DOI 10.5817/MUJLT2024-1-4

UNVEILING THE BLACK BOX: BRINGING ALGORITHMIC TRANSPARENCY TO AI

by

GYANDEEP CHAUDHARY *

Overall, algorithmic transparency is an important aspect of responsible AI development and deployment. Ensuring that AI systems are transparent and accountable will help build trust and confidence in these systems and ensure that they are used ethically and effectively. Artificial intelligence (AI) has emerged as a cutting-edge domain that is fundamentally redefining different areas of daily experiences, such as health care, transport, finance, education, and others. The systems are not created for making a judgment like human judgment of natural language, spotting patterns and problem-solving; rather AI produces machines that also have intelligence level same as that of human beings.

AI having more influence over us, it is to be considered the ethical directions of these tools and see that they operate under principles of transparency and accountability. The element regarding algorithmic transparency, which means the process of understanding the functioning and explanation of how AI systems make their decisions is the one that is most crucial. The issue of algorithm transparency is of fundamental importance for many considerations. AI systems are not only supported by fairness but also by their non-discrimination. If we do not know how a system of AI arrives at the decisions made, it becomes impossible to determine if the provided results meet equal treatment for everybody. If used in delicate areas like recruitment, credit, and legal system- where the AI-machine must make choices which are life changing, then this aspect is very important.

On top of fairness, algorithmic transparency is also an important factor for accountability. If we are ignorant about what an artificial intelligence algorithm does and what is the source of its decision-making process, we are unable to track and classify the mistakes or mishaps of the system. This has always mattered when central to the operation of systems with high stake, such as those used in

* Assistant Professor, School of Law, Bennett University, India, gyan.2889@gmail.com

self-driving vehicles or in health care. Algorithmic transparency may be reached using different instruments. The transparent AI systems can be made by a more transparent design, for example, the simple modelling tools, that use interpretable models. Another method is designing technologies and techniques that can help people why the artificial systems difficult to be decoded but easy to understand which they can utilize in making decisions.

Therefore, algorithmic transparency is a key factor of the AI made responsibly and used by the society. It is crucial that AI machines are both transparent and accountable since this will lead to people building trust in the system and accepting its ethical and practical implications. This paper examines regulation of algorithmic transparency in the EU, specifically provisions under the General Data Protection Regulation (GDPR), it aims to situate analysis of the GDPR's provisions on explainability of AI systems within broader technology ethics and policy discourse. The paper's scope is limited to EU regulations applicable to AI data processing transparency.

KEY WORDS

Artificial Intelligence, Accountability, Algorithmic Transparency, Explainability, Right to Explanation

1. INTRODUCTION

New technologies that blur the distinctions between the *"analogue world"* and the *"digital world"* will have far-fetching effects on many spheres of society, including education, industry, politics, and the arts. The notion of an emerging *"digital ecosystem"* in which decision-making centres are migrating to automated and intelligent systems is described by some as the fourth industrial revolution.[1] These new models use AI and machine learning algorithms that can process vast datasets, learn from experience, and solve complex problems that were once considered exclusively human abilities.[2] However, this shift towards algorithmic decision-making also poses risks to fundamental civil liberties, as opaque systems undermine public trust in the fairness and legality of the choices ultimately made.[3] Experts in fields such as psychology and education are working to address the challenge of making algorithmic decisions more scrutable and open to

[1] Schwab, K. (2017) *The fourth industrial revolution*. Crown Publishing Group, New York.

[2] Brynjolfsson, E. and Mitchell, T. (2017) What can machine learning do? Workforce implications. *Science*, 358(6370), pp. 1530-1534. see also, Data Guidance. (2022) *Norway - Data Protection Overview*. [online] Available from: `https://www.dataguidance.com/notes/norway-data-protection-overview` [Accessed 5 September 2023].

[3] Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Massachusetts, p.320.

public examination.[4] For instance, in February 2020, a Dutch court banned the government from using the SyRI system, which detected welfare fraud by combining various data,[5] because authorities refused to disclose the source code.[6] New technologies that blur the distinctions between the *"analogue world"* and the *"digital world"* will have far-fetching effects on many spheres of society, including education, industry, politics, and the arts. The notion of an emerging *"digital ecosystem"* in which decision-making centres are migrating to automated and intelligent systems is described by some as the fourth industrial revolution.[7] These new models use AI and machine learning algorithms that can process vast datasets, learn from experience, and solve complex problems that were once considered exclusively human abilities.[8] However, this shift towards algorithmic decision-making also poses risks to fundamental civil liberties, as opaque systems undermine public trust in the fairness and legality of the choices ultimately made.[9]

Experts in fields such as psychology and education are working to address the challenge of making algorithmic decisions more scrutable and open to public examination.[10] For instance, in February 2020, a Dutch court banned the government from using the SyRI system, which detected welfare fraud by combining various data,[11] because authorities refused to disclose the source code.[12]

---

4 Burrell, J. (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1).

5 Meek, C. (2022) *Artificial Intelligence in The Age of Algorithmic Transparency.* [online] Les Echos. Available from: `https://www.americangreetings.job.com/news/artificial-intelligence-in-the-age-of-algorithmic-transparency/` [Accessed 12 August 2023].

6 Court of The Hague. (2020, February 5). SyRI legislation in conflict with higher law. Rechtspraak.nl. [online] Available from: `https://uitspraken.rechtspraak.nl/\#!/details?id=ECLI:NL:RBDHA:2020:1878` [Accessed 1 August 2023].

7 Schwab, K. (2017) *The fourth industrial revolution*. Crown Publishing Group, New York.

8 Brynjolfsson, E. and Mitchell, T. (2017) What can machine learning do? Workforce implications. *Science*, 358(6370), pp. 1530-1534. see also, Data Guidance. (2022) *Norway - Data Protection Overview*. [online] Available from: `https://www.dataguidance.com/notes/norway-data-protection-overview` [Accessed 5 September 2023].

9 Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Massachusetts, p.320.

10 Burrell, J. (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1).

11 Meek, C. (2022) *Artificial Intelligence in The Age of Algorithmic Transparency.* [online] Les Echos. Available from: `https://www.americangreetings.job.com/news/artificial-intelligence-in-the-age-of-algorithmic-transparency/` [Accessed 12 August 2023].

12 Court of The Hague. (2020, February 5). SyRI legislation in conflict with higher law. Rechtspraak.nl. [online] Available from: `https://uitspraken.rechtspraak.nl/\#!/details?id=ECLI:NL:RBDHA:2020:1878` [Accessed 1 August 2023].

The notion of *"algorithmic accountability"* is based on the premise that understanding how machines work enables appropriate oversight.[13] The goal of algorithmic transparency is to ensure that computational processes are accurate and unbiased, which becomes increasingly difficult as algorithms become more complex.[14] However, opaque *"black box"* techniques[15] are being employed in diverse high-stakes decisions about credit, employment, education, government benefits, border-control, surveillance, and even sports stadium monitoring, often with unfair and unexplainable outcomes even to the organizations deploying them.[16]

The primary challenge is to protect the right to informational self-determination while preventing algorithmic harm to individuals and society.[17] Demands for *"traceability"* of automated decisions arise from AI's expanding real-world impacts. More broadly, transparency obligations address the *"opacity of the algorithms"*, which neither users nor designers sufficiently comprehend.[18] Privacy advocates, researchers, and policymakers have raised concerns[19] regarding the inscrutable nature of how machine learning systems categorize new inputs and derive predictions.[20]

Legal frameworks such as the EU's General Data Protection Regulation (GDPR) provide starting points for discovering applicable rules even when AI is involved in processing personal data.[21] GDPR enables algorithmic impact assessments through provisions around evaluating effects on personal information rights. It also mandates strict transparency requirements,

[13] Diakopoulos, N. (2016) Accountability in Algorithmic Decision-Making. *Communications of the ACM*, 59(2), pp. 56-62.

[14] Ananny, M., and Crawford, K. (2018) Seeing Without Knowing: Limitations of The Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society*, 20(3), pp. 973-989.

[15] Chaudhary, G. (2020), Artificial Intelligence: The Liability Paradox, *ILI Law Review*, p. 144.

[16] O'Neil, C. (2017) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers. See also Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Massachusetts, p.320.

[17] Mittelstadt, B. D. et al. (2016) The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).

[18] Burrell, J. (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1).

