Law and Business

“Just” Algorithms: Justification (Beyond Explanation) of Automated Decisions Under the General Data Protection Regulation

Gianclaudio Malgieri

Augmented Law Institute, EDHEC Business School

Abstract

This paper argues that if we want a sustainable environment of desirable AI systems, we should aim not only at transparent, explainable, fair, lawful, and accountable algorithms, but we also should seek for “just” algorithms, that is, automated decision-making systems that include all the above-mentioned qualities (transparency, explainability, fairness, lawfulness, and accountability). This is possible through a practical “justification” statement and process (eventually derived from algorithmic impact assessment) through which the data controller proves, in practical ways, why the AI system is not unfair, not discriminatory, not obscure, not unlawful, etc. In other words, this justification (eventually derived from data protection impact assessment on the AI system) proves the legality of the system with respect to all data protection principles (fairness, lawfulness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, and accountability). All these principles are necessary components of a broader concept of just algorithmic decision-making and is already required by the GDPR, in particular considering: the data protection principles (Article 5), the need to enable (meaningful) contestations of automated decisions (Article 22) and the need to assess the AI system necessity, proportionality and legality under the Data Protection Impact Assessment model framework. (Article 35).

Keywords

AI • justification • automated decision-making • explainability • Algorithmic Impact Assessment

Introduction

The regulation of automated decision-making (ADM) in the General Data Protection Regulation (GDPR) is a topic of vivid discussion. If the first commentators focused mostly on the existence of a right to an explanation in the body of the Regulation, the following discussion has focused more on *how* to reach a good level of explainability or, even better, a good level of algorithmic accountability and fairness.

This paper argues that if we want a sustainable environment of desirable ADM systems, we should aim not only at transparent, explainable, fair, lawful, and accountable algorithms, but we also should seek for “just” algorithms, that is, automated systems that include all the above-mentioned qualities (transparency, explainability, fairness, lawfulness, and accountability). This is possible through a practical “justification” statement and process through which the data controller *proves*, in practical ways, the *legality* of an algorithm with respect to all data protection principles (fairness, lawfulness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, and accountability). Indeed, as this article shows, all these principles are necessary components of a broader concept of *just* algorithmic decision-making.

This justificatory approach might be not only comprehensive of many other fragmentary approaches proposed thus far in the legal or computer science literature, but also might solve many existing problems in the artificial intelligence (AI) explanation debate, for example, the difficulty in “opening” black boxes, the transparency fallacy, and the legal difficulties in enforcing a right to receive individual explanations.

Section 2 mentions the general rules about profiling and ADM in the GDPR, while Section 2.1 refers to the debate about the interpretation of those rules. Then, Section 3 addresses the definition and limits of the concept of “explanation,” and Section 4 tries to propose some preliminary, tentative solutions to those limits—adopting a systemic accountability approach. Developing upon these first elements, Section 5 introduces the concept of Algorithm Justification, while Section 6 contextualises this

[†] Corresponding author: Gianclaudio Malgieri

E-mail: gianclaudio.malgieri@edhec.edu

Open Access. © 2021 Gianclaudio Malgieri, published by Sciendo

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

concept in the legal field and Section 7 in the GDPR field, explaining on which bases a “justificatory” approach is not only useful but also necessary under the GDPR. Further developing this challenge, Section 8 addresses how an ADM justification should be conducted, considering in particular the data protection principles in Article 5. Finally, Section 9 proposes a practical “justification test” that could serve as a first basis for data controllers who want to justify ADM data processing under the GDPR rules.

The GDPR Rules about Automated Decision-Making

The GDPR has tried to provide a solution to risks of automated decision-making through different tools: a right to receive/access meaningful information about logics, significance, and envisaged effects of automated decision-making processes (Articles 13(2), lett. f; 14(2), lett. g; and 15(1), lett. h) and the right not to be subject to automated decision-making (Article 22) with several safeguards and restrains for the limited cases in which automated decision-making is permitted.

Article 22(1) states as follows: “the data subject shall have the right not to be subject to *a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.*” This right shall apply almost always in case of sensitive data¹ (Art. 22(4)). For other personal data, ADM shall not apply in only three cases:

- a. the decision “is necessary for entering into, or performance of, a contract between the data subject and a data controller”;
- b. “is authorised by Union or Member State law to which the controller is subject and which also lays down *suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests*”; or
- c. “is based on the data subject’s explicit consent” (Art. 22(2)).

In cases (a) and (c) from the above list “the data controller shall implement *suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision*” (Art. 22(4)). In addition, recital 73 explains that such suitable safeguards “should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to *obtain an explanation* of the decision reached after such assessment and to challenge the decision.”

In sum, in case of a “decision based solely on automated processing, including profiling, which produces legal effects concerning [data subjects] or similarly significantly affects [them],” individuals have two different safeguards:

1. *The right to know the existence* of that processing and *meaningful information* about its *logic, significance, and consequences*.
2. *The right not to be subject to that processing*, unless in specific cases (precontractual processing, explicit consent of data subjects) where other *appropriate safeguards* must be provided, such as (at least):
 - a. *the right to obtain human intervention* from the controller;
 - b. *the right to express his or her point of view*;
 - c. *the right to contest the decision* (or “challenge” it, as referred at recital 73);
 - d. eventually, the *right to “obtain an explanation* of the decision reached after such assessment.” However, this right is not included in the body of Article 22, but only in the explanatory recital 71.

Debate and Interpretations

The interpretation of the GDPR rules about automated decision-making has generated a vivid debate in legal literature. Several authors have interpreted this net of provisions as a new right to algorithm explanation²; while other scholars have adopted a more sceptical approach analysing limits and constraints of the GDPR provisions³ and concluding that the data subject’s rights are more limited than expected and that there is no right to explanation.⁴ In addition, other scholars have preferred a contextual interpretation of Articles 13(2)(f), 14(2)(g), 15(1)(h), and 22 of the GDPR, suggesting that the scope of

¹ For sensitive data we refer to “special categories of personal data” according to Article 9(1). This exemption does not apply in case of point (a) or (g) of Article 9(2) (i.e., sensitive data given with explicit consent of data subject or processing necessary for reason of substantial public interest) when “suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests are in place.”

² Bryce Goodman and Seth Flaxman, 2016. “EU Regulations on Algorithmic Decision-Making and a ‘Right to Explanation’” arXiv:1606.08813 [cs, stat] <http://arxiv.org/abs/1606.08813>, accessed 30 June 2018.

³ See, e.g., Lilian Edwards and Michael Veale, “Slave to the Algorithm? Why a ‘Right to an Explanation’ Is Probably Not the Remedy You Are Looking For,” 16 *Duke Law & Technology Review* 18 (2017). Available at SSRN: <https://ssrn.com/abstract=2972855>.

