Proxy Fairness under the European Data Protection Regulation and the AI Act: A Perspective of Sensitivity and Necessity

Ioanna Papageorgiou
Leibniz University Hannover
Institute for Legal Informatics
Hannover, Germany
ioanna.papageorgiou@iri.uni-hannover.de

Abstract

This paper navigates the convergence of the European Data Protection Regulation and the AI Act within the paradigm of computational methods that operationalise fairness in the absence of demographic data, notably through the use of proxy variables and inferential techniques (*Proxy Fairness*). Particularly, it explores the legal nature of the data involved in Proxy Fairness under the European Data Protection Regulation, focusing on the legal notion of Sensitivity. Moreover, it examines the lawfulness of processing sensitive personal data for Proxy Fairness purposes under the AI Act, particularly focusing on the legal requirement of Necessity. Through this analysis, the paper aims to shed light on core aspects of the legitimacy of Proxy Fairness in the context of EU law, providing a normative foundation to this line of Fair-AI approaches.

Introduction

The increasing adoption of AI systems at high stake areas of public life along with extensive studies on the discriminatory potential of AI (Mehrabi et al. 2021), have prompted a proliferation of algorithmic methods that study and pursue fairness in AI systems (*Fair-AI*) (Ntoutsi et al. 2020; Schwartz et al. 2022; Mitchell et al. 2021). These methods are centered on the detection, mitigation and evaluation of bias across legally protected groups, and almost invariably require access to sensitive attributes, like demographics, that determine group membership. However, this often implies the processing of personal sensitive data, which is in principle prohibited according to the EU data protection law, posing challenges to the feasibility of Fair-AI approaches.

In response to this challenge, a growing line of AI research (Ashurst and Weller 2023; Centre for Data Ethics and Innovation and Department for Science, Innovation and Technology 2023; Awasthi et al. 2021; Chen et al. 2019a; Yan, te Kao, and Ferrara 2020; Zhu et al. 2023; Gupta et al. 2018) has studied computational methods that enable fairness operationalization in the absence of demographic data, notably through the use of proxy variables and inferential techniques (*Proxy Fairness*). Besides being increasingly studied, proxy fairness methods have already been widely employed across various sectors, including tax, finance, consumer protection (Elzayn et al. 2023; Baines and Courchane 2014; Consumer Financial Protection Bureau 2014; Elliott et al. 2009) and social media (Alao et al. 2021) (Belli et al. 2021), particularly in a U.S. context. However, scant attention has been given thus far to the interaction of these methods with existing data protection regulations, posing significant legal uncertainty regarding the legitimacy of these approaches.

This uncertainty intensifies in the face of ongoing regulatory developments. Particularly, the upcoming AI Act has also addressed the challenge of data scarcity in the context of Fairness, by enabling, on grounds of public interest, the processing of personal sensitive data for the purposes of bias detection and correction in high-risk AI systems. Precisely, according to the Article 10 (5) AI Act, the processing of personal sensitive data is permitted only "to the extent that it is *strictly necessary* for the purposes of ensuring bias detection and correction in relation to the high-risk AI systems..[emphasis added]". While the enabling provision appears to be method-agnostic, meaning that it's not restricted to a particular fairness approach, the stipulated necessity requirement significantly influences the choice of fairness methods, and to a greater extent, the scope of Proxy Fairness.

In light of the above, this paper aims to examine the legal implications of Proxy Fairness under the General Data Protection Regulation and the AI Act, providing a normative foundation to this line of Fair-AI approaches. Precisely, the paper provides the following contributions:

- The first section analyzes Proxy Fairness under the General Data Protection Regulation, focusing on the notion of data "*Sensitivity*". Drawing from established jurisprudence of the European Court of Justice and prominent legal scholarship, the paper scrutinizes the nature of data involved in Proxy Fairness approaches, shedding light on nuanced data protection aspects surrounding *proxy variables* and *data inferences*.
- The subsequent section analyzes Proxy Fairness within the scope of the AI Act, particularly operationalising the legal notion of "*Necessity*". Drawing upon Article 10 (5) of the AI Act, the paper examines the necessity of processing sensitive personal data in the context of proxy fairness methods, shedding light on core aspects of their

legitimacy. This examination involves a comparative assessment of proxy fairness approaches versus default alternatives, considering the necessity criteria of *intrusiveness*, *effectiveness*, and *reasonableness*.

While the analysis draws on European regulatory frameworks, its relevance transcends beyond Union borders to extra-territorial Fair-AI endeavors. Specifically, AI providers, who place high risk AI systems on the EU market or put those systems into service in the Union, fall under the scope of the Act, regardless of their establishment or location within the Union or a third country.[1] Accordingly, they are also subject to the AI Act's data governance obligations, including the obligation to examine training, testing and validation datasets for relevant biases and take appropriate measures to detect, prevent and mitigate them.[2] To the extent thus that non-EU AI providers process information related to individuals located in the EU in order to detect or correct biases in their datasets, they must adhere to the applicable EU requirements, including the requirement of necessity. [3]

Methodology

The paper fundamentally adopts a doctrinal legal research methodology. Two EU legislative pieces, the General Data Protection Regulation and the AI Act, serve as the normative framework of the analysis, while the legal notions of data 'sensitivity' and processing 'necessity' form its central concepts. However, insofar as it examines technical phenomena within this normative and conceptual framework, such as computational fairness approaches, the analysis takes on a techno-legal approach concerning the subject of inquiry.

Particularly, the paper combines descriptive, classifying, and evaluative methodological features (Kestemont 2018; Smits 2017), in order to identify the underlying legal system applicable in the context of Proxy Fairness and to steer its lawful application.

To describe the legal notion of data "sensitivity", the first section employs grammatical and systemic legal interpretation as well as interpretation on the basis of EU jurisprudence, non-binding legal sources and legal doctrine. Subsequently, the insights from the descriptive analysis, largely following deductive reasoning, inform the classification of 'proxy' and 'inference' data under the described notion of personal sensitive data, with insights from technical literature serving as an external criterion for this classification.

The second section employs the same international tools and internal criteria to describe the legal notion of "necessity" and to classify features of Proxy Fairness under necessity's components. Due to the inherent comparative element of Necessity, the section builds largely upon a comparative evaluation between Proxy fairness approaches[4] and their 'Default' counterparts[5], along the legal criteria of intrusiveness, effectiveness, and reasonableness. Insights from the literature of computer science, STS (Science, Technology, and Society) and AI ethics are incorporated as evaluation indicators, informing the legal criteria without altering though the paper's assessment framework.