[19] Lepri, B. et al. (2018) Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4), pp. 611-627.

[20] Mitrou, L. (2018) Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'? [online] *Tilburg: TILT Law & Technology Working Paper Series* Available at SSRN: `https://ssrn.com/abstract=3386914` or `http://dx.doi.org/10.2139/ssrn.3386914`[Accessed 28 August 2023]; *see also* Wischmeyer, T. (2020) Artificial Intelligence and Transparency: Opening the Black Box. In: Wischmeyer, T. et.al. (eds) *Regulating Artificial Intelligence*, Switzerland, pp. 75-101.

[21] Goodman, B. and Flaxman, S. (2016) *EU regulations on algorithmic decision-making and a "right to explanation"*. [preprint] arXiv:1606.08813.

notably that data controllers must inform people about the *"existence of automated decision-making and provide meaningful information*[22] *about the logic involved"*[23]- which implies elucidating the algorithm's basic principles in plain language rather than code.[24] Does the duty explain a computer's decision to fall within this obligation? Once a choice is made, can people whose data have been used ask for an explanation of the AI model's decision-making process? If so, what kind of information should be included in such an explanation? Whether this duty extends to explaining specific decisions of an AI model post hoc remains debated among legal experts and computer scientists.[25]

In the intricate landscape of data protection, the GDPR stands as a pivotal framework governing the handling of personal data.[26] However, it is crucial to note that GDPR primarily concerns itself with personal data, leaving a notable gap in the regulatory framework when it comes to non-personal data. The European Data Strategy, unveiled in 2020, recognizes the significance of harnessing the potential of non-personal data while underscoring the necessity for a regulatory framework to ensure responsible and fair usage (European Commission, 2020).[27] Additionally, various EU member states have initiated efforts to bridge this gap by formulating legislation specific to non-personal data, such as the French Data Protection Act[28] and the German Federal Data Protection Act.[29] These national legislations complement GDPR by extending regulatory oversight to encompass non-personal data, emphasizing the need for

[22] Wachter, S. et al. (2017) Transparent, explainable, and accountable AI for robotics. *Science Robotics*, 2(6), eaan6080.

[23] GDPR §§ Articles 13(2)(f), 14(2)(g).

[24] Meek, C. (2022) *Artificial Intelligence in The Age of Algorithmic Transparency.* [online] Les Echos. Available from: https://www.americangreetings.job.com/news/artificial-intelligence-in-the-age-of-algorithmic-transparency/ [Accessed 12 August 2023].

[25] Wachter, S. et al. (2017) Transparent, explainable, and accountable AI for robotics. *Science Robotics*, 2(6), eaan6080.

[26] European Parliament & 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). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX\%3A32016R0679 [Accessed 15 January 2024].

[27] European Commission. (2020). European Data Act. https://digital-strategy.ec.europa.eu/en/policies/data-act

[28] France. (2018). Law No. 2018-493 of June 20, 2018, on the Protection of Personal Data. https://www.legifrance.gouv.fr/eli/loi/2018/6/20/JUSX1721380L [Accessed 20 January 2024].

[29] Bundesministerium der Justiz und für Verbraucherschutz. (2017). Federal Data Protection Act (BDSG). https://www.gesetze-im-internet.de/englisch_bdsg/[Accessed 10 January 2024].

a harmonized approach to safeguard all forms of data. This nuanced evolution in legislation reflects a broader global trend, where jurisdictions are grappling with the complexities of data governance, acknowledging the pivotal role non-personal data plays in the digital era. As we navigate this intricate regulatory landscape, it becomes imperative to strike a delicate balance, ensuring the facilitation of innovation while upholding fundamental principles of data protection.

This article examines the transparency and explainability issues surrounding AI, considering the challenges and concerns raised, and situating them in the evolving regulatory landscape. The "*black box*" nature of complex AI models poses transparency and accountability challenges and ultimately, ensuring *"black box"* does not become Pandora's box will require interdisciplinary collaboration between law, computer science, and social sciences to balance innovation, ethics, and human rights. However, opacity also represents a knowledge problem- neither designers nor regulators fully grasp modern algorithmic systems. Advancing research on interpretable machine learning and auditing processes, combined with public education, provides paths to make AI ethical and accountable. This article examines regulation of algorithmic transparency in the EU, specifically provisions under the General Data Protection Regulation (GDPR) and analyses legal debates on whether GDPR mandates 'ex-post' explanations of specific AI decisions. It discusses technical and legal obstacles to transparency such as proprietary interests and data privacy. The notion of 'qualified transparency' is proposed as a nuanced approach involving disclosures tailored to diverse stakeholders. The article argues transparency is essential for accountability but must balance competing values. It recommends ongoing interdisciplinary collaboration to make algorithmic systems interpretable, auditable, and ethical while upholding innovation and human rights. Overall, this article focuses on the regulatory framework around algorithmic transparency and accountability in the EU. It aims to situate analysis of the GDPR's provisions on explainability of AI systems within broader technology ethics and policy discourse. The paper's scope is limited to EU regulations applicable to AI data processing transparency.

2. THE BLACK BOX

The term AI encompasses a broad spectrum[30] of technological advancements designed to emulate human cognition and behaviour.[31] At its core, AI enables machines to learn from data and past experiences, reason through complex problems, and make autonomous decisions.[32] One prevalent technique is deep learning, which uses multi-layered artificial neural networks loosely inspired by biological neural networks. These networks can discern intricate patterns and relationships within massive datasets by adjusting internal parameters during training.[33] Consequently, deep learning models can extract insights from new data by generalizing patterns learned from training data.[34]

The advent of 'Big Data' has amplified both the potential and complexity of AI systems. The sheer volume, velocity, and variety of digital data now available for analysis is unprecedented. However, our capacity to fully comprehend the inner workings of sophisticated AI models remains limited. Their decision-making processes can be as opaque as the human mind.[35] Experts liken unravelling the 'black box' of AI to deciphering neurobiological processes in the brain.[36] We can observe the inputs and outputs of AI systems but lack granular visibility into how interconnected nodes within neural networks produce outputs. Even developers struggle to pinpoint the factors

[30] Samoili, S., Cobo, M.L., et al. (2020) *AI Watch. Defining Artificial Intelligence, Towards an Operational Definition and Taxonomy of Artificial Intelligence*. EUR 30117 EN, Publications Office of the European Union, Luxembourg.

[31] European Commission (2018) *Communication from the Commission to the European Parliament, the European Economic and Social Committee and the Committee of the Regions*. [online] Available from: `https://commission.europa.eu/publications/communication-commission-european-parliament-council-european-economic-and-social-committee-and\_en` [Accessed 5 September 2023].

[32] Samoili, S., Cobo, M.L., et al. (2020) *AI Watch. Defining Artificial Intelligence, Towards an Operational Definition and Taxonomy of Artificial Intelligence*. EUR 30117 EN, Publications Office of the European Union, Luxembourg.

[33] European Commission (2018) *Communication from the Commission to the European Parliament, the European Economic and Social Committee and the Committee of the Regions*. [online] Available from: `https://commission.europa.eu/publications/communication-commission-european-parliament-council-european-economic-and-social-committee-and\_en` [Accessed 5 September 2023].

[34] Data Guidance. (2022) *Norway - Data Protection Overview*. [online] Available from: `https://www.dataguidance.com/notes/norway-data-protection-overview` [Accessed 5 September 2023].

[35] Bathaee, Y. (2018) The Artificial Intelligence Black Box and The Failure of Intent and Causation, *Harvard Journal of Law and Technology*, 31(2), p. 891.

[36] Council of Europe Committee of experts on internet intermediaries (MSI-NET). (2017) *Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications*. [online] Available from: `https://rm.coe.int/study-hrdimension-of-automated-data-processing-incl-algorithms/168075b94a` [Accessed 5 September 2023].

that hold the greatest weight for any given decision.[37] Thus, AI judgments often emerge from a metaphorical 'black box'[38] devoid of any interpretability or explainability.[39]

As algorithmic decision-making permeates more aspects of daily life, the need to demystify AI's black box becomes increasingly urgent.[40] We are entering an 'Algorithmic Society' where social and economic outcomes hinge on automated systems and agents.[41] AI now extends far beyond the controlled laboratory setting into consequential real-world applications. More individual and collective decisions will depend on algorithmic calculations. Properly implemented, AI can uplift human rights and democratic principles.[42] However, opacity also increases the risks of bias, discrimination, manipulation, violations of due process, and physical harm.[43] Even well-intentioned developers can engrain unfairness within models by training them in incomplete, biased, or unrepresentative data.[44] Without visibility into AI reasoning, auditing the process, remedying harms, and ensuring equitable treatment becomes challenging. Thus, transparency has emerged as an ethical imperative and prerequisite for socially responsible AI deployment.[45]

3. A RIGHT TO EXPLANATION

The critical principle called 'transparency' is one such concept on which GDPR is based.[46] The persons whose data are being used should be clearly informed by the person in charge. The main idea is that people whose data is being processed should be informed about it and, by extension, about

[37] Castelvecchi, D. (2016) Can We Open the Black Box Of AI? *Nature,* 538(7623), p. 20.

