LARGE LANGUAGE MODELS: LEGAL ISSUES UNDER GDPR AND THE AI ACT

Almujtba Izzeldin Elbashir Suliman

Student Number: s329637

Department of Regional and Urban Studies and Planning

Politecnico di Torino

04TXHTD: Data Ethics and Protection

Prof. Gianmaria Federico Ajani

Prof. Marco Giraudo

June 30, 2024

Abstract

Large Language Models (LLMs) have become game changing technologies attracting attention in the media and significant investments. However, the swift progress and implementation of these models give rise to issues especially within the regulatory framework of the European Union. This study delves into the hurdles presented by LLMs under the General Data Protection Regulation (GDPR) and the proposed Artificial Intelligence Act (AI Act). It delves into data protection matters such as consent mechanisms, transparency requirements and the right to an explanation. Additionally it assesses how the AI Acts risk based strategy impacts LLM classification and governance. By combining existing research and regulatory directives this paper contributes to discussions on striking a balance between technological progress and individual rights while safeguarding society. The research emphasizes the necessity for structures and increased collaboration, among technologists, policymakers and legal professionals to navigate the ever evolving realm of AI technologies.

Keywords: Large Language Models, LLMs, GDPR, AI Act data protection, algorithmic regulation, artificial intelligence law, digital ethics, innovation regulatory compliance, AI governance

1.0 Introduction

Large Language Models (LLMs) are among the most striking advancements in Artificial Intelligence (AI) which produce human-like text and capable of a broad range of language tasks (Brown et al., 2020). The growing use of LLMs in AI applications raises difficult legal questions about their impact, especially within the regulatory framework of the European Union (EU). This paper will analyze the legal implications arising from LLM phenomenon under two specific legislations; General Data Protection Regulation (GDPR) and the upcoming Artificial Intelligence Act (AI Act).

The GDPR was enacted in 2018 to protect personal data privacy and has particular rules on how data should be processed (European Parliament and Council, 2016). Multiple new issues around GDPR compliance arise when using very large amounts of data for training and operation of LLMs. These include:

1. Data collection and processing practices for LLM training
2. Applicability of consent mechanisms to LLM data usage
3. Integration within LLMs' rights for subjects like erasure
4. Transparency and explainability relating to its decision-making processes by means of LLMs (Wachter et al., 2017)

At the same time, the Artificial Intelligence Act (AI Act) which was proposed in 2021 has set the objectives of regulating AI systems according to their risk level (European Commission, 2021). These queries that arise from classification and regulation of LLMs under the new framework include:

1. How LLMs Should be Classified under Risk Based Approach in AI Act
2. Conformity to Transparency and Documentation Rules

3. Ensuring Human Oversight Over LLM applications
4. Dealing with Possible Biases and Discriminatory Outputs (Zuiderveen Borgesius, 2020)

The aim of this research paper is to offer an elaborate analysis of these legal concerns by examining how GDPR's existing framework and AI Act's envisaged design apply to peculiar qualities of LLMs. By focusing on these two statutes, we seek to highlight what is currently happening in terms of law and possible future challenges from a regulatory perspective (Veale & Zuiderveen Borgesius, 2021).

Section 1: LLMs Under GDPR

1.1 GDPR Principles and their Relevance to LLMs

The General Data Protection Regulation (GDPR) consists of a comprehensive approach to data protection which has been founded on some principles that bring about unique challenges as applied in the context of Large Language Models (LLMs).

This means that all personal data must be processed in this way: legal, ethical, and clear to the person whose information is being dealt with. For LLMs this poses a big problem. However, due to the sheer diversity and flexibility of this type of data used for training such models, it becomes difficult to guarantee every bit of it has been gotten and processed legally. Additionally, because LLMs are often complex processes, they do not meet the transparency requirement leading to opaqueness in decision making (Kamarinou et al., 2016).

The purpose limitation and data minimization principles provide that data shall be collected for specified, explicit, and legitimate purposes and not further processed in a

manner incompatible with those purposes. Data shall be adequate, relevant, and limited to what is necessary about the purposes. LLMs are meant to be multi-purpose and need massive datasets for effective working. Inherent tension comes about with these GDPR principles (Bietti, 2020).