While holding to a doctrinal normative approach, the paper adopts an interdisciplinary perspective, exposing the complex interdependencies between (Data Protection) Law, Computer Science, and Ethics in the context of Fair-AI and creating a conversational crossroads to be further leveraged and enriched by the respective disciplines and their methodologies.

Proxy Fairness under the GDPR: a Sensitivity Perspective

The notion of sensitive data, well-established in European data protection law, is used within the General Data Protection Regulation to denote "special categories of personal data", which due to their nature are particularly sensitive in relation to fundamental rights and thus require more extensive protection compared to "ordinary" personal data (recital 51 GDPR). Article 9 of the GDPR restrictively defines sensitive data as those revealing racial and ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, as well as genetic, biometric and health data, and permits their processing only under strict exceptions.

The AI Act provides such an exception in Article 10 (5), enabling the processing of sensitive personal data for bias detection and correction in high-risk AI systems, explicitly referencing Article 9 of the GDPR and its concept of sensitive data.[6] In order to assess thus Proxy fairness under this exception of the AI Act, it is necessary to first investigate the extent to which it involves the processing of sensitive data under the meaning of the GDPR.[7] For this purpose,

[1] See art. 2 (1) AI Act.

[2] See art. 10 (2) (g), (f) AI Act

[3] AI providers are not obliged to debias their datasets by processing sensitive information relating to individuals in the EU. However, under Article 10 (3), (4) AI Act, they are required to ensure that the training, validation, and testing datasets possess appropriate statistical properties with regards to the individuals to whom the high-risk AI system is intended to be used, while also considering the characteristics specific to the geographical setting in which the high-risk AI system will operate. This requirement may lead to the customization of the datasets used for fairness purposes to suit the Union's local context, incentivizing the utilization of EU-related datasets.

[4] The legal analysis considers a baseline paradigm of Proxy Fairness, which involves inferring the pertinent sensitive attribute from related features (proxy variables) through the use of an attribute classifier - also known as a proxy model-, in order to evaluate or control for fairness. While variations of proxy fairness approaches may influence aspects of the analysis, the underlying logic applies *mutatis mutandis* to any fairness method that deals to some extent with proxies and inferences of sensitive data.

[5] In the context of this analysis, 'Default' refers to approaches situated in a non-demographically scarce regime that directly process the "real" sensitive attributes obtained directly or indirectly from the data subjects.

[6] The terms "special categories of data" and "sensitive data" are used interchangeably within the GDPR. Cf. Recital 10 GDPR.

[7] The "processing" of data involves under Article 4 (2) GPDR a wide range of operations performed on personal data, including the data collection, organisation, structuring, storage, alteration, retrieval, use or disclosure.

the ensuing chapter distinguishes between two main data-pillars involved in Proxy Fairness, namely the *Proxy* and the *Inferred* data, and assesses them under the notion of sensitivity.[8] The legal analysis considers a baseline paradigm of Proxy Fairness, which involves inferring the pertinent sensitive attribute from related features (proxy variables) through the use of an attribute classifier - also known as a proxy model-, in order to evaluate or control for fairness. Besides providing the necessary foundation for the subsequent analysis, this section, represents a critical legal exercise. This is because inferential approaches, due to their inherent intricacies and indirect interference with sensitive data, often risk slipping through the cracks of the European Data Protection Regulation and its strict protective regime for sensitive data.

Proxy Data

Proxy-based approaches are conceptually built around the utility and use of proxy attributes, namely data that are associated with, and could be used as a substitute or stand-in for unavailable or inaccessible actual demographic traits. An individual's surname or address, for example, could serve as proxies for their ethnicity and be used to calculate the likelihood of their belonging to a specific race or ethnicity. In the context of Proxy Fairness, ordinary source data are being processed to infer the pertinent sensitive attributes, which (inferences) are then used along with standard fairness metrics for fairness evaluation or optimization. From a data protection perspective, this raises the non-trivial question of the nature of the first data block involved in Proxy Fairness, namely the Proxy Data, also referred to as inference-producing data.

To address this question one must first draw from a textual reading of the GDPR and specifically of its Article 9 (1), which defines sensitive data as "data *revealing* racial or ethnic origin..[emphasis added]". The wording of this provision explicitly allows for a broad interpretation, encompassing not just inherently sensitive data — i.e. data that by its nature contain sensitive information — but also data from which sensitive information with regard to an individual can be concluded. This is due to the key term "revealing", which describes the nature of the link between the personal data deemed as sensitive and the information content of this data (Spiecker et al. 2023).

This expansive interpretative approach has been explicitly endorsed by the Article 29 Working Party across various guidelines, and finally affirmed by the European Court of Justice (ECJ) in its recent jurisprudence (Nowak 2017). In this specific ruling, the ECJ explicitly expanded the GDPR's scope from *inherently* sensitive data to information that *indirectly* allows the inference of sensitive information by means of intellectual operations like deduction or cross-referencing. Particularly, according to the ECJ, it suffices that the data is merely capable to indirectly reveal sensitive information.

Consequently, ordinary source data may also be treated as sensitive data, when it can be shown to allow for sensitive attributes to be inferred (cf. (Wachter and Mittelstadt 2018; Quinn and Malgieri 2020; Malgieri and Comandè 2017)). This interpretation is particularly relevant in the context of Proxy Fairness, which is conceptually grounded on the capability of proxy models to infer missing sensitive attributes from ostensibly non-sensitive data. For instance, an individual's postcode may not be inherently sensitive, but when combined with other data in the context of Proxy Fairness, it can lead to the inference of the individual's actual ethnicity. Failing to consider this data as sensitive would risk, according to the established ECJ jurisprudence, compromising the effectiveness of GDPR's special protection afforded to sensitive data and in greater extent the associated fundamental rights and freedoms.

However, the ECJ implicitly dismissed with this jurisprudence a large volume of legal scholarship that has critically questioned the contemporary relevance and effectiveness of the GDPR's concept of sensitive data and has accordingly sough to narrow down its scope. Particularly concerned with an "inflation risk" of sensitive data in the face of technological developments, many scholars have proposed moderating approaches that inject additional subjective or objective factors into the definition of sensitive data (Quinn and Malgieri 2020; Wachter and Mittelstadt 2018; Solove 2024; Schiff, Ehmann, and Selmayr 2017). The subsequent paragraphs discuss the case of Proxy Fairness in light of these restrictive approaches.