⁴ Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, “Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation,” *International Data Privacy Law* 7, no. 2 (2017): 76–99, <https://doi.org/10.1093/idpl/ixp005>.

those provisions is not so limited and that they actually can provide individuals with more transparency and accountability.⁵ This last view was also partially confirmed by Article 29 Working Party, which has released some guidelines on profiling and automated decision-making.⁶

However, scholars have proposed different ways to address the issue of suitable safeguards, as indicated in Article 22(3). To mention a few examples, some scholars proposed a model of counterfactual explanations, that is, a duty to clarify for individuals targeted by automated decisions, among others, "what would need to change in order to receive a desired result in the future, based on the current decision-making model."⁷ Other scholars proposed a "legibility model" for automated decisions, where transparency and comprehensibility should⁸ exist so that individuals are able to understand autonomously (readability) the importance and implications (comprehensibility) of algorithmic data processing.⁹ Looking at the broader picture of the GDPR, some scholars proposed a more dynamic link between existing data protection rights (access, erasure, rectification, portability, etc.) in order to react to adverse effects of automated decisions¹⁰ or to also focus on the dualistic nature of the GDPR, based both on individual rights and on a multilevel design of algorithms (co-governance).¹¹ A corollary of this proposal is a system of multilayered explanations based on a data protection impact assessment of algorithms.¹²

Definition and Limits of "Explanation" of AI

Most of these proposed solutions are based on transparency, explainability, and interpretability of artificial intelligence.¹³ In general terms, *explaining* automated decision-making is a complex task. Many commentators have first questioned about the notion of explanation.¹⁴ In general terms, explaining means making an idea or a situation clear to someone by describing it in more detail or revealing relevant facts.¹⁵ In other words, the explanation is an act of spotting the main reasons or *factors* that led to a particular consequence, situation, or decision.

In the field of Computer Science, explanation of AI has been referred to as making it possible for a human being (designer, user, affected person, etc.) to understand a result or the whole system.¹⁶ Tim Miller, analysing the structure and expectations of explanations, identified four characteristics of explanations:¹⁷ (a) *contrastive*, i.e., mostly in response to some counterfactuals;¹⁸ (b) *selected*, i.e., not comprehensive, but based only on the few main factors that influenced the final decision; (c) *causal*, rather than correlational/statistical; and (d) *social* and *contextual*, i.e., depending on the specific social relations and contexts at stake. As affirmed in legal theory, an explanation attempts to render a situation or a process understandable under a *causal* or *intentional* perspective.¹⁹ The causal nature of explanation is based on the link between cause and effect (What are the causes

- 5 Andrew D. Selbst and Julia Powles, "Meaningful Information and the Right to Explanation," *International Data Privacy Law* 7, no. 4 (2017): 233–42, <https://doi.org/10.1093/idpl/ix022>; Gianclaudio Malgieri and Giovanni Comandé, "Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation," *International Data Privacy Law* 7, no. 4 (2017): 243–65, <https://doi.org/10.1093/idpl/ix019>. See also, Margot E. Kaminski, "The Right to Explanation, Explained," 34 *Berkeley Tech. L.J.* (2019): 189.
- 6 Article 29 Working Party, "Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679," WP251rev.01, adopted on 3 October 2017, as last revised and adopted on 6 February 2018; Lilian Edwards and Michael Veale, "Slave to the Algorithm?," (2017).
- 7 Sandra Wachter, Brent Mittelstadt, and Chris Russell, "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law & Technology* 31, no. 2 (2018). Available at SSRN: <https://ssrn.com/abstract=3063289> or <http://dx.doi.org/10.2139/ssrn.3063289>.
- 8 See Richard Mortier and others, 'Human-Data Interaction' in The Interaction Design Foundation (ed), *The Encyclopedia of Human-Computer Interaction*, (2nd edition, The Interaction Design Foundation 2015) <<https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed>> accessed 26 July 2021.. See also, Malgieri and Comandé, Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation," p.14. For a different meaning of *legibility*, see Luke Hutton and Tristan Henderson, "Beyond the EULA: Improving Consent for Data Mining," in Tania Cerquitelli, Daniele Quercia, and Frank Pasquale (eds.), *Transparent Data Mining for Big and Small Data* (Springer, New York 2017), 147 at 162.
- 9 Gianclaudio Malgieri and Giovanni Comandé, "Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation," *International Data Privacy Law* 7, no. 243 (2017).
- 10 Lilian Edwards and Michael Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For," 16 *Duke Law & Technology Review* 18, 2017
- 11 Kaminski, "The Right to Explanation, Explained," (2019).
- 12 Margot Kaminski and Gianclaudio Malgieri, "Multi-Layered Explanation from Algorithmic Impact Assessments in the GDPR," *FAT 2020 Proceedings* (ACM Publishing, 2020).
- 13 See e.g., in general, Andrew D. Selbst and Solon Barocas, "The Intuitive Appeal of Explainable Machines," 87 *Fordham Law Review* 1085 (2018).
- 14 Tim Miller, "Explanation in Artificial Intelligence: Insights from the Social Sciences," 267 *Artificial Intelligence* 1, (2019).
- 15 Lexico, "Explain," <https://www.lexico.com/definition/explain>.
- 16 Clément Henin and Daniel Le Métayer, 'A Framework to Contest and Justify Algorithmic Decisions' [2021] *AI and Ethics* <<https://doi.org/10.1007/s43681-021-00054-3>>.
- 17 Miller (n 14) 4.
- 18 Sandra Wachter, Brent Mittelstadt, and Chris Russell, "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law & Technology* (2018), <http://arxiv.org/abs/1711.00399>, accessed 16 September 2019.
- 19 Aulis Aarnio, *The Rational as Reasonable: A Treatise on Legal Justification* (Springer Science & Business Media 1986) 22.

behind this decision?); while its intentional nature is based on the motives of the actor and her beliefs regarding reality (What are purposes or intentions behind this decision?). Considering these two sides of the coin, the explanation is the “answer to the question of *why* something happened or why someone acted as he did.” Said in other terms, an explanation is a framework for understanding the action that has happened.²⁰

The GDPR (and, in particular, the provisions in Article 22 and recital 71) are often interpreted as referring to only “one” kind of explanation. Actually, there is no unique explanation in practice;²¹ each form of explanation highly depends on the context at issue.²² More importantly, the capability to give a fair and satisfactory explanation also depends on the possibility to show *causal* link between the input data (and, in particular, some crucial factors within the input information) and the final decision. However, this is not always possible. While for traditionally data-based decision-making it might be easier to give adequate explanations, addressing the causes, the determining factors, and the counterfactuals; in more complex AI-based decisions, it might be hard to reach this high level of explainability. Indeed, looking at the quick development of deep learning in different forms of automated decisions (even COVID-19 automated diagnosis based on, for example, lungs images), explaining the specific reasons and factors of an individual decision might be nearly impossible.²³ An explanation which is neither causal nor contextual is perhaps inadequate to show to the data subject eventual grounds for challenging the decision and then is unsuitable under Article 22(3) of the GDPR.

These last considerations may lead to an insurmountable dichotomy: either we *prohibit* more technologically advanced and inscrutable decision-making systems because they cannot comply with the GDPR explainability requirements; or we *tolerate* AI-based decision-making systems that do not formally respect the transparency duties in the GDPR.

In addition, explanations are not only sometimes problematic, but also not sufficient to make AI socially and legally “desirable.” In particular, several scholars reflected upon the “transparency fallacy” of algorithmic explanation,²⁴ that is, the risk that even a meaningful explanation could be not effectively received or understood by the data subjects (due to its technical nature or to the limited attention, interests, or—even temporary—cognitive capabilities of the data subject).²⁵

The Shift from Explanation to the Bigger Accountability Picture

To overcome the above-mentioned limits of AI explanation, a possible solution might be to look at the broader picture of the GDPR. Article 22(3) and recital 71, when mentioning the possible measures to make automated decisions more accountable, do not address only the right to an individual explanation, but several other complementary tools (e.g., the right to contest and the right to human involvement and algorithmic auditing). In particular, there are several principles and concepts that might influence the interpretation of accountability duties also in case of algorithmic decision-making: the fairness principle (Article 5(1)(a)), the lawfulness principle (Article 5(1)(a)), the accuracy principle (Article 5(1)(d)), the risk-based approach (Articles 24, 25, 35), and the data protection impact assessment model (Article 35).

