Technology and Regulation

Explaining and Contesting Judicial Profiling Systems

Beyond a Procedural Right to an Explanation

Author(s)	Ljubiša Metikoš
Contact	l.metikos@uva.nl
Affiliation(s)	University of Amsterdam, Faculty of Law, Institute for Information Law (IViR), Paul Scholten Centre for Jurisprudence and the RPA Human(e)AI

Keywords	contestation, adjudication, AI, the right to an explanation, transparency, GDPR, AI-Act, right to a fair trial, adversarial principle		
Published	Received: 26 Apr. 2024	Accepted: 3 Jul. 2024	Published: 11 Sep. 2024
Citation	Ljubisa Metikos, Explaining and Contesting Judicial Profiling Systems, Technology and Regulation, 2024, 188-208 • https://doi.org/10.26116/techreg.2024.017 • ISSN: 2666-139X		

Abstract

This paper argues that a right to an explanation can enable litigants to contest judicial profiling systems on various grounds. However, the technical opacity of certain types of systems, integrity concerns, and the rights and interests of third parties, can hinder the ability of courts to provide an explanation. To overcome these obstacles, a number of technical and organizational measures can be taken before and during the development of these systems, to ensure that they are contestable. This paper also critically interprets EU Data Protection Law, the right to a fair trial, and the AI-Act. It shows how these laws (partially) protect contestation by design, as well as their limitations and potential loopholes.

1. Introduction[1]

Increasingly, Machine Learning (ML) algorithms are used to produce AI models that assist judges across the world. ML algorithms can 'teach' themselves certain tasks by analyzing large labeled datasets. This has enabled algorithms to create models that can perform very sophisticated tasks in the field of law, which were impossible to automate previously.[2] Some notable examples include the use of predictive systems such as COMPAS in the U.S.A, which provided judges with a recidivism risk score of inmates,[3] or the '*Little Judge*

[1] Iextend my sincere gratitude for the input I received from the anonymous peer reviewers, and the editor of TechReg; Ronald Leenes. I also wish to thank Natali Helberger, Iris van Domselaar, Anna van Duin, Naomi Appelman, and my colleagues at the Research Priority Area Human(e)AI, for the insightful feedback and thoughtful comments they provided.

[2] Harry Surden, 'Machine Learning and Law' (2014) 89 *Washington Law Review* 87.

[3] Sam Corbett-Davies, Emma Pierson, Avi Feller, and Sharad Goel, 'A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not That Clear' *The Washington Post* (17 October 2016) <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/> accessed 19 October 2022.

Bao', which was used in Chinese courts to provide judges with fully written suggested verdicts.[4] In response to these developments, legal scholars are now increasingly discussing how adjudication is, or shall be, impacted by new technologies and how these systems must be regulated when used in the administration of justice.[5]

The rising use of AI in the judiciary has prompted many legal scholars to focus on the potential emergence of a '*robot judge*'; an AI system that could one day fully replace human judges.[6] In 2019, Wired reported that Estonia was developing such an autonomous 'robot judge'.[7] This article was cited by a number of researchers as an example of how human judges could be fully replaced in the near future.[8] However, the claims made by Wired were directly refuted by the Estonian Ministry of Justice in a press release titled '*Estonia does not develop AI Judge'*.[9]

This example shows us that we perhaps focus too eagerly on the idea of the autonomous robot judge. Whether or not it will ever arrive is neither here nor there. In any case, the current introduction of *supportive* systems *precedes* any widespread implementation of a robot judge.[10] Currently, the development and use of AI in the judiciary is leaning towards such support, and not toward the total replacement of human legal decision-making.[11]

In this paper, I discuss judicial support systems that substantially impact the content of a verdict by generating output that can serve as evidence in a trial. Such systems have been used in criminal, administrative, and civil procedures, for a variety of different purposes. In particular, I will look at litigant profiling systems. These systems profile a litigant based on various parameters and can provide a judge with additional information about the person in question.

[4] Jiahui Shi, 'Artificial Intelligence, Algorithms and Sentencing in Chinese Criminal Justice: Problems and Solutions' (2022) 33 *Criminal Law Forum* 121.

[5] Tania Sourdin, *Judges, Technology and Artificial Intelligence* (Edward Elgar Publishing 2021) 320; Eugene Volokh, 'Chief Justice Robots' (2019) 68 *Duke Law Journal* 1135; Katherine Quezada-Tavárez, Plixavra Vogiatzoglou, and Sofie Royer, 'Legal Challenges in Bringing AI Evidence to the Criminal Courtroom' (2021) 12 *New Journal of European Criminal Law* 531–551; Nídia Andrade Moreira, 'The Compatibility of AI in Criminal System with the ECHR and ECtHR Jurisprudence' in Goreti Marreiros and others (eds), *Progress in Artificial Intelligence* (Springer International Publishing 2022) 356–365; Jasper Ulenaers, 'The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?' (2020) 11 *Asian Journal of Law and Economics* 1–20.