To begin with, subjective factors center around "intentionality" and essentially look at the intentions and stated purposes of data controllers with respect to the inference-generation, excluding thereby data that accidentally or co-incidentally reveal sensitive information. Given that the use of proxy data in the context of Proxy Fairness conceptually involves the intentional or affirmative inference of relevant sensitive information, upon which the detection and correction of bias would draw, proxy data seem to pass the sensitivity test under suggested purposeful approaches. This is plausible, given that the sensitive data is typically the desired result of the classifier's analytical process in standard proxy methods.[9]

Furthermore, Proxy Fairness seems to satisfy most of the objective criteria of sensitivity, which consider the processing context and include inter alia the costs and the amount of time required for the inference, the available technology at the time of processing (Malgieri and Comandè 2017) and the ease or the reliability of inference (Malgieri and Comandè 2017; Wachter and Mittelstadt 2018). Firstly, proxy models claim to be today a relatively simple and easily de-

[8] Given that the classification of data as "sensitive" within the meaning of the GDPR essentially implies its classification as "personal", aspects of the personal link of data —especially those relevant to the context of proxy fairness— are inevitably included. However, the analysis primarily assumes the existence of personal data under the meaning of the GDPR, with questions regarding identifiability and anonymity falling outside the scope of this paper.

[9] This factor may vary in approaches where proxy data are not treated as such, but rather as a very rough signal about potential unfairness, without drawing conclusions or inferring the sensitive attribute itself (Andrus et al. 2021).

ployable solution, without implying significant complexities or prohibitive costs (Centre for Data Ethics and Innovation and Department for Science, Innovation and Technology 2023; Ashurst and Weller 2023). Particularly, a wide range of proxy tools is currently available as open source [e.g. predictrace (Kaplan 2023), ZRP (ZestAI 2024)], or commercial products (Namsor 2024), with proxy methodologies for race and ethnicity such Bayesian Improved Surname Geocoding (BISG) already being widely employed in the US context for fairness purposes (Elliott et al. 2009; Consumer Financial Protection Bureau 2014). In the same vein, academic research has emphatically suggested that sensitive data, and particularly ethnicity data, can be *readily* inferred from a variety of other seemingly innocuous data, such as location and surname, doctors' notes, or even Facebook likes (Solove 2024; Kosinski, Stillwell, and Graepel 2013).

However, the question whether proxy fairness methods are also *reliably* estimating the pertinent sensitive attribute appears more controversial. Several commentators (Wachter and Mittelstadt 2018; Gola and Heckmann 2022; Finck 2021) point out that ordinary proxy data should only be reclassified if they provide a reliable or statistically significant basis for inference generation, although the ECJ itself does not specify any threshold level required for the proxy data's revealing "capability".[10] In the examined context of Proxy Fairness, although currently available or employed proxy methods often (self-) report a high level of accuracy in predicting the sensitive attribute in question, reported accuracy rates frequently prove unreliable and inconsistent across demographic groups, especially in the presence of concept and model drift (Centre for Data Ethics and Innovation and Department for Science, Innovation and Technology 2023; LeClair, Parker, and Young 2023). Overall, assessing Proxy Fairness against the factor of reliability, would condition the "sensitivity" of proxy data on the level of accuracy of the attribute classifier : the higher the predictive accuracy of the attribute classifier the better the indication of data sensitivity.

Inferred Data

The second data block involved in in Proxy Fairness, namely the generated inferences, have also been subject to intense academic debate, particularly due to their *artificial* and *probabilistic* nature. Given that inferences are algorithmic byproducts of an analytical process on other data (proxy), rather than directly observed or collected from the data subject, they raise the following preliminary question: should they be regarded personal data under the GDPR, distinct from the data from which they were inferred?

The GDPR's definition of personal data as "*any information...*" and the available guidelines of the Article 29 Working Party (Article 29 Data Protection Working Party 2007, 2017, 2016) call for a wide interpretation which is not constrained to a specific type or form of information and may thus include not only collected but also inferred data.[11]

The ECJ has also indirectly addressed this question in (Nowak 2017), extending the scope of personal data to encompass both "objective" and "subjective" information such as opinions or assessments, as long as it relates to data subjects. Through means of legal analogy and by resembling inferences to subjective opinions or types of assessments, several scholars (Hallinan and Zuiderveen Borgesius 2020; Wachter and Mittelstadt 2018) have employed this jurisprudence to include algorithmically generated inferences under the scope of the GDPR. Accordingly, inferences generated in the context of Proxy Fairness can be viewed as probabilistic "opinions", which emerge on the back of factual datasets (Proxy Data) subjected to an analytical and interpretative framework (the classification algorithm) in order to generate new, probable conclusions regarding the individuals represented in the datasets (e.g. their ethnicity). As such, they would fall under the GDPR scope of personal data based on the aforementioned jurisprudence. Moreover, according to the approach followed by the ECJ and the Article 29 Working Party the inferences generated in the context of Proxy Fairness would relate to the data subjects by virtue of their "content"[12], representing information *about* the data subjects (i.e. their ethnic origin) and directly describing them.[13]

Importantly, the generated inferences should be classified as sensitive personal data regardless of their accuracy or validity, meaning whether they accurately predict the actual ethnic origin of the corresponding data subject. As stated by the Article 29 Working Party (Article 29 Data Protection Working Party 2007), information does not need to be true or proven to be considered personal data under the meaning of the data protection regulation, highlighting that data protection rules already envisage the possibility of incorrect information. That is particularly relevant in the context of Proxy Fairness, as the predicted inferences rely on probabilistic correlations rather than causation, lacking often scientific validity.[14] Finally, by explicitly including subjective information in the form of opinions or assessments under the GDPR, the European Court of Justice has also indicated that personal data do not relate only to true or verifiable facts. Accordingly, utilizing data inferred by proxy models for evaluating or correcting bias would constitute processing of sensitive data, irrespective of the accuracy of the proxy

[10]Cf. (Wachter and Mittelstadt 2018), that interprets the General Court's decision in (Egan and v European Parliament 2012), as an affirmation that reliability is an essential attribute of sensitive data.

[11]Cf. also the statement of the Article 29 Working Party in the context of biometric data: "*the extraction of information from the samples is collection of personal data, to which the rules of the Directive applies* (Art29WP 2012)."

[12]According to the threefold approach followed by the Article 29 Working Party (Article 29 Data Protection Working Party 2007) and ECJ jurisprudence, information can 'relate' to an individual "by reason of its content, purpose or effect". In that regard, attributes inferred in the context of Proxy Fairness differ from other types of inferences, for example, those generated in the context of profiling, which relate to data subjects by virtue of their "purpose" or "effects" (Wachter and Mittelstadt 2018).