[38] Chaudhary, G. (2020), *Artificial Intelligence: The Liability Paradox*, *ILI Law* Review, p. 144.

[39] Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Massachusetts, p.320.

[40] Ibid.

[41] Balkin, J. (2017) The Three Laws of Robotics in The Age of Big Data, Ohio State Law Journal, 78, p. 1218.

[42] Council of Europe Committee of experts on internet intermediaries (MSI-NET). (2017) *Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications*. [online] Available from: `https://rm.coe.int/study-hrdimension-of-automated-data-processing-incl-algorithms/168075b94a` [Accessed 5 September 2023].

[43] Ibid.

[44] CNIL. (2021) *How Can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence*. [online] Available from:`https://www.cnil.fr/sites/default/files/atoms/files/cnil\_rapport\_ai\_gb\_web.pdf` [Accessed 5 September 2023].

[45] Ibid.

[46] European Commission. (2018) *Guidelines on Automated Individual Decision-Making and Profiling for The Purposes of Regulation 2016/679 (Wp251rev.01)*. [online] Available from: `https://ec.europa.eu/newsroom/article29/items/612053`[Accessed 5 September 2023].

the risks that come with it so that they can make intelligent decisions and protect their interests.

The data subject has every right to be informed about the data collection being done by the third party or otherwise directly as per the regulations contained within the GDPR, most notably Article 13 and 14. Data controllers must warn individuals about *"the existence of automated decision-making"* and offer *"meaningful information about the logic involved and the expected consequences of such processing"* in accordance with Articles 14(2)(g) and 13(2)(f). Furthermore, people have the right to access their own details and personal data under Article 15(1)(h). Article 22(3) also states that when automated decisions are made, data controllers must *"take appropriate measures to protect the rights, freedoms, and legitimate interests of the data subject, including at least the right to get human intervention from the controller, to express his or her point of view, and to contest the decision"*. Finally, individuals enjoy the right to "specific information" and *"the right to get an explanation of the decision reached after such an [automated] assessment",* all of which are required safeguards for automated processing, under Recital 71 of GDPR.

As per the above discussion, it is clear enough to state that a right to know about the outcome of the correct model or weight and to consider the data used in this situation vests upon the data subject. However, contrary to common opinion, the right to an explanation is not part of natural control. There is a linguistic barrier between the preamble and the articles. The preamble is not binding and cannot provide a right to an explanation in and of itself. Then, is the 'right to explanation' in the GDPR equivalent to the 'right to information'?

First, we must define what it means to 'explain' an automated judgement before deciding whether the GDPR affords persons the right to an explanation.[47] Researchers distinguish between discussing how a system operates in general and explaining how a particular choice was made by or through an AI system. Furthermore, researchers say that to explain how an automated system for making decisions works, one must explain its "*logic, significance, expected consequences, and general functionality*". In contrast, addressing individual decisions necessitates elaborating on the "*reasons and individual circumstances of a specific automated decision*", such as the elements considered, and their relative importance or the case-specific

[47] Burt, A (2020) *Is there a 'right to explanation' for machine learning in the GDPR?* [online] International Association of Privacy Professionals. Available from: `https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/` [Accessed 5 September 2023].

decision rules defined by the machine.[48] It is also possible to differentiate between explanations by examining the order in which they are given during decision-making. A preliminary declaration is made before automatic selection. Hence, it is only logical that it describes how the system operates. Conversely, ex-post descriptions are provided after an automated decision has already been made and detail the procedure's inner workings and the rationale behind a given conclusion.[49] As a result, the second justification offered for decisions after they have already been made may be the only kind to which a meaningful right to explanation applies.

Part of the theory of law has been critical of the right to explanation.[50] Opponents have made a big deal because this right was intentionally excluded of the final draft of GDPR. Considering the most recent drafts of the GDPR and the feedback received during the trialogue negotiations, the original versions of the draft incorporated stringent protections for profiling and automated decision-making. However, the right to a legally binding explanation of one's decisions was abandoned.[51] In addition, the idea that recitals are legally binding has been questioned. A renowned group of researchers stated that *"Recitals have no positive effect of their own and cannot give rise to legitimate expectations"*.[52] *"In principle, the ECJ does not give effect to recitals written in normative terms"*, experts argue, supporting this view. Recitals can assist in explaining why and how a normative instrument was developed. They can also be used to clarify issues in the legislation to which they pertain, but they lack independent legal authority.[53] The European Court of Justice (ECJ) precedents were used to show that it is not the job of data protection law to determine whether a particular set of findings and evaluations is correct. To prove this, we used the ECJ

[48] Ferretti, A. et.al. (2018) Machine Learning in Medicine: Opening the New Data Protection Black Box. *European Data Protection Law Review,* 4, p322(2018).

[49] Wachter, S. et.al. (2017) Why A Right to Explanation of Automated Decision Making Does Not Exist in The General Data Protection Regulation. *International Data Private Law,* 7, p.81.

[50] European Commission. (2018) *Guidelines on Automated Individual Decision-Making and Profiling for The Purposes of Regulation 2016/679 (Wp251rev.01*). [online] Available from: `https://ec.europa.eu/newsroom/article29/items/612053` [Accessed 5 September 2023].

[51] Wachter, S. et.al. (2017) Why A Right to Explanation of Automated Decision Making Does Not Exist in The General Data Protection Regulation. *International Data Private Law,* 7, p.81.

[52] Klimas, T. and Vaitiukait, J. (2008) The Law of Recitals in European Community Legislation. *ILSA Journal of International and Comparative Law*, 15(1), pp. 65-93.

[53] Baratta, R.14 (2014) Complexity of EU law in the domestic implementing process. In: 19th Quality of Legislation Seminar EU Legislative Drafting: Views from those applying EU law in the Member States. *Brussels: European Commission Service Juridique - Quality of Legislation Team*, 3 July. Available from: `https://ec.europa.eu/dgs/legal\service/seminars/20140703\baratta\speech.pdf` [Accessed 21 June 2023].

case law as an example.[54] The construction methodology of AI systems validates these assertions. For example, explaining how intricate algorithmic decision-making systems operate and the reasons behind the judgments they make is a complex problem from a technical standpoint. The use of such explanations is called into question because it is likely that the data subjects would not receive a significant amount of beneficial information from them.

Today, however, it appears that most academics think this logic is incorrect. Therefore, it would be overly formalistic to dismiss the concept of the right to an explanation only because recitals are not legally enforceable, given the ECJ's consistent treatment of recitals as interpretative aids in its case law.[55] According to experts, recitals, are intended to clarify the interpretation of a legal norm. Even though they cannot act as such a rule, they are given a grey area of the law and are not enforceable. However, they are generally accepted as definitive interpretations of the GDPR in cases of uncertainty. To better understand how the standards of the GDPR should be implemented, consider reading the accompanying recitals. Often, they include language that goes well beyond GDPR due to political compromises made during negotiations. Recitals cannot be used to create new legal requirements; however, it can be challenging to determine what constitutes a legal interpretation of a new law and what does not.[56] Recital 71 is thus not considered superfluous but instead serves a clear purpose in facilitating interpretation and contributing to the determination of positive law.[57] Because the GDPR is collaborative and evolving, researchers who debate the normative character of the recitals risk cutting themselves off from potentially valuable sources of clarification for data subjects as the legislation advances (differentiating between harsher and softer legal instruments). This is because scholars who disagree with the Recitals' normative status argue on a technicality: the need to distinguish between tougher and softer legal instruments.[58]

[54] Wachter, S. et.al. (2017) Why A Right to Explanation of Automated Decision Making Does Not Exist in The General Data Protection Regulation. *International Data Private Law,* 7, p.81.

