AI and GDPR

Margonis Phevos A. f3352317
Tsirmpas Dimitris f3352315

3rd April 2024

Athens University of Economics and Business
MSc in Data Science
Legal and Ethical Issues in Data Science

Contents

1 Introduction

The advent of AI technologies has led to AI systems gradually permeating daily life. From search engine indexing, to recommendation algorithms and more recently, to powerful generative AI such as Large Language Models (LLMs) like ChatGPT, humans have constantly found new use cases for AI systems. These uses however have come at the cost of constant need for more data, some of which have been sourced from citizens, with questionable motives and means.

The General Data Protection Regulation (GDPR) is a European Union law designed to protect the personal data of individuals and regulate their processing by organizations. Within public perception, it is a regulation closely associated with privacy notices in websites and user tracking via cookies, however since GDPR regulates data often used by AI systems for both training and use after deployment, the regulation inadvertently impacts those models in multiple ways.

In this short report, we explore both the impact of AI on the lives of European citizens, and the degree with which GDPR addresses that impact in a variety of cases. We begin with a short introduction to GDPR's basic interaction with AI systems within the legislation itself in Section 2.1. We then explore the impact of AI and GDPR in matters of privacy (Section 2.2), democracy (Section 2.3), discrimination (Section 2.4), and the public and private sectors (Section 2.5). We also explore the role of Transparency (Section 2.6) and Privacy-by-design (Section 2.7). Finally, we close with a comprehensive overview of this report's analysis and our recommendations in Section 3.

2 GDPR and AI

2.1 GDPR's relationship with AI systems

GDPR aims to protect the fundamental rights and freedoms of individuals, and in particular their right to the protection of personal data. It also aims to establish a uniform legal framework for the protection of personal data throughout the Union, ensuring the free movement of such data within it.

More specifically, GDPR comprises a set of principles that guide the relationship between the data controller and the data subject. The *controller* is an entity (individual, organization, or public authority) that decides why and how personal data is processed, either on its own or in partnership with others. A *subject* is an individual who can be identified, either directly or indirectly, by unique identifiers like a name, ID number, location data, or specific aspects related to their physical, mental, economic, cultural, or social identity.

It is imperative to note that **GDPR does not specifically address AI**, since the legislation was designed to be "technologically neutral". The legislation ideally will cover current and future uses of personal data and AI systems, irrespective of technology (hence why it is not supplied with a "sunset clause"). It does however greatly impact its use by regulating the data it is allowed to access, since personal data powers AI, which in turn produces new data (Committee (2018); Buttarelli (2016) as cited by Mitrou (2019)), as well as the kind of processing it is allowed to perform on data which can impact the fundamental human rights of the citizens of the European Union.

For example, the term "algorithm", essential for defining AI and data processing systems, is defined in various ways across reports and legal documents. Committee of Ministers (2020) define "algorithmic systems" as "applications that, often using mathematical optimisation techniques, [that] perform one or more tasks such as gathering, combining, cleaning, sorting, classifying and inferring data, as well as selection, prioritisation, the making of recommendations and decision making". Wagner et al. (2018) on the other hand, refer to the much more liberal definition used by Gillespie (2013): "encoded procedures for transforming input data into a desired output, based on specified calculations. The procedures name both a problem and the steps by which it should be solved". This interpretation does not confine algorithms to machine-related tasks.

GDPR does **not** contain a standard definition of algorithms, and as an extension to AI systems, only defining data processing procedures. These are defined as a set of operations performed on personal data or sets of personal data, whether automated or not, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction.

2.2 AI and personal data

The use of AI becomes problematic in regard to personal privacy in cases where the amount of data collected as input to a model can lead to tracking and profiling an individual. Profiling and tracking involve the collection, storage, and analysis of personal data, often generated automatically through internet-connected devices, to identify patterns and categorize individuals' online activities and behaviours Dinant et al. (2008); Information Commissioner (2023) as cited by Mitrou (2019). These procedures are in the modern day, unsurprisingly, carried out by AI models Kamarinou et al. (2016). Moreover, the data are often collected without the knowledge or consent of the individual.