[13]This link may be disputed in approaches, where inferences are generated at the group aggregated level (e.g. "X percent of this dataset is women/native") and not linked to an individual (Centre for Data Ethics and Innovation and Department for Science, Innovation and Technology 2023; Andrus et al. 2021).

[14]Cf. subsection "Reasonableness".

model in question.

Finally, in terms of sensitivity, it seems straightforward that inferences drawn for fairness evaluations or interventions would be of a sensitive nature when they directly disclose or inherently represent a category of data that is, as such, highly protected by the GDPR, as in the case of "ethnicity".

Proxy Fairness under the AI Act: a Necessity Perspective

The preceding section has analyzed how Proxy Fairness intersects with the European Data Protection Regulation, focusing particularly on the Regulation's special regime for sensitive data. To the extent that they involve the processing of sensitive personal data for bias detection and correction, proxy fairness approaches fall under the scope of article 10 (5) of the AI Act.

As mentioned above, this provision outlines an exception to the GDPR's prohibition on processing sensitive personal data, opening up this possibility for bias detection and correction purposes. While this exception applies to any type of fairness method that involves sensitive data processing, be it Default or Proxy, the prescribed necessity requirement significantly impacts the choice of fairness method in a specific case. Particularly, according to art. 10 (5) AI Act the processing of sensitive data is permitted only "to the extent that it is strictly *necessary* for the purposes of ensuring negative bias detection and correction in relation to the high-risk AI systems [emphasis added]", i.e. only under the requirement of Necessity.

The necessity principle, which has been a recurrent condition to the processing of personal data, essentially dictates that data processing is permissible only to the extent that there is not a *less intrusive* but *similarly effective* alternative available, which can *reasonably* achieve the objective at hand (European Data Protection Supervisor 2023; Schantz and Wolff 2017). According to ECJ Jurisprudence (Meta Platforms and Others 2023; TK v Asociaţia de Proprietari bloc M5A-ScaraA 2018), it must be interpreted in a manner that fully reflects the objective of the data protection regulation and, importantly, in conjunction with the *data minimisation* principle enshrined in Article 5 (1) (c) GDPR.

In that regard, AI providers seeking to rely on the exception of the AI Act and process sensitive personal data for bias detection and correction must conduct a necessity test, which involves comparing available alternatives based on their levels of a) *intrusiveness*, b) *effectiveness* and c) *reasonableness*. Particularly, according to the Article 10 (5) (f) of the final version of the AI Act, AI providers are explicitly required to draw up a specific justification as part of record-keeping, where they explain that the processing operation was in compliance with the necessity principle. Accordingly, the utility of a fairness method alone can not justify the processing of sensitive personal data.

Conducting this necessity test is a highly complex task for AI providers as it demands not only a comprehensive grasp of the state-of-the-art in Fair-AI but also the application of vague, open-ended legal notions. This complexity is intensified by the fact that methods for bias detection and correction are still in their infancy, with limited real-world application, yet rapidly evolving. Consequently, this poses challenges for conducting a rigorous and evidence-based assessment of the availability, efficacy, and trade-offs of different fairness methods for a specific case.

On the other hand, given that the necessity requirement is intrinsically tied to the legitimacy of data processing, a wrongful interpretation and application risk not only compromising the data subjects' fundamental right to data protection but also exposing AI providers to serious consequences, ranging from reputational damage to severe financial penalties.

The following paragraphs examine proxy fairness approaches under the necessity requirement, particularly by comparing them with default approaches that directly collect and use "real" sensitive attributes[15], along the necessity axes of intrusiveness, effectiveness, and reasonableness.

Given that the application of necessity is inherently context-dependent and involves a level of discretion on the part of AI providers, this contribution cannot provide conclusive answers regarding the overall legitimacy of the examined methods. Instead, by shedding light on nuanced aspects of the necessity requirement, it aims to support the lawful processing of sensitive personal data in the context of Fairness, and provide practitioners, researchers and AI providers with the necessary tools for interpreting and implementing necessity across various fairness contexts.

Intrusiveness

To adhere to the necessity principle, AI providers are primarily required to assess whether the desired fairness objective could be attained with less intrusive means, namely with means that interfere less with data protection rights (Gola and Heckmann 2022). Some of the criteria that have been deemed relevant for evaluating the severity of the interference include, inter alia, the *volume* and *type* of data processed, *linkage possibilities* and the *risks of data misuse* (Schantz and Wolff 2017).

Since proxy fairness methods require processing individuals' 'related' features (e.g., address or surname) to infer the one missing attribute of interest (e.g., ethnicity), they *de facto* involve the processing of a larger volume of data compared to default approaches, effectively adding the extra 'data block' of proxy information. Moreover, to the extent that proxy data and generated inferences are classified as sensitive data under the GDPR[16], Proxy Fairness *de jure* implies the processing of more data of a sensitive nature, thereby increasing severity under the criterion of data type.

Similar findings emerge when considering the principle of data minimization, which implies i.a. minimizing the number of data and its uses to the greatest extent possible (Gola and Heckmann 2022). Particularly, this principle aims at reducing not just the quantity of data, but rather the relation

[15]By "real" sensitive attributes the paper refers to those that are directly or indirectly obtained from the data subjects, as opposed to new data that is inferred from already available data

[16]See subsection "Inferred Data".

of the data to a natural person, i.e. the ease with which the data can be connected to an individual (Spiecker et al. 2023; Panel for the Future of Science and Technology, EPRS — European Parliamentary Research Service, Scientific Foresight Unit (STOA) 2020). Accordingly, it can be argued that the collection and processing of various categories of data related to a data subject in the context of Proxy Fairness increase not only the possibilities of data linkage but to a greater extent the ease of the subject's identifiability, with certain types of common proxy data, such as surnames, being particularly identifying.

Thus far, it becomes apparent that the outcome of the intrusiveness assessment in a specific case would strongly depend on the type of model used and particularly on the number and type of input features it requires for the attribute and/or fairness estimation. In that respect, approaches that study fairness under the use of a small amount of proxies and especially under "weak" (i.e. inaccurate) proxies (Zhu et al. 2023) become increasingly relevant.

