



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE DELLA RICERCA

Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

The Role of the Right to Explanation and Its Safeguards in the Realization of Trustworthy AI

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Availability:

This version is available at: <https://hdl.handle.net/11585/909982> since: 2022-12-19

Published:

DOI: <http://doi.org/>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Yalcin, O.G. (2023). **The Role of the Right to Explanation and Its Safeguards in the Realization of Trustworthy AI**. In: de la Iglesia, D.H., de Paz Santana, J.F., López Rivero, A.J. (eds) *New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence*. DiTTEt 2022. Advances in Intelligent Systems and Computing, vol 1430. Springer, Cham.

The final published version is available online at https://doi.org/10.1007/978-3-031-14859-0_16

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

The Role of the Right to Explanation and Its Safeguards in the Realization of Trustworthy AI*

Orhan Gazi Yalcin¹[0000-0001-7990-6531]

University of Bologna, Via Galliera, 3, 40121, Bologna, Italy
Polytechnic University of Madrid, Av. de Montepríncipe, 28223 Madrid, Spain
orhangazi.yalcin2@unibo.it
<https://last-jd-rioe.eu/>

Abstract. This paper presents a relationship timeline diagram between the GDPR safeguards introduced to secure data subjects' right to explanation and the ethical principles of the Trustworthy AI framework laid out by the High-Level Expert Group. To create the desired output, we initially analyze the articles of the GDPR that establishes the foundation of the right to explanation. Then, we cover the relevant safeguards enabled to secure the right to explanation that should be regarded as an umbrella concept. We analyze the seven ethical principles required for the realization of trustworthy AI and associate them with the relevant safeguards. Finally, a relationship timeline diagram is presented in which the relationship between the safeguards, the articles creating these safeguards, and the corresponding ethical principles protected with these safeguards are demonstrated.

Keywords: The Right to Explanation, GDPR, Trustworthy AI, Explainability, XAI

1 Introduction, Background, Scope, and Definition

In today's data-driven society, previously unknown issues such as profiling and algorithmic decision-making have become an everyday reality. While some fields do not raise concerns, using these technologies in sensitive fields such as law, finance, military, law enforcement, and human resources causes human rights and privacy concerns [10].

Amid growing concerns about automated decision-making systems and profiling of data subjects, in May 2018, European Union's new General Data Protection Regulation, or *Regulation 2016/679*, came into effect, replacing the Data Protection Directive of 1995, or *Directive 95/46/EC* [11]. Although Data Protection Directive was an essential step toward data privacy and protection, it did not provide EU-wide direct enforceability since it was enacted as a directive rather than a regulation. The GDPR - on the other hand- is enacted as a regulation and, therefore, does not require a secondary procedure to be implemented at the national level [6]. Thus, the GDPR embodies direct EU-wide enforceability.

* This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ITN EJD "Law, Science and Technology Rights of Internet of Everything" grant agreement No 814177.

One of the novelties that came with GDPR is much disputed “the right to explanation.” Although the Data Protection Directive created the preliminary version of a right to explanation, GDPR took it to a whole new level. In a narrow definition, the right to explanation refers to a data subject’s right to receive information from the data controller in relation to automated decision-making or profiling [15]. However, the actual scope of the right to explanation is not limited to receiving basic-level information about an AI system. In contrast, it should be regarded as an umbrella right with several safeguards to enhance the trustworthiness of the entirety of the AI systems.

In this paper, we will first review the GDPR articles relevant to the right to explanation and automated decision making -including profiling-, then analyze the suitable GDPR safeguards to ensure the realization of Trustworthy AI, and understand the relationship between these safeguards and the associated ethical principles within the GDPR’s right to explanation framework.