The principles of data protection legislation, emphasize the importance of obtaining explicit consent from individuals before collecting or processing their personal data. This consent must be informed, meaning that individuals should be aware of the extent and purpose of the data collection, and it must be freely given, indicating that there should be no repercussions if an individual chooses not to consent. However, in practice, the consent mechanism is often criticized for being inadequate in the face of complex data ecosystems driven by AI and Big Data. The traditional "notice and consent" model struggles to provide meaningful control to individuals over their personal data due to the opaque nature of AI algorithms and the vast scale of data processing activities. Furthermore, the ubiquity of "consent fatigue" — where users habitually agree to terms and conditions without fully understanding them — undermines the efficacy of consent as a tool for personal data protection (Mantelero (2018)).

The existence and use of personalized information deeply concerns the individual on multiple aspects. Firstly, it may influence the person's informational self-determination (the shaping of individuals' behaviours, choices, and opportunities based on the data collected about them), their personality, and the information they are exposed to (de Andrade (2010); Conrad (2017); Hildebrandt and Gutwirth (2008) as cited by Mitrou (2019)). One concrete example could be a social media platform using AI algorithms to analyse users' browsing history, interactions, and preferences to create detailed profiles of their personalities, interests, and behaviour patterns without their explicit consent. These profiles may then be used to target them with personalized advertisements or content, changing their online experiences and potentially influencing their decisions without their awareness or control.

GDPR clearly outlines the prerequisites under which data may be processed. Under Article 6 of GDPR, lawful grounds for processing include consent of the data subject, performance of a contract, exercise of public authority, compliance with legal obligations, legitimate interests, or protection of vital interests. These purposes must be established and communicated at the latest in the phase of data collection.

Of notable importance to the above is Recital 32, which outlines that for electronic means, consent must be conscious, explicit, opt-in, and must not be disruptive to the offered service, regardless of the paid or free access to that service. Challenges include lengthy and overly-technical privacy notices, the constant advancement of systems using the data provided by the users, opaque and non-explainable models. "Big data processing" especially suffers from these challenges, since its premise is the collection of as much data as possible, and the later application of algorithms and AI models to discover patterns and correlations. Thus, informed consent may become unrealistic.

More generally, GDPR enforces compliance of the above with the **Purpose limitation**, **Data Minimization** and **Accuracy** Principles.

The **Purpose Limitation Principal** (Article 5(1)(b) GDPR) is concerned with ensuring that data is collected for specified, explicit, and legitimate purposes and not further processed in a manner incompatible with those purposes. However, AI and Big Data systems tend to reuse data that were collected for purposes different from those originally stated. For example, the use of customer contact details can be leveraged for targeted advertisements. As such, sufficient compatibility between new and original purpose is constituted when both purposes are relevant, their processing does not veer from the subjects' expectations and the proper safeguards remain in place. A more complicated situation arises when there is strong ambiguity in the relevance of the purposes. To this extent, European Parliamentary Research Service (2020) also questions the vagueness that surrounds this repurposing period, and whether the subject should be informed for every consequent reuse. Another concern is the inclusion of the subject's data in training datasets, which puts them in a vulnerable position that data breaches or general misuse entail when they are not properly anonymized or deleted once the model is constructed. Even with anonymization, this inclusion may seem inconsequential when the volume of such datasets is concerned, but they form the basis for potential discriminations or exclusion for the individual or the entire group that shares the same characteristics inferred by the ML models, thus facilitating 'Profiling'.

Complementing the Purpose Limitation, the **Data Minimization Principle**, in Article 5(1)(c), mandates that personal data collection should be adequate, relevant, and limited to what is necessary for processing purposes. This, again, contradicts the need of Big Data Analytics which often involve analyzing large datasets to uncover new correlations. Two main considerations are proposed to alleviate this tension. Firstly, the principle of minimization is suggested to be considered alongside proportionality. It doesn't prevent adding more personal data if the benefits outweigh the risks to data subjects. Future utility of processing, alongside adequate security measures like pseudonymization, could justify data retention, aligning with minimization principles. Secondly, data processed purely for statistical purposes may face less stringent requirements. Recital 162 of the GDPR highlights that statistical processing aims to produce aggregated data, not personal data, hence not intended for categorizing or making decisions about individuals. The focus is on the collective interest rather than individual data subjects.