The accuracy principle states that personal data must be kept in a way that is accurate and up-to-date. For LLMs, which are trained on static datasets, ensuring continuous accuracy presents a significant technical challenge. Since the fundamental world changes, knowledge in the model becomes more outdated and leads to potentially misleading or flawed outputs (Stalla-Bourdillon & Knight, 2017).

1.2 The Data Collection Puzzle

LLMs require massive amounts of data to train on, a feat primarily done through web-scraping or the use of enormous text corpora. This undertaking presents several challenges within the GDPR framework. First, there is difficulty in establishing a lawful basis for processing such vast amounts of data. While 'legitimate interests' may be argued to apply, data collection about the scale of LLM training could be considered disproportionate and possibly not balance the test required for this basis. The second point involves the indiscriminate nature of data collection for LLMs. Inevitably, they process particular category data that includes information regarding people's racial or ethnic origin, political opinions, and health, among others. The GDPR raises the bar for processing such sensitive data—requiring explicit consent or fulfilling specific conditions. This is extremely hard, or just practically impossible, in large-scale web scraping during LLM training. Also, the data minimization principle goes against the data-hungry nature of LLMs. Since they profit from more data, the systems encourage collecting as much data as possible, even if it is

unnecessary —an evident contradiction with the GDPR's tenet of minimal data use: more and more vs. less data (Bietti, 2020).

1.3 The Transparency and Consent Paradox

The GDPR places a significant emphasis on transparency and consent in data processing. The requirements are onerous for LLMs. For one, valid consent to process data in training LLMs is inherently problematic in many ways. The GDPR requires that consent should be specific, informed, and unambiguous. However the very scope of LLM objectives, and most often, its inherently unpredictable outputs, does not allow clear and specific information at the time of data collection about how exactly the data will be used in the future. The possible uses of an LLM are so varied and developing that it is almost impossible to provide data subjects with the information they need to appreciate the scope of processing fully (Selbst & Powles, 2017). In addition, the GDPR provides for the possibility of a person to withdraw consent at any time. Technically, this is difficult when data has already been included in the training set of an LLM.

The fact that it is interconnected means that the model cannot have specific data points removed without retraining the entire model in most cases or, if not, is almost impossible (Villaronga et al., 2018). Transparency is another area where LLMs find it hard to meet the laid-down requirements of the GDPR. According to the regulation, data controllers must provide clear information on how personal data are used. However, the complexity and sometimes opaque character of LLM decision-making processes search for true transparency, a challenging endeavor (Wachter et al., 2017).

1.4 The Right to Explanation: A Square Peg in a Round Hole

The GDPR safeguards concerning automated decision-making and the so-called "right to explanation" poses numerous challenges in terms of suitability to LLMs. LLMs are highly complex and opaque, meaning it is often impossible to give meaningful explanations of their decision-making processes. Usually, these models work so opaquely that even the creators of such models may not understand how they result in specific outputs. This type of opacity is opposed to the GDPR's requirement that such systems should provide "meaningful information about the logic involved" in automated processing (Kaminski, 2019).

Additionally, it is not always clear whether the outputs of LLMs routinely fall under the definition of "automated individual decision-making" according to the GDPR. More precisely, LLMs serve, in most cases, as tools intended to assist human beings in making decisions rather than to make decisions on their own. This does blur the boundary between human and machine decision-making, applying the right to explanation a bit more complexly. (Selbst & Powles, 2017).

1.5 Data Subject Rights: A Technical Nightmare

The right to implement the rights of the data subjects is particularly complex when viewed from the standpoint of LLMs.

Of all the rights, none is more problematic than the right to access that allows individuals to know if their personal information is being processed and get a copy of it. It’s technically hard if not impossible to identify specific personal data within an LLM’s parameters. The model used becomes so mixed up and transformed that extracting individual points becomes impractical (Veale et al., 2018).

Again, both rectification and erasure rights pose enormous technical challenges too. Correcting wrong personal information or deleting particular points from trained large language models (LLMs) without affecting overall performance is not easy at all. In some cases, entire retraining may be required which consumes resources and could be detrimental for model performance as well (Villaronga et al., 2018).

Similarly, the right to portability cannot be easily implemented in relation with LLMs because they are designed based on interconnectedness between datasets among other things which makes isolation and exporting single data points meaningless.