Also looking at these last provisions, a *justification* of automated decisions taken is not only more *feasible* but also more *useful* and *desirable* than an explanation of the algorithm.²⁶ Justifying a decision means not merely explaining the logic and the reasoning behind it, but also explaining why it is a legally acceptable (correct, lawful, and fair) decision, that is, why the decision complies with the core of the GDPR and is, thus, based on proportional and necessary data processing, using pertinent categories of data and relevant profiling mechanisms.

This justification process will be addressed in the next section. However, at this moment we can already affirm that justification and explanation are not necessarily in conflict with each other. When explanations are not satisfactory or feasible, the data controller should implement some complimentary accountability tools.²⁷ In a previous paper, the author and a co-author

²⁰ Id..

²¹ Miller, (n 14); Selbst and Barocas, (n 13).

²² Henin and Le Métayer, “A Multi-Layered Approach for Interactive Black-Box Explanations,” 38.

²³ See Ronan Hamon and others, ‘Impossible Explanations? Beyond Explainable AI in the GDPR from a COVID-19 Use Case Scenario’, *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2021) <<https://doi.org/10.1145/3442188.3445917>> accessed 27 May 2021.

²⁴ Edwards and Veale, (n 3); Lilian Edwards and Michael Veale, “Enslaving the Algorithm: From a ‘Right to an Explanation’ to a ‘Right to Better Decisions’?” 16 *IEEE Security & Privacy* 46 (2018).

²⁵ See e.g., Gianclaudio Malgieri and Jędrzej Niklas, “The Vulnerable Data Subject,” 37 *Computer Law & Security Review* (2020).

²⁶ Sandra Wachter and Brent Mittelstadt, “A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI,” *Columbia Business Law Review* 2 (2019) available at <https://papers.ssrn.com/abstract=3248829>; Margot Kaminski, Binary Governance: Lessons from the GDPR’s Approach to Algorithmic Accountability, 92 S. Cal. L. Rev. 1529 (2019):12–17, available at <https://scholar.law.colorado.edu/articles/1265>.

²⁷ Edwards and Veale, “Enslaving the Algorithm,” (n 24).

proposed to disclose meaningful information about a Data Protection Impact Assessment (DPIA) on the algorithmic decision-making system—the DPIA, as mentioned in Article 35 of the GDPR, is a process to assess and mitigate the impact of data processing operations on fundamental rights and freedoms of data subjects.²⁸ This paper, in addition to that proposal, introduces a practical description of a possible *justification test* on the algorithm, where the data controller explains why the algorithm (analysed on the aggregated final effects on different data subjects, but also analysed in its purposes, intentions, etc.) is not unfair, unlawful, inaccurate, beyond the purpose limitation, and so forth.

Justification Beyond Explanation of ADM

Before describing the practicalities of a possible justification model and before exploring the advantages of this approach, it is useful to understand what justification means in general as well as in the legal field, particularly with regard to data protection. In general terms, a justification is an action to prove or show something (a person, an action, opinion, etc.) to be *just*, right, desirable, or reasonable.²⁹ Actually, the meaning of justification acquires different shadows in different fields. For example, in theology the justification is the action of declaring or making "righteous" in the sight of God.³⁰ Similarly, in philosophical terms, the justification of decision-making that affects human agents and human societies means proving (under a utilitarian or deontological basis) whether a theory or an opinion reaches desirable goals according to the accepted values (utilitarian, deontological, etc.).³¹ In scientific terms, justifying means proving that a theory or a statement is correct and verified through the scientific method.³²

While the explanation, as mentioned above, aims to make humans understand why a decision was taken, a justification aims at convincing that that decision is "just" or "right" (following the different benchmarks of rightness in different fields).³³ In different terms, while explanations are *descriptive* and *intrinsic* because they only depend on the system itself, justifications are *normative* and *extrinsic* because they are grounded on external references, namely a "norm" according to which we can assess the validity of the decision.³⁴ This means that a justification requires two elements: (1) the reference *norm* and (2) the *proof* that that case or decision applies to that norm.

The proof can follow logical reasoning standards, while the "norm" depends on the specific context at issue. As shown above, the norm can be based on theological, philosophical (utilitarian, deontological, etc.), scientific (scientific method) and, of course, legal grounds. Indeed, in legal terms, justification means proving that a certain action or act respects the current law and, more generally, the *legality* principle.³⁵

Actually, as Loi and colleagues argue,³⁶ the two-dimensional justification that we mention above (norm and proof) should be of a hybrid nature. In particular, the norms also can be from different sources (e.g., utilitarian and legal): a decision-maker can justify a decision on her "primary goals" based on utilitarian norms, (i.e., business objectives), but she also is asked to justify her decision on "constraining goals" imposed by law and, thus, based on legal norms (or other ethical values), such as privacy, fairness, and so forth.³⁷ Justifying a decision on the primary goals aims to show that the decision is not morally arbitrary, while justifying it on the constraining goals aims to prove the legality of that decision.

28 Kaminski and Malgieri, (n 12).

29 Lexico, "Justification," <https://www.lexico.com/definition/justification> (accessed on 25 November 2020).

30 Lexico, "Justification," <https://www.lexico.com/definition/justification> (accessed on 25 November 2020).

31 See, in general, Larry Alexander and Michael Moore, "Deontological Ethics" in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020, Metaphysics Research Lab, Stanford University 2020) <https://plato.stanford.edu/archives/win2020/entries/ethics-deontological/> (accessed 1 December 2020).

32 Paul K Moser, "Justification in the Natural Sciences" *The British Journal for the Philosophy of Science* 39 (1991): 557–75. <<I confirm this note>>

33 Or Biran and Courtenay V. Cotton, "Explanation and Justification in Machine Learning: A Survey," *paper/Explanation-and-Justification-in-Machine-Learning-%3A-Biran-Cotton/02e2e79a77d8aabc1af1900ac80ceebac20abde4*, accessed 26 November 2020.

34 Clément Henin and Daniel Le Métayer, 'A Framework to Contest and Justify Algorithmic Decisions' [2021] *AI and Ethics* <<https://doi.org/10.1007/s43681-021-00054-3>>.

35 Aarnio, (n 19) 8; Mireille Hildebrandt, *Law for Computer Scientists and Other Folk* (Oxford University Press 2020), 267.

36 Michele Loi, Andrea Ferrario, and Eleonora Viganò, "Transparency as Design Publicity: Explaining and Justifying Inscrutable Algorithms" in *Ethics and Information Technology*, <https://doi.org/10.1007/s10676-020-09564-w>, accessed 30 November 2020.

37 Id., at 8.

The Legal Approach to Justification

Returning to the notion of *legal* justification, scholars proposed different approaches to it,³⁸ in particular observing judgments and the reasoning behind judicial acts, which have a function for appeal.³⁹ In general terms, there are legal positivist approaches (the valid law in itself is a sufficient justification) and more balanced ones (a justification lies on a balance between the letter of the law and other grounds having significance in the decision-making).⁴⁰ A more balanced approach might better solve different issues related to the law's open nature and the defeasible nature of legal justification (if additional information is taken into account, the status of a conclusion can change).⁴¹ These considerations also are evident in criminal law, where the "justification" is an exception to the prohibition of committing certain offenses that renders a nominal violation of the criminal law lawful and therefore exempt from criminal sanctions. In doing so, such a justification balances a general legal norm with other contextual interests at issue.⁴²

In sum, while an explanation tends to clarify only why a decision was taken (on which "primary goals," and on which practical interests and needs it was taken),⁴³ a "legalistic" justification usually tends to focus on the mere written law, without a contextual consideration of the balance of interests.