Bolstering the above, Article 5(1)(d) emphasizes the **Accuracy Principle**, mandating that personal data must be accurate and, when necessary, updated to correct any inaccuracies. This principle is crucial, especially when personal data are used as input for AI models, where inaccuracies can lead to misrepresentation or harm to individuals by making incorrect inferences or decisions about them. In the context of Machine Learning, a distinction arises between the use of personal data solely within training sets for identifying general patterns, and their application in profiling algorithms for individual assessments. The potential misuse of training data for individualized inferences exemplifies the importance of employing anonymization or pseudonymization techniques alongside robust security measures to mitigate risks associated with data inaccuracy.

2.3 AI and democracy

Democratic procedures are susceptible to influence by AI systems. AI systems which do not follow basic constitutional principles may lead to the public disregarding democracy and the rule of law (Nemitz (2018) as cited by Mitrou (2019)).

Wagner et al. (2018) cite numerous present and future threats to democratic functions. Firstly, information mining can be used for large-scale surveillance both from government and private entities Joo and Kwon (2023). Additionally, the way information is indexed, ranked and presented by search engines may influence the public's access to information. This, combined with social media recommendation algorithms, can create "filter bubbles" or "echo chambers", which can compromise freedom of expression (see Rubel (2016) for an exploration of echo chambers, and Borgesius et al. (2016) for a rebuttal of their perceived danger). Echo chambers, when applied to political discussions, can create "ideological bubbles" which trap individuals and lead to polarization and subsequent poisoning of public dialogue.

Even more concerning is recent research and empirical examples indicating the ability of algorithms being able to meaningfully manipulate election results. Bond et al. (2012) show how voters can be persuaded to vote by simply telling them that their friends had already voted. Zittrain (2014) conducted a similar study, increasing turnout directly by 60,000 and indirectly by an additional 340,000 votes. It is evident thus, that large private entities and individual actors using them, have the ability to influence elections. This is further reinforced by platforms such as Facebook selling political advertisement to candidates, possibly restricting the ability of less-known candidates with less financial resources to get elected.

Another evolving danger for democracy is the advent of new Generative AI, especially in the context of Natural Language Processing, where models can automatically generate believable text. These algorithms can be used by "social bots", defined by Wagner et al. (2018) as "algorithmically controlled accounts that emulate the activity of human users but operate at much higher pace [. . .] while successfully keeping their artificial identity undisclosed". Generative AI has also been used to produce fake videos ("deep fakes"), which, besides undermining democratic norms and functions, and spreading disinformation and hate speech (Pawelec (2022)), can also been utilized as part of interstate conflict, as seen by examples from the Russo-Ukrainian War (Wakefield (2022); Abo Marq (2024)).

Despite these ever-present and extremely serious dangers, GDPR does not apply to most of the issues outlined above. While similar problems have been faced in the cases of radio and TV broadcasting, and largely solved by legislation, the current AI landscape remains without effective legislation in the European Union (Wagner et al. (2018)).

2.4 AI and discrimination

Since most AI models operate on discovering patterns in data, these patterns may relate to undesirable patterns and biases which human operators would consciously be aware of. Since the decisions of these models may negatively impact citizens in cases such as employment opportunities, criminal punishment or unjust treatment, biased AI systems can violate the fairness principle, the right to a fair trial, freedom from discrimination and many other fundamental rights of the European Convention on Human Rights.

Bias can be introduced to models in various stages of the AI system's lifecycle, including in its design phase (e.g. models allowed to access information such as race, age, and gender), in the data provided during training (a model learning discriminatory patterns from human decisions) or in reinforcement learning during the system's deployment (such as Microsoft's failed "Tay" Experiment (Kraft (2016))).