[55] Brkan, M (2019) Do Algorithms Rule the World? Algorithmic Decision-Making in The Framework of The GDPR And Beyond. *International Journal of Law and Information Technology*, 27(2), pp. 91-121; *see also* Wischmeyer, T (2020) Artificial Intelligence and Transparency: Opening the Black Box. In: Wischmeyer, T. and Rademacher (eds.) *Regulating Artificial Intelligence.* Switzerland: Springer International Publishing, pp. 75-101.

[56] Kaminski, M. (2019) The Right to Explanation, Explained Berkeley Technology Law Journal, 34, p.194.

[57] Selbst, A. and Powles, J. (2017) Meaningful Information and The Right to Explanation. *International Data Privacy Law*, 7, p.235.

[58] Kaminski, M. (2019) The Right to Explanation, Explained Berkeley Technology Law Journal, 34, p.194.

The Article 29 Working Party Guidelines interpret and intend Recital 71 in the same way, stating that it is a necessary and sufficient condition *"conditio sine qua non"* to safeguard the rights of the data subject.[59] GDPR provides an individual with a form of algorithmic due process in the form of a hearing, as explained in the Guidelines.[60] According to the Guidelines, controllers must take necessary precautions to maintain the legitimate interests, freedom, and rights of data subjects,[61] *"including a mechanism for human intervention in defined cases, such as providing a link to an appeals process at the time an automated decision is communicated to the data subject, with agreed timescales and a named contact"*.

Data protection authorities have also raised this human-in-the-loop methodology. The UK's Information Commissioner's Office (ICO) has acknowledged that the rights to intervention and acquire human explanation present new encounters to developers and industry, urging *"Big Data organisations to exercise caution before relying on machine learning decisions that cannot be rationalised in human-understandable terms"*.[62] According to the French Commission (CNIL), *"What seems to matter is the ability to comprehend the general logic underlying the algorithm's operation"*. This emphasis on understanding the algorithm's logic comes at the expense of transparency. Because it must be communicated in words rather than code, this justification needs to be easily understood. The most crucial factor is not that the code is clear but that we understand the algorithm's inputs, outputs, and purpose. This needs to be made clear.[63]

The potential of automated decision-making processes has been acknowledged by the European Parliament as having the potential to revolutionize the data industry in its resolution of new digital services such as chatbots and virtual assistants.[64] However, the Parliament clarifies that

[59] European Commission. (2018) *Guidelines on Automated Individual Decision-Making and Profiling for The Purposes of Regulation 2016/679 (Wp251rev.01)*. [online] Available from: `https://ec.europa.eu/newsroom/article29/items/612053` [Accessed 5 September 2023].

[60] Kaminski, M. (2019) The Right to Explanation, Explained Berkeley Technology Law Journal, 34, p.194.

[61] Art. 22 GDPR and Recital 71.

[62] Information Commissioner's Office (2017) *Big Data, Artificial Intelligence, Machine Learning and Data Protection*. [online] Available from: `https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf` [Accessed 5 August 2023].

[63] *Ibid.*

[64] European Parliament. (2020) *European Parliament Resolution Of 12 February 2020 on Automated Decision-Making Processes: Ensuring Consumer Protection and Free Movement of Goods and Services*. [online] Available from: `https://www.europarl.europa.eu/doceo/document/TA-9-2020-0032\_EN.html` [Accessed 25 June 2023].

"in light of the significant impact it can have on consumers, one should be properly informed about how a system that automates decision-making operates, how to reach a human with decision-making authority, and how the system's decisions can be reviewed and corrected". The resolution emphasizes that these systems must employ high-quality, objective datasets and clear, explicable, and accurate algorithms. Concurrently, automated choice procedures for detecting and correcting flaws should be developed. The Parliament's official stance is that *"humans must always be ultimately responsible for and able to override decisions made using automated decision-making processes"*.

In a technical report published by the Joint Research Centre in a similar vein but with more urgent implications, the significance of 'explainability-by-design' in AI systems may endanger users' fundamental rights.[65] This report asserts that for human oversight to be effective, algorithmic processing must be understandable to the person conducting the evaluation.

GDPR is not the first recent law to approach the issue of algorithmic accountability and transparency. By requiring that the regulations that define such action and its key features be provided to those who seek them, the French Act for the Digital Republic[66] ensures that those impacted by administrative algorithmic choices can obtain an explanation of such decisions. Furthermore, administrative organizations are obligated to report the type and extent of algorithmic processing used in decision-making, the treatment parameters used, and, if applicable, the weights assigned to those considerations.

4. ALGORITHMIC TRANSPARENCY: A PRACTICAL ISSUE

People's right to be informed under the GDPR involves several practical challenges, such as explaining what information should be disclosed and the AI-based decision-making process. Theoretically, the difficulties of the rationalization process of artificial intelligence, particularly unsupervised models,[67] have been emphasised. It is generally agreed that the inherent complexity of the data volume, algorithm modularity, iterative processing, and randomised tiebreaking may pose a formidable cognitive obstacle.[68] Furthermore, the dynamic character of several algorithms seems to contradict the static nature of transparency. Continuous updates and modifications are

---

65 *Ibid.*

66 *Digital Republic Act 2016*, Law No. 2016-132117. France. In French.

67 Wang, P. (2012) Theories of Artificial Intelligence—Meta-Theoretical Considerations. *Atlantis Thinking Machines*, 9, pp. 305–323.

68 Z C Lipton, Z.C. (2018) The Mythos of Model Interpretability. In Machine Learning, The Concept of Interpretability Is Both Important and Slippery. *ACMQueue*, 16(3), p.13.

made to the algorithms, although any transparency disclosure refers only to the current algorithm.[69] Decontextualization is also a technical barrier that, occurs when algorithmic models initially used for one purpose are repurposed for a different purpose and context.[70]

Machine learning, however, is not a monolithic idea; it includes various methodologies, from the tried-and-true (such as decision tree algorithms and linear regression) to the cutting edge (such as various forms of neural networks). The difficulty of establishing an ex-post causal relationship between a particular input and output varies significantly among different methods.[71] Improved accuracy can be seen across the board with Bayesian classifiers, additive models, decision trees, and spare linear models, the likelihood that they will provide models that people can comprehend. These algorithms frequently employ several internal features (i.e., paths, controls, or characteristics) to adequately track and explain their results. Deep learning algorithms build high-dimensional input-based applications, such as speech recognition, picture identification, and natural language processing by using intricate networks across network layers to develop extremely nonlinear correlations among inputs and outputs.[72] As the number of nonlinear parameters a system considers while making a decision grows, it becomes harder for humans to understand the model.

We need to consider at the legal and regulatory obstacles to algorithmic transparency in addition to the technical ones posed by algorithms' inherent flexibility and unpredictability. It is understandable to want to limit the amount of detail that can be provided about models and procedures to protect proprietary information and intellectual property.[73] Data controller competition and security requirements could limit algorithm access.[74] Because it constitutes a non-transferable competitive advantage, companies are unwilling to disclose information about their assets. Similarly, privacy and security professionals stress the dangers of revealing sensitive information about an organisation's inner workings, which could increase

[69] Wischmeyer, T. (2020) Artificial Intelligence and Transparency: Opening the Black Box. In: Wischmeyer, T. et.al. (eds) *Regulating Artificial Intelligence*, Switzerland, pp. 75-101.

[70] Donovan, J., Matthews, J., et al. (2018) *Algorithmic Accountability: A Primer*. [online] Data & Society. Available from: `https://datasociety.net/output/algorithmic-accountability-a-primer/` [Accessed 1 September 2023].

[71] Wischmeyer, T. (2020) Artificial Intelligence and Transparency: Opening the Black Box. In: Wischmeyer, T. et.al. (eds) *Regulating Artificial Intelligence*, Switzerland, pp. 75-101.

[72] Rai, A. (2020) Explainable AI: From Black Box to Glass Box. *Journal of the Academy of Marketing Science*, 48, pp. 137–141.

[73] GDPR. EU (n.d.) *Recital 63, Right of access*. [online] Available from: `https://gdpr.eu/recital-63-right-of-access/` [Accessed 15 July 2023].