Both these approaches appear incomplete to our purposes (justification of algorithmic decisions). A desirable justification should not merely show the compliance with the "law," but with the *core* or essence of the legal principles, that is, with the *legality* principle.⁴⁴ As we will argue below, the core of data protection in the GDPR is summarised in the data protection principles in Article 5. Accordingly, justifying an automated decision-making under the data protection *goals* and *norms* means, at the least, showing the respect with the principles of data protection in Article 5.

Justification in the GDPR: On Which Basis It Might Be Requested (or Encouraged)

In the GDPR we observe several references to justification of data processing in general, and of automated decision-making in particular. In different parts of the GDPR, when there is a prohibition (e.g., the prohibition to repurpose the data processing, as stated in Article 5(1)(b); the prohibition to process sensitive data, as stated in Article 9(1); the prohibition to conduct automated decision-making, as stated in Article 22(1); the prohibition of transferring data outside the European Union, as mentioned in Article 44, etc.), there is always a list of exceptions, often accompanied by some safeguards to protect fundamental rights and freedoms of the data subject. This combination of exceptions and safeguards is the basis of what we can consider a *justification*. In addition, in these cases the GDPR often refers to the "principles of data processing" as the overarching norm or goal with which the data controller needs to comply in order to *justify* the legality of some nominally illegal acts (see, e.g., recital 72 about profiling or recital 108 about data transfer).

We might observe another strong example of justification in the GDPR: it is the case of high-risk data processing (Article 35). Under the Data Protection Impact Assessment (DPIA) model, data controllers must prove the legal proportionality and necessity of the data processing, and thus the legal necessity and proportionality of eventual automated decisions taken (Article 35(7)(d)). This may constitute a form of *justification of data processing* on the basis of legality and legitimacy, aiming at the "essence" of data protection.⁴⁵

In addition, the Article 29 Working Party Guidelines on profiling recommend that data controllers (in order to comply with Articles 13–15) explain the pertinence of categories of data used and the relevance of the profiling mechanism.⁴⁶ Assessing whether the data used are pertinent and the profile is relevant for a decision, as well as assessing the necessity and proportionality of the data processing in an automated decision-making system, seems to constitute a call for justification. The purpose of such

38 See, in general, Aarnio, (n 19); Arno R. Lodder, *Dialaw: On Legal Justification and Dialogical Models of Argumentation* (1999^o edizione, Kluwer Academic Publishers, 1999).

39 Aarnio, (n 19) 6.

40 Id., at 8.

41 Lodder, (n 38).

42 J. C. Smith, *Justification and Excuse in the Criminal Law* (Stevens 1989); Donald L Horowitz, "Justification and Excuse in the Program of the Criminal Law," *Law and Contemporary Problems* 49 (1986): 109.

43 See also Kiel Brennan-Marquez, " 'Plausible Cause': Explanatory Standards in the Age of Powerful Machines," 70 *Vanderbilt Law Review* 70, no. 53 (2017)<<Au: Pls supply date>>: 1288; Kaminski, (n 26) 1545.

44 On the link between legality and justification, see Hildebrandt, (n 35) 267.

45 Dariusz Kloza et al., "Data Protection Impact Assessment in the European Union: Developing a Template for a Report from the Assessment Process," (LawArXiv 2020) DPiALab Policy Brief 29 available at <https://osf.io/7qrpf>, accessed 1 December 2020.

46 See Article 29 Working Party, Opinion on Automated Individual Decision-Making, Annex, p. 30.

assessment is not only transparency about the technology and its processes, but a justification about the lawfulness, fairness, necessity, accuracy, and legitimacy of certain automated decisions.⁴⁷

Interestingly, empirical research revealed that justification of algorithms (defined as showing the fairness of goals and rationales behind each step in the decision) is the most effective type of explanation in changing users' attitudes toward the system.⁴⁸

The Grounds for Algorithmic Justifications in the GDPR: The Principles in Article 5

While some scholars have already addressed the need for justification of automated decision-making (rather than a mere need for explanation), very few authors have tried to clarify *what* this ADM justification should be and *how* it should be conducted under the GDPR rules. This article argues that, considering the meaning of "legal justification," as mentioned in the previous sections, justifying an algorithmic decision should *prove the legality of that decision*. For "legality," we mean not just lawfulness, but also accountability, fairness, transparency, accuracy, integrity, and necessity.

In recent years, scholars have called for fair algorithms,⁴⁹ or accountable algorithms,⁵⁰ or transparent algorithmic decisions,⁵¹ or, again, for lawful, accurate, and integrous automated decisions. Justifying ADM means calling for algorithmic decision processes that prove to have *all* the aforementioned characteristics and respect the *essence* or the core of data protection.⁵² The author argues that the essence of data protection in the GDPR consists of the data protection *principles* cited in Article 5. Accordingly, *justifying* automated decisions means proving that they comply (or adjusting them in order to comply) with the data protection principles in Article 5.

Interestingly, the principles of data protection seem to lead to the desirable characteristics of automated decision-making, as mentioned above. We will now analyse them one by one, contextualising them to the case of algorithmic decision-making.

Article 5(1)(a) refers to lawfulness, transparency, and fairness. As regards to *lawfulness*, automated decision-making should be lawful, that is, it should have a legal ground and respect fundamental rights and freedoms. Such a legal basis can be found not only in Article 6(1) (or in Article 9(2) in case of special categories of personal data), but also in Article 22. Since Article 22(1) is interpreted as a *prohibition* of automated decision-making,⁵³ in order to make it lawful it is necessary to prove that one of the exceptions in Article 22(2) (consent, contract, European Union, or national law) applies, with the related requirements in Article 22(3) (suitable measures to safeguard the data subject's rights, including at least the right to human intervention, to express his or her point of view, and to contest the decision). This part of "justification" is the most formal one: The controller needs to *justify* why an activity which is apparently unlawful (profiling individuals or making significant decisions on automated bases) is instead lawful. In this sense, this part of justification refers to the legal justification in criminal law, as mentioned above.⁵⁴

As regards to *fairness* justification, the data controller should prove that the decision-making processing is fair, that is, it is nondiscriminatory, unbiased, nonmanipulative, and that, in general, it does not exploit a significant imbalance between

47 Kaminski and Malgieri (n 12) 8.

48 Biran and Cotton (n 33); Kaminski (n 26) 1548; Tom R. Tyler, "Procedural Justice, Legitimacy, and the Effective Rule of Law," *Crime and Justice* 30, no. 283, (2003): 317–18, according to whom a justification based on a fair process is more likely to make people accept a decision.

49 Future of Privacy Forum, "Unfairness by Algorithm: Distilling the Harms of Automated Decision-Making," (2017) available at <https://fpf.org/2017/12/11/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/>, accessed 8 February 2020; Sainyam Galhotra, Yuriy Brun, and Alexandra Meliou, "Fairness Testing: Testing Software for Discrimination," *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering—ESEC/FSE 2017*, (ACM, 2017) <http://dl.acm.org/citation.cfm?doid=3106237.3106277>, accessed 31 May 2019; Andrew D. Selbst et al., "Fairness and Abstraction in Sociotechnical Systems," *Proceedings of the Conference on Fairness, Accountability, and Transparency* (ACM, 2019) <http://doi.acm.org/10.1145/3287560.3287598>, accessed 16 September 2019.

