The unfair side of Privacy Enhancing Technologies: addressing the trade-offs between PETs and fairness

Alessandra Calvi
d.pia.lab, LSTS, Vrije Universiteit Brussel; ETIS lab, UMR 8051, CY Cergy Paris Université / ENSEA / CNRS
alessandra.calvi@vub.be

Gianclaudio Malgieri
eLaw, Leiden University; Brussels Privacy Hub, Malgieri
g.malgieri@law.leidenuniv.nl

Dimitris Kotzinos
ETIS lab, UMR 8051, CY Cergy Paris Université / ENSEA / CNRS
Dimitrios.Kotzinos@cyu.fr

ABSTRACT

Data sharing in the European Union (EU) has gained new momentum, among others for machine learning (ML) and artificial intelligence (AI) training purposes. By enabling models' training whilst preserving the privacy of data, Privacy Enhancing Technologies (PETs) have therefore gained popularity, especially among policymakers. So far, computer science research has focused on advancing state-of-the-art privacy engineering and exploring trade-offs between privacy and accuracy. Meanwhile, legal scholarship began investigating the challenges arising therefrom. Yet, few works have delved into the fairness implications of PETs. Further research is essential to both prevent the propagation of bias and discrimination and to limit the accumulation of market power within very few economic entities suitable to undermine fair competition and consumer rights. In our work, we will address this knowledge gap by adopting a legal and computer science point of view. After scoping our understanding of possible unfair sides of PETs based on technical and socio-legal understandings of fairness (Section 2), we provide an overview of PETs mostly relevant for ML and AI training (Section 3). We then discuss fairness-related challenges arising from their use (Section 4) and we suggest possible technical and regulatory (e.g., impact assessment, new rights) solutions to address the shortcomings identified (Section 5). We finally provide conclusions and ideas for future research (Section 6).

CCS CONCEPTS

• **Security and privacy**; • **Privacy-preserving protocols**; • **Law**;

KEYWORDS

AI, competition, data protection, fairness, PETs, privacy

ACM Reference Format:
Alessandra Calvi, Gianclaudio Malgieri, and Dimitris Kotzinos. 2024. The unfair side of Privacy Enhancing Technologies: addressing the trade-offs between PETs and fairness. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24), June 03–06, 2024, Rio de Janeiro, Brazil.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3630106.3659024

FAccT '24, June 03–06, 2024, Rio de Janeiro, Brazil
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0450-5/24/06
https://doi.org/10.1145/3630106.3659024

1 INTRODUCTION

In the past years, data policies in the European Union (EU) have seemingly shifted their focus. Since the protection of personal data and their free flow within the European Economic Area (EEA) are somehow granted thanks to the General Data Protection Regulation (GDPR), data sharing and re-use, also towards third countries, have gained new momentum. To accelerate the development of artificial intelligence (AI) and machine learning (ML) and support the creation of Common European data spaces, the EU Commission's priorities are now fostering data sharing between public authorities; and promoting the sharing and use of public sector information by businesses, of privately-held data by other companies and of privately-held data by government authorities. Thus, in addition to the Regulation on the free flow of non-personal data and the Open Data Directive (ODD), new rules were adopted or are about to be adopted, such as the Data Governance Act (DGA), the Data Act (DA) and the AI Act (AIA).

Consequently, Privacy Enhancing Technologies (PETs), also known as privacy-preserving methods, increased their popularity, as their application is deemed an essential precondition to foster trust in data sharing and reuse [72]. To date, a commonly agreed definition of PETs is lacking. Some consider PETs as "a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system" [12]. Such definition appears rather biased towards data minimisation [49]. Others understand them as a "collection of digital technologies, approaches and tools that permit data processing and analysis while protecting the *confidentiality*, and in some cases also the *integrity* and *availability*, of the data" [72]. Indeed, information privacy is often perceived as part of information security, defined as a composite of those three elements [48]. But this definition does not capture the breadth of the related issues with PETs.

In any event, the importance and complexity of PETs for data protection purposes are such that in 2014 the European Data Protection Supervisor (EDPS) established the IPEN - Internet Privacy Engineering Network initiative to advance the state-of-the-art of privacy engineering. Then, in 2016 the European Union Agency for Cybersecurity (ENISA) began to work on a Maturity Assessment Repository for PETs. PETs are recalled in the EU data legal framework in multiple instances, as well. For instance, Recital 7 of the DGA advocates for the use of "state-of-the-art privacy-preserving methods that could contribute to a more privacy-friendly processing of data", such as anonymisation, differential privacy, generalisation, suppression and randomisation, synthetic data. PETs could also foster compliance with e.g., Art. 5 GDPR on data minimisation and security, Art. 25 GDPR on data protection by design and by default, Art. 32 GDPR on processing security. Even the economic implications generated by the deployment and adoption of PETs are not negligible: privacy and data protection are said to generate a competitive advantage for businesses, which are pushed towards offering consumers products and services with the latest privacy or data security solutions [32]. Similarly, in the United States (US), Biden's Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence insists on the research, development and use of privacy-preserving methods, whilst advancing equity and civil rights.

Yet, what does the application of PETs entail for AI and ML? Specifically, what are the possible fairness implications thereof? Still very little is known about the unfair side of PETs, as well as how to address this problem. Despite both computer scientists and legal scholars being aware of PETs' limitations, the EU data legal framework and policy initiatives portray PETs as effective solutions to address the privacy challenges brought about by data sharing. Even the narrative advanced by big tech companies, among the few entities capable of affording the high costs in terms of resources, time, know-how, etc. related to PETs implementation [5], corroborates optimism towards these solutions. Yet, the reality is much more nuanced.

Our goal is to contribute to filling this knowledge gap. Building on a desk analysis of relevant EU and US legal and computer science literature, and of EU legislation and policy documents, we will provide a fairness-inspired critique of PETs by merging legal and computer science profiles. Whereas some of the legal and technical challenges of PETs have already been explored separately, few contributions adopt an interdisciplinary angle [49, 69, 79]. Concerns expressed by legal scholars on PETs are not expressly linked to the AI fairness debate [87, 89]. With some exceptions, existing works largely focus on the US context [1, 22, 25, 31, 71, 92]. Regardless of the overlaps between competition, data protection and consumer law [41], as also witnessed in the Digital Markets Act (DMA), market imbalances generated by big tech companies are underexplored by EU data protection legal scholars. Despite becoming relevant in the context of PETs [4], the profiles of fair competition among companies (linked e.g., to the prominent role played by big tech companies in advancing privacy engineering) and consumer rights (e.g., preserving their freedom of choices) are often neglected in AI fairness scholarship or rather addressed in different terms (i.e., to what extent algorithms can directly harm consumers or distort competition [28]).

With this study, we do not claim to be exhaustive about the possible unfair sides of PETs or solutions to address them. Rather, we aim to help legal and computer science experts and/or enthusiasts familiarize themselves with the main fairness-related challenges existing in both domains and offer some sources of reflection on how to address them. Meanwhile, we deem it essential that scholars, practitioners, activists, regulators and policymakers with different backgrounds are aware that PETs may have undesirable effects (e.g., generate bias and discrimination against individuals and groups, distortions of competition). This way, it would be easier to map and tackle these challenges. In our work, we also clarify some contentious terminological points (e.g., privacy, data protection, sensitive data) that contribute to generating confusion across disciplines and jurisdictions. Despite our focus on the EU context, we believe that our contribution remains relevant for other territories. After all, the interest in PETs is growing globally [5] and some of the challenges and solutions we identify appear sufficiently generalizable.

The structure of our paper is the following. First, we briefly scope our understanding of what the unfair side of PETs could be, based on a socio-technical understanding of fairness (Section 2). We then provide an overview of PETs mostly relevant for ML and AI training purposes (Section 3) and we challenge them by adopting a fairness perspective (Section 4). We then suggest possible technical and regulatory solutions to address the shortcomings identified, like data protection impact assessments (DPIAs), transparency and new rights for people (Section 5). We finally provide conclusions and ideas for future research (Section 6).

2 SCOPING THE FAIRNESS CHALLENGES BROUGHT ABOUT BY PET

Lacking a common definition of fairness, we briefly recall some of its main understandings across disciplines to clarify the meaning of the *unfair side of PETs* in this contribution. In computer science, fairness is a mathematical property for algorithms [42] connected to the problem of bias. (Technical) bias occurs when computer systems "systematically and unfairly discriminate against certain individuals or groups of individuals in favour of others", by for instance denying opportunities or assigning undesirable outcomes [36]. It may manifest at the pre-, in- or post-processing stages [16, 26, 70, 86]. Thus, technical bias entails a performance failure, in so far as an automated system does not have the same level of accuracy across different individuals and groups [90]. Yet, as the impacts of technical bias in the real world can be significant depending on the sector where an automated system is deployed (e.g., job or study opportunities can be denied or people can be wrongly flagged as criminals) computer scientists have been investigating strategies (e.g., fairness metrics, organisational measures) to address this issue. And whilst focusing exclusively on technical bias without addressing the inequalities deeply rooted in society where automated systems are deployed is insufficient, arguably fairness metrics could contribute making automated systems fairer even from a socio-legal point of view [19].

By contrast, for EU data protection law, fairness is a core principle of personal data processing (Art. 5(1)(a) GDPR) informing the relationship between data controllers (i.e., the entities deciding purposes and means of processing) and data subjects (i.e., the people to whom the information refers). European regulators relate fairness to a procedural principle (guaranteeing the respect of the expectations and concrete interests of the data subjects beyond the mere compliance to the other data protection principles and rules) and to a substantive principle to safeguarding the autonomy of data subjects; not discriminating against them nor exploiting their needs and vulnerabilities; towards a better power balance between controllers and data subjects. According to some scholars, having fair algorithms means that controllers need to regularly assess whether algorithms are functioning in line with the purposes and adjust them to mitigate uncovered biases [33].