[74] Veale, M. et.al. (2018) Algorithms That Remember: Model Inversion Attacks and Data Protection Law. *Philosophical Transactions of the Royal Society A*, 376 (2133).

cyberattacks.[75] Another legal basis for limiting access to information is protecting state secrets and public interests, which must be protected from disclosure to the public.[76] Access to information is often seen as a key issue in the regulation of AI-based systems because it is necessary for external parties, such as regulatory authorities and auditors, to be able to assess the performance and risks of these systems. Without access to information, it is difficult for these parties to understand how the systems work and identify any potential issues or risks. In some cases, access to information may be restricted because of concerns about confidentiality, intellectual property, or national security. In these cases, it may be necessary to find ways to balance the need for access to information with these other considerations. There are also technical challenges that can make it difficult to provide access to information about AI-based systems. For example, some AI systems may be complex and have many components that are difficult to understand or analyse. In addition, there may be issues with data privacy and security that need to be addressed when providing access to information. Overall, access to information is an important issue in the regulation of AI-based systems, and finding ways to ensure that regulatory authorities and other external parties have the necessary access to information will be crucial to the effective oversight and regulation of these systems.[77]

The potential consequences of this new right for the AI sector and the advancement of AI, in general, have also been mentioned as a source of concern. Access to algorithms alone is not sufficient to effectively clarify and comprehend a decision-making process. Therefore, companies require time and expertise to conduct this type of assessment.[78] Owing to the interrelated nature of algorithms and datasets in complex information systems and the potential for errors and biases in models and data to become concealed over time, *"explainability may prove especially disruptive for data-intensive industries"*. Some have argued that the GDPR threatens one of AI's most

[75] Wischmeyer, T. (2020) Artificial Intelligence and Transparency: Opening the Black Box. In: Wischmeyer, T. et.al. (eds) *Regulating Artificial Intelligence*, Switzerland, pp. 75-101.

[76] Burrell, J. (2016) How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data Society*, p.9.

[77] Meek, C. (2022) *Artificial Intelligence in The Age of Algorithmic Transparency.* [online] Les Echos. Available from: `https://www.americangreetings.job.com/news/artificial-intelligence-in-the-age-of-algorithmic-transparency/` [Accessed 12 August 2023].

[78] Ananny, M. and Crawford, K. (2016) Seeing Without Knowing: Limitations of The Transparency Ideal and Its Application to Algorithmic Accountability. *New Media Society*, 20(3), p.975.

beneficial uses by restricting the usage of AI's most desirable characteristics: automation and autonomy.[79]

5. TOWARDS A QUALIFIED TRANSPARENCY

Unlocking the 'black box' does not have to be done just because people are curious. The data subject must comprehend the reasoning behind decisions to pursue undesirable outcomes and what, if anything, could be done differently in the future considering the current decision-making model.[80] Technical data and in-depth analyses of the algorithms may not be helpful.[81] In the framework of meaningful transparency, data subjects may be given access to information on several parts of the algorithmic process at any time.[82] Human participation details, input/output details, data quality (how training data were collected/labelled, source reliability, precision, timeliness), algorithm model/architecture/variables/weights/inference process details are all examples of such things (including the margin of error predicted).[83]

Furthermore, the ability to criticise a decision based on the facts presented is not necessarily related to the need for openness and explanation. These protections are integral to the principles of fairness and responsibility and are essential for creating unbiased and robust AI systems. Thus, algorithmic accountability is a part of algorithmic transparency, the idea that an algorithmic system should employ numerous checks and balances to ensure that the system functions as intended by the human operator. The undesirable results can be pinpointed and fixed.[84] Data controllers are responsible for enforcing specific measures inside their organisations to guarantee adherence to data protection obligations by the principle of accountability. These steps may include using a privacy-by-design system architecture or setting up data protection impact assessments.

[79] Mitrou, L. (2018) Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'? [online] *Tilburg: TILT Law & Technology Working Paper Series* Available at SSRN: https://ssrn.com/abstract=3386914or http://dx.doi.org/10.2139/ssrn.3386914 [Accessed 28 August 2023].

[80] Wachter, S. et.al. (2017) Why A Right to Explanation of Automated Decision Making Does Not Exist in The General Data Protection Regulation. *International Data Private Law*, 7, p.81.

[81] European Commission. (2018) *Guidelines on Automated Individual Decision-Making and Profiling for The Purposes of Regulation 2016/679 (Wp251rev.01).* [online] Available from: https://ec.europa.eu/newsroom/article29/items/612053 [Accessed 5 September 2023].

[82] The Council of Europe (2019).

[83] Kamarinou, D. et.al. (2017) Machine Learning with Personal Data. In: Leenes, R. et.al., (eds) *Data Protection and Privacy: The Age of Intelligent Machines.* Hart: Oxford University Press.

[84] New, J. and Castro, D. (2018) *How Policymakers Can Foster Algorithmic Accountability. [online]* Centre for Data Innovation. Available from: https://datainnovation.org/2018/05/how-policymakers-can-foster-algorithmic-accountability/ [Accessed 18 May 2023].

Decision-makers with access to sensitive information must ensure that no group or individual is subjected to a disproportionate share of the risks or rewards associated with using data-driven decisions. Unless adequate governance structures are developed, there is a rising fear that the opaque nature of algorithmic systems could result in circumstances in which individuals are negatively impacted without resorting to a profound explanation and a rectification procedure.[85]

To meet these stringent standards, the decision-making process, the development of AI systems, and the justification for their deployment must be communicated to stakeholders, documented, and audited.[86] Working Party Guidelines on Article 29, call for a system that makes decisions based on algorithms to be constantly tested and given feedback to stop mistakes, inaccuracies, and unfair treatment. Source code, databases, and technical data may not be accessible to individuals but are accessible to regulatory authorities and other parties.[87] This occurs because the concept of transparency may vary depending on the circumstances. The system's manufacturer or operator performs testing to guarantee that it is accurate and fair. In addition to the aforementioned uses, they also allow the testing of whole subsystems by authorised users, the explanation of algorithmic or operational methods by computer scientists and managers, and the submission of findings to regulatory bodies.[88] None of the parts of an algorithmic system should be treated equally with respect to transparency. The algorithmic system's unique characteristics, the complexity of the situations needing governance, and the goals of the governing body all call for diverse applications of this principle.[89]

Kaminski makes a good point when he says, "*It seems that the GDPR is the closest to creating what Frank Pasquale has called 'qualified transparency'* ",

[85] European Parliament Think Tank. (2019) *EU guidelines on Ethics in Artificial Intelligence: Context and implementation*. [online] Available from:

`https://www.europarl.europa.eu/thinktank/en/document/EPRS\_BRI(2019)640163` [Accessed 5 September 2023].

[86] European Commission. (2018) *Guidelines on Automated Individual Decision-Making and Profiling for The Purposes of Regulation 2016/679 (Wp251rev.01)*. [online] Available from:

`https://ec.europa.eu/newsroom/article29/items/612053`[Accessed 5 September 2023].

[87] European Data Protection Board. (2018) Article 29 Working Party. [online] Available from:

`https://edpb.europa.eu/about-edpb/more-about-edpb/article-29-working-party\_en` [Accessed 1 September 2023].

[88] *Ibid*.

[89] New, J. and Castro, D. (2018) *How Policymakers Can Foster Algorithmic Accountability. [online]* Centre for Data Innovation. Available from: `https://datainnovation.org/2018/05/how-policymakers-can-foster-algorithmic-accountability/` [Accessed 18 May 2023].

which is a scheme of targeted disclosures with *"different levels of depth and scope that are meant for different people"*. In practise, transparency does not just mean telling the public what is happening. It also integrates internal company oversight, regulatory oversight, and communication with affected parties. Each of these disclosures may have a distinct character or level of depth. For instance, a board of directors may have access to the full source code, whereas individuals may only have quick, uncomplicated summaries of the information.[90]

The pursuit of transparency can stem from simple curiosity in some cases. However, its true value lies in empowering individuals to comprehend the reasoning behind decisions that negatively impact them. This understanding allows them to question the decision advocate for change and contribute to the development of more accountable and transparent AI systems. Nevertheless, merely granting access to technical details and complex algorithms might overwhelm and prove unhelpful for most people.[91]

As a result, the notion of "qualified transparency" emerges as a nuanced approach. It suggests offering varying degrees of transparency based on the needs and comprehension levels of stakeholders. This approach aligns with calls, for 'meaningful transparency', which goes beyond mere technical disclosures and focuses on providing users with actionable insights.[92]

Tailored Transparency for Diverse Stakeholders:

- **Data Subjects:** Individuals directly affected by algorithmic decisions should have access to clear explanations of the outcome, the factors that influenced it, and the potential for bias. This could involve summaries of the data used, the decision-making process, and the associated risks and limitations.