50 Joshua Kroll et al., "Accountable Algorithms," *University of Pennsylvania Law Review* 165 (2017): 633.

51 Bruno Lepri et al., "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes," *Philosophy & Technology* 31 (2018): 611; Bilyana Petkova and Philipp Hacker, "Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers," *Lecturer and Other Affiliate Scholarship Series* available at <https://digitalcommons.law.yale.edu/ylas/13> (2016); Mireille Hildebrandt, "Profile Transparency by Design? Re-Enabling Double Contingency," available at https://works.bepress.com/mireille_hildebrandt/63/, accessed 3 January 2019.

52 Here *essence* is used in a general sense; we don't refer to the essence of the fundamental right to data protection as interpreted by the European Union Court of Justice in application of Article 52 of the Charter. For a specific analysis about this topic, see Maja Brkan, "The Essence of the Fundamental Rights to Privacy and Data Protection: Finding the Way Through the Maze of the CJEU's Constitutional Reasoning," *German Law Journal* 20 (2019): 864.

53 WP29, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, 23: "Article 22(1) sets out a general prohibition on solely automated individual decision-making with legal or similarly significant effects, as described above. This means that the controller should not undertake the processing described in Article 22(1) unless one of the following Article 22(2) exceptions applies (. . .)." See Michael Veale and Lilian Edwards, "Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling," *Computer Law & Security Review* 34 (2018): 398.

54 See, e.g., Smith (n 42).

the controller and the subject in particular contexts (vulnerable individuals).⁵⁵ In general, the algorithmic processing should not violate the expectations of the subjects,⁵⁶ and its effects should not impair human dignity, autonomy, safety, or other fundamental rights set out in the EU Charter of fundamental rights.⁵⁷

As regards to *transparency* justification, the data controller should prove that the algorithmic processing is legible⁵⁸ in the sense that, at least, meaningful information about the logic, the significance, and envisaged consequences of the decision-making are communicated to the subject at the beginning of the data processing (Articles 13(2)(f) and 14(2)(g)) and, upon request, after the processing has started (Articles 15(1)(h)). As argued in another article,⁵⁹ there are at least three levels of possible transparency: general (or “global”) information, group-based explanation, or individual (or “local”) explanation (implementing recital 71). Each level of transparency should depend on the level of risk of that algorithmic decision-making process.⁶⁰ This multilayered approach already has been discussed and endorsed in the field of computer science.⁶¹

Article 5(1)(b) refers, then, to *purpose limitation*. According to this principle, the justification should also prove that the ADM system is based only on data collected for the specific (licit and declared) purpose of obtaining an automated decision affecting the data subject. Under a broader perspective, the purpose limitation justification also should clarify that the algorithm was not originally developed for other purposes (military, commercial, etc.) and then eventually repurposed for the processing at stake.⁶² This would help to prevent algorithmic biases based on a decontextualisation of algorithms.⁶³

Article 5(1)(c) mentions the principle of *data minimisation*. Under this principle, the justification of the data controller should prove that the ADM is based on the processing of only data that are adequate, relevant, and limited to what is necessary for the purpose of taking that automated decision. To present an example, if the controller is an employer who needs to hire a new employee and she declares that the automated decision-making processing has the purpose of selecting the worthiest candidate, any information about, for example, the sexual orientation, ethnic origin, religion, or the possibility of taking maternity leave (fertility, marital status, etc.), are unnecessary and should not be collected. This might be a way also to prevent intentional discrimination⁶⁴ hidden through “masking”⁶⁵—when the data controller tries to cover intentional discrimination behind the shield of data analytics. In those cases, the data minimisation justification could be helpful. At the same time, it is helpful when the processed data are not explicitly about protected categories of information but also could reveal information that might potentially lead to discrimination.⁶⁶

Article 5(1)(d) refers to data *accuracy*. When justifying ADM, accuracy is also fundamental. The data controller should prove that the algorithmic decision is correct and accurate. Recital 71 (addressing ADM) requires data controllers to make sure “that factors which result in *inaccuracies* in personal data are corrected and the risk of errors is minimised.”⁶⁷ Indeed, accuracy (of input data and of the final product decision) has generally been considered one of the main elements to justify the use of certain algorithms.⁶⁸ WP29 has referred to inaccuracy as one of the main issues of automated decision-

55 Damian Clifford and Jef Ausloos, “Data Protection and the Role of Fairness,” *Yearbook of European Law* 37 (2018): 130; Gianclaudio Malgieri, “The Concept of Fairness in the GDPR: A Linguistic and Contextual Interpretation,” *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2020) available at <https://doi.org/10.1145/3351095.3372868>, accessed 29 January 2020.

56 See, e.g., Information Commissioner’s Office, *Big Data, Artificial Intelligence, Machine Learning and Data Protection*, 2017, 22. See also Michael Butterworth, “The ICO and Artificial Intelligence: The Role of Fairness in the GDPR Framework,” *Computer Law & Security Review* 34 (2018): 257.

57 See, in particular, European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics, and related technologies (2020/2012(INL)), Annex I, Article 5.

58 Malgieri and Comandé (n 9).

59 Kaminski and Malgieri (n 12).

60 Margot E. Kaminski and Gianclaudio Malgieri, “Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations,” 19–28 U of Colorado Law Legal Studies Research Paper available at <https://papers.ssrn.com/abstract=3456224>, accessed 28 October 2019.

61 Karthikeyan Natesan Ramamurthy et al., “Model Agnostic Multilevel Explanations,” available at <https://arxiv.org/abs/2003.06005v1>, accessed 25 March 2020; Henin and Métayer (n 22).

62 See the importance of purposes in the legibility test proposed in Malgieri and Comandé (n 9) 259–60.

63 Jonida Milaj, “Privacy, Surveillance, and the Proportionality Principle: The Need for a Method of Assessing Privacy Implications of Technologies Used for Surveillance,” *International Review of Law, Computers & Technology* 30 (2016): 115, 116; on the contextual nature of decision-making processes (and of explanation) see Miller (n 14) 5.

64 Pauline T. Kim, “Data-Driven Discrimination at Work,” 58 Wm. & Mary L. Rev. (2017), 857.

884.<<Au: Pls verify citation: 857 is the first page, 884 is the page I want to refer to>>

65 See this definition in Solon Barocas and Andrew D. Selbst, “Big Data’s Disparate Impact,” *California Law Review* 104 (2016): 671, 692 referring to employers using data analytics to intentionally discriminate as a shield or a mask to better discriminate against a protected group; see also Cynthia Dwork and Deirdre K. Mulligan, “It’s Not Privacy, and It’s Not Fair” (2013) 66 Stanford Law Review 6. See also Kroll et al., (n 50) 682.

66 Sandra Wachter, “Affinity Profiling and Discrimination by Association in Online Behavioural Advertising,” (Social Science Research Network 2019) SSRN Scholarly Paper ID 3388639 <https://papers.ssrn.com/abstract=3388639>, accessed 2 June 2019.