2 Right to Explanation in the European Union

There has been a heated discussion in academia regarding the existence of an effective right to explanation in the EU GDPR. While the predominant stand is on the existence of a right to explanation in the text of GDPR, some scholars claim that the “restrictive, unclear, or even paradoxical” nature of the GDPR makes it unfeasible to trigger any explanation-related right [4]. While a right to explanation is not explicitly stated in the binding articles of GDPR [13], the legal framework articulated by the GDPR embodies several adjacent rights and safeguards, which together may constitute a right to explanation [10]. While Art. 13-15 aims to regulate a right to explanation in case of automated decision making, Art 22 limits the use cases of automated decisions and creates several safeguards in the event of their use.

GDPR Art. 13(2)(f), 14(2)(g), and 15(1)(h) are the provisions that define similar rights and obligations for different scenarios, which create the first part of the right to explanation. These provisions are also the main battlefield between two groups who claim and oppose the existence of a right to explanation in GDPR. While a healthy discussion on this issue is essential to find the best application of the legal framework, the discussion, which initially started between two papers, seems to pay little attention to the text of the relevant GDPR provisions [2]. While the first paper, by Goodman and Flaxman, is for the existence of a groundbreaking and powerful right to explanation, the second paper, by Wachter, Mittelstadt, and Floridi, opposes this idea and asserts that GDPR does not articulate a right to explanation and claims that GDPR sets out other safeguards and rights to protect data subjects [11].

On the other hand, Art. 22(1) gives data subjects the right not to be subject to a decision based solely on automated processing except if the decision is (a) necessary for a contractual relationship, (b) is authorized by the Union or Member State law that lays down suitable safeguard measures to protect data subject’s rights, freedoms, and legitimate interests, and finally, (c) based on the explicit consent of the data subject. For the application of Art. 22(1)(a) and (c), the data subject must be provided with

suitable safeguards to obtain human intervention, express his or her point of view, and finally, contest the decision [10].

Recital 71 goes one step further and extends the safeguards with two additional rights: (i) the right to challenge the decision and (ii) the right to obtain an explanation of the decision reached after assessment. Although Recital 71 is not directly enforceable [5], Article 15(1)(h) implicitly creates the right to obtain an ex-post explanation [10].

When read together, according to Articles 13(2)(f), 14(2)(g), 15(1)(h), and Article 22, in the event where a data subject is subject to a “decision based solely on automated processing”, “which produces legal effects concerning him or her or similarly significantly affects him or her”, he or she has a right to “meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject” [11], as demonstrated in **Fig. 1**:

| GDPR Article | Scope |
|---------------|---------------------------------------------------------------------------------------------|
| Art 13(2) (f) | Information to be provided where personal data are collected from the data subject |
| Art 14(2)(f) | Information to be provided where personal data have not been obtained from the data subject |
| Art 15(1)(h) | Right of access by the data subject |
| Art 22(3) | Automated individual decision-making, including profiling |
| Recital 71 | Profiling |

Fig. 1. The GDPR Articles Relevant to the Right to Explanation

There are two main requirements to trigger a right to receive meaningful information: (i) the decision should be made based solely on automated processing and (ii) this decision should produce a legal effect or affect the data subject significantly. When interpreting the first requirement, the “solely” component should not be interpreted too rigid to avoid causing the right to be ineffective. An insignificant level of human intervention should not deem this right ineffective. Additionally, the second requirement clearly states that a decision does not have to create a legal effect on the data subject, in the event where the decision creates a significant economic or social effect concerning the data subject, a right to explanation can be triggered. When these requirements are fulfilled, a data subject can exercise his or her right to receive meaningful information. But a right to explanation should also be meaningful and impactful [6].

According to Selbst and Powles, there are four components to having a meaningful and impactful right to explanation. First, the “meaningful information” should be interpreted subjectively based on the data subject since Art. 13-15 particularly aims at data subjects. Second, meaningful information should either have instrumental or intrinsic value or enable a possible action. Thirdly, meaningful information should provide enough functionality to guarantee the data subject’s exercise of rights. Finally, meaningful information requirements should be interpreted based on the facts of the case to avoid hampering innovation and R&D efforts in this field [11]. While the first

three components are in favor of data subjects ((a) requiring explanations with protective interpretation, (b) providing instrumental or intrinsic value or enabling a possible action, and (c) having enough functionality to guarantee the exercise of rights), the fourth component tries to balance the impact of the first three to ensure the continuation of innovation in this field which leads to long-term prosperity and wealth. Only by actively managing and respecting these components, we can truly achieve responsive competitiveness [3].