- **Regulators and Auditors:** Regulatory bodies tasked with overseeing algorithmic fairness and compliance require deeper access to technical details, including algorithm architecture, training data quality, and testing methodologies. This enables them to effectively assess potential risks and ensure adherence to regulations.

[90] Kaminski, M. (2019) The Right to Explanation, Explained *Berkeley Technology Law Journal*, 34, p.194.

[91] Chaudhary, G. (2023) Explainable Artificial Intelligence (xAI): Reflections on Judicial System. *Kutafin Law Review*, 10(4), pp. 872-889. https://doi.org/10.17803/2713-0533.2023.4.26.872-889.

[92] Rai, A. (2020) Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science*, 48, pp. 137–141. https://doi.org/10.1007/s11747-019-00710-5

- **Internal Stakeholders:** Developers, engineers, and managers responsible for designing and maintaining the algorithm need comprehensive access to the inner workings of the system. This allows them to identify and address issues, improve performance, and ensure responsible development practices.

Transparency Mechanisms:

Several mechanisms can be implemented to achieve qualified transparency:

- **Explainable AI (XAI) techniques:** These techniques can provide human- understandable explanations of how algorithms arrive at decisions, making them more interpretable for non-technical audiences.
- **Interactive dashboards and visualizations:** Interactive interfaces can allow users to explore data, understand how different factors influence outcomes, and identify potential biases.
- **Algorithmic impact assessments:** Conducting regular assessments can help identify and mitigate potential negative impacts of algorithms on specific groups or individuals.[93]
- **Clear and accessible communication:** Providing clear and concise communication to users about how their data is used, what decisions are made based on it, and how they can exercise their rights is crucial for building trust and transparency.

While transparency is essential for fostering trust and accountability in AI systems, it must be balanced with other important values such as privacy, security, and intellectual property. For instance, disclosing sensitive trade secrets or user data could have negative consequences. Therefore, it is crucial to carefully consider the potential risks and benefits of transparency before implementing any specific measures.[94]

Qualified transparency, achieved through targeted disclosures and appropriate mechanisms, is not just about satisfying curiosity but about empowering individuals, ensuring fairness, and fostering responsible AI development. By providing the right information to the right stakeholders, we can build AI systems that are not only effective but also accountable and trustworthy.

---

[93] Selbst, A.D. (2021) An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology*, 35(1).

[94] Katyal, S.K. (2022) Democracy & Distrust in an Era of Artificial Intelligence. *Daedalus*, 151(2), pp. 322-334. doi:10.1162/daed_a_01919.

6. THE WAY FORWARD

There is little doubt that this discussion regarding AI transparency and explainability will continue for a considerable time, as AI systems still need to tackle numerous difficulties. The biggest obstacle is switching from 'black box' to 'glass box' models without halting creativity. Every person or group, whether private or public, plays an integral part in this process. Many scientific projects are ongoing in Explainable AI (XAI).[95] Computer scientists have been focussing a lot of their recent work on figuring out the reasons behind decisions made by artificial intelligence, investigating techniques, and developing built-in tools that can perform these tasks and explain them in a way that humans can understand.

Moreover, data processors and controllers must employ particular organisational and technical safeguards[96] to ensure compliance with GDPR standards. It is anticipated that implementing data protection impact assessments (DPIAs) in high-risk activities will dramatically affect AI research and application.[97] The goal of adopting a "*risk-based approach*" to data protection—which includes DPIAs—is to shift the focus from managing data processing to managing risks associated with that processing.[98] Even if the term "*high-risk threshold*" is not precise, most AI and ML applications likely fall under the processing category requiring a DPIA.[99] As a result, DPIAs should be conducted by both the private and public sectors before creating and implementing AI systems and computerised decision-making methods to foresee and prepare for potential risks to human beings. It is also crucial to determine what national supervisory agencies or courts will say about these DPIAs and how they will rule, and how data controllers in different business sectors will interpret and implement the principles of openness and explainability in their DPIAs.

---

[95] Data Guidance. (2022) *Norway - Data Protection Overview*. [online] Available from: https://www.dataguidance.com/notes/norway-data-protection-overview [Accessed 5 September 2023].

[96] Information Commissioner's Office (2017) *Big Data, Artificial Intelligence, Machine Learning and Data Protection*. [online] Available from: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf [Accessed 5 August 2023].

[97] Mitrou, L. (2018) Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'? [online] *Tilburg: TILT Law & Technology Working Paper Series* Available at SSRN: https://ssrn.com/abstract=3386914 or http://dx.doi.org/10.2139/ssrn.3386914 [Accessed 28 August 2023].

[98] *Ibid.*

[99] Information Commissioner's Office (2017) *Big Data, Artificial Intelligence, Machine Learning and Data Protection*. [online] Available from: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf [Accessed 5 August 2023].

On the other hand, regulators and policymakers are anticipated to play an essential role in developing AI. Already, there are calls for policymakers to get involved and technology-specific laws to be enacted. In addition, it is planned to create a regulatory body for algorithmic decision-making whose job will be to develop the standards by which we can distinguish between safe and harmful AI systems. Algorithmic decision-making system providers should also be held to strict and transparent obligations, such as publicising the source code of their systems.[100]

In this approach, regulation is crucial for developing AI because it promotes transparency and openness, reduces disparities and errors, and delivers legal certainty to individuals. On the other hand, many rules or oversights could add to bureaucracy, slow down the development of technology, and make it take longer for artificial intelligence products to be commercially sold. Governments must strike a balance between stifling creativity and the digital revolution, protecting citizens' rights, and addressing unintended consequences. To achieve this objective, policymakers must abandon conventional regulatory frameworks, revaluate current methods, and explore alternative models such as collaborative, hybrid, outcome-based self-regulation, and co-regulation.

6.1. EUROPEAN UNION'S AI ACT

As part of its digital strategy, the European Union has enacted pioneering legislation regulating AI to promote responsible development and adoption of this transformative technology. The new AI Act establishes a risk-based framework that imposes varying obligations on AI providers and users. Although many systems present minimal risk, assessment is required. Adoption of the AI Act represents a momentous decision, constituting the first regulatory regime governing much-discussed AI innovations promising to revolutionize society. Passage was uncertain until the final days, as the French, German, and Italian governments advocated substituting the legislation with a less stringent AI code of conduct. Their rationale was that minimizing compliance burdens for European companies would better position them to compete internationally. However, legislators rejected this path, judging that balanced regulation would also compel global firms to meet the Act's standard's as well. In their assessment, this would enable fairer market competition. With this trailblazing law, Europe asserts leadership in directing AI toward ethical evolution and alignment with societal priorities.

[100] European Parliament Think Tank. (2019) *Understanding algorithmic decision-making: Opportunities and challenges.* [online] Available from:

`https://www.europarl.europa.eu/thinktank/en/document/EPRS\_STU(2019)624261` [Accessed 28 August 2023].

In line with this objective, the Act's definition of AI systems aligns with internationally recognized criteria from OECD guidelines, which characterize such systems as follows:[101]

> "Machine-based systems that, based on explicit or implicit objectives, make inferences from received inputs to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."

The expansive breadth of AI's potentially disruptive reach underscores the need for judicious governance, with the exception of certain domains such as national defence, military, and scientific research, which require tailored policies that balance innovation against ethical risks.

A fiercely debated exception to the AI Act's broad regulatory ambit pertains to systems built on free and open-source software. However, the tightly circumscribed scope of the said exception renders it applicable almost solely to private, non-commercial AI applications. Specifically, the free and open-source software waiver does not apply

- **1.** If the AI system is either
  (a) for high-risk use-case,
  (b) falls under prohibited uses, or
  (c) is a use-case with transparency requirements,[102] and
- **2.** If the free and open-source software licensed system furnishes extensive documentation of its model architecture, training methodology, and other technical particulars, it limits its legal duties to providing said summaries and adhering to copyright strictures. However, the exemption becomes void upon the system's commercial deployment or professional commissioning i.e., it is "made available on the market" or "put into service".