Yet, the concept of fairness exists in other areas of law and it is linked to the ideas of equity and non-discrimination [7, 8, 11, 13, 91]. To overcome the limitations of the antidiscrimination law discourse (e.g., treating discrimination as a single axis –and not intersectional– matter; overlooking how the design of socio-economic institutions could perpetuate discrimination [50, 93]), fairness has then been associated with the transformative concept of social justice [27, 56], namely "the fair treatment and equitable status of all individuals and social groups within a state or society" [19, 29]. However, fairness is also an overarching principle in EU consumer and competition law [41]. As such, it is connected, again, to the autonomy of individuals and their free decision-making, as the goal of EU consumer law is empowering individuals to make well-informed autonomous choices [41]. In turn, EU competition law aims to ensure that markets are kept competitive so that consumers have freedom of choice [41]. Accordingly, EU competition law strives to grant equal opportunities to businesses (through, for instance, the removal of barriers to entry) to let them compete in the market based on their merits (and not just on price, quantity, quality, choice and innovation) [41, 64]. Unfair and anti-competitive practices being contextual, exactly defining fairness in competition law is impossible. Yet, this principle can be considered as an inherent objective of competition law and enforcement [41, 64]. Besides the EU Treaties, fair competition is recalled in the EU data legal framework: e.g., Art. 12(f) DGA refers to the need to safeguard it and mandates data intermediaries to offer access to their services under "fair, transparent and non-discriminatory conditions".

In light of the above, our investigation of the unfair side of PETs will merge technical and socio-legal profiles, which are different but interconnected. Specifically, we will investigate the problems of bias, discrimination, social injustices and market power imbalances that may arise from the application of PETs mostly relevant to AI and ML. AI and ML requirements for massive data sets to learn from, many times come in conflict with the targets of PETs to provide specific (privatized) views on data and with the capacity to learn, aka predict information that PETs might try to keep hidden. We will evaluate for instance if PETs are equally effective on all individuals and groups, regardless of their (protected) characteristics and if the application of PETs could undermine bias discovery investigations. PETs indeed may fail to provide fair privacy protection, being possible that the likelihood and/or the cost of a privacy failure affect users differently, depending on protected characteristics [31]. We will discuss whether the application of PETs could otherwise undermine individual autonomy; and whether the deployment of PETs could affect competition and consumers' autonomy. Whereas privacy may indeed generate a competitive advantage, claims to protect privacy and personal data may hide anticompetitive practices [4]. Although, admittedly, such a list does not exhaustively cover what the unfair sides of PETs could be, we deem it a sufficiently comprehensive starting point to trigger further interdisciplinary discussions and research on this subject.

3 OVERVIEW OF PET

3.1 An introduction to AI, ML and privacy engineering

Generally speaking, AI is a field of computer science whereby automated systems are designed using algorithmic techniques to perform tasks (e.g., recognise patterns, cognitive learning, problem solving) in such a way to mimic the human behaviour or eventually the human brain [37]. ML is a subset of AI, whereby, using algorithmic building blocks, computers are given the ability to learn without being explicitly programmed. In other words, a programmer, instead of coding every decision-making scenario, leaves the system to identify rules by itself [37] and decide appropriate actions. Yet, such distinction is not necessarily acknowledged in the EU legal framework[1].

Before deploying a model into the real world, a number of steps, constituting the typical AI/ML lifecycle, are required. They are: (1) problem definition, i.e., what an algorithm will predict or estimate and how to measure it; (2) data collection, essential in so far as an algorithm is only as good as the data used to train it (so-called principle of "garbage in-garbage out"); (3) data cleaning, to address missing and inaccurate values that could affect the quality of prediction; (4) summary statistics review, to remove outliers and address overfitting (occurring when an algorithm considers random correlation as legitimate) and lack of generalisability (occurring when certain variables take on, non-randomly, very high or low values); (5) data partitioning, consisting of splitting a dataset into a training part and into a test part to evaluate how an algorithm trained on one dataset will perform on another; (6) model selection and (7) model training, namely the process of running an algorithm on the dataset and doing feature selection, pattern extraction, tuning and assessment [65].

Meanwhile, privacy engineering refers to a field of study aimed at implementing the principle of privacy-by-design in IT systems across their life-cycle, like any other functional requirement [49]. Even before the advent of AI/ML, scholars and practitioners have been investigating ways to safeguard content privacy (to prevent the reidentification of a person) and interaction privacy (to protect users interaction from eavesdroppers) in IT systems [98]. Privacy design strategies have been distinguished into data-oriented, aimed to *minimise* the data collected; *hide* data and their interrelationships from plain sight; *separate* data processing into different compartments, whenever possible; *aggregate* data; and process-oriented, connected to the need to comply with legal requirements, aimed to *inform* individuals about data processing and ensure that they have *control* over the processing of their information; *enforce* privacy policies and *demonstrate* compliance with legal requirements [49]. PETs gained prominence in the context of software development and engineering as concrete techniques aimed at implementing certain privacy design patterns [49]. Nowadays, PETs can be operationalised also individually and not only as part of a software

[1] For instance, the 19th April 2024 version (Corrigendum) of Art. 3(1) AIA blurs the line between AI and ML by defining an AI system as a "machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical
or virtual environments"

development process, through privacy enhancing algorithms and systems.

Several actors may be involved in the AI lifecycle and the development of privacy preserving automated systems, such as public and private entities willing to use AI/ML for different purposes but without technical know-how; tech providers, either private or research institutes, who have the technical know-how for AI, ML and PETs development; persons affected by the automated decisions; persons to whom the data collected and then used for AI/ML belong; data curators, namely the entities that take care of collecting, store and release (anonymised) datasets, that can then be used for different purposes; users of datasets, that include also attackers [98].

Some of these actors are defined under the EU legal framework. For instance, *providers* under Art. 3(3) AIA are the entities that develop an AI system or a general-purpose AI model; or that have an AI system, or a general-purpose AI model developed and placed it on the market; or put the AI system into service (i.e., supply an AI system for first use) under own name or trademark. *Deployers* are the entities using an AI system under their authority (except for personal activities) (Art. 3(4) AIA). Despite not formally defining them, the AIA mentions the *affected persons* and introduces for them, in some situations, a right to explanation of individual decision-making taken by the deployer on the basis of the output from a high-risk AI system (Art. 86 AIR). Under Art. 4(1) GDPR, personal data is information relating to an identified or identifiable person (i.e., the *data subject*). *Controllers* are the entities determining the purposes of and means of personal data processing (Art. 4(7) GDPR, whilst *processors* process data on behalf of the controllers (Art. 4(8) GDPR). It is possible that depending on the situation and processing the same entity is differently classified and that overlaps between the AIA and GDPR occur.

3.2 Types of PETs mostly relevant for AI and ML

In relation to AI/ML, privacy engineering becomes relevant to extract knowledge from data whilst preserving privacy (so-called privacy-preserving data mining) [79]. PETs may serve many goals, ranging from secure data sharing between parties to training AI models and generating anonymous statistics [51]. Privacy as a technical concept has been operationalised through e.g., anonymisation and de-identification; semantic security, typical for encryption schemes; formal privacy models [69]. PETs can be combined among themselves and used together with organisational and legal tools [72]. Like in the case of fairness, privacy is technically operationalised through metrics. Thus, contrary to normative understanding of privacy (see 4.1), privacy in computer science is defined with mathematical precision, which does not however entail automatically compliance with legal requirements [69]. Whilst multiple classifications remain possible, we consider that such privacy metrics belong to two main categories: anonymity set size metrics and entropy-based metrics. The former builds on the assumption that the larger the set of indistinguishable entities, the lower the probability of identifying any one of them. The latter assumes that the system entropy decreases with attribute disclosures [24]. In this contribution, we divide PETs into three broad categories, namely *data obfuscation*; *encrypted data processing tools* and *federated and distributed analytics*, which we deem the most relevant for ML and AI training [72]. However, multiple classifications remain possible [24, 34, 48].

Data Obfuscation These PETs (encompassing for instance anonymization, pseudonymization, synthetic data, differential privacy) build upon data alteration [3]. They *obfuscate* information by removing or replacing identifiers or adding noise [72]. They are relevant for AI and ML because they aim to remove personal information from datasets that are then used to train models.

Anonymisation – Process of removing, manually or automatically, direct (e.g., first name, last name, address, social security number) and indirect identifiers (e.g., socio-economical information) from data to prevent the re-identification of a person [3, 72].

Pseudonimisation – Process of replacing data elements, such as identifiers, with other types of information, that despite their fictitiousness preserve the usability of data [3]. The possibility to re-identify a person by relying on other information stored somewhere else is usually preserved [72].

Through tokenisation, typically random strings of numbers and letters, identifiers are replaced with a token, but since one entity maintains the key that matches the tokens, the process is reversible [3]. By contrast, masking, despite relying on random strings of numbers and letters as well, applies typically to data in use and lacks reversibility [3]. Finally, generalisation, as hinted by the name, replaces specific information with generalised ones (e.g., age ranges vs exact age) [3]. Within generalisation, a specific privacy metric is k-anonymity, defined "as a property or a requirement for databases that must not leak sensitive private information". Assuming that a database stores two kinds of attributes, namely identifying information and sensitive information, a database table is considered anonymous when every search for identifying information results in a group of at least k candidate records, with k being the privacy parameter determining the minimum group size [34].

Synthetic data – Process of generating (fully or partially) artificial data maintaining similar statistical properties to the original source [72], through for example ML. Real data are fed to ML algorithms which identify patterns and replicate them in synthetic data [3].