[6] John Morison and Adam Harkens, 'Re-Engineering Justice? Robot Judges, Computerised Courts and (Semi) Automated Legal Decision-Making' (2019) 39 *Legal Studies* 618; Richard M Re and Alicia Solow-Niederman, 'Developing Artificially Intelligent Justice' (2019) 22 *Stanford Technology Law Review* 242 (n 6); Ric Simmons, 'Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System' (2018) 52 *U.C. Davis Law Review* 1067; Tania Sourdin and Richard Cornes, 'Do Judges Need to Be Human? The Implications of Technology for Responsive Judging' in Tania Sourdin and Archie Zariski (eds), *The Responsive Judge: International Perspectives* (Springer 2018) 53–73; Ulenaers, *(n 5)*; SGC van Wingerden and others, *Artificial Intelligence and Sentencing: Humans against Machines* (Oxford University Press 2022) 230–251; Tim Wu, 'Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems' (2019) 119 *Columbia Law Review* 2001.

[7] Eric Niiler, 'Can AI Be a Fair Judge in Court? Estonia Thinks So' (25 March 2019) Wired <https://www.wired.com/story/can-ai-be-fair-judge-court-estonia-thinks-so/> accessed 27 August 2021.

[8] Rebecca Crootof, 'A Meaningful Floor for "Meaningful Human Control"' (2016) 30 *Temple International & Comparative Law Journal* 53; Sourdin (n 5); Re and Solow-Niederman (n 6); Ulenaers (n 5); Monika Zalnieriute and Felicity Bell, 'Technology and the Judicial Role' in Gabrielle Appleby and Andrew Lynch (eds), *The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Austra*lia(Cambridge University Press 2021) 116-142.

[9] Justiitsministeerium, 'Estonia Does Not Develop AI Judge' (16 February 2022) <https://www.just.ee/en/news/estonia-does-not-develop-ai-judge> accessed 15 October 2022.

[10] Morison and Harkens (n 6); Tania Sourdin, 'Judge v. Robot: Artificial Intelligence and Judicial Decision-Making' (2018) 41 *University of New South Wales Law Journal* 1114; Sourdin, (n 5); Surden (n 2); Wu (n 6).

[11] CEPEJ, *CEPEJ European Ethical Charter on the Use of Artificial Intelligence (AI) in Judicial Systems and Their Environment* (European Commission for the Efficiency of Justice (CEPEJ), 2018) <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment> accessed 14 October 2022; Tania Sourdin, *The Role and Function of a Judge: The Adoption and Adaptation of Technology by Judges* (Edward Elgar Publishing 2021); Tania Sourdin, *The Role and Function of a Judge: The Adoption and Adaptation of Technology by Judges* (Edward Elgar Publishing 2021) 25-48.

Examples of judicial profiling systems include the Colombian PretorIA system, which was used to profile and suggest legal guardians, or the Mexican EXPERTIUS system, which determined the eligibility of a litigant for social security.[12] One of the more controversial uses of such profiling systems have been recidivism risk prediction systems, used by criminal courts. These systems help judges to predict the risk for future criminal behavior of a litigant. Examples include COMPAS in the U.S.A and HART in the U.K.[13]

This paper focuses on these profiling systems as they can provide intrusive analyses of the character and behavior of litigants in high-stake judicial settings. As judicial proceedings commonly provide individuals with ample opportunity for contestation,[14] litigants could desire an explanation of how the system functions, so that they might contest this output.[15] The grounds on which these profiling systems arrive at a certain conclusion about a litigant can, however, be inscrutable. Several different technical and organizational obstacles can hinder courts from explaining how these profiling systems operate, which obstructs the ability of litigants to contest such algorithmically generated evidence.[16] Arguably, the inscrutability of these systems, whose findings can preponderantly influence a verdict, hinders litigants from participating in their own trial. Consequently, I argue that to safeguard contestation, litigants require *a right to an explanation*.[17]

We can define explanations as information that aims to help a litigant to understand how a system came to a certain output, so that they might be better able to contest that output.[18] In this paper, I further expand on this definition, by formulating concrete epistemic grounds based on which a litigant could potentially contest a judicial profiling system. These epistemic grounds show what kind of information a litigant could reasonably need to contest a profiling system. This helps us to connect what the needs are of explainability for litigants, with the different technical and legal approaches to explainability.[19]

To safeguard contestation, I argue, however, that a right to an explanation cannot be merely provided *post-hoc,* after an opaque model already has been developed and used. Rather, a variety of technical and organizational measures need to be implemented during its development process, to ensure that the algorithm is contestable *by design*.[20] To this end, I discuss the need to implement a variety of technical of organizational measures.