In summary, GDPR has given a good foundation for protecting our data but we need better understanding about how these laws can work together with AI technology like Large Language Models (LLMs). This comes because there are many conceptual limitations when applying such rules in practice especially considering such factors unique characteristic nature complex processing mechanisms high volumes involved difficulty giving explanations inputs provided outputs obtained from them etcetera. Therefore, people need to keep talking among themselves across different fields including legal experts, technologists and policymakers who should come up with solutions that are useful for this kind of area since LLMs have become part of every walk of life nowadays .

Section 2: LLMs under the AI Act

2.1 The AI Act: A New Paradigm for AI Regulation

The proposed artificial intelligence act (AI Act) is the European Union's attempt to create a legal framework for regulating artificial intelligence systems throughout its member states. GDPR was designed with data protection in mind but this new legislation takes a

wider view by considering other aspects such as safety which can be threatened by AI (Veale & Zuiderveen Borgesius, 2021).

The Act adopts a risk-based approach where different levels of risks are identified and specific measures put across each level. The four categories used in classifying AI systems under this act include unacceptable risk, high risk, limited risk and minimal risk. The levels are meant to ensure that any harm caused by an AI system corresponds with the level it falls under thus imposing relevant obligations on users or developers depending on their position relative to these categories (Ebers, 2021).

2.2 Classifying LLMs within the AI Act Framework

Categorizing large language models within the framework laid out by the EU's proposed artificial intelligence act (AI Act) can be quite challenging given their diverse applications across various sectors. There are many instances where such models may fall into high-risk category especially if they are employed in critical areas like health care delivery systems, educational institutions etcetera

For example, if an LLM is used to help diagnose medical conditions or plan treatments, it will probably be regarded as high-risk because it could seriously affect patients' health and safety. Likewise, an LLM employed in schools for student assessment or during recruitment processes to sift through applicants may also raise the level of risk associated with this technology since it has the power to significantly impact people's life chances (Smuha et al., 2021).

Nonetheless, things get trickier when it comes down to general purpose LLMs that can be applied across different fields. The current wording of the AI Act does not offer any

hints on how these systems should be classified leaving space for interpretation and potential regulatory holes (Veale & Zuiderveen Borgesius, 2021).

2.3 Transparency and Documentation Requirements

As for AI systems that are classified as high-risk, the AI Act imposes strict transparency and documentation requirements. These duties are very difficult for LLMs because of their complexities and being often opaque.

High-risk AI system providers must produce and keep detailed technical documentation as required by the AI Act. Such documentation should give information on how the system was developed, training methods applied, performance metrics or evaluation methods used among other things relating to it. In this regard, LLMs can be challenging to provide such a level of comprehensive documentation due to their complex architectures and large-scale data sets (Ebers, 2021).

The Act also requires that logs be automatically recorded during operation by high-risk AI systems. For LLMs, which consume huge volumes of data and generate results based on complicated internal representations, there is a great deal of complexity in making this logging meaningful or interpretable.

Moreover, high-risk AI systems must offer capabilities enabling automatic event logging while operating as per the Act. However, in case of LLMs which process extensive amounts of data and generate outputs grounded in intricate internal representations; implementation of such logging capabilities that have meaning and provide interpretability comes with notable technical barriers.

This requirement is consistent with the growing demand for explainable AI but poses specific problems for LLMs under the AI Act. Oftentimes, simple answers cannot account for

highly sophisticated decision-making processes within these models hence giving clear information to users about how results are generated becomes quite difficult (Weidinger et al., 2021).

2.4 Human Oversight and Control

According to the AI Act human oversight is vital when considering high risk artificial intelligent systems therefore they ought to be designed in a manner than humans can effectively oversee them (Bommasani et al., 2021). The purpose behind this requirement is avoiding automation bias while allowing humans ability to disengage or take control over an AI system in case of necessity.

Due to the autonomous nature of LLMs and their ability to produce output very fast, it is hard to implement effective human oversight. Most advanced LLMs possess a wealth of knowledge and have complex reasoning capacities that may sometimes surpass those of human supervisors making it difficult for humans to effectively monitor or intervene in the system's operations (Bommasani et al., 2021).

Additionally, there is an issue raised by the requirement in the Act about a proper balance between AI autonomy and human control under such circumstances. When these LLMs are being deployed as decision-support tools, how much involvement should be given to humans and what mechanisms should be put in place for them can interfere (Smuha et al., 2021).

2.5 Addressing Bias and Discrimination

According to the AI Act, preventing biases and discriminations from being perpetuated or amplified by AI systems is more important than other things. This is critical

for LLMs especially since they have been found to mirror and sometimes exaggerate societal biases inherent in their training data.