[101] OECD. AI-Principles overview - OECD.AI. The OECD Artificial Intelligence Policy Observatory - *OECD.AI*. [online] Available from: `https://oecd.ai/en/ai-principles` [Accessed 3 February 2024].

[102] European Parliament (2023) EU AI Act: first regulation on artificial intelligence. [online] Available from: `https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence` [Accessed 5 December 2023].

Regardless of whether an AI system qualifies for free and open source software exemptions, those exemptions no longer apply if the system is categorized as GPAI with systemic risks.[103]

A governance framework is implemented for general purpose artificial AI systems and foundation models. General-purpose AI refers to an AI system capable of adaptable functionality across multiple domains. Governance considerations also extend to the integration of general-purpose AI capabilities into supplementary high-risk architectures. Foundation models constitute expansive AI architectures adept at undertaking a diverse range of tasks including video, text, and image generation, lateral natural language processing, mathematical computation, and computer code synthesis.[104]

The AI Act emphasizes safeguarding fundamental rights and promoting transparency, mandating human rights impact evaluations for high-risk AI architectures, including those deployed in the insurance and banking sectors. General purpose AI systems carrying systemic risk implications must meet additional requirements[105]

i. *Risk Management*: Entities must conduct rigorous model assessments harnessing state-of-the-art audit protocols and instruments.

ii. *Red Teaming*: Exhaustive adversarial evaluations must be undertaken and documented thoroughly to unveil and mitigate systemic hazards.

iii. *Cybersecurity*: Robust cybersecurity defences for both the AI model and the supporting physical infrastructure must be instituted.

iv. *Energy Consumption*: Obligatory tracking, logging, and public disclosure of actual or projected energy consumption by the model.

In addition, providers must adhere to Union copyright legislation, integrate technological solutions as necessary, and furnish a comprehensive inventory detailing training data used for model development. Presumably,

[103] Gibney, E. (2024) What the EU's tough AI law means for research and ChatGPT. *Nature*. https://doi.org/10.1038/d41586-024-00497-8.

[104] Süme, O. (2023) The proposed regulation of AI Foundation models and General Purpose AI under the draft European AI Act. [online] *Fieldfisher*. Available from: https://www.fieldfisher.com/en/insights/the-proposed-regulation-of-ai-foundation-models-and-generative-purpose-ai [Accessed 5 February 2024].

[105] Article 6, Classification Rules for High-Risk AI Systems. *EU AI Act* [Online] Available from: https://www.euaiact.com/article/6\#:\textasciitilde:text=An\%20AI\%20system\%20intended\%20to,that\%20product\%20pursuant\%20to\%20above [Accessed 17 February 2024].

the AI Act's recitals shall explicitly delineate that the requisite training data inventories need not enumerate discrete data points, as this would prove excessively onerous.

In addition, compulsory registration in a European database is mandated, which in tandem with the disclosure of materials used for AI system development noted previously, could engender substantial litigation. Specifically, copyright and privacy laws may provide grounds for legal challenges by right holders of content leveraged by AI architectures regarding usage impropriety.

Regarding governance and conformity, the AI Act establishes a European AI Office for oversight of sophisticated AI architectures. A scientific panel and advisory forum will be constituted to assimilate diverse stakeholder insights, enabling continuously informed, contemporary regulatory approaches attuned to AI progress.

However, delegating authority between centralized and localized entity incubates the debate. Specifically, national and local bodies may resist ceding influence per GDPR precedent. While the AI Office may mitigate inconsistent EU-wide approaches, political friction between disparate local authorities persists as a risk.

Finally, these models must approach artificial intelligence that prioritises humans. This implies that they must place human values at the centre of the design, deployment, use, and monitoring of AI systems. These systems will ensure the protection of all fundamental human rights. Respect for human dignity, which claims that every person possesses a distinct and unchangeable moral standing, is the foundation of all these rights.[106] Given recent technological advances with as-yet-unknown or unclear effects for individuals and society, our ethical and legal mission is to find a mechanism to cast light on 'black boxes' in such a system following the Protagorean dictum "*man is the measure of all things*".[107]

7. CONCLUSION

Algorithmic transparency is an indispensable element of the responsible AI development and also very effective usage. With AI still making an impact across many areas of the life, the critical question remains of how

[106] European Parliament Think Tank. (2019) *EU guidelines on Ethics in Artificial Intelligence: Context and implementation*. [online] Available from: `https://www.europarl.europa.eu/thinktank/en/document/EPRS\_BRI(2019)640163` [Accessed 5 September 2023].

[107] Protagoras. (2016.) *Testimonia, Part 2: Doctrine*. [online] Available at: `https://www.loebclassics.com/view/protagoras-doctrine/2016/pb\_LCL531.43.xml` [Accessed 12 February 2022].

these systems decide. Accountability and fairness are underpinned by the understanding the decision making. The crucial aspect of transparency in the AI systems is very much especially noticed in the sensitive areas like employment and also criminal justice, where unfair and biased decisions may cause a huge consequent.

With the GDPR in place, a data controller is now required to prioritize its transparency and structure while giving the people a right to explain the how automated decisions are made. Nevertheless, such transparency mechanism should be in place, and the explanation of the AI system should be defined by the whom is the target audience. GDPR is a landmark in the road for the ensuring the transparency in AI systems, but it is not the whole answer. There is a need for a country-wide accountability regime for algorithms, which entails the data controllers putting in checks and balances to make sure that all the data processing and algorithm systems adhere to the provisions of the applicable data privacy regime.

One of the crucial problems in creating and implementing AI systems is the so-called 'black box issue'. The opaqueness and the unintelligibility of AI systems may result in a lot of bias, discrimination, and also other disadvantageous behaviours. The responsibility for the establishment of meaningful checks and balances lies on the data controllers to be adhered to all the data processing and also algorithmic systems that will conform to the applicable privacy standards. Transparency and explainability of the algorithms is about a fairness for the users, it is not only an individual issue, but a part of a larger accountability framework for the algorithms.

The chances of getting a fair outcome without algorithmic transparency are slim in fields like employment and criminal justice. Racial and unfair decisions could be devastating to individuals and the entire society. Transparency issue in the AI systems can also create the problem of untrustworthiness in the AI systems, which can consequently, hold back the development and use of these systems.

The lawmakers and regulators need to develop means of privacy safeguards for citizens while not hampering technological advancements. The 'black box' issue solves differently as technology advances and will remain the main concern. Hence, it is necessary to make sure that more efforts are made to support the concepts of transparency and accountability of such systems to make sure that they are used in the right manner.

Transparency of algorithms is one of the most essential features of the responsible AI design and application. It is of paramount importance to understand the need for transparency in AI in these sensitive areas, where the implications of biased or discriminatory decisions can have grave

consequences, like employment and criminal justice. The GDPR is a very important soft measure to usher in transparency in AI systems, but not the entire solution on its own. There should be a wider accountability regime for algorithms, in which data controllers would have the right to ensure that there are check and balance mechanisms for all data processing and algorithmic systems that they control, and that these align with the privacy framework. The black box problem is the greatest issue while developing the AI systems and deploying the AI systems which should be always paid attention to and we should improve the transparency and accountability of AI systems in order to be sure that they are being used ethically and effectively.

LIST OF REFERENCES

[1] Ananny, M., and Crawford, K. (2018) Seeing Without Knowing: Limitations of The Transparency Ideal and Its Application To Algorithmic Accountability. *New Media & Society*, 20(3), pp. 973-989.

[2] Balkin, J. (2017) The Three Laws of Robotics in The Age of Big Data, Ohio State Law Journal, 78, p. 1218.

[3] Baratta, R.14 (2014) Complexity of EU law in the domestic implementing process. In: 19th Quality of Legislation Seminar EU Legislative Drafting: Views from those applying EU law in the Member States. *Brussels: European Commission Service Juridique - Quality of Legislation Team*, 3 July. Available from: `https://ec.europa.eu/dgs/legal\_service/seminars/20140703\_baratta\_speech.pdf` [Accessed 21 June 2023].

[4] Bathaee, Y. (2018) The Artificial Intelligence Black Box and The Failure of Intent and Causation, *Harvard Journal of Law and Technology*, 31(2), p. 891.