Finally, shifting focus to the criterion of *misuse risks* yields somewhat contradictory results. Data misuse typically describes cases where collected data are unlawfully used for other purposes, particularly in ways that could harm or damage individuals. A risk, on the other hand, describes the existence of the possibility of an event with harmful consequences occurring (van Dijk, Gellert, and Rommetveit 2016). In particular, under the GDPR, the notion of risk encompasses two dimensions, the *severity* of the harm to the rights of data subjects and the *likelihood* of the harmful event and the harm occurring (Recital 75 GDPR, (DSK Datenschutzkonferenz 2018)). Comparing Proxy and Default methods in terms of harmful risks is not an easy exercise, given that each approach comes with unique trade-offs and risks to fundamental rights.

First of all, recital 75 GDPR identifies cases where data subjects are prevented from exercising control over their personal data as potential instances of detrimental use of data bearing risks for their rights and freedoms. In the same vein, recital 7 GDPR states that in the face of rapid technological developments "natural persons should have control of their own personal data". Data processing in the context of inference-based Fairness is inherently more prone to such detrimental risks, as it hiders the participation and control of data subjects over the generation of information related to sensitive aspects of their identity. Particularly, treating sensitive identity characteristics such as race, religion, or sexual orientation as qualities that can be predicted externally, produces new forms of control over an individual's agency to define themselves and, to a further extent, over their autonomy (cf. (Keyes 2018)). Accordingly, the application of proxy fairness intrudes more on the right to privacy and data protection, particularly on their rationales of protecting *autonomy* and *informational self-determinaton*, respectively. Indeed, while legal scholars disagree on many aspects of privacy, there is a greater consensus that privacy plays an important role in protecting an individual's identity and autonomy and that informational self-determination constitutes a core rationale of the GDPR (Puri 2021; Thouvenin 2021).

On the other hand, recital 75 GDPR describes also discrimination cases as potential detrimental uses of sensitive data, while the GDPR itself, and particularly its sensitive data regime, is set to protect the fundamental right of non-discrimination (Spiecker et al. 2023). Evidently, abuses of non-discrimination principles stand out as a significant misuse risk linked to the processing of sensitive data. Numerous examples attest to the way the collection and storage of population data have facilitated human rights abuses and the prosecution of various groups based on racial or religious classifications (Seltzer and Anderson 2001). In that respect, and drawing on the dimension of 'likelihood' contained within the GDPR's notion of risk, we could argue that default approaches entail greater risks to non-discrimination rights. Particularly, given that quick reference to group membership records facilitates misuse and that inherently sensitive data de facto allows for quicker reference than proxies, it can be reasoned that the direct collection and storage of inherently sensitive data entail a higher risk of misuse compared to that of proxy data. To elaborate, while proxy data are capable of revealing sensitive attributes and shall be highly protected as such, it should not be overlooked that an additional level of deduction is still required. This implies a certain level of inferential effort and infrastructure in the event of unauthorized access or data leakage, which reduces the probability of their misuse for racial discrimination purposes.[17]

Centering the necessity assessment around concrete risks and harms rather than types of data is in line with the increasing scholarly voices arguing for a risk-oriented and contextual approach to the protection of sensitive data (Solove 2024; Ohm 2015; Simitis, Hornung, and Spiecker 2019). Daniel Solove (Solove 2024), for example, argues that *to be effective, privacy law must focus on use, harm, and risk rather than on the nature of personal data*, while, *harm and risk depend upon the situation*, and, *can rarely be determined outside of a context.*[18]

While this scholarship has so far explored risk and context as means to evaluate varying degrees of data sensitivity, the present contribution focuses on on the notion of risk as a means to assess the intrusiveness of data processing and, to a greater extent, its necessity. Due to its inherently flexible and contextual nature, the concept of necessity provides a suitable framework to incorporate circumstantial and risk-related factors, without interfering with the established ECJ approach concerning the "natural" sensitivity of fixed categories of data.

Finally, while risks and harms associated with autonomy loss are not negligible, it could be argued that focusing on discrimination risks aligns more with the rationale behind articles 10 (5) AI Act and 9 (2) GDPR as well as with the systematics of necessity in the GDPR. Particularly, even

[17]Crucial in this regard would be the type of technical and organisational measures in place to safeguard data security, including, for example, segregating the proxy datasets from the inferred ones.

[18]Similarly, according to (Spiecker et al. 2023), the intrusive intimate nature of certain personal data is situational and defining a set of cases that are entitled to a higher level of protection is one dimension of social reality.

though the heightened protection of sensitive data in articles 9 GDPR and 10 (5) AI Act aims to target risks to all fundamental rights, those risks are predominantly perceived in terms of an elevated probability of discrimination.[19] Moreover, while most scholars agree that informational self-determination is a rationale of the GDPR, it is not per se prominently reflected in the concept of necessity. Particularly, necessity serves as a requirement in these legal grounds for data processing that rely less on exercising informational self-determination, such as public interest or controller's legitimate interest, as opposed to those based on consent or the publication of data by the data subject itself.[20]

Effectiveness

Compliance with the requirement of necessity does not require prioritizing any kind of milder alternative, but only those milder alternatives that can attain the pursued objective in a comparably effective manner.

In a second step, AI providers must compare the identified alternatives with respect to their effectiveness in detecting and correcting bias, by relying on theoretical and/or empirical evidence regarding the utility and limitations of the fairness methods under consideration. This includes qualitative and quantitative arguments about the way relevant demographic groups would be better served by the planned intervention, such as performance and fairness metrics, accuracy of fairness estimates, and associated trade-offs. Considerations based solely on convenience or cost-effectiveness in terms of operational costs or organizational resources fall short of satisfying the effectiveness criterion of necessity. In the context of the present analysis, the consideration of effectiveness prompts a high-level comparison between Default and Proxy Fairness approaches based on theoretical and empirical evidence presented in the literature with respect to their effectiveness in detecting and correcting (racial) bias.

On one hand, simple proxy methods have already been employed in several domains[21], thereby demonstrating, in practical terms, a level of efficacy in evaluating and even accounting for racial disparities, without directly collecting or relying on real ethnicity data (Bogen, Rieke, and Ahmed 2020; Andriotis and Ensign 2015). Scientific research such as (Diana et al. 2022) has also demonstrated that it is possible to efficiently train proxies which can stand in for missing sensitive features in order to effectively train downstream classifiers subject to a variety of demographic fairness constraints. Moreover, it has been argued (Andrus et al. 2021), that inferred demographics might sometimes offer more objective and accurate insights compared to self-reported or labeled demographic data, making them more effective for specific bias detection tasks, such as in cases of bias linked to perceived race. However, the accuracy and utility of common proxy methods, which rely on a variety of assumptions, has been heavily disputed in academic research due to their tendency to overestimate demographic disparities and introduce errors and systematic biases (Ashurst and Weller 2023; Chen et al. 2019a; LeClair, Parker, and Young 2023; Imai et al. 2023).