Differential privacy – Process that ensures privacy by adding random noise to a dataset. In other words, it promotes the protection of privacy by introducing some error to a dataset, thus affecting the accuracy thereof. This technique makes it possible to gather statistically significant insights from said dataset whilst at the same time preventing the re-identification of individuals therein. By looking at the output (namely, the new differentially private dataset), it would not be possible to tell whether any individual's data were included in the original dataset or not [31, 71]. The amount of noise added depends on the parameter ϵ [24]. Differential privacy was for instance used by Google to produce mobility reports during the pandemic [17].

Encrypted data processing

These PETs enable data processing on encrypted data, meaning that data are not altered but are never visible or disclosed during the processing [72]. Encryption is a reversible process that converts data to an unintelligible form, from plaintext to ciphertext (i.e., random strings of characters), preventing this way users from accessing plaintext unless they have a key [3]. When the same key is used to encrypt and decrypt data, this process is called symmetric or private key cryptography; when the keys are different, because each user has a pair of keys (one public, usually to encrypt, and one private, usually to decrypt), the process is called asymmetric or public key cryptography [3, 24]. Whilst the former is usually employed on data at rest, the latter is used to protect data in transit [3].

They entail shielding data and they encompass homomorphic encryption, multi-party computation, private set intersection and trusted execution environments [3, 72]. These techniques are relevant for ML and AI in so far as they enable using data held by different entities that do not necessarily trust each other without actually disclosing them in order to train models.

Homomorphic encryption – Process enabling the computation of encrypted data, whereas the encryption key is held by the person and not the organisation processing data [72]. It enables the shielding of data in use [3]. The results of the calculation on encrypted data are encrypted, too, but the decrypted result of a calculation remains the same as if the calculation had been performed on the unencrypted data [24]. There are three types of homomorphic encryption, namely full (when any function can be computed), somewhat (when a predefined set of functions can be computed) and partial (when only addition or multiplication operations can be performed, but not both) [51]. Use cases range from smart grids to healthcare and were used also in contact tracing apps during the Covid-19 pandemic [72]. It enables for instance cloud services to conceal user content from cloud providers [6].

Multi-party computation – A set of tools (systems and architecture) enabling participating parties to jointly compute a function over their input data whilst keeping them private [72]. At the end of the computation, all parties learn the final result whilst ignoring each other parties' inputs [24]. For instance, an existing dataset, instead of being centralised, is split into multiple parts so that multiple parties can interact with data without revealing the complete underlined information, meaning that even if one party is compromised the full dataset is not put at risk [3]. Secure multi-party computation was the basis of a VAT fraud detection system developed for the Estonian Tax and Customs Board [84].

Private set intersection – It enables organisations to keep their datasets private whilst highlighting the common elements of respective datasets [72]. Thus, participating database owner do not reveal the entirety of the values in their databases [24]. Practical applications include contact tracing apps provided by e.g., Apple and Google during the Covid-19 pandemic [72].

Trusted execution environments – Separate areas on computer processors that store data that remain inaccessible to the operating system. They have been implemented by major chip manufacturers and software providers (e.g., Intel, Qualcomm, Samsung) [72].

Federated and distributed analytics

These PETs allow performing analytical tasks (such as training models) upon data that are not visible to those executing these tasks, who have access exclusively to the summary statistics [72]. They encompass federated learning and distributed analytics [72].
Federated learning – Process that enables pre-processing of data at the data source, or device, level and that transfers only the summary statistics [72]. It enables the training of ML models on multiple devices, although the computation is then transferred back to a primary server, which means that such computation may reveal primary data on the device [3]. When the model aggregation is centralized, the process is called centralised federated learning, whilst when model aggregation occurs in a peer-to-peer manner this is decentralised federated learning [66]. Federated learning was for instance used to improve word recommendation of the Google Android keyboard and speech recognition of the Google Assistant [20].
Distributed analytics – Process according to which data are stored centrally whereas the training occurs across different nodes [72].

3.3 Legal and technical challenges raised PETs

Despite their commendable aim to increase the privacy-friendliness of data processing, in the past decade, PETs have been extensively criticised by computer scientists and legal scholars. From a technical perspective, PETs themselves were originally largely conceptualised and developed by security engineers [44], thus in a different domain than AI/ML. PETs are challenged in the age of big data analytics [98], for several reasons. For instance, due to the new possibilities of privacy attacks given by AI/ML: AI itself can turn into an adversary for privacy. In the context of image recognition, it was noted how certain models are capable of re-identify e.g.., blurred or pixeled images [97].Then, considering the amount of data available, re-identification is getting easier and easier [98].

Computer science literature focused then on the trade-offs between privacy and the accuracy/performance of models. Considering that PETs often rely on de-identification techniques reducing details or distorting information, this intuitively leads to a loss of performance [21]. Works [99] on differential privacy pointed out how the more the noise, the less accurate the data analytics is. On unstructured data (e.g., images and videos), its applicability is not trivial and training can become difficult or wrong [97]. And since noise is added at the individual level, large datasets are needed to preserve the ability to obtain accurate aggregate statistics [3]. Furthermore, when noise is added after an entity receives data, the process can be reversed when such entities are provided with a key or reference indicating which data were added [3]. Older computer science works pinpointed the costs of privacy against utility [15], although more recent literature confirmed the possibility of reconciling the different goals at stake [21]. Other scholars noted how training a model with differential privacy can increase *predictive multiplicity*, a phenomenon occurring when, for the same input, in a prediction task, multiple models achieve comparable levels of accuracy yet output drastically different predictions, which translates into arbitrary decisions [62]. Moreover, it has also been noted how computer science literature on PETs remains quite detached from practices of software architectures, development methods and scope of designer control [61]. The operationalisation of PETs requires a level of expertise and economic resources not available everywhere. Then, PETs assume knowledge of a desired level of privacy. Yet, this is very difficult to establish ex ante. Furthermore, PETs may be at different stages of maturity [3].

In the EU, the question of anonymisation is particularly thorn for legal scholars considering that, when data are allegedly anonymous, meaning that the data subject is not or no longer identifiable, the GDPR and all the safeguards contained therein are not applicable (Recital 26 GDPR). By contrast, the Regulation remains applicable when data undergo pseudonymisation, defined as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures [. . .]" (Art. 4(5) GDPR). And yet, the operationalisation of these two concepts remains difficult, considering also that perfect anonymisation is technically impossible to achieve [74].

Confusion exists about the legal and technical understandings of anonymisation and pseudonymisation, and the uncertain relationships thereof with the risk-based approach permeating the GDPR [74, 75, 82]. Anonymisation and pseudonymisation in a GDPR sense can be encompassed within –but do not exhaustively illustrate– the broader technical concept of de-identification techniques, and this confusion hampers the dialogue between legal and technical experts [75]. Some scholars recommend that the law ought to focus rather on processes aimed at minimising the risks of reidentification and sensitive attribute disclosure [74]. Others suggest understanding anonymisation in a context and proposed a dynamic approach thereto, arguing that the characterisation of data as personal or not does not depend on their intrinsic properties but on the environment. Indeed, as technology and computing power evolve, what used to be anonymous information could become identifiable again [82].

Another area where legal and technical challenges merge is machine unlearning, aimed at removing the influence of undesirable data and associated model capabilities from the pre-trained models [14, 67]. Unlearning is a problematic task, among others, due to difficult estimation about how each data point impact a model and the fact that training is incremental [14], but when data underwent through PETs it may be even more difficult to engage with this exercise. Considering that machine unlearning may be relevant for the operationalisation of the right to be forgotten [14], this demonstrates once again how PETs may be operationalised in such a way to clash with data subjects rights [89]. It was then noted how the very same knowledge of the inclusion of a person a database could be considered a privacy violation (e.g., where a persons participated in a study about the efficacy of interventions for substance abuse, a potential employer may infer that this person has a history of substance abuse) [69].

As anticipated, literature relating the challenges brought about by PETs to the AI fairness debate is growing, especially in the computer science domain [1, 22, 46, 71, 92, 96]. Someone may argue that PETs were created in the context of security engineer, meaning that they are not concerned with AI fairness in a technical sense. Yet, PETs do not operate in a *vacuum*. The fact that they aim to address privacy losses does not entail that they are free from other undesirable effects. And being aware of them is essential to protect people and prevent harm. The point is not just ensuring the protection of information, machines or networks but mainly the humans that use them [78].

To ground our fairness-inspired critique of PETs and better understand them in the EU context, we will first differentiate *privacy*, *confidentiality* and *data protection*, and question the narrow understanding of data minimisation that has been used to operationalised PETs [72]. Then, we will delve into the fairness-related challenges brought about by PETs.

4 A FAIRNESS-INSPIRED CRITIQUE OF PET

4.1 Questioning the narrow operationalisation of PETs

Although they are sometimes used interchangeably, the concepts of *privacy* and *confidentiality* ought to be kept separate [95]. Whilst, again, finding an exact definition of these elusive concepts proves to be difficult, confidentiality was said to be about privileged communications and information. It refers to the trust that information that is about to be used to take decisions has not be seen by unauthorised people [95] Meanwhile, privacy refers to the reasonable expectations that information stays within a place and space that individuals can control [95].

However, this conceptualisation does not coincide with privacy in a legal sense. Privacy is indeed a multidimensional concept, whose importance can be easily grasped but whose exact scope is debated [73]. Even if overlaps exist, the US and EU have different understandings thereof, depending on their sociopolitical traditions [60, 80]. To simplify, the former conceives privacy predominantly as liberty against (state) intrusions, whilst the latter understands privacy in terms of protection of personal dignity (including e.g., rights to informational self-determination, one's image and reputation) [94]. Recent elaborations comparing Western legal traditions about privacy classify privacy into bodily, intellectual, spatial, decisional, communicational, associational, proprietary and behavioral [60]. Informational privacy, "typified by the interest in preventing information about one-self to be collected and in controlling information about one-self that others have legitimate access to" overlaps, but does not coincide, with the other types [60].