[12] Miriam Stankovich, *Global toolkit on AI and the rule of law for the judiciary* (Toolkit, CI/DIT/2023/AIRoL/01, UNESCO 2023) <https://unesdoc.unesco.org/ark:/48223/pf0000390781> accessed 17 August 2023.

[13] Łukasz Grad and Katarzyna Koprowska, 'Chapter 10 Story COMPAS: Recidivism Reloaded' in XAI Stories, Case Studies for eXplainable Artificial Intelligence (2020) <https://pbiecek.github.io/xai_stories/story-compas.html> accessed 7 November 2023.

[14] Jeremy Waldron, 'The Rule of Law and the Importance of Procedure' (2011) 50 *Nomos* 3.

[15] Mireille Hildebrandt, 'Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning' (2019) 20 *Theoretical Inquiries in Law* 83.

[16] Emre Bayamlıoğlu and others (eds), *Being Profiled: Cogitas Ergo Sum : 10 Years of 'Profiling the European Citizen'* (Amsterdam University Press 2018); Mireille Hildebrandt, 'Algorithmic Regulation and the Rule of Law' (2018) 376 Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 20170355; Daniel McQuillan, 'Data Science as Machinic Neoplatonism' (2018) 31 Philosophy & Technology 253; Daniel McQuillan, 'Non-Fascist AI' in Propositions for Non-Fascist Living Tentative and Urgent, Maria Hlavajova and Wietske Maas (eds) (MIT Press 2019), 'Power, Process, and Automated Decision-Making' (2019) 88 Fordham Law Review 613.

[17] Emre Bayamlıoğlu, 'Contesting Automated Decisions' (2018) 4 *European Data Protection Law Review* 433-446; Claudio Sarra, 'Put Dialectics into the Machine: Protection against Automatic-Decision-Making through a Deeper Understanding of Contestability by Design' (2020) 20 *Global Jurist* 1-23.

[18] Clément Henin and Daniel Le Métayer, 'Beyond Explainability: Justifiability and Contestability of Algorithmic Decision Systems' [2021] AI & Society 2; Clément Henin and Daniel Le Métayer, 'A Framework to Contest and Justify Algorithmic Decisions' (2021) 1 AI and Ethics 463.

[19] Madeleine Fink and Michèle Finck, 'Reasoned A(I)Dministration: Explanation Requirements in EU Law and the Automation of Public Administration' (2022) 47 *European Law Review* 376; Christoph Molnar, Interpretable Machine Learning <https://christophm.github.io/interpretable-ml-book/> accessed 4 December 2023

[20] Marco Almada, 'Human Intervention in Automated Decision-Making: Toward the Construction of Contestable Systems', *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (Association for Computing Machinery 2019) 2.

In addition, I analyze whether the current regulatory framework can safeguard such a right in the context of adjudication. In this regard, I discuss art. 6 of the European Convention on Human Rights (ECHR), the General Data Protection Regulation (GDPR),[21] the Law Enforcement Directive (LED),[22] and the AI Act (AIA),[23] to see whether these laws can safeguard contestability *by design,* through a right to an explanation that requires ex-ante technical and organizational measures to be taken by judiciaries and the developers of profiling systems.

In short, this paper aims to add to the existing literature on the right to an explanation by providing a thorough conceptual and legal analysis of how this right can be applied to safeguard litigant contestation of judicial profiling systems. To this end, this paper further builds upon the concept that a right to an explanation can enable contestation.[24] It specifies how contestation is impacted in adjudicatory proceedings by opaque profiling systems and applies this analysis to different legal frameworks. Moreover, the paper also expands upon the idea that the right to an explanation is not merely a procedural right that can only be invoked *ex post*.[25]

The paper also provides new perspectives to the legal-doctrinal debate on the right to an explanation. Currently, legal-doctrinal research on this topic has mostly revolved around the GDPR.[26] This paper shows, however, that different decision-making contexts, such as that of judicial decision-making, necessitate that the discussion of the right to an explanation should also include other relevant legal instruments, such as the right to a fair trial, the LED, and the AIA. Moreover, this paper also shows the limits of traditional transparency and due process safeguards such as those present in the right to a fair trial, when judicial profiling systems are used in the courtroom.

2. Conceptualizing the right to an explanation

Arguably, adjudicatory proceedings should provide a space where litigants can effectively contest all evidence and observations presented before a judge, enabling them to meaningfully participate in their own trial.[27] However, various scholars have argued that contestability is harmed by the opacity and inscrutability of certain AI systems.[28] One of the possible goals of explainability measures then, is to make

[21] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) OJL119/1.