This issue can gravely affect whole communities or societal groups. One such example would be predictive policing, where data analysis and algorithms are used to anticipate where crime is likely to occur, aiming to deploy resources more efficiently. Predictive policing relies on historical crime data, which can reflect existing biases and disparities in law enforcement practices (Wagner et al. (2018)). Algorithms trained on this data may inadvertently perpetuate and exacerbate these biases, leading to disproportionate surveillance and enforcement in marginalized communities. Consequently, certain groups, such as racial minorities or low-income neighbourhoods, may be unfairly targeted and subjected to increased police presence and scrutiny. This can erode trust between law enforcement and the community, exacerbate tensions, and perpetuate cycles of suspicion and fear.

Another case is that of the use of AI for decision-making in judicial decisions. AI systems can be used for predicting the chance of an individual recommitting a crime after incarceration (recidivism) and aiding judges by predicting the result of a trial (although the accuracy of those systems is still fairly low (Wagner et al. (2018))). As stated above, AI systems are vulnerable to bias, lack transparency, and can lead to errors, resulting in unfair and inaccurate rulings. An example of this would be the COMPAS controversy (Larson et al. (2016)), where a system predicting the chance of recidivism was found to be both lacking useful insights, and discriminating against people of colour in the US, leading to a large amount of wrongful, biased sentences.

Additionally, after a string of terrorist attacks in the US and Europe, many politicians have demanded social media platforms to publish predictions on individuals likely to commit terrorist attacks, which may infringe on Articles 5 (arbitrary deprivation of liberty), 6 (presumption of innocence) and 7 (no punishment without law) of the ECHR (Wagner et al. (2018)).

The common denominator in almost all cases is the public perception of AI models as fair, objective and with much fewer limitations than in reality. In reality, most AI systems relying on Machine Learning techniques are designed to identify patterns in data and do not have cognition or notions of "bias" and "discrimination". These models often perpetuate human patterns, some of which may be linked to false information, stereotypes, and intolerance.

2.5 AI in the public and private sectors

AI, and therefore the collection and processing of personal data, are essential in the way that numerous fields of the European and global economy operate. This is acknowledged by Committee of Ministers (2020) where it is stated that AI (and generally algorithmic) models have revolutionized fields like medical diagnostics and transportation, enabling global information sharing and new forms of cooperation and are, as a result, deeply integrated into various aspects of modern human life. However, they also pose significant human rights challenges, noting the reliance on data aggregation and analysis for system functionality.

According to the same report, AI systems need not necessarily operate on personal data in order to negatively affect humans with their use (hence possibly limiting GDPR's effectiveness in these cases). Additionally, AI systems may become the target of adversarial or cyber attacks, which are enabled by the scale and nature of the data that are being processed. The potential of these systems to infringe on human rights and threaten democratic values and processes are enough justification for EU member states to outline and enforce obligations on the private sector. One such obligation, according to the report, is the protection of individuals who take steps to evade capture by automated data collection systems.

Private sector actors are also faced with difficult dilemmas. A practical problem born from the vast amount of data and users on popular platforms is the issue of automatic complaint and appeal systems.

To adequately describe the problem we must emphasize that current machine learning approaches lead to the development of algorithms that by design cannot be understood by the human mental model. Deep learning techniques are often characterized by opaqueness, complexity, and lack of explainability, making assessing accountability and risk especially problematic. Indeed, whether these systems can be made explainable to humans is an activate debate in the academic community Wagner et al. (2018).

Thus, automated decision-making processes pose challenges for individuals seeking effective remedies due to the opacity of decisions, lack of consent regarding data usage, and limited awareness of decision impacts. While historically human decision-makers allowed for reasoned and individual decisions, the increasing reliance on algorithmic techniques in complaints presents questions about the observance of Article 13 of the European Convention Wagner et al. (2018), which cites that those whose rights have been violated shall have an effective remedy before a national authority. The right to an effective remedy necessitates a judicial challenge option, but suggestions for government supervision in consumer-corporation negotiations have also emerged. Automated techniques in privacy matters may also raise issues of secret surveillance and notification absence, hindering the effectiveness of remedies according to the European Court of Human Rights.