Meanwhile, despite *privacy* and *personal data protection* being sometimes used interchangeably, this generates confusion [39]. The EU legal system distinguishes the rights to privacy (Art. 7 Charter of Fundamental Rights of the European Union, CFR) and personal data protection (Art. 8 CFR). Remarkably, the GDPR is about personal data protection and does not deal directly with privacy [87]. Whilst the relationship between the two rights is ambiguous [40], EU scholarship tried to elaborate on their differences. Some propose considering privacy as an *opacity tool*, aimed at shielding individuals from interferences into their private sphere, thus against illegitimate and excessive uses of power. In turn, personal data protection is a *transparency tool*, with a channelling function aimed at empowering people by giving them the possibility to control what others can do with their personal information [45]. Yet, since the risks arising from personal data processing do not depend exclusively on data disclosure, equating data protection with informational self-determination would be insufficient [87].

Whereas such discussion may sound abstract to non-experts, awareness of the different connotations of privacy and the interrelations between privacy and personal data protection has practical relevance. On the one hand, it facilitates interdisciplinary dialogue across different jurisdictions. These differences explain why, in the US, there is a tendency to consider *sensitive* all "data whose unauthorised access would make the subject of the data feel uncomfortable or could imply negative consequences for the subject", thus requiring protection against unauthorised disclosure [24]. By contrast, under EU data protection law, the safeguards remain applicable regardless of the publicity of data [87] and the scope of sensitive information is much more limited[2]. Furthermore, the GDPR protects both direct and indirect identifiers. The former refers to unique attributes that clearly identify individuals (e.g., names, ID numbers), whilst the latter to attributes that could potentially enable the reidentification of an individual when grouped together (e.g., age, career, location) [98].

On the other hand, it enables us to better understand PETs in the EU context. More recent elaborations on PETs stress how their operationalisation in terms of data minimisation and information security is misleading. PETs may (and ought to) be operationalised in such a way to comply with the purpose limitation principle, to prevent scope creep [85]. Furthermore, the very same data minimisation is not just about data collection and retention, but it may be operationalised in different ways, such as: minimising the risks (likelihood and severity) of privacy breaches; minimising the need for trust in other parties to fulfil the functionality of a system; minimising disclosure (to third parties) and replication (occurring when data is processed in multiple entities); minimising centralisation, to avoid single points of failure, as well as linkability, to limit inferences [44].

Giving prominence to the privacy-enhancing, or shielding function, of PETs may lead to overlook that in the EU they are also data protection instruments, responding to an empowering logic for data subjects. Indeed, it was noted how privacy design can be also process-oriented [49]. Thus, they need to be operationalised consistently with the regulatory framework in which they are embedded. The application of PETs does not indeed automatically ensure compliance with data protection laws; by contrast, clashes with other legal provisions (e.g., data subjects' right to access) may occur [89]. The same is true concerning competition law and consumer protection, whereby businesses' claims to protect privacy and personal data may hide anticompetitive practices [4].

4.2 Fairness-related challenges of PETs

Hereafter, we will illustrate how PETs themselves can both contribute generating or amplifying biases and discrimination when realised as privacy enhancing algorithms. Indeed, as other types of algorithms, they are subject to such risk. But also how they can be used as a veneer for algorithms that would otherwise cause the same unfairness.

PETs and bias discovery

The very same idea of anonymisation through the removal of identifiers may clash with bias discovery investigations. At the early stages of AI fairness research, computer scientists believed that they could ensure AI fairness through unawareness of protected characteristics, that were stripped away from data [54]. Similarly, anonymisation builds upon the premise that direct and indirect identifiers, including sensitive information, need to be stripped away to prevent re-identification. Yet, it was demonstrated that fairness through unawareness builds upon erroneous premises by overlooking how protected characteristics may affect access to resources and contribute to further stigmatising marginalised groups [54]. This entails that when a dataset is anonymised to protect privacy concerns and then shared and/or used to train AI or ML systems, it could embed bias. Whereas, admittedly, more recent PETs are aware of the limitations of old approaches and thus aim to achieve privacy in different ways, and bias can be tackled at the post-processing stage, anonymisation could undermine earlier detections of bias. By contrast, through pseudonymisation, sensitive information remains available,

Anonymisation could be problematic also from a legal compliance point of view. The AIA is expected to require high-risk AI systems providers, namely the entities that develop or have an AI system developed and place them on the market or put it into service, to comply with a series of obligations requiring bias investigations [19]. For example, Art. 10(2)(f) and (g) AIA require providers to examine datasets "in view of possible biases [. . .]" and identify appropriate measures to address them. Art. 10(3) AIA states that datasets shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used. Then, Art. 10(5) AIA expressly allows (though does not oblige) providers to process special categories of data for the purposes of ensuring bias monitoring, detection and correction in relation to high-risk AI systems. If datasets are anonymised by different entities than the providers, the necessary information to comply with these provisions may not be available. Admittedly, recent work in computer science demonstrated how anonymisation does not necessarily increase statistical bias: differences between original and anonymised datasets may be small, bias-wise [59]. Yet, for a correct interpretation of results, it is essential to be transparent about the anonymisation process and data pre-processing steps before sharing information across the AI lifecycle [59]. To our knowledge, such transparency requirements are currently insufficient under EU law.

Theoretically, in case of encrypted data processing tools, the possibility of carrying out fairness investigation is technically preserved. They do not alter data but they ensure that information is never visible or disclosed. Yet, some caveats as to secure multi-party computation remain: it was demonstrated how, when applied to federated model training process, centralised multi-party computation could contribute to introducing bias [66].

At a technical level, federated and distributed analytics raise bias concerns, too. Current state-of-the-art techniques on federated learning were not designed to face challenges that may arise, for instance, from having parties with heterogeneity in distribution and amounts of data that, due to privacy concerns, cannot be freely shared [1]. Then, the participation may change throughout the training process [1]. Studies on group fairness have shown how this PET, compared to centralised training, propagates bias. Indeed, bias from a few parties can influence all parties in the network, hence aggravating the fairness problem globally. Specifically, biased parties negatively influence other parties via aggregation throughout the training [22]. Some authors explained how biased parties encode their bias into the local updates by increasing the signal of a few parameters steadily throughout the training process.

[2] They are generally associated exclusively with those suitable to reveal racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data (for the purpose of uniquely identifying a natural person) data concerning health or data concerning a natural person's sex life or sexual orientation, and data concerning criminal convictions and offences (Artt. 9 and 10 GDPR).

This is then propagated to the global model via aggregation and, ultimately, to other parties [22].

PETs and people belonging to protected groups

Studies proved how, when multiple data sources are available, the effectiveness of anonymisation is questionable, de-anonymisation attacks being easier to carry out against people belonging to protected groups [3, 31]. This is also true in the context of ML, prone to membership inference attacks, whereby an attacker can understand if a given data record (which may be personal information) was part of the training data [63].

By studying both synthetic and realistic settings, some scholars addressed the issue of disparate vulnerability to membership inference attacks across different subgroups, occurring when the model behaves differently across subgroups [63]. They demonstrated, among others: how vulnerability to membership inferences arises when the distribution of a model property is different for samples in and out of the training dataset, meaning that the lack of distributional generalisation is a necessary and sufficient condition for these attacks to succeed; how, in case of small subgroups, the estimation of the magnitude of disparate vulnerability is not a trivial task; the importance of studying the consequences of privacy attacks for subgroups, and not individuals, when evaluating the privacy risks of deployment; and how satisfying algorithmic fairness constraints can decrease disparate vulnerability to limited classes of attackers [63].

As to pseudonimysation techniques, despite their maturity, they remain prone to similar limitations as anonymisation. Drawing from metadata, re-identification remains possible, whilst masking and k-anonymity work best on large datasets [3]. These techniques bring about specific challenges as well. For instance, the choices made for generalisation (e.g., the scope of the age range, disaggregating data based on a binary conceptualisation of gender) may hide some concerns.

By contrast, the fairness implications of synthetic data are twofold. They have the advantage of enabling training models or testing new systems without relying on protected characteristics [3]. Rather, even if real data, including sensitive information, still need to be collected and fed to the model, they can be then deleted after the model is trained [9, 18]. Then, synthetic data may mitigate the problem of predatory inclusion of marginalised individuals and groups in AI systems, whereby publicly available information of traditionally underrepresented categories of people in datasets (e.g., racialised individuals, women, non-binary folks) are scraped from the internet and fed to AI systems to debias them [93]. Art. 10(5) AIA expressly states how the use of anonymised or synthetic data ought to be preferred to the processing of special categories of personal data. However, AI tools may remain oppressive (e.g., aimed at increasing surveillance against minorities) whilst debiased [93] and misuse of synthetic information, not being linked to an individual, may escape court scrutiny. Furthermore, whilst promising, synthetic data generation processes nevertheless currently lack maturity, meaning that reversing synthetic data to original data remains possible and bias existing in original data could be replicated [3].

Studies on differential privacy demonstrated that, when strict privacy settings apply, and even when privacy mechanisms add equivalent noise to independent populations, significant disparities in the outcomes of the algorithm occur, and thus some populations are more affected than others [71]. Furthermore, whilst training a model with differential privacy can address privacy concerns, it has the potential to skew the influence towards the majority subgroups, exacerbating already existing inequities in the data collection processes, whilst affecting the accuracy of individual decisions [23, 83]. Depending on the policy sectors (e.g., allocations of benefits to schools for children in need or linguistic assistance in the context of political elections, health sector, autonomous vehicles) the consequences can be extremely severe, as they could affect the enjoyment of fundamental rights [23, 71, 83].

For prediction models, it was noted how training with differential privacy could lead to arbitrary (opposite) decisions for the same query, due to the predictive multiplicity phenomenon. Whilst this *prima facie* appears (and could be) a disadvantage, it was noted how such multiplicity could nevertheless enable to satisfy multiple properties beyond accuracy, such as (individual) fairness [62].