3 Safeguards around Right to Explanation

As mentioned earlier, an explanation for an automated decision should either have instrumental or intrinsic value or enable a possible action. In case of enabling a possible action, suitable safeguards mentioned in Art 13-15, Art 22(4), and Recital 71 have become important references for interpretation [7]. While Art 13, 14, and 15 define the right to obtain information about automated decision-related processing as a common safeguard, Article 22 explicitly mentions -in a non-exhaustive wording [8]- three safeguards regarding automated decision making: (i) the right to obtain human intervention on the part of the controller, (ii) the right to express one's point of view, and (iii) to contest the automated decision. Additionally, Recital 71 further expands the list of suitable safeguards with (a) the right to challenge the automated decision and (b) the right to obtain an explanation of the automated decision. Therefore, we can create a non-exhaustive list of safeguards – as shown in **Fig. 2**- to ensure fairness and transparency of data processing where there is automated decision-making.

| Safeguard | Legal Basis |
|---------------------------------------------------------------------|-------------------------------------------|
| The right to obtain information about automated decisions | GDPR Art 13(2)(f), 14(2)(f), and 15(1)(h) |
| The right to contest/challenge the automated decision | GDPR Art 22(3), Recital 71 |
| The right to express one's point of view | GDPR Art 22(3) |
| The right to obtain human intervention | GDPR Art 22(3) |
| The right to obtain an explanation of the decision after assessment | GDPR Recital 71 |

Fig. 2. The Safeguards relevant to the Right to Explanation and Their Legal Basis in the GDPR

3.1 The right to obtain information about automated decisions.

The first and most important safeguard for the right to explanation is the right to obtain information about automated decision-making. This safeguard requires data controllers to provide information on (i) the existence of automated decision-making, (ii) meaningful information about its logic, and finally (iii) the significance and the envisaged consequences of such automated decisions. This safeguard does not require data controllers to provide a specific explanation for a particular decision. Instead, it requires an explanation of the general mathematical logic used in the decision-making

process [13]. In literature, scholars often limit the scope of the right to explanation merely to this safeguard [8]. However, in a broader sense, this safeguard does not cover the true boundaries of the right to explanation and secure the relevant ethical principles.

3.2 The right to contest/challenge the automated decision

While Article 22(3) mentions the right to “contest” a decision, Recital 71 takes a step further and mentions the right to “challenge” the automated decision. While contesting simply refers to adjusting or reviewing a decision, challenging a decision refers to requesting to identify the inadequateness of the decision to deem it ineffective. Despite the claims made by some scholars about the existence of the right to explanation, some argue that the mere existence of a right to contest a decision as a safeguard requires a right to explanation [1]. Without prejudice to the unbinding nature of Recital 71, we can assume that when there is an automated decision that creates a significant or legal effect, the data subject is provided with the right to contest/challenge the automated decision [10].

3.3 The right to express one’s point of view

Expressing one’s opinion is itself a safeguard as well as a component of another safeguard, the right to contest/challenge an automated decision. Expression of one’s opinion has a different significance for contesting and challenging a decision. In contesting a decision, data subjects may express their opinion to change the outcome of the decision whereas expressing an opinion during the challenge of a decision can make the decision null and void. Therefore, the importance of expressing one’s point of view can create different legal results.