On the other hand, the direct use of true demographic data, provided it is of high quality, evidently offers the most accurate estimates of group fairness metrics, allowing for a deeper analysis of group disparities (Ashurst and Weller 2023).[22] Contemporary research suggests that fairness methods which rely on proxies, even if well-trained, should still be regarded as a secondary solution in terms of efficiency (Ashurst and Weller 2023; Chen et al. 2019b). That seems plausible since proxy methods, by definition, strive to reproduce the default alternative in constrained settings where the relevant data are either inaccessible or not available at all stages of the development-deployment pipeline.

To sump up, the effectiveness comparison, like the entire necessity test, needs to be conducted on a case-by-case basis, and the outcome will inevitably vary depending on the targeted fairness metric or concept as well as the type of model and sector under consideration.

Reasonableness

Finally, the required assessment of alternatives is conditioned by a third element, this of reasonableness for the actor responsible for the data processing. This aspect of necessity is explicitly provided in recital 39 of the GDPR, which states that "personal data should be processed only if the purpose of the processing could not *reasonably* be fulfilled by other means [emphasis added]".

To elaborate, according to the prevailing opinion (Schantz and Wolff 2017; Information Commissioner's Office Accessed 2023-11-11), necessity does not imply that the data processing at hand is absolutely indispensable due to technical or other reasons for attaining the purpose at hand, nor does it suggest that it becomes unattainable without it. It "solely" implies that there is not an effective milder alternative available which is *reasonable* in terms of personal, operational or financial feasibility (Gola and Heckmann 2022). In particular, nothing practically impossible, prohibitively costly or illegal can be demanded from AI providers for complying with necessity. The criterion of reasonableness largely appears to accommodate utility-driven considerations, including in the case of Proxy Fairness the ease or feasibility of its implementation along with the associated costs and organizational resources.

[19] Cf. Recital 71 GDPR; Guidelines for the Regulation of Computerized Personal Data Files (Joinet, on Prevention of Discrimination, and of Minorities. Special Rapporteur on the Study of the Relevant Guidelines in the Field of Computerized Personal Files 1988), which justify a further protection for sensitive data on the premise that such data are "likely to give rise to unlawful or arbitrary discrimination"; Also (Kühling and Buchner 2020), who sees article 9 GDPR as a normative specification of Article 21 of the Charter of Fundamental Rights, i.e. the fundamental right to non-discrimination, as opposed to (Albers and Veit 2020), who find too narrow this teleological reduction. For more on the rationales for protecting sensitive data refer to (Quinn and Malgieri 2020).

[20] Cf. art. 6 (1) and 9 (2) GDPR and (Thouvenin 2021).

[21] Cf. paper's introduction.

[22] However, self-reported data are also potentially prone to inaccuracy (Andrus et al. 2021), limiting as well the effectiveness of Default Fairness.

Firstly, proxy methods are often portrayed as the sole avenue for practitioners seeking to offer quantitative insights for in-depth analysis, given the unavailability and irretrievability of "real" sensitive attributes within existing datasets. Proxy fairness, on the other hand, relies potentially on data already held by AI providers, thereby eliminating the time and resources required to directly collect sensitive demographic data. In this regard, proxies offer a relatively simple and implementable alternative, making Proxy Fairness an immediately feasible, practically usable and low-cost strategy that can facilitate various types of algorithmic fairness (Andrus and Villeneuve 2022; Centre for Data Ethics and Innovation and Department for Science, Innovation and Technology 2023).[23] This poses the critical question of whether — provided it is less intrusive — it would also be reasonable to mandate AI providers to collect demographic data "from scratch", directly from data subjects, despite the possible inconvenience, time or resource consumption, and the potential resulting obsolescence of a large corpus of existing datasets.

However, by requiring "*strict*" necessity, the Article 10 (5) of the AI Act raises the legal threshold of necessity in the case of bias detection and correction, surpassing the GDPR standard for other cases of public interest (see art.9 (1) (g)). This additional condition could impact the approach and rigor of the necessity assessment, suggesting either an absolute imperative necessity for processing sensitive data for fairness purposes - in a sense of a "conditio sine qua non"- or a less stringent standard when exploring alternative options. Satisfying necessity in the first scenario would require that it is absolutely impossible, based on the state of the art, to ensure bias detection and correction without the processing of sensitive personal data, be it through inference or direct use.

Nevertheless, the assessment of reasonableness shall not be reduced to a net utility calculus. As mentioned above, illegal alternatives would obviously fail to meet the reasonableness criterion. The same should according to the author apply *mutatis mutandis* for unethical alternatives. Particularly, inference-based approaches have garnered significant criticism from AI ethicists due to concerns over their lack of scientific validity and the ethical implications associated with predicting inherently sensitive identity qualities. These types of approaches have been criticized, among other things, for resembling physiognomy (Engelmann et al. 2022), echoing colonial practices (Scheuerman, Pape, and Hanna 2021), and notably reducing individuals' agency and identity. Such ethical concerns are backed by the majority of contemporary philosophical theories on personal identity, supporting the idea that "*being free in interpreting one's self is a constitutive element of the conceptual boundaries of personal identity*" (Engelmann and Grossklags 2019).

Under the necessity assessment, this line of research on critical ethics may take on a normative dimension, where the generation of morally objectionable inferences, such as inferring sexual orientation from facial features, could be deemed as unreasonable processing of sensitive data under the GDPR and AI Act. Similarly, default approaches that directly collect sensitive personal data on the grounds of Article 10 (5) AI ACT, albeit through unethical means, shall fail the reasonableness criterion.

[23] However, correcting bias with the use of more complex proxy methods or manual ascription would typically also imply additional expertise and excessive resources especially in the case of large datasets (Andrus and Villeneuve 2022; Ashurst and Weller 2023).

Conclusion

In the face of the increasing popularity of proxy fairness approaches within the Fair-AI realm and the lack of a thorough corresponding legal framework, this paper explored aspects of Proxy Fairness under the General Data Protection Regulation and the AI Act. The legal notions of "sensitivity" and "necessity" provided the conceptual framework for the analysis.