Moreover, the Act prescribes that training, validation, and testing datasets should be relevant/representative, error-free, as well as complete. However, complying with these requirements can be extremely difficult for LLMs because they are often trained on large corpora of Internet text. It remains virtually impossible to thoroughly vet or curate all this data due its scale while carefully curated datasets can become outdated or biased overnight courtesy of language changes among other dynamic characteristics associated with society (Bender et al., 2021).

There is also a provision in the Act stipulating that high-risk AI systems should be developed and designed in such a way that they do not issue any form of discriminatory outputs or at least minimise them. When it comes to LLMs, which can create human-like text on almost everything, achieving non-discrimination across all possible results is no small task. It entails both careful data curation and the development of robust debiasing techniques as well as continuous monitoring and adjustment of model outputs (Weidinger et al., 2021).

2.6 Compliance Challenges and Innovation Concerns

With respect to LLM developers and deployers, the AI Act's stringent requirements present significant compliance challenges. Given this factor, these models' development, training, and deployment may require substantial changes due to extensive documentations required as well as transparency and oversight demands placed by the Act.

However, it is feared that these regulatory burdens might stifle progress in AI development with the smaller firms and research institutions in AI not having enough resources to fully meet the requirements of the Act. This would potentially mean that LLM

development may get concentrated among a few major technology companies, hence narrowing down diversity and competition among them (Ebers, 2021).

On the other hand, advocates for AI act argue that these regulations are important in ensuring responsible creation of AI technologies as well as protection of fundamental rights. According to Smuha et al. (2021), through defining explicit rules and standards, the Act may actually spur innovation by providing an environment for trusted and steady growth of AI.

To summarize, the AI Act poses challenges but also provides prospects for LLMs' development and deployment. While aligning with key ethical principles such as transparency, oversight, non-discrimination etc., nevertheless practical compliance with these conditions will require tremendous effort and creativity. Therefore, it is crucial for policymakers, industry stakeholders engaged in Artificial Intelligence (AI) activities as well as researchers in this field to engage in continuous dialogue since this field is growing at a very high speed.

Section 3: Rethinking (Analysis)

3.1 GDPR versus AI Act Approaches to LLMs

The General Data Protection Regulation (GDPR) and the proposed Artificial Intelligence Act (AI Act) represent two significant regulatory frameworks that impact the development and deployment of Large Language Models (LLMs) in Europe. Both laws aim at safeguarding individuals' rights while promoting responsible AI approaches; however there are differences between them.

In terms of content focus GDPR deals mainly with data protection and privacy where it promotes guidance on principles like lawfulness, fairness or transparency relating to processing personal data (Voigt & Von dem Bussche, 2017). Contrarywise, AIAct takes on a

larger risk-based model when regulating artificial intelligence by categorizing AI systems into different ones depending on the safety and fundamental rights they may affect (Veale & Zuiderveen Borgesius, 2021).

For LLMs, this difference in focus creates a complex regulatory landscape. The GDPR is mainly concerned with issues like data collection practices, consent mechanisms and rights of data subjects. However, the AI Act introduces additional elements such as transparency in AI decision-making processes, human oversight and bias mitigation.

3.2 Potential Conflicts and Gaps Between Regulatory Frameworks

Simultaneous application of both GDPR and AI Act regarding LLMs illustrate potential inconsistencies between them.

One significant area of tension arises from the approach to transparency and explainability. Article 22's right to explanation within the GDPR has seen arguments about its scope as well as how it can be implemented practically (Wachter et al., 2017). Conversely, for high risk AI systems, more specific requirements have been introduced by the AI Act concerning transparency and explainability. This variance might puzzle LLM developers regarding what explanations are permissible under these regimes.

Data minimization is another potential conflict area. Under the GDPR principle of data minimization (Article 5(1)(c)), only that which is necessary for the purpose should be collected. However, LLMs tend to perform better with a bigger data set, which might appear to contradict this principle. Finally, the AI Act's focus on high-quality training data for high-risk AI systems can potentially lead to more extensive data collection requirements and consequently clashing with GDPR's data minimization rule (Ebers, 2021).