3.4 The right to obtain human intervention

Obtaining human intervention is a safeguard explicitly stated in GDPR Art. 22(3) and many EU Member States laws [10]. This safeguard ensures that when a decision is successfully contested or challenged by the data subjects, they will not be subject to another almost identical automated decision. Besides, this safeguard should be enforced with the right to express one’s point of view since, without the data subject’s point of view, the chance to eliminate algorithmic bias may be difficult, which corresponds to damaging the fairness element of the automated decision-making process explicitly laid out in Articles 13(2) and 14(2). It is important to note here that Art 22 applies to cases where a decision is based solely on automated processing. Although some scholars argue that using spurious human involvement limits the applicability of this safeguard and the other safeguards mentioned in Art 22 [5], adding a layer of ineffective human oversight should not be regarded as an effective medium to bypass this safeguard.

3.5 The right to obtain an explanation of the decision after assessment

Apart from the right to explanation in a general sense, Recital 71 of the GDPR also mentions an extended right to explanation after other safeguards (e.g., requesting human intervention, contesting/challenging a decision, and expressing an opinion regarding a particular decision) are triggered. There are claims that since the term right to obtain explanation is used under non-binding Recital 71, some scholars claim that the GDPR does not create a binding right to explanation in Articles 13-15 and Article 22 [14]. However, the right to obtain an explanation under Recital 71 is merely a subcategory of the right to explanation in question. This subcategory only covers the explanation about a decision after a safeguard triggers an assessment of this decision.

In addition, a valid assessment will require a review of the architecture and implementation of the algorithm. Therefore, this safeguard will require an explanation of the architecture and implementation of the algorithm, which is usually referred to as the “ex-ante” explanation in addition to the specific explanation about a particular decision.

4 Ethical Principles on Explainability and Right to Explanation

Apart from the legal response to the issues brought about by the mass adoption of AI systems, scholars and policymakers often refer to digital ethics principles, which contribute to the protection of fundamental rights and freedoms. Almost every major institution along with big tech companies published their frameworks to address today and tomorrow’s ethical problems [12]. While many of these frameworks contain common themes, this paper primarily considers the Trustworthy AI principles laid out by the European Commission’s High-Level Expert Group. The High-Level Expert Group Report offers guidelines designed to guide the AI community towards lawful, ethical, and robust AI practices [3].

One of the main arguments of the High-Level Expert Group for a system to be trustworthy is that we should be able to understand why it behaved in a certain manner and how it output the decisions. Explainable AI (i.e., XAI) is the up-and-coming AI subfield that tries to understand how AI systems behave in general or are related to individual decisions [3]. XAI can contribute to the realization of Trustworthy AI, particularly by helping the satisfaction of some of the seven key principles.¹ Although explainability is not one of these seven principles in the Trustworthy AI framework, some of these principles are directly associated with the explainability of automated decision-making systems and the relevant GDPR safeguards.

Some of these principles are explicitly mentioned under the relevant GDPR articles. Article 13(2) and 14(2) clearly states that data subjects should be provided with some information about the automated decision-making process to ensure “fairness” and “transparency”. Therefore, we can safely assert that two of the ethical goals with the

¹ The seven requirements are (i) human agency and oversight, (ii) technical robustness and safety, (iii) privacy and data governance, (iv) transparency, (v) diversity, non-discrimination, and fairness, (vi) societal and environmental wellbeing, and (vii) accountability.

right to explanation are to guarantee fairness and transparency of the data processing related to automated decision-making.

Transparency is the most relevant Trustworthy AI principle that has a direct link to explainability. In the High-Expert Group report, explainability is placed under the transparency principle along with traceability and communication. Furthermore, the transparency principle requires transparency in all three pillars of AI systems (i.e., the data, the system, and the business model) [3]. Therefore, ideally, a transparent automated decision-making system should provide explanations about the data it uses, its technical logic, and finally, its business model.