Another similar problem comes from content moderation. Most major private actors have invested in AI systems which locate and delete content that violates their platform's Terms of Service (ToS). When used excessively and at scale, however, automatic content removal can violate the individual's freedom of expression, especially since AI systems are inadequate to differentiate between actual hate speech and cultural context or humour. Thus, these systems may remove speech that are healthy for public debate Wagner et al. (2018). Again, these concerns are generally not targeted by GDPR.

GDPR however does address a very important issue, that being the targeting of employees within the private sector with AI systems. Wagner et al. (2018) note that human resource management processes which utilize AI systems for decision making and evaluating employees can lead to automated feedback loops, discrimination over race (Bertrand and Mullainathan (2003)), and class and gender (Altonji and Blank (1999); Goldin and Rouse (1999)). Additionally, monitoring and ranking employees violate worker's rights under Articles 8 and 10 (Voorhoof et al. (2013)). GDPR prohibits the collection of data for these processes (e.g. GDPR 9(2)(c)) with only essential exceptions related directly to the purposes declared by the private sector entity.

The situation becomes more complicated on the use of AI on the public sector, where these processes are increasingly used by States and government services (van Haastert (2016)) and where the expectation for public accountability is higher. The State may use AI systems for public services and policy delivery, for which the citizen may not have the possibility of opting out without serious negative consequences. Furthermore, this responsibility may not only be on the State, but on private sector providers, if the system is outsourced, who in turn depend on other providers. Even in the organizations themselves, the development of AI systems occupies a significant and diverse group of stakeholders outlined as "software designers, programmers, data sources, data workers, proprietors, sellers, users or customers, providers of infrastructure, and public and private actors and institutions" (Committee of Ministers (2020)).

Lastly, it is worth noting that, unlike other systems, suddenly withdrawing established AI systems may lead to significant disruptions, depriving, in the worst case, essential services to individuals and societal groups. Therefore, Committee of Ministers (2020) encourage States to place contingencies to ensure the availability of these systems irrespective of commercial success.

Legislative restrictions lead to many believing that compliance with GDPR incurs significant costs and complexities, requiring resources for data management and algorithm transparency. They state that limitations on data use may hinder AI effectiveness, impacting innovation and competitiveness, while uncertainty and legal risks surrounding GDPR enforcement add further obstacles for businesses. The Committee of Ministers Committee of Ministers (2020) rebukes this idea, stating that despite claims of rationalization and accuracy gains, AI systems are prone to errors and biases, with larger datasets not necessarily guaranteeing higher accuracy rates, thus increasing the risk of false positives and negatives and further interfering with human rights. Thus, effective regulation may not harm the development of AI in the majority of cases. Wagner et al. (2018) point to similar restrictions were placed on industries such as slot machines and quality control and assurance frameworks in the production and manufacturing industry for cases of successful employment of such regulations. They note however the difficulty of implementing similar schemes to automated data processing techniques and AI systems.

2.6 Transparency and accountability

Studying the personal data protection laws under GDPR and their intersection with AI, a critical tension emerges between the data-intensive requirements of AI and the need for transparency and accountability.

Transparency is a multifaceted term. In the context of understandability, it means that the controller is responsible to present any information relevant to the public concisely, easily accessible and in sufficiently simple terms (Article 5(1)(a) Recital 58). In Articles 13 and 14, it is defined as the right of the individual to know the reasons and the legality behind the processing of their personal data. In the context of explainability, it is defined as one or more sentences that describe the process followed and the reasons behind a final decision Preece (2018). Recent advancements in AI have provided more efficient and cost-effective ways to automate decision-making processes, increasing the reliance on such systems. These automated decisions, that exclude any human intervention, may come in the form of predictions, classifications, or recommendations and can impact a subject's life especially when they concern sensitive information such as monetary, professional, or medical. Machine Learning (ML) algorithms, utilized to derive these automated decisions, are often so complex that act as 'black boxes' even to their designers. This exemplifies the need to develop a 'scientific' model that describes the inner functioning of AI. Additionally, Tsakalakis et al. (2021) propose that simply explaining the lines of code is not sufficient, but other factors such as the training data and the deployment methods should be disclosed to achieve full transparency.