3.3 Challenges in Simultaneous Compliance

Development and deployment of LLMs that comply with both GDPR and AI Act pose significant challenges to business people. Such difficulties include:

1. Data Governance: Striking a balance between the GDPR's stringency as regards protection of personal information and the AI act's insistence on clear quality training records in relation to high-risk artificial intelligence systems.

2. Consent Mechanisms: Harmonising detailed consent demands of GDPR with broader societal implications of the AI Act.

3. Transparency and Explainability: Satisfying both the "right to explanation" under the GDPR and wider transparency obligations under the AI Act for high-risk AI systems.

4. Risk Assessment: Carrying out risk assessments that address data protection under the GDPR as well as societal implications under the AI Act.

5. Cross-Border Data Transfers: Complying with the AI Act's stipulations on high-risk AI systems while navigating GDPR constraints on international data transfers (Tsamados et al., 2021).

3.4 Impact on LLM Development and Deployment in the EU

It is likely that issues resulting from GDPR combined with AI Act will have far-reaching effects on LLM development and deployment within European Union:

1. Increased Compliance Costs: The costs associated with building and running LLMs might increase due to the need to follow both regulatory frameworks, which may ultimately favor well-capitalized firms (Ebers, 2021).

2. Innovation Constraints: Stricter rules could slow down the pace of LLM innovation in Europe, putting businesses located in Europe at a disadvantage compared to those outside it.

3. Focus on Ethical AI: The emphasis placed by these two regulations on fairness, transparency and accountability can encourage more research and development of ethical AI leading to credible LLMs (Floridi & Cowls, 2019).

4. Market Fragmentation: Non-uniform interpretations as well as enforcements of these legislations among EU member states may result in a fragmented market for LLMs within UE borders.

5. Global Influence: Similar to GDPR, requirements placed on LLMs by the AI Act might become global standards influencing how such models are developed worldwide (Veale & Zuiderveen Borgesius, 2021).

In summary, while both the GDPR and AI Act aim at promoting responsible development of Artificial Intelligence (AI), applying them simultaneously to LLMS makes the compliance landscape complex. To navigate it successfully requires consideration of data protection principles as well as broader AI ethics concerns. Ongoing dialogue among policy makers, industry stakeholders, and researchers will be essential in striking a balance between protectionism and innovation for LLMS with an evolving regulatory framework.

4.Conclusion

This paper has examined the intricate regulatory environment surrounding Large Language Models (LLMs) in the European Union with particular emphasis on how they interact with General Data Protection Regulation (GDPR) and the proposed Artificial

Intelligence Act (AI Act). The analysis has established significant findings and their implications for LLM stakeholders.

4.1 Summary of Key Findings

1. Regulatory Complexity: Such overlapping regulations may create a complex regulatory environment for LLMs as they are both applicable at the same time. Although GDPR focuses mostly on protection of individuals' data rights and privacy, the AI Act introduces a risk-based approach towards AI regulation resulting in potential overlaps and conflicts (Veale & Zuiderveen Borgesius, 2021).

2. Compliance Challenges: According to Ebers (2021), companies developing or using LLMs face major challenges in complying with these two sets of rules. They include finding a balance between data minimization principles and need for large training sets, implementing effective consent mechanisms, meeting diversity of explanations/transparency requirements among others.

3. Ethical AI Development: Both legislations highlight fairness, transparency and accountability to be key values in designing AI systems which could lead to improvements in ethical AI practices (Floridi & Cowls, 2019).

4. Market Dynamics: Given that these strict regulatory compliance measures favour bigger firms having more resources who can dominate market competition within the EU's artificial intelligent industry sector (Tsamados et al., 2021).

4.2 Implications for Stakeholders

4.2.1 For LLM Developers:

1.Increased investment into compliance measures as well as ethical AI developments shall be required.

2.There might be important roles played by lawyers or ethicists in this regard.

3.New approaches to data governance, model transparency along with bias mitigation will also be demanded by such models

4.2.2 For Policymakers:

1. The need for continuous development of rules to face new challenges during the LLM progress and deployment remains inevitable.

2. Striking a balance in supporting innovation while maintaining security is key.

3. Global collaboration may be required to prevent regulatory divergence and keep up with the EU's competitive edge.

4.2.3 For Users:

1. LLM operations with increased transparency and accountability may result in trustworthy AI systems.

2. Users might need to be more conscious of their rights and how LLMs can affect their own privacy as well as self-determination process.