Although the wording of GDPR Art 13-15 uses the term ‘fairness’, the High-Level Expert Group expands this principle to ‘Diversity, Non-Discrimination, and Fairness.’ By expanding the inclusiveness and diversity throughout the entire AI system’s life cycle, we can contribute to the realization of trustworthy AI. Discrimination and bias can occur at the algorithm level as well as the data and the business level. From bad data collection to inadvertent historic bias, incompleteness, and bad governance models, there is a variety of possibilities that unintentional bias is embedded into automated decision-making systems. Additionally, one may also encounter intentional discrimination [3]. At each stage of the life cycle, all five safeguards must be enabled and actionable as they are all suitable for the reinforcement of the fairness element of the automated decisions.

The accountability principle consists of audibility and minimization of negative impact. Therefore, to strengthen the accountability of automated decision-makers, we need to enable suitable safeguards that can be useful to detect negative issues such as algorithmic bias and discrimination. The assessment of algorithms, data, and design processes is particularly important when data subjects contest/challenge an automated decision [3]. Additionally, after a particular automated decision, the data subject’s right to obtain an explanation requires this explanation to provide an adequate level of accountability feature. Apart from the assessment of the processes, if a potential issue is detected, the data controllers must enable measures to minimize its negative impact.

Finally, human agency is another important principle that interacts with the safeguards reviewed in this article. AI systems should enable mechanisms to receive external feedback about their performance on fundamental rights and human autonomy. In a trustworthy AI system, the data subject should be able to make informed decisions about the AI system. The right not to be subject to a decision based solely on automated processing in case of legal or significant consequences is a consequence of this principle. Apart from encouraging human autonomy, an AI system should also enable safeguards to allow for human oversights. Therefore, we can say that the right to obtain human intervention is a suitable tool to ensure human oversight of AI systems. On the other hand, all the safeguards support human agency; therefore, contribute the human autonomy in automated decisions and the consequences.

Fig. 3 shows the trustworthy AI principles matched with the safeguards defined in the GDPR articles relevant to the right to explanation:

| Transparency | Diversity, Non-Discrimination, and Fairness | Accountability | Human Agency and Oversight |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> The right to obtain information about automated decisions The right to obtain an explanation of the decision after | <ul style="list-style-type: none"> The right to obtain information about automated decisions The right to obtain human intervention The right to express one's point of view The right to contest/challenge the automated decision Right to obtain an explanation of the decision after assessment | <ul style="list-style-type: none"> The right to contest/challenge the automated decision The right to obtain an explanation of the decision after assessment | <ul style="list-style-type: none"> The right to obtain human intervention |

Fig. 3. The Right to Explanation Safeguards that directly affect Trustworthy AI Principles

While four of the seven principles are directly associated with the safeguards reviewed in this paper and therefore, with the right to explanation, the remaining three principles are not within the direct scope of the safeguards in question. However, other safeguards and technical & organizational measures are still effective mediums to ensure (i) technical robustness and safety, (ii) privacy and data governance, and (iii) societal and environmental wellbeing where there is an AI system and automated decision-making.

5 The Effect of Right to Explanation on Trustworthy AI

In the light of the above explanation, we can finally create a relationship timeline diagram of the GDPR articles, the safeguards, and ethical principles that have an association with the right to explanation due to automated decision-making as shown in Fig. 4:

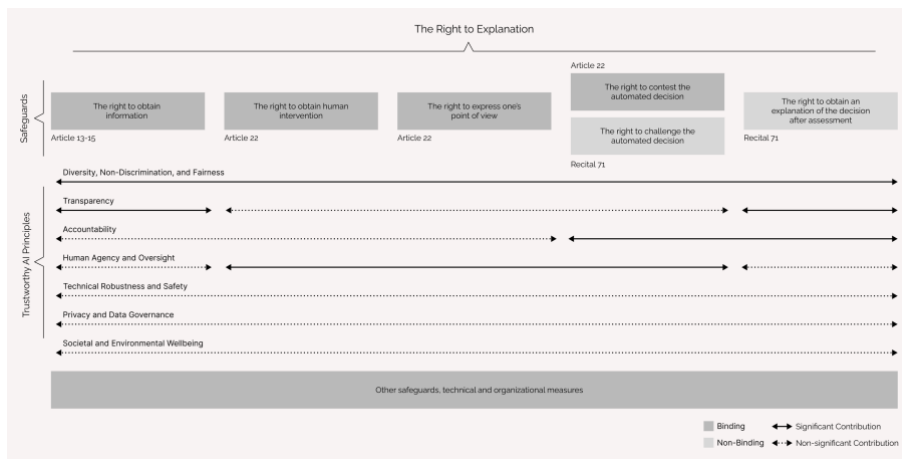


Fig. 4. The relationship between GDPR articles, safeguards, and trustworthy AI principles

In Figure 1, dark gray boxes represent the safeguards that are mentioned in the binding articles of the GDPR whereas the light gray boxes represent the non-binding safeguards mentioned in Recital 71 of the GDPR. While dashed arrowed lines represent the non-

significant contribution of the safeguards to the principle, solid arrowed lines represent a significant contribution. It is important to note that dashed lines may mean still mean some contribution, but the level of the contribution of the safeguard on the ethical principle in question would be limited. Additionally, GDPR lays out several safeguards and technical & organizational measures that can strengthen all seven principles of realizing Trustworthy AI. However, they are not included in **Fig. 4** since they do not directly aim at automated decisions and the right to explanation.

The order of the safeguards is created in a timeline structure. Although data subjects may skip one or more of them, in an automated decision-making process, the safeguards are likely to be triggered from the left to the right. We enter the realm of the right to explanation with the right to obtain information laid out in Art. 13-15. After receiving the initial information from the data controller, if the data subjects are not satisfied with the information and the decision, they can trigger the human intervention, express opinions, and contest the decision safeguards mentioned in Art. 22. Art. 22 lists these three safeguards and does not limit the potential safeguards with these three by using the term 'at least'. Recital 71 further expands these safeguards with the right to challenge a decision and the right to obtain an explanation. Although Recitals are not binding, combined with the non-exhaustive listing of Art 22 safeguards, we can safely assume that Recital 71 safeguards will be taken very seriously by the administrative authorities and judiciary when the binding Art. 13-15 and Art. 22 are interpreted.

We can also see to what extent safeguards contribute to the seven principles of Trustworthy AI. All five safeguards are effective measures to strengthen the 'diversity, non-discrimination, and fairness' principle as they require a considerable amount of reasoning, human interaction, and model assessment. The transparency principle is mainly strengthened by the right to obtain information defined in Art 13-15 and the right to obtain an explanation defined in Recital 71. They require data controllers to provide information from the existence of automated decisions to the inner logic of the models, and by some interpretations, even to specific explanations about a particular decision, which enhance the transparency property of the AI systems. These safeguards also force data controllers to use explainable models from the beginning to comply with them. Although each step contributes to the accountability feature, the right to contest/challenge a decision and the right to obtain an explanation may open a direct channel to administrative or judicial bodies, which increases the accountability of the AI systems. Finally, obtaining human intervention, expressing one's point of view, and contesting/challenging a decision will certainly contribute to strengthening the human agency and overview principle. The other principles are not directly affected by these safeguards, but when we interpret the provisions of GDPR (primarily, Art. 13-15, Art. 22, and Recital 71), we should take them into consideration. For example, adopting a too rigid interpretation for explanation requirements can damage the technical robustness of the AI systems. In addition, when designing solutions for compliance with these safeguards, data controllers should also consider the 'privacy and data governance' and 'societal and environmental wellbeing' principles for a sustainable business model.