Drawing upon article 9 (1) of the GDPR, the paper investigated the nature of data involved in Proxy Fairness, shedding light on nuanced data protection aspects surrounding proxy variables and data inferences. In doing so, it demonstrated that inferential methods are in principle not exempt from the reach of the GDPR and its extensive regime for sensitive data. Subsequently, the paper examined the lawfulness of processing sensitive data for Proxy Fairness under article 10 (5) of the AI Act. By applying the necessity requirement and comparing Proxy against Default fairness approaches with respect to their intrusiveness, effectiveness, and reasonableness, the paper delved into fundamental aspects of the legitimacy of bias detection and correction approaches that deal with sensitive data.

Ensuring legal compliance while navigating the fairness landscape, particularly when sensitive personal data are at stake, presents evidently a challenging task for AI providers and practitioners. Yet, it remains of utmost importance, given the looming threat of regulatory fines and, particularly, the risks for individuals' fundamental rights. By shedding light on the regulatory nuances involved in Proxy Fairness and guiding the lawfulness of processing sensitive data in this context, this paper sought to assist AI providers in regulatory compliance and safeguard the data protection rights of data subjects. Last but not least, the conducted analysis laid the groundwork for further scientific research at the intersection of data protection law, ethics and Fair-AI, which goes beyond an adversarial conceptualisation of fairness versus privacy.

Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Actions (grant agreement number 860630) for the project: NoBIAS - Artificial Intelligence without Bias. Furthermore, this work reflects only the author's view, and the European Research Executive Agency (REA) is not responsible for any use that may be made of the information it contains.

References

Alao, R.; Bogen, M.; Miao, J.; Mironov, I.; and Tannen, J. 2021. How Meta is working to assess fairness in relation to race in the U.S. across its products and systems.

Albers, W.; and Veit, B. 2020. *BeckOK DatenschutzR*. C.H.BECK München 2020, 37 edition. DS-GVO Art. 9 Rn. 4, beck-online.

Andriotis, A.; and Ensign, R. L. 2015. U.S. Government Uses Race Test for $80 Million in Payments.

Andrus, M.; Spitzer, E.; Brown, J.; and Xiang, A. 2021. What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, 249–260. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383097.

Andrus, M.; and Villeneuve, S. 2022. Demographic-Reliant Algorithmic Fairness: Characterizing the Risks of Demographic Data Collection in the Pursuit of Fairness. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*. ACM. ISBN 978-1-4503-9352-2.

Art29WP, A. . W. P. 2012. Article 29 Data Protection Working Party Opinion 3/2012 on Developments in Biometric Technologies.

Article 29 Data Protection Working Party. 2007. Opinion on the concept of personal data.

Article 29 Data Protection Working Party. 2016. Guidelines on the Right to Data Portability. *Data Protection Working Party*, (16/EN): 9–11. On file with the Columbia Business Law Review.

Article 29 Data Protection Working Party. 2017. Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679. Document No. 17/EN, WP251rev.01.

Ashurst, C.; and Weller, A. 2023. Fairness Without Demographic Data: A Survey of Approaches. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. New York, NY, USA: Association for Computing Machinery. ISBN 9798400703812.

Awasthi, R.; Beutel, A.; Kleindessner, M.; Morgenstern, J.; and Wang, X. 2021. Evaluating Fairness of Machine Learning Models Under Uncertain and Incomplete Information. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 206–214.

Baines, A. P.; and Courchane, M. J. 2014. Fair lending: Implications for the indirect auto finance market. Study, American Financial Services Association.

Belli, L.; Yee, K.; Tantipongpipat, U.; Gonzales, A.; Lum, K.; and Hardt, M. 2021. County-level algorithmic audit of racial bias in Twitter's home timeline. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 368–378.

Bogen, M.; Rieke, A.; and Ahmed, S. 2020. Awareness in practice: tensions in access to sensitive attribute data for antidiscrimination. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20, 492–500. New York, NY, USA: Association for Computing Machinery. ISBN 9781450369367.

Centre for Data Ethics and Innovation and Department for Science, Innovation and Technology. 2023. Enabling responsible access to demographic data to make AI systems fairer. Research and analysis report. Published on 14 June 2023.

Chen, J.; Kallus, N.; Mao, X.; Svacha, G.; and Udell, M. 2019a. Fairness Under Unawareness: Assessing Disparity When Protected Class Is Unobserved. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, 339–348. New York, NY, USA: Association for Computing Machinery. ISBN 9781450361255.

Chen, J.; Kallus, N.; Mao, X.; Svacha, G.; and Udell, M. 2019b. Fairness Under Unawareness: Assessing Disparity When Protected Class Is Unobserved. In *FAT*, 339–348. ACM.

Consumer Financial Protection Bureau. 2014. Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity: A Methodology and Assessment. Technical report, Consumer Financial Protection Bureau.

Diana, E.; Gill, W.; Kearns, M.; Kenthapadi, K.; Roth, A.; and Sharifi-Malvajerdi, S. 2022. Multiaccurate Proxies for Downstream Fairness. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. Association for Computing Machinery. ISBN 978-1-4503-9352-2.

DSK Datenschutzkonferenz. 2018. Risiko für die Rechte und Freiheiten natürlicher Personen.

Egan, K.; and v European Parliament, M. H. 2012. Egan and Hackett v Parliament. ECLI:EU:C:2019:1064. Judgment of the General Court (Fifth Chamber) of 28 March 2012.

Elliott, M. N.; Morrison, b. A.; Fremont, A.; McCaffrey, D. F.; Pantoja, P.; and Lurie, N. 2009. Using the Census Bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Services and Outcomes Research Methodology*, 9(2): 69–83.

Elzayn, H.; Smith, E.; Hertz, T.; Ramesh, A.; Goldin, J.; Ho, D. E.; and Fisher, R. 2023. Measuring and Mitigating Racial Disparities In Tax Audits. Technical report, Stanford Institute for Economic Policy Research (SIEPR).

Engelmann, S.; and Grossklags, J. 2019. Setting the Stage: Towards Principles for Reasonable Image Inferences. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, UMAP'19 Adjunct, 301–307. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367110.

Engelmann, S.; Ullstein, C.; Papakyriakopoulos, O.; and Grossklags, J. 2022. What People Think AI Should Infer From Faces. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. Seoul, Republic of Korea: ACM.

European Data Protection Supervisor. 2023. Assessing the necessity of measures that limit the fundamental right to the protection of personal data: A Toolkit.

Finck, M. 2021. *Hidden Personal Insights and Entangled in the Algorithmic Model: The Limits of the GDPR in the Personalisation Context*, 95–107. Cambridge University Press.