67 Italics added.

68 Kroll et al., (n 50) 684. See the importance of accuracy (compared to interpretability) in Cynthia Rudin, “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead,” *Nature Machine Intelligence* 1 (2019): 206, 207. See also Zachary C. Lipton, “The Myths of Model Interpretability,” *Communications of the ACM* 61 (2018): 36.

making, since these errors in data or in the ADM process itself might result in "incorrect classifications; and assessments based on imprecise projections that impact negatively on individuals."⁶⁹ To make a practical example, the European Bank Authority, in its report on advanced analytics, has given great importance to data accuracy for justifying algorithms in the bank sector and has developed that concept through different subconcepts: accuracy and integrity, timeliness, consistency, and completeness of data.⁷⁰ The accuracy justification should result not only in proving the accuracy of input data, but also proving that the chosen algorithm is fit-for-purpose, that is, produces accurate results. Indeed, often discriminatory decisions also are inaccurate and incorrect.⁷¹ Empirical studies also confirm that the "usefulness" of an algorithmic decision is a key component in its social acceptance.⁷²

Article 5(1)(e) mentions the principle of *storage limitation*. Although in the field of ADM this principle seems not so pertinent, its function is important. This principle requires that data should be stored for no longer than necessary for the purpose of the processing: This time limitation also should apply to algorithmic decision-making. In other words, ADM should not be based on data that are no longer necessary (e.g., outdated data) for the purpose and the context of the decision. At the same time, controllers should not use algorithms that are no longer necessary for the declared purposes.

Article 5(1)(f) mentions the principle of *integrity and confidentiality*. In the context of ADM, it is central that algorithmic decisions are integrous and do not lead to cybersecurity risks that could adversely affect the safety (or any other fundamental right or freedom) of the data subject. Recital 71 also indirectly refers to these "risks" when mentioning automated decisions. However, cybersecurity, safety, and integrity are central elements to consider when justifying algorithms. A "just" algorithm is based on and produces integrous data, and does not endanger the (digital or physical) safety of the data subject.⁷³ The EU Parliament approved a resolution according to which any high-risk artificial intelligence and related technologies, including "algorithms and data used or produced by such technologies, shall be developed, deployed and used in a manner that ensures that they are: developed, deployed and used in a resilient manner so that they ensure an adequate level of security by adhering to minimum cybersecurity baselines proportionate to identified risk, and one that prevents any technical vulnerabilities from being exploited for malicious or unlawful purposes."⁷⁴

The last principle in Article 5 is *accountability* (Article 5(2)). Accountability of ADM is an overarching goal that is considered the final objective of legally desirable AI, in particular in the data protection framework.⁷⁵ This is a "meta-principle," that is, a methodology to apply and implement all the other data protection principles in Article 5. We can identify two perspectives of accountability justification in the GDPR: a practical perspective and a methodological one. The practical accountability justification should demonstrate that the data controller has proactively implemented some suitable ADM measures under Article 22(3) and recital 71,⁷⁶ that she is ready to make data subjects exercise their ADM-related rights (within and beyond Article 22), and that those rights are effective—the right to contest the algorithm, for example, should be made effective through clear information about the system⁷⁷ and the decision, and there should be concrete technical or organisational steps to take into account the eventual data subjects' contestation, to comply with it or to explain why such a request is unreasonable.⁷⁸

On the other hand, the methodological perspective of accountability indicates *how* the justification should be conducted, that is, how the justificatory auditing should be carried out (see below) and what the *legal approach* to justification should be. In

69 Article 29 Working Party, Opinion on Profiling, 23.

70 See European Bank Authority, EBA Report on Big Data and Advanced Analytics, Annex I. See also, more specifically, Benjamin T. Hazen et al., "Data Quality for Data Science, Predictive Analytics, and Big Data in Supply Chain Management: An Introduction to the Problem and Suggestions for Research and Applications," *International Journal of Production Economics* 154 (2014): 72.

71 Julia Dressel and Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism," *Science Advances* 4 (2018): eaao5580.

72 Theo Araujo et al., "In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence," *AI & Society* 35 (2020): 611, 616.

73 European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics, and related technologies (2020/2012(INL)), § 26: "Stresses that the protection of networks of interconnected AI and robotics is important and strong measures must be taken to prevent security breaches, data leaks, data poisoning, cyber-attacks and the misuse of personal data, (. . .); calls on the Commission and Member States to ensure that Union values and respect for fundamental rights are observed at all times when developing and deploying AI technology in order to ensure the security and resilience of the Union's digital infrastructure."

74 European Parliament resolution of 20 October 2020, Annex I, Article 8(1)(a).

75 See, e.g., Bryce Goodman, "A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection" (2016); Dillon Reisman et al., "Algorithm Impact Assessment: A Practical Framework for Public Agency Accountability," (AI Now Institute: 2018); Sonia K. Katyal, "Private Accountability in the Age of Artificial Intelligence," *UCLA Law Review* 66 (2019): 88; Kroll et al., (n 50); Kaminski (n 26); Lepri et al., (n 51).

76 See, e.g., Antoni Roig, "Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)," *European Journal of Law and Technology* 8 (2018) available at <http://ejlt.org/article/view/570>, accessed 15 January 2019.

77 See Article 29 Working Party, Opinion on Profiling, 27, arguing, e.g., that "The data subject will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis." See also Margot Kaminski and Gianclaudio Malgieri (n 12) 4.

78 See, e.g., the UK example of "procedural safeguards" about contestations in Gianclaudio Malgieri, "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, no. 105327 (2019): 9–11.

particular, the accountability principle—as Article 5(2) indicates—put the burden of proving data processing compliance on the data controller.⁷⁹ This means that there is a rebuttable presumption (*praesumptio iuris tantum*) that the data processing activity at stake—and, thus, any ADM processing as well—is not compliant with the data protection principles. The burden of proof about legality is on the data controller.⁸⁰ In other terms, we should consider that algorithmic decisions are illegal by default, unless the data controller *justifies* them through a valid justification process, meant both as a process of justificatory auditing and an eventual final justification statement.

The ADM Justification Test

After having explained what the ADM should be, in terms of content and approach, this section is going to present a possible practical example of a justification test (and the related justification statement for data subjects and Data Protection Authorities). In a previous paper, the author and a co-author proposed a “legibility test” for automated decisions under the GDPR.⁸¹ Other scholars have proposed an Algorithmic Impact Statement⁸² for accountable decision-making on personal data. This paper, with the aim to include those different experiences in a broader perspective, proposes an algorithmic *justification statement*, based on a justification test.

