address privacy when dealing with technology. Typically, one single approach does not prevent individuals privacy along the data value chain, so they must be combined. Though it is important to note that algorithms and systems are becoming more and more robust and effective.

AI has been also used to overcome some of its own limitations. Some efforts are already being made to address the "black box" nature of some AI systems. As for example, *explainable AI* intends to explain the process and the reasons behind the results, improving the interpretability of those systems [9, 13].

c. Case study

To illustrate the dichotomy between technology and privacy, one can mention, as a recent example, the smartphone apps developed in the context of COVID-19 for contact tracing. The core idea of these apps is to replace the manual contact tracking, performed by humans, by a digital contact tracking, performed by technology. These apps have been quite controversial though. Many concerns have been raised, especially in the field of privacy, namely related to the access, the storage, the use of data and data-sharing. While in some countries legal provisions regarding privacy have been removed in the interest of public health, in other countries developers have incorporated some techniques to comply with data protection legislations. Though, it is important to mention that these apps are not all the same, as they may deal with privacy differently. Different approaches have been used to preserve privacy along the process, some more effective than others.

IV. What does the future bring?

As more complex technologies are developed, the role of the human is increasingly replaced by that of the machines, amplifying machines' power. In an extreme scenario, decision-making will be up to machines, without human involvement. The technology is moving in the direction of the development of machines with complete autonomy and capacity to "think like a human", where machines will be able to make decisions, without human intervention nor supervision. The ability of AI to replicate or even surpass human intelligence is commonly known as *strong AI* or *artificial general intelligence* [3, 4, 5].

Most literature focuses on *narrow AI*. *Strong AI* is still seen as fiction for most people, though others may consider it a possibility in the long term but not in the near future. Implications of AI are already controversial at this stage of its development, arising a lot of privacy, security and ethics concerns, in particular regarding the interactions between people and AI. For those reasons, privacy implications of *strongAI* are not quite explored.

Attending the data-intensive nature of AI, one would expect that, in a scenario with *strong AI*, the collection of data would be intensified, probably in a way that all information would be identifiable. Individuals would be object of deep and mass scrutiny and surveillance by technology, by constantly being tracked, traced and monitored. In this regard, technology would be highly invasive and intrusive to individuals, being impossible to individuals to control their personal data in such scenario.

The deployment of a *strong AI* would probably create new potential challenges and risks, that may not be safeguarded in the current legal and regulatory frameworks. Privacy, security and ethics concerns will certainly continue to be raised and eventually be more prominent with *strong AI*. But, the truth is that the implications of a *strong AI* are still an open question: no one really knows in what extent a *strong AI* will affect privacy. Its consequences regarding privacy will depend on the way AI is developed and used. That is why it is relevant to guarantee that privacy and security are embedded in the process of development of AI and preserved in each phase of the data value chain, so that privacy and security threats are mitigated. If data privacy, protection and security are not somehow addressed at this phase of AI development, when human can intervene, it cannot be expected that those fundamental values and rights for individuals will be safeguarded in the future. And even if they are implemented, there is no guarantee that they are flawless and completely secure, therefore there is no guarantee, at this stage, that there will be no privacy breaches in the future.

V. Conclusions

The human element has been appointed as the main source of privacy breaches when there is a human in the loop. As technology evolves, the human role diminishes. In many situations, we're already in a scenario where there is no human in the loop. But, in today's reality, that does not mean that the human element is completely absent.

In general, algorithms, being programmed by humans, typically follow human rules. So, algorithms can be manipulated by humans. Even if there is no intentional manipulation, the algorithms will be influenced by humans' subjective decisions, which may also compromise privacy. So, in this context, it is still possible to breach privacy when there is no human in the loop, attending that there is still a way that the human being may interfere with privacy along the data lifecycle.

No one doubt of the potential offered by AI to society, business and individuals, although its implications, especially in the field of privacy and ethics, are not yet clear. AI algorithms are becoming increasingly complex and subjective, not always acting the way they were programmed and, therefore, not always understood by humans. With the advances of technology, the human risk diminishes, but other risks arise. AI is so powerful that it may potentially offer new ways to affect privacy, creating new types and levels of risk. Even though the development of AI challenges privacy, it is important to recognize that AI also provides tools to enhance privacy. Risks will always exist, but AI can be used to manage the magnitude of those risks.

Traditional methods of privacy protection will probably fail as technology evolves. New approaches will be needed to proactively identify and curb harms that may arise. So, governance, policy, regulatory and legal solutions will need to be adapted, in order to address potential challenges, as an attempt to mitigate some risks to data privacy and security. Removing completely the human element from the decision-making process will probably bring even more challenges and risks. But nobody knows the real implications of a scenario where technology dictates the rules.

AI already has a significant impact on people's lives So, nor ignoring nor impeding AI development is a viable approach. However, it is important to bear in mind that "(…) *the ability or inability of societies to master technology, and particularly technologies that are strategically decisive in each historical period, largely shapes their destiny* (…)" [1]. Therefore, there should be a balance in the adoption of data practices, algorithms and applications, not only considering the effects of technology, but also respecting fundamental individuals' rights and ethical principles, in order to maximise benefits while minimising harms. We have the responsibility to promote innovation that respects key values for individuals and societies, such as privacy, security and ethics. Otherwise potential adverse outcomes that might arise from the deployment and consequent use of AI and other technologies are not responsibility of machines, but human responsibility, even if there is no human in the loop.

References