This lack of clarity and the reliance on automated systems without adequate explanation undermine trust, which is foundational to the acceptance and ethical use of AI technologies. Trust, in this context, is based on openness and the accessibility of information regarding the use of personal data, the risks involved, and the safeguards in place. However, the drive for transparency must be balanced against the need to protect trade secrets, intellectual property rights, and state secrets, highlighting the complexities of achieving transparency in practice. Although there is a great debate whether, and to what extent, AI systems are explainable, recent studies indicate that there is potential for interpretability (Främling (2020); Tsakalakis et al. (2021)). By making the workings of AI systems more transparent, responsibility for decision-making failures or biases can be more easily traced and assigned, thus promoting accountability (House of Commons Science and Technology Committee (2016)).

This accountability framework is essential for building public trust in AI technologies and ensuring that they are deployed in a manner that respects individuals' rights and freedoms. Data controllers are mandated to implement specific measures to adhere to data protection requirements, with accountability tools like Data Protection Impact Assessments (DPIA) and privacy-by-design system architectures.

DPIA is a process required (under GDPR, Article 35) to identify and minimize the data protection risks of a project. This approach encompasses data protection by design and default, the appointment of data protection officers, data breach notification protocols, and prior consultation requirements. DPIAs are particularly important in AI and machine learning applications where processing — of often sensitive data — is likely to pose a high risk to individuals' fundamental rights and freedoms (such as profiling and discrimination).

Stakeholder engagement is also required by GDPR, to improve the DPIA process by including input from data subjects and their representatives to identify risks and ensure data processing aligns with their expectations. Additionally, the involvement of data processors (agents which process personal data on behalf of the controller) especially when processing is outsourced, is crucial in supporting controllers with DPIA, underlining the need for cooperation in addressing privacy and rights concerns in data processing. However, a significant limitation of DPIAs is the inherent difficulty in accurately predicting how new technologies will be applied and the potential delay in fully understanding their implications.

2.7 AI and privacy-by-design

Privacy by Design (PbD) refers to the proactive embedding of privacy into the design, operation, and management of IT systems, networks, and business practices. PbD advocates for privacy as a default setting and encourages the minimization of personal data collection, ensuring that privacy measures are built into technology from the beginning, rather than retrofitted as an afterthought (Schaar (2010)).

Two fundamental approaches form the intersection on which AI and privacy by design meet, the right-based and the risk-based approach. The right-based approach aims to defend the fundamental rights to privacy and data protection (like dignity, freedom of thought, choice, and expression) on the societal level. On the individual level, GDPR grants practical control over the flow of information that concerns the subjects. The risk-based approach, on the other hand, aims to safeguard the public against biases and fallacies introduced by automated decisions. While this approach of potential over-regulation could have a stifling effect on innovation, particularly for smaller organizations, the EU suggests that risk-prevention measures should be proportional to the size of the controller.

In designing an AI system that integrates 'Data protection by design and by default' as stipulated in GDPR's Article 25, several key legal obligations must be considered. Article 25(1) mandates that controllers implement mechanisms compliant with data protection regulations from the outset of the development process. Article 24 requires the controller to be in a position to prove that these mechanisms are present, and they are reviewed and updated as necessary, since they concern AI systems which change and evolve regularly. The principles of data minimization are emphasized in Article 25(2), which dictates the scope, variety, and retention period of personal data utilization. Complementary to these articles, codes of conduct and certification mechanisms can aid in the demonstration of compliance.

Article 35 mandates that a DPIA be conducted prior to any endeavor that may fall under the 'high-risk' category. In the case of a positive result, the controller is obliged to contact the supervisory authority for advice. The authority, in return, must identify the discrepancies, order their correction, and in the event that compliance is impossible, temporarily or permanently ban the system. To further mitigate such eventualities, DPOs are assigned when operation or monitoring of sensitive data is performed on a large scale, which is a very common phenomenon in AI.