4.3 Areas for Future Research

1. Technical Solutions: Advanced techniques in model interpretability, data minimization in LLMs, and effective bias mitigation strategies are subjects of study.

2. Regulatory Impact Assessment: Longitudinal studies on the effects of GDPR and AI Act on innovation of LLMs in Europe.

3. Cross-border Implications: How do EU regulations influence global LLM development? Possible Strategies for global LLM regulatory harmonization.

4. Ethical Framework: Development of comprehensive ethical guidelines specifically tailored to LLMs, considering their unique capabilities and risks.

4.4 Personal Opinion on the Future of LLM Regulation in the EU

To me it appears that regulating LLMs will revolve around finding a delicate balance between encouraging technological advancements while effectively safeguarding individual rights as well as societal values from harm? However, this does not mean that they are complete opposites; rather they share commonalities such as an emphasis on risk assessment framed through a juridical language that helps make this distinction more pronounced but at times blurs those lines when one stops to think about what it means by “high risk.” Notwithstanding, as technology continues to transform rapidly into newer versions so should governance frameworks switch into new forms too?

I expect we could observe specific guidelines or amendments provided centering on peculiarities and challenges of LLMs. These may comprise clearer benchmarks on LLM transparency, more sophisticated tactics for data minimization attuned to the requirements of training LLMs as well as improved frameworks for identifying and correcting biases in LLMs.

Furthermore, considering that AI is a global phenomenon, Europe’s approach to regulating LLM will maintain its international significance. This may result in attempts to

globally cooperate on regulation thereby leading to the creation of globally harmonized standards for the development as well as deployment of LLMs.

Ultimately the success of the regulation of ML in the EU will depend on ongoing dialogue and collaboration between policy-makers, industry actors, researchers as well as civil society. Therefore, if this cooperative approach is encouraged by the European Union (EU) it might position itself at the forefront as one with substantive issues for AI development that is responsible and innovative besides protecting its own citizens.

5. References

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.
- Bietti, E. (2020). From ethics washing to ethics bashing: A view on tech ethics from within moral philosophy. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 210-219.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. arXiv preprint arXiv:2005.14165.
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. Science and Engineering Ethics, 24(2), 505-528.
- Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B., & Sen, P. (2020). A survey of the state of explainable AI for natural language processing. arXiv preprint arXiv:2010.00711.
- Ebers, M. (2021). Regulating AI and Robotics: Ethical and Legal Challenges. In Algorithms and Law (pp. 37-99). Cambridge University Press.

- European Commission. (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence. Brussels, 21.4.2021 COM(2021) 206 final 2021/0106 (COD).
- European Parliament and Council. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).
- Hao, K. (2021). We read the paper that forced Timnit Gebru out of Google. Here's what it says. MIT Technology Review.
- Kamarinou, D., Millard, C., & Singh, J. (2016). Machine Learning with Personal Data. Queen Mary School of Law Legal Studies Research Paper, (247/2016).
- Kaminski, M. E. (2019). The right to explanation, explained. Berkeley Tech. LJ, 34, 189.
- Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. International Data Privacy Law, 7(4), 233-242.
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. International Journal of Human–Computer Interaction, 36(6), 495-504.
- Smuha, N. A., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R., & Yeung, K. (2021). How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act. SSRN Electronic Journal.

- Stalla-Bourdillon, S., & Knight, A. (2017). Data analytics and the GDPR: friends or foes? A call for a dynamic approach to data protection law. In Data Protection and Privacy: The Age of Intelligent Machines (pp. 175-197). Hart Publishing.
- Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The ethics of algorithms: key problems and solutions. AI & SOCIETY, 1-16.
- Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97-112.
- ·Veale, M., Binns, R., & Edwards, L. (2018). Algorithms that remember: model inversion attacks and data protection law. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180083.
- Villaronga, E. F., Kieseberg, P., & Li, T. (2018). Humans forget, machines remember: Artificial intelligence and the Right to Be Forgotten. Computer Law & Security Review, 34(2), 304-313.
- Voigt, P., & Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR). A Practical Guide, 1st Ed., Cham: Springer International Publishing.
- Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law, 7(2), 76-99.
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Irving, G. (2021). Ethical and social risks of harm from Language Models. arXiv preprint arXiv:2112.04359.
- Zuiderveen Borgesius, F. (2020). Strengthening legal protection against discrimination by algorithms and artificial intelligence. The International Journal of Human Rights, 24(10), 1572-1593.