[5] Brkan, M (2019) Do Algorithms Rule the World? Algorithmic Decision-Making in The Framework of The GDPR And Beyond. *International Journal of Law and Information Technology*, 27(2), pp. 91-121

[6] Brynjolfsson, E. and Mitchell, T. (2017) What can machine learning do? Workforce implications. *Science*, 358(6370), pp. 1530-1534. see also, Data Guidance. (2022) *Norway - Data Protection Overview*. [online] Available from: `https://www.dataguidance.com/notes/norway-data-protection-overview` [Accessed 5 September 2023].

[7] Burrell, J. (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1).

[8] Burt, A (2020) *Is there a 'right to explanation' for machine learning in the GDPR?* [online] International Association of Privacy Professionals. Available from: `https://iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/` [Accessed 5 September 2023].

[9] Castelvecchi, D. (2016) Can We Open the Black Box Of AI? *Nature*, 538(7623), p. 20.

[10] Chaudhary, G. (2020), Artificial Intelligence: The Liability Paradox, *ILI Law Review*, p. 144.

[11] Chaudhary, G. (2023) Explainable Artificial Intelligence (xAI): Reflections on Judicial System. *Kutafin Law Review*, 10(4), pp. 872-889. https://doi.org/10.17803/2713-0533.2023.4.26.872-889.

[12] CNIL. (2021) *How Can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence*. [online] Available from: https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf [Accessed 5 September 2023].

[13] Council of Europe Committee of experts on internet intermediaries (MSI-NET). (2017) *Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications*. [online] Available from: https://rm.coe.int/study-hrdimension-of-automated-data-processing-incl-algorithms/168075b94a [Accessed 5 September 2023].

[14] Court of The Hague. (2020, February 5). SyRI legislation in conflict with higher law. Rechtspraak.nl. [online] Available from: https://uitspraken.rechtspraak.nl/\#!/details?id=ECLI:NL:RBDHA:2020:1878 [Accessed 1 August 2023].

[15] Data Guidance. (2022) *Norway - Data Protection Overview*. [online] Available from: https://www.dataguidance.com/notes/norway-data-protection-overview [Accessed 5 September 2023].

[16] Diakopoulos, N. (2016) Accountability in Algorithmic Decision-Making. *Communications of the ACM*, 59(2), pp. 56-62.

[17] *Digital Republic Act 2016*, Law No. 2016-132117. France. In French.

[18] Donovan, J., Matthews, J., et al. (2018) *Algorithmic Accountability: A Primer*. [online] Data & Society. Available from: https://datasociety.net/output/algorithmic-accountability-a-primer/ [Accessed 1 September 2023].

[19] European Commission (2018) *Communication from the Commission to the European Parliament, the European Economic and Social Committee and the Committee of the Regions*. [online] Available from: https://commission.europa.eu/publications/communication-commission-european-parliament-council-european-economic-and-social-committee-and_en [Accessed 5 September 2023].

[20] European Commission. (2018) *Guidelines on Automated Individual Decision-Making and Profiling for The Purposes of Regulation 2016/679 (Wp251rev.01)*. [online] Available from: https://ec.europa.eu/newsroom/article29/items/612053 [Accessed 5 September 2023].

[21] European Data Protection Board. (2018) Article 29 Working Party. [online] Available from:

[22] European Parliament Think Tank. (2019) *EU guidelines on Ethics in Artificial Intelligence: Context and implementation*. [online] Available from: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2019)640163 [Accessed 5 September 2023].

[23] European Parliament Think Tank. (2019) *Understanding algorithmic decision-making: Opportunities and challenges*. [online] Available from:

[24] European Parliament. (2020) *European Parliament Resolution Of 12 February 2020 on Automated Decision-Making Processes: Ensuring Consumer Protection and Free Movement of Goods and Services*. [online] Available from: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0032_EN.html [Accessed 25 June 2023].

[25] Ferretti, A. et.al. (2018) Machine Learning in Medicine: Opening the New Data Protection Black Box. *European Data Protection Law Review*, 4, p.322.

[26] GDPR §§ Articles 13(2)(f), 14(2)(g).

[27] GDPR. EU (n.d.) *Recital 63, Right of access*. [online] Available from: https://gdpr.eu/recital-63-right-of-access/ [Accessed 15 July 2023].

[28] Goodman, B. and Flaxman, S. (2016) *EU regulations on algorithmic decision-making and a "right to explanation"*. [preprint] arXiv:1606.08813.

[29] Information Commissioner's Office (2017) *Big Data, Artificial Intelligence, Machine Learning and Data Protection*. [online] Available from: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf [Accessed 5 August 2023].

[30] Kamarinou, D. et.al. (2017) Machine Learning with Personal Data. In: Leenes, R. et.al., (eds) *Data Protection and Privacy: The Age of Intelligent Machines*. Hart: Oxford University Press.

[31] Kaminski, M. (2019) The Right to Explanation, Explained Berkeley Technology Law Journal, 34, p.194.

[32] Katyal, S.K. (2022) Democracy & Distrust in an Era of Artificial Intelligence. Daedalus, 151(2), pp. 322-334. doi:10.1162/daed_a_01919.

[33] Klimas, T. and Vaitiukait, J. (2008) The Law of Recitals in European Community Legislation. *ILSA Journal of International and Comparative Law*, 15(1), pp. 65-93.

[34] Lepri, B. et al. (2018) Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4), pp. 611-627.

[35] Meek, C. (2022) *Artificial Intelligence in The Age of Algorithmic Transparency.* [online] Les Echos. Available from: https://www.americangreetings.job.com/news/artificial-intelligence-in-the-age-of-algorithmic-transparency/ [Accessed 12 August 2023].

[36] Mitrou, L. (2018) Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) 'Artificial Intelligence-Proof'? [online] *Tilburg: TILT Law & Technology Working Paper Series* Available at SSRN: https://ssrn.com/abstract=3386914 or http://dx.doi.org/10.2139/ssrn.3386914 [Accessed 28 August 2023].

[37] Mittelstadt, B. D. et al. (2016) The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2).

[38] New, J. and Castro, D. (2018) *How Policymakers Can Foster Algorithmic Accountability. [online]* Centre for Data Innovation. Available from: https://datainnovation.org/2018/05/how-policymakers-can-foster-algorithmic-accountability/ [Accessed 18 May 2023].

[39] O'Neil, C. (2017) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers.

[40] Pasquale, F. (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Massachusetts, p.320.

[41] Protagoras. (2016.) *Testimonia, Part 2: Doctrine*. [online] Available at: https://www.loebclassics.com/view/protagoras-doctrine/2016/pb_LCL531.43.xml [Accessed 12 February 2022].

[42] Rai, A. (2020) Explainable AI: from black box to glass box. Journal of the Academy of Marketing Science, 48, pp. 137–141. https://doi.org/10.1007/s11747-019-00710-5

[43] Samoili, S., Cobo, M.L., et al. (2020) *AI Watch. Defining Artificial Intelligence, Towards an Operational Definition and Taxonomy Of Artificial Intelligence*. EUR 30117 EN, Publications Office of the European Union, Luxembourg.

[44] Schwab, K. (2017) *The fourth industrial revolution*. Crown Publishing Group, New York.

[45] Selbst, A. and Powles, J. (2017) Meaningful Information and The Right to Explanation. *International Data Privacy Law*, 7, p.235.

[46] Veale, M. et.al. (2018) Algorithms That Remember: Model Inversion Attacks and Data Protection Law. *Philosophical Transactions of the Royal Society A*, 376 (2133).

[47] Wachter, S. et al. (2017) Transparent, explainable, and accountable AI for robotics. *Science Robotics*, 2(6), eaan6080.

[48] Wachter, S. et.al. (2017) Why A Right to Explanation of Automated Decision Making Does Not Exist in The General Data Protection Regulation. *International Data Private Law,* 7, p.81.

[49] Wang, P. (2012) Theories of Artificial Intelligence—Meta-Theoretical Considerations. *Atlantis Thinking Machines*, 9, pp. 305–323.

[50] Wischmeyer, T (2020) Artificial Intelligence and Transparency: Opening the Black Box. In: Wischmeyer, T. and Rademacher (eds.) *Regulating Artificial Intelligence.* Switzerland: Springer International Publishing, pp. 75-101.

[51] Z C Lipton, Z.C. (2018) The Mythos of Model Interpretability. In Machine Learning, The Concept of Interpretability Is Both Important and Slippery. *ACMQueue*, 16(3), p.13.