3 Conclusion

In this report, we studied the interactions between AI and GDPR, highlighting how AI's data-intensive nature conflicts with GDPR's regulations. It examines AI's influence on democracy, emphasizing its potential to undermine democratic values by affecting public opinion and elections. Additionally, the issue of discrimination stemming from the use of AI is addressed, showing how biases within AI can infringe on human rights. Our analysis also covers the implications of AI in both public and private sectors, emphasizing that its ethical, legal, and social aspects should be considered prior to use, as well as the importance of transparency and accountability in the development of AI, which advocates for privacy-by-design to embed privacy in AI's architecture from the outset, thereby aligning technological applications with GDPR's regulatory framework.

We find that although GDPR is not an AI-specific legislation, its requirements significantly affect how AI systems can process personal data, enabling the protection of individual rights and privacy in the context of AI. However, there are significant gaps in their protection from the negative impacts of AI, such as the influence of democratic functions, in discriminatory processes and surveillance applications, highlighting the need for a more thorough examination of their intersection.

4 Bibliography

W. Abo Marq. Deepfake video of ukraine's national security secretary alleges ukraine involvement in moscow attack, March 2024. URL https://misbar.com/en/factcheck/2024/03/25/deepfake-video-of-ukraines-national-security-secretary-alleges-ukraine-involvement-in-moscow-attack.

J. G. Altonji and R. M. Blank. Chapter 48 race and gender in the labor market. volume 3 of *Handbook of Labor Economics*, pages 3143–3259. Elsevier, 1999. doi: https://doi.org/10.1016/S1573-4463(99)30039-0. URL https://www.sciencedirect.com/science/article/pii/S1573446399300390.

M. Bertrand and S. Mullainathan. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. Working Paper 9873, National Bureau of Economic Research, July 2003. URL http://www.nber.org/papers/w9873.

R. Bond, C. Fariss, J. Jones, et al. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489:295–298, 2012. doi: 10.1038/nature11421. URL https://doi.org/10.1038/nature11421.

F. J. Z. Borgesius, D. Trilling, J. Moeller, B. Bodó, C. H. de Vreese, and N. Helberger. Should we worry about filter bubbles? *Information Systems: Behavioral & Social Methods eJournal*, 2016. URL https://api.semanticscholar.org/CorpusID:52211897.

G. Buttarelli. Privacy in an age of hyperconnectivity. Keynote speech at the Privacy and Security Conference 2016, Rust am Neusiedler See, Austria, November 2016.

C. C. Committee. Council of europe consultative committee report on artificial intelligence and data protection, September 2018.

Committee of Ministers. *Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems.* April 2020. Adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies.

C. Conrad. Künstliche intelligenz — die risiken für den datenschutz. *Datenschutz und Datensicherheit - DuD*, 41: 740–744, 12 2017. doi: 10.1007/s11623-017-0870-4.

N. N. G. de Andrade. Data protection, privacy and identity: Distinguishing concepts and articulating rights. In *PrimeLife*, 2010. URL https://api.semanticscholar.org/CorpusID:17381091.

J.-M. Dinant, N. Lefever, C. Lazaro, Y. Poullet, and A. Rouvroy. Application of convention 108 to the profiling mechanism: some ideas for the future work of the consultative committee (t-pd): final version. 2008. URL https://api.semanticscholar.org/CorpusID:115046048.

European Parliamentary Research Service. The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence. Study, European Parliament, 2020. URL http://www.europarl.europa.eu/stoa/. PE 641.530, QA-QA-02-20-399-EN-N.

K. Främling. Explainable ai without interpretable model, 2020. URL https://doi.org/10.48550/arXiv.2009.13996.

T. Gillespie. *The Relevance of Algorithms.* 01 2013. ISBN 9780262525374. doi: 10.7551/mitpress/9780262525374.003.0009.

C. Goldin and C. E. Rouse. Orchestrating impartiality : The impact of “ blind ” auditions on female musicians by. 1999. URL https://api.semanticscholar.org/CorpusID:261258493.

M. Hildebrandt and S. Gutwirth. Profiling the european citizen, cross-disciplinary perspectives. 2008. URL https://api.semanticscholar.org/CorpusID:44785133.

House of Commons Science and Technology Committee. Robotics and Artificial Intelligence: Fifth Report of Session 2016–17. Report HC 145, House of Commons, London, 10 2016. URL https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf.