PETs, individual autonomy and market imbalances

The fairness concerns raised by encrypted data processing tools are largely socio-legal. Cryptography currently suffers from a paradoxical situation. In principle, the shielding function of cryptography in the digital world is so important that some authors consider the *right to cryptography* and *freedom of choice of encryption methods* a corollary of the right to privacy, essential to protect people against surveillance [35]. Ideally, cryptography eliminates the need to trust service providers, by empowering users to protect their privacy themselves against untrusted third parties. Yet, cryptographic tools are nowadays largely implemented by the very same service providers these technologies purportedly protect users from. Usually, end-to-end encryption by messaging services like WhatsApp is not implemented by users directly by relying on an independent client, but by the provider itself [6]. In other words, current implementations of cryptography aim rather to protect users against external threats but not the very service providers' threats [6]. This discourse may be transposed at the ML level, whereas tech developers who perform ML are the same as those providing encryption techniques.

This operationalisation of cryptography is problematic at a twofold level. First, it clashes with the autonomy of data subjects, which fair data processing and consumer law ought to pursue. Whilst admittedly data subjects may lack awareness and competencies to implement cryptography themselves, it is also true how technical restrictions (e.g., impossibility to add encrypting plug-ins or extensions, need to have the cooperation of the service provider) would prevent them from doing so [6]. Secondly, this paradigm affects also market power dynamics, as it increases both the power that providers have over consumers and possible competitors, by forcing users to adhere to their cryptographic choices to benefit from a service and deny interoperability.

Then, whilst encryption may increase data security, it does not automatically grant informational self-determination. Indeed, encrypted data may be passed securely to a database, and then accessed by unauthorised persons [47]. This becomes relevant in the context of ML whereby a data subject may not be aware of the use of their information.

As regards federated and distributed analytics, rather than guaranteeing the inaccessibility of data, they are focused on the structure of the processing [87]. By enabling the training across different devices, they challenge the paradigm of data accumulation within a sole entity [87]. Yet, these PETs may be problematic in terms of socio-legal fairness, too. Data subjects are often unaware of their participation in this type of analytics, and even when they gain awareness they lack the capacity to technically object to that [88]. This severely jeopardises their autonomy. Then, despite federated and distributed analytics having this commendable potential to reduce data accumulation within a sole entity, they also respond to a logic of liability avoidance, considering the costs in terms of legal compliance efforts that data processing has nowadays [87]. However, federated and distributed analytics have also an impact on competition. Building on what has already been noted in the context of microtargeting [87], big tech companies, like Google or Apple, who have already control over devices would push for the implementation of these solutions, from which they could benefit the most whilst losing the least. In addition, other big tech companies, who might have access to big (training) data and huge data analytics capabilities (e.g. Meta), might easily implement PETs solutions exploiting their data-power competitive advantage and keeping high profits from the exploitation of (pseudonymized or even anonymized) data. The paradoxical result would be that the data controllers who would pose the highest risks for data subjects due to their huge power imbalance and to the highly detrimental effects that they can have on the autonomy of individuals (e.g., their possibility to manipulate limit individuals' behaviours) would be also the data controllers who would easily anonymize their data through PETs solutions and escape the GDPR rules and other EU legal obligations. In other terms, PETs might be the trojan horse of the whole system of laws limiting power concentrations and unfairness in the digital markets. The more powerful players will also be the ones with higher chances to use PETs (to the contrary of smaller competitors, which lack the technical and data infrastructures to train PETs solutions) and escape the law.

In sum, all the PETs previously described raise fairness-related concerns, either from a socio-legal or technical perspective or a combination of both. Consequently, to properly address these concerns, it is essential to come up with different kinds of solutions.

5 TECHNICAL AND REGULATORY SOLUTIONS

The analysis above demonstrates the necessity of a shift in the narrative about PETs. So far, businesses having the means to invest in them have leveraged the competitive advantage that these technologies offer. Similarly, policymakers and regulators have been increasingly advocating for them. EU legal and policy documents on data law portrayed them as mitigation measures for risks arising from processing. Yet, it needs to be clear that PETs are also sources of risk themselves. They can cause harm to people and markets. They do not automatically ensure compliance with the EU data legal framework, nor they are fair by default. This does not entail interrupting research about privacy engineering or denying its importance for AI and ML purposes. Rather, to increase the efforts to ensure the fairness of these technologies as well as their coherence with the EU legal and regulatory framework. Again, synergies between technical and socio-legal sectors are essential.

From a technical point of view, solutions have been proposed to reconcile PETs and fairness, whilst maintaining a satisfactory level of accuracy that would enable ML on privacy-protected data. Some authors suggest customise privacy mechanisms, targeting performance on specific assignment problems, although the practical feasibility of this is uncertain [71]. Others argued to have found a balance between accuracy, privacy and model fairness, by applying local and global differential privacy to federated learning whilst quantifying the level of fairness based on the constraints of three definitions of fairness, including demographic parity, equal odds, and equality of opportunity [43]. It was noted how through multi-party computation it would be possible, whilst maintaining cryptographic privacy of sensitive information, to (i) certify and sign a model as fair, (ii) learn a fair model and (iii) verify that a fair-certified model has indeed been used [55]. Other technical solutions to mitigate unfairness when applying differential privacy have been proposed in specific contexts [63, 71]. Consider also that technical fairness metrics themselves may be helpful to guide PETs' evaluation. For instance, demographic parity and equalised odds were used for the evaluation of membership inferences attacks across subgroups in a dataset [63]. Yet, further research is needed, as those solutions were applied only in specific situations and considering also the misalignment between PETs academic literature and practices of software production [61].

The application of PETs can be functional to overcome individual prejudice. For instance, the Lighthouse project by Airbnb aims at uncovering and addressing disparities in how people of colour experience the platform, among others by eliminating guest profiles photo prior to booking, which was demonstrated to be slightly beneficial for guests who are perceived as Black by hosts [2]. When PETs are operationalised not just in accordance with a strict interpretation of data minimisation and information security, but in such a way to comply with e.g., the purpose limitation principle, to prevent scope creep [85], they arguably hint towards a greater compatibility with socio-legal understandings of fairness.

Some legislative updates in the EU may contribute to improving PETs' fairness from a technical point of view. Among the requirements for providers of high-risk AI systems that the AIA will probably include, is the need to establish a risk management system running throughout the entire AI lifecycle (Art. 9 AIA) and comply with data and data governance measures aimed, among others, at debiasing datasets (Art. 10 AIA). We mentioned how this provision may be difficult to comply with when AI systems are trained on anonymised datasets. However, in so far as PETs can be classified as (part of) high-risks AI systems (e.g., privacy-enhancing algorithms), providers will have to abide by these rules. Art. 13 AIA contains also transparency requirements aimed at enabling deployers to understand and use an AI system appropriately, which could include information on the de-identification processes.

EU legislation may contribute to limiting anti-competitive practices under the pretext of protecting privacy, too. Despite not addressing directly PETs for ML purposes, the DMA contains some relevant provisions. The Regulation assumes that, due to e.g., their size or types of services provided, certain (big-tech) platforms, the so-called gatekeepers, may engage with practices suitable to negatively affect data protection and competition. It therefore introduces some rules to prevent this. Among the obligations listed in Chapter III DMA, Art. 7 DMA mandates gatekeepers to, progressively, make interoperable end-to-end messaging, calls and sharing of images, voice messages, videos and other attached files between individuals and groups of end users. Although derogations from this rule remain possible, this provision could still enable overcoming the current paradigm whereby the providers of services are the same implementing cryptography solutions. Coordination between different regulatory authorities could improve the coherence of overlapping sectors, such as competition and data protection [4].

New rights could contribute to empower users, at least to some extent. In the case of federated learning and distributed analytics, as well as secure multi-party computation, the first step would be strengthening rights to information about the existence of this type of processing whilst ensuring opt-out options [88]. This could also promote fair competition in so far as it would limit the power of those big techs who own devices [87]. Consistently with the DMA provision previously mentioned, people may be given the right to choose their cryptographic means [35]. However, it must be clear that such individual rights are just a part of the solution. It was noted how current data protection rights, especially under the GDPR, put an excessive burden on individuals, who may lack the necessary expertise to exercise them, whereby some privacy issues are systemic and ought to be addressed not at an atomistic level [81].

Yet, for rights to be effective, individuals need to be aware of them, have enough resources to exercise them and be able to rely on an enforcement mechanism in case of non-compliance [38]. Rights such as cryptography would entail a level of digital literacy that may be unavailable to laypersons. That is why increasing digital education and enabling collective representations and actions is essential to ensure that people who lack resources are granted justice.

In the meantime, a possible means to ensure that trade-offs between PETs and fairness issues emerge and are addressed in the practice is through impact assessments. Indeed, to be effective, PETs need to be combined with other organisational measures [51]. Like impact assessment enables the operationalisation and contextualisation of fairness [19], it could enable the operationalisation and contextualisation of privacy. Hereafter, we will focus on a specific form of impact assessment, which is the Data Protection Impact Assessment (DPIA), regulated by Art. 35 GDPR. DPIAs can be defined as processes to support the informed decision-making of controllers, legally requested to self-evaluate the consequences of personal data processing likely to result in a high risk for the rights and freedoms of natural persons [58]. They root, among others, in Privacy Impact Assessments (PIAs), that were operationalised in such a way as to center on data quality and data security [68]. Accordingly, even nowadays, many DPIAs are focused on data security profiles, despite scholars challenging this trend [52, 57]. Whereas in the US DPIAs are reconducted in self-regulatory environments, in the EU they belong to the regulatory framework [10]. DPIAs are, or ought to be, expressions of collaborative governance, with collaborative processes, participation of different entities, local experimentation, public/private partnerships and flexible policy formation, implementation and monitoring [76]. They could enable collaborations between e.g., legal scholars, computer scientists and even data subjects, contributing to awareness of the risks for rights and freedoms raised by data processing, thus even by the application of PETs. DPIAs have been considered potentially extremely relevant in the context of ML design [30] and their importance is recalled in Recital 7 DGA stating that the application of PETs "together with comprehensive data protection impact assessments and other safeguards, can contribute to more safety in the use and re-use of personal data and should ensure the safe re-use of commercially confidential business data for research, innovation and statistical purposes". It was already noted how DPIA could be useful to consider trade-offs between data protection by design and data subjects' rights [89] and plays a role in the operalisation of fairness [53]. Yet, a DPIA could support decision-makers in choosing the best PET depending on the context of use. Even PETs themselves, like privacy-enhancing algorithms, could be assessed through a DPIA.