Gola, P.; and Heckmann, D. 2022. *Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO / BDSG*. C.H. Beck, 3 edition.

Gupta, M. R.; Cotter, A.; Fard, M. M.; and Wang, S. L. 2018. Proxy Fairness. *CoRR*, abs/1806.11212.

Hallinan, D.; and Zuiderveen Borgesius, F. 2020. Opinions Can Be Incorrect! In Our Opinion. On the Accuracy Principle in Data Protection Law. *International Data Privacy Law*, ipz025.

Imai, K.; McCartan, C.; Goldin, J.; and Ho, D. E. 2023. Estimating Racial Disparities when Race is Not Observed.

Information Commissioner's Office. Accessed 2023-11-11. A Guide to Lawful Basis.

Joinet, L.; on Prevention of Discrimination, U. S.; and of Minorities. Special Rapporteur on the Study of the Relevant Guidelines in the Field of Computerized Personal Files, P. 1988. Guidelines for the Regulation of Computerized Personal Data Files: Final Report.

Kaplan, J. 2023. predictrace: Predict the Race and Gender of a Given Name Using Census and Social Security Administration Data. https://github.com/jacobkap/predictrace. https://jacobkap.github.io/predictrace/.

Kestemont, L. 2018. *Handbook on Legal Methodology: From Objective to Method*. Intersentia Ltd. Published online by Cambridge University Press.

Keyes, O. 2018. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW).

Kosinski, M.; Stillwell, D.; and Graepel, T. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the national academy of sciences*, 110(15): 5802–5805.

Kühling, J.; and Buchner, B. 2020. *Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO / BDSG Kommentar*. C.H. Beck, 3 edition.

LeClair, B.; Parker, W.; and Young, A. 2023. Frazer Nash Report - Algorithmic Bias: A technical study on the feasibility of using proxy methods for algorithmic bias monitoring in a privacy-preserving way. Technical report, Frazer Nash Consultancy.

Malgieri, G.; and Comandè, G. 2017. Sensitive-by-distance: quasi-health data in the algorithmic era. *Information & Communications Technology Law*, 26(3): 229–249.

Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; and Galstyan, A. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.*, 54(6).

Meta Platforms and Others. 2023. Judgment of the Court (Grand Chamber) of 4 July 2023 (request for a preliminary ruling from the Oberlandesgericht Düsseldorf – Germany)– Meta Platforms Inc., formerly Facebook Inc., Meta Platforms Ireland Limited, formerly Facebook Ireland Ltd. ECLI:EU:C:2023:537, para 109.

Mitchell, S.; Potash, E.; Barocas, S.; D'Amour, A.; and Lum, K. 2021. Algorithmic Fairness: Choices, Assumptions, and Definitions. *Annual Review of Statistics and Its Application*, 8(1): 141–163.

Namsor. 2024. NamSor: Origin of Names. https://namsor.app/. Accessed: 2024-12-04.

Nowak. 2017. Judgment of the Court (Second Chamber) of 20 December 2017, Court of Justice of the European Union C-434/16.

Ntoutsi, E.; et al. 2020. Bias in data-driven artificial intelligence systems - An introductory survey. *WIREs Data Mining Knowl. Discov.*, 10(3).

Ohm, P. 2015. Sensitive Information. *Southern California Law Review*, 88.

Panel for the Future of Science and Technology, EPRS — European Parliamentary Research Service, Scientific Foresight Unit (STOA). 2020. The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence: Study.

Puri, A. 2021. A theory of group privacy. *Cornell Journal of Law and Public Policy*, 30: 477–538.

Quinn, P.; and Malgieri, G. 2020. The Difficulty of Defining Sensitive Data – The Concept of Sensitive Data in the EU Data Protection Framework. *German Law Journal*. (Forthcoming).

Schantz, P.; and Wolff, H. A. 2017. *Das neue Datenschutzrecht: Datenschutz-Grundverordnung und Bundesdatenschutzgesetz in der Praxis*. C.H.BECK.

Scheuerman, M. K.; Pape, M.; and Hanna, A. 2021. Autoessentialization: Gender in automated facial analysis as extended colonial project. *Big Data & Society*, 8(2): 20539517211053712.

Schiff, A.; Ehmann; and Selmayr. 2017. *Datenschutz-Grundverordnung DS-GVO Kommentar*. C.H.BECK, 3. auflage edition.

Schwartz, R.; Vassilev, A.; Greene, K.; Perine, L.; Burt, A.; and Hall, P. 2022. Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. Technical Report 1270, NIST Special Publication.

Seltzer, W.; and Anderson, M. 2001. The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses. *Social Research*, 68(2): 481–513.

Simitis; Hornung; and Spiecker. 2019. *Nomos Kommentar Datenschutzrecht DSGVO mit BDSG*. Nomos.

Smits, J. M. 2017. What is Legal Doctrine? On the Aims and Methods of Legal-Dogmatic Research. 207–228. Maastricht European Private Law Institute Working Paper No. 2015/06.

Solove, D. J. 2024. Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data. *Northwestern University Law Review*, 118: 1081. GWU Legal Studies Research Paper No. 2023-22, GWU Law School Public Law Research Paper No. 2023-22.

Spiecker; Papakonstantinou; Hornung; and Hert, D. 2023. *General Data Protection Regulation: GDPR Article-by-Article Commentary*. C.H.BECK. ISBN 978-3-406-74386-.

Thouvenin, F. 2021. Informational Self-Determination: A Convincing Rationale for Data Protection Law? *JIPITEC*, 12(4): 2021.

TK v Asociaţia de Proprietari bloc M5A-ScaraA. 2018. Judgment of the Court (Third Chamber) of 11 December 2019. ECLI:EU:C:2019:1064, para 47.

van Dijk, N.; Gellert, R.; and Rommetveit, K. 2016. A risk to a right? Beyond data protection risk assessments. *Computer Law and Security Review*, 32(2): 286–306.

Wachter, S.; and Mittelstadt, B. 2018. A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. *Columbia Business Law Review*, 2019(2). October 5, 2018.

Yan, S.; te Kao, H.; and Ferrara, E. 2020. Fair Class Balancing: Enhancing Model Fairness Without Observing Sensitive Attributes. In *Proceedings of the 29th ACM*.

ZestAI. 2024. zrp: Zest Race Predictor. https://github.com/zestai/zrp. Accessed: March 28, 2024.

Zhu, Z.; Yao, Y.; Sun, J.; Li, H.; and Liu, Y. 2023. Weak proxies are sufficient and preferable for fairness with missing sensitive attributes. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.