6 Conclusion

In today's digital society, automated decisions are becoming more widely used in every part of our lives. Companies, institutions, and governments take advantage of AI systems to increase their revenue, profitability, and service quality. Although some of these systems do not pose threats to the individuals whose data have been processed and used for automated decision-making and profiling, some of these systems can output automated decisions that can produce significant social, economic, or legal outcomes. In such events, GDPR lays out several safeguards specifically aimed at addressing the issues that may surface due to automated decisions along with many other general-purpose safeguards that can also be used for the protection of fundamental rights and freedoms. These specific safeguards and the rights created around these safeguards create a comprehensive framework for the right to explanation. This ecosystem of safeguards also serves for the realization of Trustworthy AI by strengthening the ethical principles described by the High-Level Expert Group of the European Commission. In this paper, we analyze GDPR's legal framework which contains the abovementioned rights and safeguards around the right to explanation. Then, we cover the ethical principles that are highlighted by the High-Level Expert Group. As a result, we create a relationship timeline diagram to show the timeline of an automated decision process with the relevant safeguards suitable at each stage and their contribution to the seven principles of Trustworthy AI.

References

1. Bayamlioğlu, E. (2021). The right to contest automated decisions under the General Data Protection Regulation: Beyond the so-called "right to explanation." *Regulation & Governance*. <https://doi.org/10.1111/REGO.12391>.
2. Casey, B., Farhangi, A., & Vogl, R. (2019). Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise. *Berkeley Technology Law Journal*, 34, 145. <https://doi.org/10.15779/Z38M32N986>.
3. Directorate-General for Communications Networks, C. and T. (European C. (n.d.). Ethics guidelines for trustworthy AI. In *Publications Office of the EU*. Retrieved April 24, 2022, from <https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>.
4. Edwards, L., & Veale, M. (2017). Slave to the Algorithm? Why a Right to Explanation is Probably Not the Remedy You are Looking for. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2972855>.
5. Edwards, L., & Veale, M. (2018). Enslaving the Algorithm: From a "right to an Explanation" to a "right to Better Decisions"? *IEEE Security and Privacy*, 16(3), 46–54. <https://doi.org/10.1109/MSP.2018.2701152>.
6. Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision making and a "right to explanation." *AI Magazine*. <https://doi.org/10.1609/aimag.v38i3.2741>.
7. Hamon, R., Junklewitz, H., Malgieri, G., Hert, P. de, Beslay, L., & Sanchez, I. (2021). Impossible explanations?: Beyond explainable AI in the GDPR from a COVID-19 use case

- scenario. *FACCT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 549–559. <https://doi.org/10.1145/3442188.3445917>.
8. Kaminski, M. E., Bertolini, A., Brennan-Marquez, K., Comandé, G., Cushing, M., Helberger, N., van Drunen, M., van Eijk, N., Eskens, S., Malgieri, G., Price, N., Sax, M., & Selbst, A. (2019). The Right to Explanation, Explained. *BERKELEY TECHNOLOGY LAW JOURNAL*, 34, 189. <https://doi.org/10.15779/Z38TD9N83H>.
 9. Kim, T. W., & Routledge, B. R. (2022). Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach. *Business Ethics Quarterly*, 32(1), 75–102. <https://doi.org/10.1017/BEQ.2021.3>
 10. Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other “suitable safeguards” in the national legislations. *Computer Law & Security Review*, 35(5), 105327. <https://doi.org/10.1016/J.CLSR.2019.05.002>.
 11. Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, 7(4), 233–242. <https://doi.org/10.1007/s13347-017-0263-5>.
 12. Siau, K., & Wang, W. (2020). Artificial intelligence (AI) Ethics: Ethics of AI and ethical AI. *Journal of Database Management*, 31(2), 74–87. <https://doi.org/10.4018/JDM.2020040105>.
 13. Winikoff, M., & Sardelic, J. (2021). Artificial Intelligence and the Right to Explanation as a Human Right. *IEEE Internet Computing*, 25(2), 108–112. <https://doi.org/10.1109/MIC.2020.3045821>.
 14. 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*, 1–47. <https://doi.org/10.2139/SSRN.2903469>
 15. Zingales, N. (2021). Right to Explanation. In *Glossary of Platform Law and Policy Terms* (pp. 279–281). <https://hdl.handle.net/10438/31365>.