Admittedly, some caveats remain. Being regulated under the GDPR, DPIAs apply exclusively to personal data, and it can be argued that trained models do not encompass them. Yet, when models are vulnerable to de-anonymisation attacks, they ought to be treated as personal data [63]. Then, only controllers are legally obliged to perform DPIAs. Determining who the controller is in the ML learning lifecycle may be a challenge. Finally, consultation of data subjects and other entities during a DPIA is a best practice but not a legal obligation (except the consultation of the data protection officer (DPO), meaning that key features of this process remain demanded to the goodwill of data controllers. Furthermore, such entities need to be put in the condition of meaningfully participating in the process, to prevent risks of participation washing [77].

Whilst this list of solutions is far from being exhaustive, we deem it sufficiently comprehensive to demonstrate how both technical and legal or otherwise regulatory instruments could be functional to tackle the problem of the unfair side of PETs.

6 CONCLUSIONS

In this paper, we shed some light on the underexplored topic of the unfair side of PETs, building on computer science and legal literature of mainly EU and US scholars. We explained how, due to the multidimensionality of the notion of fairness, even the fairness implications of PETs are multidimensional, ranging from strictly technical concerns connected to bias to broader socio-legal challenges, including market power imbalances. After clarifying how our understanding of the possible unfair side of PETs (Section 2) and providing an overview of PETs relevant to ML and AI training purposes, divided into data obfuscation, encrypted data processing and federated and distributed analytics (Section 3), we engaged with our fairness-inspired critique thereof (Section 4).

To facilitate the dialogue between legal and computer science experts and better understand PETs in the EU context, we first provided some terminological clarification as to the notions of privacy, confidentiality and data protection and challenged a narrow operalisation of privacy, based on a strict interpretation of the principle of data minimisation and information security. We then entered into details as to the fairness challenges brought PETs, We noted how anonymisation based on stripping away identifiers may undermine bias discovery and how the reversibility of this technique is affected by protected characteristics. Likewise, the effectiveness of differential privacy appears to change across groups with protected characteristics. Similar concerns as to re-identification exist also for pseudonymisation. Synthetic data could represent a promising solution as they limit the need to rely on sensitive information, but they are not exempted from bias propagation risks, nor do they ensure that otherwise debiased AI tools are not oppressive. Risks of bias propagation exist also for federated and distributed analytics and multi-party computation. As to fairness in a socio-legal sense, we noted how, through federated and distributed analytics, data subject autonomy may be undermined in so far as they do not have the power of opting out of these processes. Meanwhile, big tech which own devices may exploit their market power. As to encryption tools, the current implementation of cryptographic solutions both limits the choice of data subjects and is suitable to undermine competition.

To remedy this situation, both technical and regulatory solutions appear possible (Section 5). We mentioned computer science studies that managed to reconcile, in some cases, privacy, accuracy and model fairness. We reflected on the role of EU regulatory instruments, such as theAIA and the DMA, in improving PETs fairness from both technical and socio-legal points of view. We nevertheless called for the creation of new rights of information and opt-out concerning federated and distributed analytics, as well as a right to choose our own cryptography. Yet, we highlighted how, in parallel, people need to be put in condition to exercise their rights and called for introducing collective rights. We proposed to use DPIAs to support practitioners in both evaluating PETs and choosing the most appropriate solution for the specific context, whilst aware of the limitations of this instrument.

Whereas with this work we sketched possible unfair sides of PETs and solutions to them, this subject ought to be further investigated. The technical fairness challenges arising from PETs, as well as possible remedies to them, need to be further explored in multiple contexts of use. The situation is further complicated by the fact that PETs can be combined. As privacy engineering evolves and technology matures, new challenges may emerge. Meanwhile, even the legal panorama is constantly evolving, both in the EU and at a global level. Whether the legal and regulatory instruments previously discussed would be able to effectively change the practices of tech developers is still uncertain. Meanwhile, due to the underrepresentation of technical experts among EU legislators, risks of mismatches between regulatory requirements and technical possibilities remain. Overlaps of different legal domains, like data protection, competition, consumer and non-discrimination, are more and more frequent, so a need for comparative legal research and or cooperation among regulatory authorities in the attempt to preserve the consistency of EU law. Last but not least, more efforts are required to increase digital literacy to increase individual autonomy in choosing, e.g., cryptographic means or opting out of federated analytics. Enhanced literacy would also be functional to meaningful participation of laypersons, and not just experts, in the DPIA process and more in general in the AI lifecycle.

ACKNOWLEDGMENTS

This work by Alessandra Calvi has been funded under the EUTOPIA PhD co-tutelle programme 2021. Award number: EUTOPIA-PhD-2021-0000000127 OZRIFTM5. .

REFERENCES

[1] Annie Abay, Yi Zhou, Nathalie Baracaldo, Shashank Rajamoni, Ebube Chuba, and Heiko Ludwig. 2020. Mitigating Bias in Federated Learning. 1–16. Retrieved from http://arxiv.org/abs/2012.02447

[2] Airbnb. 2022. Fighting discrimination and building inclusion - Project Lighthouse. 2022. Retrieved from https://www.airbnb.com/against-discrimination

[3] Kaitlin Asrow and Spiro Samonas. 2021. Privacy Enhancing Technologies: Categories, Use cases, and Considerations. Retrieved from https://www.frbsf.org/economic-research/wp-content/uploads/sites/4/Privacy-Enhancing-Technologies-Categories-Use-Cases-and-Considerations.pdf

[4] Autorité de la concurrence and Commission nationale de l'informatique et des libertés (CNIL). 2023. Competition and personal data: a common ambition.

[5] Maria Badillo. 2023. Navigating Privacy-Enhancing Technologies: Key Takeaways from the Inaugural Meeting of the Global PETs Network. Future of Privacy Forum Blog, 1–5. Retrieved from https://fpf.org/blog/navigating-privacy-enhancing-technologies-key-takeaways-from-the-inaugural-meeting-of-the-global-pets-network/

[6] Ero Balsa, Helen Nissenbaum, and Sunoo Park. 2022. Cryptography, Trust and Privacy: It's Complicated. In Proceedings ofthe 2022 Symposium on Computer Science and Law (CSLAW'22), 13.

[7] Solon Barocas. 2014. Data Mining and the Discourse on Discrimination. In Data Ethics Workshop, Conference on Knowledge Discovery and Data Mining, 6. Retrieved from https://pdfs.semanticscholar.org/abbb/235fcf3b163afd74e1967f7d3784252b44fa.pdf

[8] Solon Barocas and Andrew D. Selbst. 2016. Big Data's Disparate Impact. California Law Review 104, 671 (2016), 671–732.

[9] Marvin Van Bekkum and Frederik Zuiderveen Borgesius. 2022. Using sensitive data to prevent discrimination by artificial intelligece: Does the GDPR need a new exception? Computer Law & Security Review 48, (2022), 105770. DOI:https://doi.org/10.1016/j.clsr.2022.105770

[10] Colin J. Bennett and Charles D. Raab. 2020. Revisiting the governance of privacy: Contemporary policy instruments in global perspective. Regulation and Governance 14, 3 (2020), 447–464. DOI:https://doi.org/10.1111/rego.12222

[11] Reuben Binns. 2017. Fairness in Machine Learning: Lessons from Political Philosophy. In Machine Learning Research (Conference on Fairness, Accountability, and Transparency 2018), 1–11. DOI:https://doi.org/https://doi.org/10.48550/arXiv.1712.03586

[12] G. W. Van Blarkom, John Balking, and H. Verhaar. 2003. PET. In Handbook of Privacy and Privacy-Enhancing Technologies - The case of Intelligent Software Agents. 33–54.

[13] Frederik Zuiderveen Borgesius. 2018. Discrimination, artificial intelligence and algorithmic decision making. Retrieved from https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73

[14] Lucas Bourtoule, Varun Chandrasekaran, Christopher A. Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, and Nicolas Papernot. 2021. Machine Unlearning. In 2021 IEEE Symposium on Security and Privacy (SP), IEEE, 141–159. DOI:https://doi.org/10.1109/SP40001.2021.00019

[15] Justin Brickell and Vitaly Shmatikov. 2008. The cost of privacy: Destruction of data-mining utility in anonymized data publishing. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2008), 70–78. DOI:https://doi.org/10.1145/1401890.1401904

[16] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of Machine Learning Research 81, (2018), 77–91.

[17] Alessandra Calvi. 2021. Differentially private algorithms. Privacy Laws & Business, 11–12.

[18] Alessandra Calvi. 2023. Exploring the Synergies between Non-Discrimination and Data Protection: What Role for EU Data Protection Law to Address Intersectional Discrimination? European Journal of Law and Technology 14, 2 (2023), 1–34.

[19] Alessandra Calvi and Dimitris Kotzinos. 2023. Enhancing AI fairness through impact assessment in the European Union: a legal and computer science perspective. In ACM Conference on Fairness, Accountability, and Transparency (FAccT '23), 1229–1245. DOI:https://doi.org/10.1145/3593013.3594076

[20] Alessandra Calvi and Gianclaudio Malgieri. 2021. Federated Learning: a useful privacy-enhancing technique? Privacy Laws & Business International Report, 30–32.