Such a justification test might act as an initial framework for conducting Algorithmic Auditing under a legal (and not merely technical) perspective. This test also might be the basis for conducting DPIA on automated decision-making, in particular for what concerns the vague reference to the “assessment of the necessity and proportionality of the processing operations in relation to the purposes” (Article 35(7)(b)). The WP29 Guidelines on DPIA have proposed some criteria for an acceptable DPIA. These guidelines explain that the “assessment of the necessity and proportionality of the processing operations” also implies a good and comprehensive implementation of—inter alia—data protection principles in Article 5 (namely the lawfulness, purpose limitation, data minimisation, and storage limitation principles).⁸³

Following the structure of Article 5 and the framework discussed in the previous section, a possible *ADM Justification test* might be as follows:

1. *ADM lawfulness justification:*
 - a. Does the ADM data processing have a lawful basis under Article 6(1)?
 - b. Is the ADM based on or produce special categories of data?
 - i. If yes, does the ADM data processing have a lawful basis under Article 9(2)?
 - c. Is the ADM based on one exception in Article 22(2)?
 - d. Is the ADM equipped with suitable safeguards as required in Article 22(3) and with the safeguards eventually required by Member States legislation?
2. *ADM fairness justification:*
 - a. Is the ADM nondiscriminatory? How can the controller ensure the nondiscriminatory nature of the ADM result? Does the controller employ anti-discriminatory auditing on a regular basis (e.g., on a statistical basis)?
 - b. Is the ADM nonmanipulative? How can the controller ensure the nonmanipulative nature of the ADM result? Does the controller employ anti-manipulative auditing on a regular basis (e.g., on a statistical basis)?
 - c. Does the ADM exploit individual vulnerabilities through excessive power imbalance? How can she ensure that this does not happen?
3. *ADM transparency justification:*
 - a. Has the controller provided meaningful information about the logic, significance, and the envisaged consequences of the ADM?
 - b. Is the controller ready to provide a clear general, group-based, or individual explanation to enable the data subject to challenge the decision?

⁷⁹ Information Commissioner's Officer, “Accountability and Governance,” (1 October 2020) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/>, accessed 29 November 2020.

⁸⁰ Raluca Oprișiu, “Reversal of ‘the Burden of Proof’ in Data Protection | Lexology,” available at <https://www.lexology.com/library/detail.aspx?g=e9e8c734-23d9-41bb-a723-5d664b3c86cc>, accessed 29 November 2020.

⁸¹ Malgieri and Comandé (n 9) 264.

⁸² Andrew D. Selbst, “Disparate Impact in Big Data Policing,” *Georgia Law Review* 52 (2018): 109; Reisman et al., (n 76); Kaminski and Malgieri, (n 61).

⁸³ Article 29 Working Party, Guidelines on DPIA, Annex II, 21.

4. *ADM purpose-limitation justification*:
 - a. Is the purpose of the decision-making processing licit, clearly determined, and declared to the subject?
 - b. Is the ADM processing based on data collected solely for that declared purpose?
 - c. How does the data controller ensure that re-purposing of data is avoided in that ADM system?
 - d. Was the ADM developed for other different purposes?
 - e. Was the ADM eventually trained on data collected for other purposes?
 - f. If the answer to points d) and/or e) is yes, how does the controller ensure that the ADM processing has been adjusted in order to avoid biases?
5. *ADM data minimisation justification*:
 - a. Is the ADM based solely on data that are adequate and necessary for the declared purpose of the automated decision?
 - b. Does the ADM system produce decisions that are strictly necessary for the declared purpose and the context?
6. *ADM accuracy justification*:
 - a. Is the ADM based on accurate data?
 - b. Does the ADM produce accurate results?
 - c. How can the data controller ensure that points a and b are respected (e.g., on a regular basis)?
7. *ADM storage limitation justification*:
 - a. Is the ADM based solely on data that are still necessary (e.g., not outdated data) for the purpose and context and the decision?
 - b. Is the ADM processing based on algorithms that are still necessary for the declared purposes?
8. *ADM integrity and confidentiality justification*:
 - a. Is the ADM based on integrous data?
 - b. Is the ADM processing resilient enough to protect digital and physical safety of data subjects?
 - c. Are cybersecurity risks adequately assessed and mitigated?
9. *ADM accountability justification*:
 - a. Are all data protection safeguards and rights (related to ADM) adequately implemented?
 - b. Are these rights made effective? For example, is the right to challenge the decision enabled by clear explanation to the subject? Are there organisational steps to “put a human in the loop”? Are there organisational steps to comply with a challenge request?

Conclusion

In recent years, legal scholars and computer scientists have discussed widely *how* to reach a good level of AI explainability, algorithmic accountability, and fairness. This paper argues that, in the field of data protection, the GDPR already proposes a sustainable environment of desirable ADM systems, which is broader than any ambition to have “transparent,” “explainable,” “fair,” “lawful,” or “accountable” ADM: We should aspire to *just* algorithms, that is, *justifiable* automated systems that include all the above-mentioned qualities (fairness, lawfulness, transparency, accountability, etc.).

This might be possible through a practical “justification” process and statement through which the data controller *proves*, in practical ways, the *legality* of an algorithm with respect to all data protection principles (fairness, lawfulness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, and accountability). This justificatory approach might also be a solution to many existing problems in the AI explanation debate, for example, the difficulty of “opening” black boxes, the transparency fallacy, and the legal difficulties in enforcing a right to receive individual explanations.

After an overview of the GDPR rules (Section 2) and of the definition and limits of AI explanations (Section 3), this article has proposed a wider systemic approach to algorithmic accountability (Section 4). In order to do that, the concept of justification is introduced and analysed both in general (Section 5), and in the legal and data protection fields (Sections 6 and 7). This article argues that the justificatory approach is already required by the GDPR rules. Accordingly, Sections 8 and 9 explain why and how the data protection principles could be a meaningful basis to justify ADM systems.