Information Commissioner. *Information Commissioner's Annual Report and Financial Statements 2022/23*. Number 1440. HC, July 2023.

M.-H. Joo and H.-Y. Kwon. Comparison of personal information de-identification policies and laws within the eu, the us, japan, and south korea. *Government Information Quarterly*, 40(2):101805, 2023. ISSN 0740-624X. doi: https://doi.org/10.1016/j.giq.2023.101805. URL https://www.sciencedirect.com/science/article/pii/S0740624X23000059.

D. Kamarinou, C. Millard, and J. Singh. Machine learning with personal data. *ERN: Other European Economics: Microeconomics & Industrial Organization (Topic)*, 2016. URL https://api.semanticscholar.org/CorpusID:63785275.

A. Kraft. Microsoft shuts down ai chatbot after it turned into a nazi, March 2016. URL https://www.cbsnews.com/news/microsoft-shuts-down-ai-chatbot-after-it-turns-into-a-nazi/.

J. Larson, S. Mattu, L. Kirchner, and J. Angwin. How we analyzed the compas recidivism algorithm. *ProPublica*, May 2016. URL https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

A. Mantelero. Ai and big data: A blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review*, 34(4):754–772, 2018. ISSN 0267-3649. doi: https://doi.org/10.1016/j.clsr.2018.05.017. URL https://www.sciencedirect.com/science/article/pii/S0267364918302012.

L. Mitrou. *AI and GDPR Study Prof MItrou Mar19 FINAL*. 04 2019.

P. Nemitz. Constitutional democracy and technology in the age of artificial intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376, 2018. URL https://api.semanticscholar.org/CorpusID:53501707.

M. Pawelec. Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital society*, 1(2):19, 2022.

A. Preece. Asking 'why' in ai: Explainability of intelligent systems – perspectives and challenges. *Intelligent Systems in Accounting, Finance and Management*, 25(2):63–72, 2018. doi: https://doi.org/10.1002/isaf.1422. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/isaf.1422.

A. Rubel. The black box society: The secret algorithms that control money and information, by frank pasquale. cambridge: Harvard university press, 2015. 320 pp. isbn 978–0674368279. *Business Ethics Quarterly*, 26:568 – 571, 2016. URL https://api.semanticscholar.org/CorpusID:151540271.

P. Schaar. Privacy by design. *Identity in the Information Society*, 3(2):267–274, 2010. ISSN 1876-0678. doi: 10.1007/s12394-010-0055-x. URL https://doi.org/10.1007/s12394-010-0055-x.

N. Tsakalakis, S. Stalla-Bourdillon, L. Carmichael, T. D. Huynh, L. Moreau, and A. Helal. The dual function of explanations: Why it is useful to compute explanations. *Computer Law & Security Review*, 41:105527, 2021. ISSN 0267-3649. doi: https://doi.org/10.1016/j.clsr.2020.105527. URL https://www.sciencedirect.com/science/article/pii/S0267364920301321.

H. van Haastert. Government as a platform: Public values in the age of big data. Paper for the OII IPP 2016 Conference, 2016.

D. Voorhoof, P. Humblet, F. Dorssemont, K. Lörcher, and I. Schömann. The right to freedom of expression in the workplace under article 10 echr. 2013. URL https://api.semanticscholar.org/CorpusID:155829806.

B. Wagner, W. Schulz, K. Turk, B. de la Chapelle, J. Hörnle, T. Kersevan-Smokvina, M. Kettemann, D. Nieland, A. Nedyak, P. Podvinskis, T. Schneider, S. Stalla-Bourdillon, and D. Voorhoof. *Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications.* Council of Europe, 2018.

J. Wakefield. Deepfake presidents used in russia-ukraine war, March 2022. URL https://www.bbc.com/news/technology-60780142.

J. L. Zittrain. Engineering an election. *Harvard Law Review Forum*, 127:335, 2014. Harvard Public Law Working Paper No. 14-28, 7 Pages Posted: 23 Jun 2014 Last revised: 26 Jul 2015.