[21] Tânia Carvalho, Nuno Moniz, Pedro Faria, and Luís Antunes. 2023. Towards a Data Privacy-Predictive Performance Trade-off. Expert Systems with Applications 223, 1 (2023), 119785. Retrieved from http://arxiv.org/abs/2201.05226

[22] Hongyan Chang and Reza Shokri. 2023. Bias propagation in federated learning. In The Eleventh International Conference on Learning Representations, 1–26.

[23] Victoria Cheng, Vinith M Suriyakumar, Natalie Dullerud, Shalmali Joshi, and Marzyeh Ghassemi. 2021. Can You Fake It Until You Make It?: Impacts of Differentially Private Synthetic Data on Downstream Classification Fairness. In Conference on Fairness, Accountability, and Transparency (FAccT'21), 149–160. DOI:https://doi.org/10.1145/3442188.3445879

[24] Peter Christen, Thilina Ranbaduge, and Rainer Schnell. 2020. Linking Sensitive Data. DOI:https://doi.org/10.1007/978-3-030-59706-1

[25] Cynthia Dwork and Deirdre K. Mulligan. 2013. It's Not Privacy, and It's Not Fair. Stanford Law Review Online 66, 2000 (2013), 35. Retrieved from http://www.stanfordlawreview.org/online/privacy-and-big-data/its-not-privacy-and-its-not-fair%5Cnhttp://www.stanfordlawreview.org/sites/default/files/online/topics/DworkMullliganSLR.pdf

[26] Catherine D'Ignazio and Lauren F. Klein. 2020. "What Gets Counted Counts." In Data Feminism. MIT Press, Cambridge, 1–34. DOI:https://doi.org/10.7551/mitpress/11805.003.0006

[27] Catherine D'Ignazio and Lauren F. Klein. 2020. Introduction: Why Data Science Needs Feminism. In Data Feminism. MIT Press. DOI:https://doi.org/10.7551/mitpress/11805.003.0002

[28] Ambroise Descamps, Timo Klein, and Gareth Shier. 2021. Algorithms and competition: the latest theory and evidence. Competition Law Journal 20, 1 (2021), 32–39.

[29] Brian Duignan. 2023. Social Justice. Britannica. Retrieved from https://www.britannica.com/topic/social-justice

[30] Lilian Edwards and Michael Veale. 2017. Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for. Duke Law & Technology Review 16, 1 (2017), 19–84. DOI:https://doi.org/10.2139/ssrn.2972855

[31] Michael D. Ekstrand, Rezvan Joshaghani, and Hoda Mehrpouyan. 2018. Privacy for All: Ensuring Fair and Equitable Privacy Protections. In Proceedings of Machine Learning Research, 35–47. Retrieved from https://proceedings.mlr.press/v81/ekstrand18a.html

[32] European Commission. 2020. Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition - two years of application of the General Data Protection Regulation. 1–18.

[33] European Data Protection Board. 2020. Guidelines 4/2019 on Article 25 Data Protection by Design and by Default v.2.0.

[34] Simone Fischer-Hubner and Stefan Berthold. 2017. Privacy-Enhancing Technologies. In Computer and Information Security Handbook. Elsevier Inc., 759–778. DOI:https://doi.org/10.1016/B978-0-12-803843-7.00053-3

[35] Andreas Fisher-Lescano. 2016. Struggles for a global Internet constitution: protecting global communication structures against surveillance measures. Global Constitutionalism 5, 2 (2016), 145–172. DOI:https://doi.org/10.1017/s204538171600006x

[36] Batya Friedman and Helen Nissenbaum. 1996. Bias in Computer Systems. ACM Transactions on Information Systems 14, 3 (1996), 330–347. DOI:https://doi.org/10.1145/230538.230561

[37] Future of Privacy Forum. 2018. The Privacy Expert's Guide To Artificial Intelligence and Machine Learning.

[38] Gloria González Fuster, Jef Ausloos, Damian Bons, Lee A Bygrave, Barbara Rosa, Laura Drechsler, Olga Gkotsopoulou, Christopher Hristov, and Kristina Irion. 2022. The right to lodge a data protection complaint: OK, but then what?

[39] Gloria González Fuster and Bella. 2013. European Data Protection and the Haunting Presence of Privacy. Novática (2013), 17–22.

[40] Gloria González Fuster and Hielke Hijmans. 2019. The EU rights to privacy and personal data protection: 20 years in 10 questions.

[41] Inge Graef, Damian Clifford, and Peggy Valcke. 2018. Fairness and enforcement: bridging competition, data protection, and consumer law. International Data Privacy Law 8, 3 (2018), 200–223.

[42] Ben Green and Lily Hu. 2018. The myth in the methodology: Towards a recontextualization of fairness in machine learning. In 35th International Conference on Machine Learning.

[43] Xiuting Gu, Zhu Tianqing, Jie Li, Tao Zhang, Wei Ren, and Kim Kwang Raymond Choo. 2022. Privacy, accuracy, and model fairness trade-offs in federated learning. Computers and Security 122, (2022), 102907. DOI:https://doi.org/10.1016/j.cose.2022.102907

[44] Seda Gürses, Carmela Troncoso, and Claudia Diaz. 2015. Engineering privacy by design reloaded. In Amsterdam Privacy Conference, 1–21. Retrieved from https://iapp.org/resources/article/engineering-privacy-by-design-reloaded/

[45] Serge Gutwirth and Paul De Hert. 2021. Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power. Direito Público 18, 100 (2021), 500–549. DOI:https://doi.org/10.11117/rdp.v18i100.6200

[46] Faisal Hamman, Jiahao Chen, and Sanghamitra Dutta. 2022. Can Querying for Bias Leak Protected Attributes? Achieving Privacy With Smooth Sensitivity. In ACM Conference on Fairness, Accountability, and Trans- parency (FAccT '23), 1358–1368. DOI:https://doi.org/10.1145/3593013.3594086

[47] Christopher Heinz, Nigel Wall, Alexander H. Wansch, and Christoph Grimm. 2020. Privacy, GDPR, and Homomorphic Encryption. In IoT Platforms, Use Cases, Privacy, and Business Models. 165–184. DOI:https://doi.org/https://doi.org/10.1007/978-3-030-45316-9_8

[48] Johannes Heurix, Peter Zimmermann, Thomas Neubauer, and Stefan Fenz. 2015. A taxonomy for privacy enhancing technologies. Computers and Security 53, (2015), 1–17. DOI:https://doi.org/10.1016/j.cose.2015.05.002

[49] Jaap-Henk Hoepman. 2014. Privacy Design Strategies. In 29th IFIP International Information Security Con ference (SEC), Springer, 446–459.

[50] Anna Lauren Hoffmann. 2019. Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. Information Communication and Society 22, 7 (2019), 900–915. DOI:https://doi.org/10.1080/1369118X.2019.1573912

[51] Information Commissioner's Office. 2022. Chapter 5: Privacy Enhancing Technologies (PETs).

[52] Margot E. Kaminski and Gianclaudio Malgieri. 2021. Algorithmic impact assessments under the GDPR: producing multi-layered explanations. International Data Privacy Law 11, 2 (August 2021), 125–144. DOI:https://doi.org/10.1093/idpl/ipaa020

[53] Atoosa Kasirzadeh and Damian Clifford. 2021. Fairness and Data Protection Impact Assessments. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, ACM, New York, NY, USA, 146–153. DOI:https://doi.org/10.1145/3461702.3462528

[54] Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich. 2021. Fairness as Equality of Opportunity: Normative Guidance from Political Philosophy. (June 2021). Retrieved from http://arxiv.org/abs/2106.08259

[55] Niki Kilbertus, Adria Gascon, Matt Kusner, Michael Veale, Krishna P. Gummandi, and Adrian Weller. 2018. Blind Justice: Fairness with Encrypted Sensitive Attributes. In Proceedings of the 35th International Conference on Machine Learning, 1–10.

[56] Thierry Kirst, Olivia Tambou, Virginie Do, and Alexis Tsoukiàs. 2022. Fairness and Explainability in Automatic Decision-Making Systems. A challenge for computer science and law. Cahier du Lamsade 402, May (2022), 1–53.

[57] Dariusz Kloza, Alessandra Calvi, Simone Casiraghi, Sergi Vazquez Maymir, Nikolaos Ioannidis, Alessia Tanas, and Niels van Dijk. 2020. Data protection impact assessment in the European Union: developing a template for a report from the assessment process. d.pia.lab Policy Brief, VUB (2020), 1–52. DOI:https://doi.org/10.31228/osf.io/7qrfp

[58] Dariusz Kloza, Niels van Dijk, Simone Casiraghi, Sergi Vazquez Maymir, and Alessia Tanas. 2021. The concept of impact assessment. In Border Control and New Technologies, Peter J. Burgess and Dariusz Kloza (eds.). ASP, Brussels, 31–48. DOI:https://doi.org/10.46944/9789461171375.2

[59] Carolin E. M. Koll, Sina M. Hopff, Thierry Meurers, Chin Huang Lee, Mirjam Kohls, Christoph Stellbrink, Charlotte Thibeault, Lennart Reinke, Sarah Steinbrecher, Stefan Schreiber, Lazar Mitrov, Sandra Frank, Olga Miljukov, Johanna Erber, Johannes C. Hellmuth, Jens-Peter Reese, Fridolin Steinbeis, Thomas Bahmer, Marina Hagen, Patrick Meybohm, Stefan Hansch, István Vadászc, Lilian Krist, Steffi Jiru-Hillmann, Fabian Prasser, Jörg Janne Vehreschild, and NAPKON Study Group. 2022. Statistical biases due to anonymization evaluated in an open clinical dataset from COVID-19 patients. Nature - Scientific Data 7, 776 (2022), 1–15. DOI:https://doi.org/10.1038/s41597-022-01669-9

[60] Bert Jaap Koops, Bryce Clayton Newell, Tjerk Timan, Ivan Škorvánek, Tomislav Chokrevski, and Maša Galič. 2017. A typology of privacy. University of Pennsylvania Journal of International Law 38, 2 (2017), 483–575.