References

- Aarnio, Aulis. *The Rational as Reasonable: A Treatise on Legal Justification* (Springer Science & Business Media 1986): 22.
- Alexander, Larry, Moore, Michael. "Deontological Ethics." In *The Stanford Encyclopedia of Philosophy* edited by Edward N. Zalta (Winter 2020, Metaphysics Research Lab, Stanford University 2020).
- Araujo, Theo, et al. "In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence." *AI & Society* 35 (2020): 611, 616.
- Barocas, Solon, Selbst Andrew D. "Big Data's Disparate Impact." *California Law Review* 104 (2016): 671, 692.
- Biran Or, Cotton, Courtenay. "Explanation and Justification in Machine Learning: A Survey." /paper/Explanation-and-Justification-in-Machine-Learning- %3A-Biran-Cotton/02e2e79a77d8aabc1af1900ac80ceebac20abde4.
- Brennan-Marquez, Kiel. "Plausible Cause': Explanatory Standards in the Age of Powerful Machines." *Vanderbilt Law Review* 70, no. 53 (2017).
- Brkan, Maja. "The Essence of the Fundamental Rights to Privacy and Data Protection: Finding the Way Through the Maze of the CJEU's Constitutional Reasoning." *German Law Journal* 20 (2019): 864.
- Butterworth, Michael. "The ICO and Artificial Intelligence: The Role of Fairness in the GDPR Framework." *Computer Law & Security Review* 34 (2018): 257.
- Clifford, Damian, Ausloos, Jeff. "Data Protection and the Role of Fairness." *Yearbook of European Law* 37 (2018): 130.
- Dressel, Julia, Farid, Hany. "The Accuracy, Fairness, and Limits of Predicting Recidivism." *Science Advances* 4 (2018): eaao5580.
- Dwork, Cynthia, Mulligan, Deirdre K. "It's Not Privacy, and It's Not Fair." *Stanford Law Review* 6 (2013): 66.
- Edwards, Lillian, Veale, Michael. "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For," 16 *Duke Law & Technology Review* 18 (2017).
- Edwards, Lillian, Veale, Michael. "Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'?" 16 *IEEE Security & Privacy* 46 (2018).
- Galhotra, Sainyam, Brun, Yuri, Meliou, Alexandra. "Fairness Testing: Testing Software for Discrimination." Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering— ESEC/FSE 2017, (ACM, 2017) <http://dl.acm.org/citation.cfm?doid=3106237.3106277>.
- Goodman, Bryce, Flaxman, Seth. "EU Regulations on Algorithmic Decision-Making and a 'Right to Explanation'" arXiv:1606.08813 [cs, stat] <http://arxiv.org/abs/1606.08813>, accessed 30 June 2018.
- Goodman, Bryce. *A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection*. (2016).
- Hamon, Ronan and others. "Impossible Explanations? Beyond Explainable AI in the GDPR from a COVID-19 Use Case Scenario." *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2021).
- Hazen, Benjamin T., et al. "Data Quality for Data Science, Predictive Analytics, and Big Data in Supply Chain Management: An Introduction to the Problem and Suggestions for Research and Applications." *International Journal of Production Economics* 154 (2014): 72.
- Henin, Clément, Le Métayer, Daniel. "A Framework to Contest and Justify Algorithmic Decisions." [2021] *AI and Ethics*.
- Hildebrandt, Mireille. *Law for Computer Scientists and Other Folk* (Oxford University Press 2020): 267.
- Hildebrandt, Mireille. "Profile Transparency by Design? Re-Enabling Double Contingency." available at https://works.bepress.com/mireille_hildebrandt/63/.
- Horowitz, Donald L. "Justification and Excuse in the Program of the Criminal Law." *Law and Contemporary Problems* 49 (1986): 109.
- Hutton, Luke, Henderson, Tristan. "Beyond the EULA: Improving Consent for Data Mining," In *Transparent Data Mining for Big and Small Data* edited by Tania Cerquitelli, Daniele Quercia, and Frank Pasquale (eds.), (Springer, New York 2017): 147 at 162.
- Kaminski, Margot E. "The Right to Explanation, Explained." 34 *Berkeley Technology Law Journal* (2019): 189.
- Kaminski, Margot E., Malgieri, Gianclaudio. "Multi-Layered Explanation from Algorithmic Impact Assessments in the GDPR." *FAT 2020 Proceedings* (ACM Publishing, 2020).
- Kaminski, Margot E., Malgieri, Gianclaudio. "Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations." 19–28 *University of Colorado Law Legal Studies Research Paper* available at <https://papers.ssrn.com/abstract=3456224>.
- Kaminski, Margot. "Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability." 92 *Southern California Law Review* 1529 (2019):12-17.
- Katyal, Sonia K. "Private Accountability in the Age of Artificial Intelligence." *UCLA Law Review* 66 (2019): 88.
- Kim, Pauline T. "Data-Driven Discrimination at Work." 58 *Wm. & Mary L. Rev.* (2017): 857.
- Kloza, Dariusz, et al. "Data Protection Impact Assessment in the European Union: Developing a Template for a Report from the Assessment Process." (LawArXiv 2020) *DPiLab Policy Brief* 29 available at <https://osf.io/7qrpf>.
- Kroll, Joshua et al. "Accountable Algorithms." *University of Pennsylvania Law Review* 165 (2017): 633.
- Lepri, Bruno, et al., "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes." *Philosophy & Technology* 31 (2018): 611.
- Lipton, Zachary C. "The Mythos of Model Interpretability." *Communications of the ACM* 61 (2018): 36.

- Lodder, Arno R. *Dialaw: On Legal Justification and Dialogical Models of Argumentation* (1999° edizione, Kluwer Academic Publishers, 1999).
- Loi, Michele, Ferrario, Andrea, Viganò, Eleonora. "Transparency as Design Publicity: Explaining and Justifying Inscrutable Algorithms." In *Ethics and Information Technology*, <https://doi.org/10.1007/s10676-020-09564-w>.
- Malgieri, Gianclaudio, Comandé, Giovanni. "Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation." *International Data Privacy Law* 7, no. 4 (2017): 243-65.
- Malgieri, Gianclaudio. "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, no. 105327 (2019): 9-11.
- Malgieri, Gianclaudio. "The Concept of Fairness in the GDPR: A Linguistic and Contextual Interpretation." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2020) available at <https://doi.org/10.1145/3351095.3372868>, accessed 29 January 2020.
- Malgieri, Gianclaudio, Niklas, Jędrzej. "The Vulnerable Data Subject." 37 *Computer Law & Security Review* (2020).
- Milaj, Jonida. "Privacy, Surveillance, and the Proportionality Principle: The Need for a Method of Assessing Privacy Implications of Technologies Used for Surveillance." *International Review of Law, Computers & Technology* 30 (2016): 115, 116.
- Miller, Tim. "Explanation in Artificial Intelligence: Insights from the Social Sciences." 267 *Artificial Intelligence* 1, (2019).
- Mortier Richard and others. "Human-Data Interaction." In *The Interaction Design Foundation (ed), The Encyclopedia of Human-Computer Interaction*, (2nd edition, The Interaction Design Foundation 2015).
- Moser, Paul K "Justification in the Natural Sciences." *The British Journal for the Philosophy of Science* 39 (1991): 557-75.
- Oprîșu, Raluca. "Reversal of 'the Burden of Proof' in Data Protection | Lexology." available at <https://www.lexology.com/library/detail.aspx?g=e9e8c734-23d9-41bb-a723-5d664b3c86cc>.
- Petkova, Bilyana, Hacker, Philipp. "Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers." Lecturer and Other Affiliate Scholarship Series available at <https://digitalcommons.law.yale.edu/ylas/13> (2016).
- Ramamurthy, Karthikeyan Natesan, et al. "Model Agnostic Multilevel Explanations." available at <https://arxiv.org/abs/2003.06005v1>, accessed 25 March 2020.
- Reisman, Dillon, et al. *Algorithm Impact Assessment: A Practical Framework for Public Agency Accountability*. (AI Now Institute: 2018).
- Roig, Antoni. "Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)." *European Journal of Law and Technology* 8 (2018).
- Rudin, Cynthia. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1 (2019): 206, 207.
- Selbst, Andrew D., Powles, Julia. "Meaningful Information and the Right to Explanation." *International Data Privacy Law* 7, no. 4 (2017): 233-42.
- Selbst, Andrew D. "Disparate Impact in Big Data Policing." *Georgia Law Review* 52 (2018): 109; Reisman et al., (n 76).
- Selbst, Andrew D., et al. "Fairness and Abstraction in Sociotechnical Systems." *Proceedings of the Conference on Fairness, Accountability, and Transparency* (ACM, 2019) <http://doi.acm.org/10.1145/3287560.3287598>.
- Selbst, Andrew D., Barocas, Solon. "The Intuitive Appeal of Explainable Machines." 87 *Fordham Law Review* 1085 (2018).
- Smith, J. C. *Justification and Excuse in the Criminal Law* (Stevens 1989).
- Tyler, Tom R. "Procedural Justice, Legitimacy, and the Effective Rule of Law." *Crime and Justice* 30, no. 283, (2003): 317-18.
- Veale, Michael, Edwards, Lilian. "Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling." *Computer Law & Security Review* 34 (2018): 398.
- Wachter, Sandra, Mittelstadt, Brent, Floridi, Luciano. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7, no. 2 (2017): 76-99.
- Wachter, Sandra, Mittelstadt, Brent, Russell, Chris. "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR." *Harvard Journal of Law & Technology* 31, no. 2 (2018).
- Wachter, Sandra. *Affinity Profiling and Discrimination by Association in Online Behavioural Advertising* (Social Science Research Network 2019) SSRN Scholarly Paper ID 3388639 <https://papers.ssrn.com/abstract=3388639>.
- Wachter, Sandra, Mittelstadt, Brent. "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI." *Columbia Business Law Review* 2 (2019).