[61] Blagovesta Kostova, Seda Gürses, and Carmela Troncoso. 2020. Privacy Engineering Meets Software Engineering. On the Challenges of Engineering Privacy By Design. (2020). Retrieved from http://arxiv.org/abs/2007.08613

[62] Bogdan Kulynych, Carmela Troncoso, Hsiang Hsu, and Flavio du Pin Calmon. 2023. Arbitrary Decisions are a Hidden Cost of Differentially Private Training. In Proceedings of ACM Conference on Fairness, Accountability, and Transparency (FAccT'23), Association for Computing Machinery, 1–14. DOI:https://doi.org/10.1145/3593013.3594103

[63] Bogdan Kulynych, Mohammad Yaghini, Giovanni Cherubin, Michael Veale, and Carmela Troncoso. 2022. Disparate Vulnerability to Membership Inference Attacks. In Proceedings on Privacy Enhancing Technologies, 460–480. DOI:https://doi.org/10.2478/popets-2022-0023

[64] Alfonso Lamadrid De Pablo. 2017. Competition Law as Fairness. Journal of European Competition Law & Practice 8, 3 (2017), 147–148. DOI:https://doi.org/10.1093/jeclap/lpx003

[65] David Lehr and Paul Ohm. 2017. Playing with the Data: What Legal Scholars Should Learn About Machine Learning. University of California Davis Law Review 51, 653 (2017), 653–717.

[66] Fengxia Liu, Zhiming Zheng, Yexuan Shi, Yongxin Tong, and Yi Zhang. 2024. A survey on federated learning: a perspective from multi-party computation. Frontiers in Computer Science 18, 1 (2024), 181336.

[67] Sijia Liu, Yuanshun Yao, Jinghan Jia, Stephen Casper, Nathalie Baracaldo, Peter Hase, Xiaojun Xu, Yuguang Yao, Hang Li, Kush R Varshney, Mohit Bansal, Sanmi Koyejo, and Yang Liu. 2024. Rethinking Machine Unlearning for Large Language Models. 2019 (2024), 1–15. Retrieved from arxiv:2402.08787v3

[68] Alessandro Mantelero. 2022. Beyond Data - Human Rights, Ethical and Social Impact Assessment in AI. Springer.

[69] Kobbi Nissim and Alexandra Wood. 2018. Is privacy privacy? Phylosophical Transactions of the Royal Society A 376, 20170358 (2018), 1–17.

[70] Evaggelia Pitoura, Kostas Stefanidis, and Georgia Koutrika. 2021. Fairness in rankings and recommenders: Models, methods and research directions. In International Conference on Data Engineering, 2358–2361. DOI:https://doi.org/10.1109/ICDE51399.2021.00265

[71] David Pujol, Ryan McKenna, Satya Kuppam, Michael Hay, Ashwin Machanavajjhala, and Gerome Miklau. 2020. Fair decision making using privacy-protected data. In FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 189–199. DOI:https://doi.org/10.1145/3351095.3372872

[72] Christian Reimsbach-Kounatze, Taylor Reynolds, and Clarisse Girot. 2023. Emerging Privacy Enhancing Technologies - Current regulatory and policy approach.

[73] Carlotta Rigotti and Alessandra Calvi. 2022. Privacy. Elgar Encyclopedia of Law and Data Science, 275–280.

[74] Ira Rubinstein and Woodrow Hartzog. 2015. Anonymization and Risk. Washington Law Review 91, 15 (2015), 704–760.

[75] Stefan Schiffner, Bettina Berendt, Triin Siil, Martin Degeling, Robert Riemann, Florian Schaub, Kim Wuyts, Massimo Attoresi, Seda Gürses, Achim Klabunde, Jules Polonetsky, Norman Sadeh, and Gabriela Zanfir-Fortuna. 2018. Towards a roadmap for privacy technologies and the general data protection regulation: A transatlantic initiative. In Annual Privacy Forum. DOI:https://doi.org/10.1007/978-3-030-02547-2_2

[76] Andrew D. Selbst. 2021. An Institutional view of algorithmic impact assessment. Harvard Journal of Law & Technology 35, 1 (2021), 117–191.

[77] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2020. Participation is not a Design Fix for Machine Learning. In 37th International Conference on Machine Learning, Vienna, 119–125. Retrieved from http://arxiv.org/abs/2007.02423

[78] Julia Slupska. 2019. Safe at Home: Towards a Feminist Critique of Cybersecurity. St Antony's International Review 15, 1 (2019), 83–100.

[79] Ana Sokolovska and Ljupco Kocarev. 2018. Integrating Technical and Legal Concepts of Privacy. IEEE Access 6, (2018), 26543–26557. DOI:https://doi.org/10.1109/ACCESS.2018.2836184

[80] Daniel J. Solove. 2008. Understanding privacy. Harvard University Press.

[81] Daniel J Solove. 2023. The Limitations of Privacy Rights. Notre Dame Law Review 98, 975 (2023), 975–1036.

[82] Sophie Stalla-Bourdillon and Alison Knight. 2016. Anonymous data v. personal data - A false debate: an EU perspective on anonymization, pseudonymization and personal data. Wisconsin International Law Journal 34, 2 (2016), 285–322.

[83] Vinith M Suriyakumar, Nicolas Papernot, Anna Goldenberg, and Marzyeh Ghassemi. 2021. Chasing Your Long Tails: Differentially Private Prediction in Health Care Settings. In Conference on Fairness, Accountability, and Trans parency (FAccT'21), 723–734. DOI:https://doi.org/10.1145/3442188.3445934

[84] The Open Data Institute. 2022. Privacy Enhancing Technologies: Market Readiness, Enabling and Limiting Factors in the UK public sector. Retrieved from https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/Privacy-Enhancing-Technologies-Market-Readiness-Enabling-and-Limiting-Factors.pdf

[85] Carmela Troncoso, Dan Bogdanov, Edouard Bugnion, Sylvain Chatel, Cas Cremers, Seda Gürses, Jean-Pierre Hubaux, Dennis Jackson, James R. Larus, Wouter Lueks, Rui Oliveira, Mathias Payer, Bart Preneel, Apostolos Pyrgelis, Marcel Salathé, Theresa Stadler, and Michael Veale. 2022. Deploying Decentralized, Privacy-Preserving Proximity Tracing. Communications of the ACM 65, 9 (2022), 48–57. DOI:https://doi.org/10.1145/3524107

[86] Nicol Turner Lee, Paul Resnick, and Genie Barton. 2021. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Retrieved from https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

[87] Michael Veale. 2023. Some Commonly-Held but Shaky Assumptions about Data, Privacy and Power. In Research Handbook on Competition Law and Data Privacy, Maria Ioannidou and Despoina Mantzari (eds.). Edward Elgar.

[88] Michael Veale. 2023. Rights for Those Who Unwillingly, Unknowingly and Unidentifiably Compute! In The Person and the Future of Private Law, Hans-Wolfgang Micklitz and Giuseppe Vettori (eds.). Hart Publishing. Retrieved from https://osf.io/preprints/socarxiv/4ugxd/

[89] Michael Veale, Reuben Binns, and Jef Ausloos. 2018. When data protection by design and data subject rights clash. International Data Privacy Law 8, 2 (May 2018), 105–123. DOI:https://doi.org/10.1093/idpl/ipy002

[90] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2021. Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law. West Virginia Law Review 123, 3 (2021), 1–51. DOI:https://doi.org/10.2139/ssrn.3792772

[91] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2021. Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. Computer Law and Security Review 41, 105567 (2021), 1–31. DOI:https://doi.org/10.1016/j.clsr.2021.105567

[92] Ganghua Wang, Ali Payani, Myungjin Lee, and Ramana Kompella. 2023. Mitigating Group Bias in Federated Learning: Beyond Local Fairness. (2023), 1–19. Retrieved from http://arxiv.org/abs/2305.09931

[93] Lindsay Weinberg. 2022. Rethinking Fairness: An Interdisciplinary Survey of Critiques of Hegemonic ML Fairness Approaches. Journal of Artificial Intelligence Research 74, (2022), 75–109. DOI:https://doi.org/10.1613/jair.1.13196

[94] James Q. Whitman. 2004. The two western cultures of privacy: Dignity versus liberty. Yale Law Journal 113, 6 (2004), 1151–1221. DOI:https://doi.org/10.2307/4135723

[95] Michael Wills. 2022. The Official (ISC)2 SSCP CBK Reference. SYBEX - Wiley.

[96] Victor Y. Wu, Helen Webley-Brown, Jennifer King, and Daniel E. Ho. 2023. The Privacy-Bias Tradeoff: Data Minimization and Racial Disparity Assessments in U.S. Government. In 2023 ACM Conference on Fairness, Accountability, and-Transparency (FAccT' 23), 492–505. DOI:https://doi.org/10.1145/3593013.3594015

[97] Hanyu Xue, Bo Liu, Ming Ding, Tianqing Zhu, Dayong Ye, Li Song, and Wanley Zhou. DP-IMAGE: Differential Privacy for Image Data in Feature Space. Retrieved from arxiv:2103.07073v2

[98] Shui Yu. 2016. Big Privacy: Challenges and Opportunities of Privacy Study in the Age of Big Data. IEEE Access 4, (2016), 2751–2763. DOI:https://doi.org/10.1109/ACCESS.2016.2577036

[99] Ying Zhao and Jinjun Chen. 2022. A Survey on Differential Privacy for Unstructured Data Content. ACM Comput. Surv. 54, 10s, Article 207 (January 2022), 28 pages. https://doi.org/10.1145/3490237