[22] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA (Law Enforcement Directive) OJL119/89.

[23] European Parliament, 'Legislative Resolution of 13 March 2024 on the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts' (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) [2024] OJ C90/01 <https://op.europa.eu/en/publication-detail/-/publication/2fcad39a-e777-11ee-9ea8-01aa75ed71a1> accessed 21 August 2024.

[24] Sarra (n 17); Henrietta Lyons, Eduardo Velloso and Tim Miller, 'Conceptualising Contestability: Perspectives on Contesting Algorithmic Decisions' (2021) 5 Proceedings of the ACM on Human-Computer Interaction 1; Emre Bayamlıoğlu, 'The Right to Contest Automated Decisions under the General Data Protection Regulation: Beyond the so-Called "Right to Explanation"' (2022) 16 Regulation & Governance 1058; Almada (n 20); Margot E. Kaminski and Jennifer M Urban, 'The Right to Contest AI' (2021) 121 Columbia Law Review 1957.

[25] Bayamlıoğlu (n 17); Almada (n 20).

[26] Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a Right to Explanation Is Probably Not the Remedy You Are Looking For' (2017) 18 *Duke Law & Technology Review* 67; Margot E. Kaminski, 'The Right to Explanation, Explained' (2019) 34 *Berkeley Technology Law Journal* 189; Gianclaudio Malgieri, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations' (2019) 35 *Computer Law & Security Review* 105327; Andrew D. Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233; 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' (2017) 7 *International Data Privacy Law* 76; Sarra (n 17); Bayamlıoğlu (n 17); Paul B. de Laat, 'Algorithmic Decision-Making Employing Profiling: Will Trade Secrecy Protection Render the Right to Explanation Toothless?' (2022) 24 *Ethics and Information Technology* 17.

[27] Waldron (n 14).

[28] Sarra (n 19); Almada (n 20); Bayamlioglu (n 17).

it possible for individuals to contest such systems.[29] But how does opacity impact contestation? And how can explanations mitigate this impact? This section provides a conceptualization of how opacity harms the practice of contestation, how a right to an explanation can mitigate this opacity, and how we can safeguard contestation through the implementation of a variety of different technical and organizational measures during the development of profiling systems.

2.1 Contesting profiling systems

Contestation, as a concept in academic and regulatory debates on AI, has seen many different conceptualizations.[30] For the purposes of the analysis of this paper, we focus on the practice of contestation in adjudication. In this regard, Waldron describes contestation as the ability of the litigant to express themselves, and to provide their views on all evidence and observations presented in a trial.[31]

Similarly, explainability has seen many different conceptualizations as well, from the perspective of many different scholars.[32] As such, what an explanation entails depends on a variety of different conditions, such as the person who receives the explanation, as well as the goal that the explanation strives toward.[33] In this paper, we focus on explanations that can enable a litigant to contest the output of judicial profiling systems. Such explanations, which aim to explain how a specific output of a system has been created, as opposed to how the system works in general, have also been called '*subject-centric*', '*local*', or '*strong*' explanations.[34] In the next sections, I discuss how and why opacity hinders contestation, as well as on what epistemic grounds a litigant might contest the output of a profiling system.

2.2 The contestability of profiling systems

In the development process of profiling systems, developers often use ML algorithms. These algorithms rely on a range of different computer science techniques to analyze large amounts of data, often with the help of some kind of human supervisor, during its creation process. Currently, algorithms are able to ascertain correlations and patterns between various data points to produce highly complex and powerful models that can be used in the field of law.[35] There is a belief that by applying ML techniques to data concerning individuals, we can ascertain '*ground truths*' about the nature and behavior of humans.[36] However, the models generated by these algorithms are heavily influenced by a number of contestable '*epistemic and normative presumptions*' that are embedded in it during its design.[37]

ML developers make a number of '*abstraction decisions*' to create a model capable of transforming input data into the desired output data. To this end, developers transform an individual's behavior and identity into commensurable and quantifiable datapoints that can be easily analyzed.[38] However, developers make a number of assumptions as to what data points are, or are not, of importance to analyze an individual.[39] Subsequently, based on these assumptions, ML algorithms can then start to identify patterns in the training data and ascertain the most effective ways to transform input data on the litigant into the desired output data.[40] However, neither these design choices, nor the patterns that the algorithm finds, necessarily provide an accurate reflection of the individual litigant's behavior or identity.

[29] Sarra (n 17); Bayamlioglu (n 17).

[30] Lyons, Velloso and Miller (n 24).

[31] Waldron (n 14).

[32] Alejandro Barredo Arrieta and others, 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI' (2020) 58 *Information Fusion* 82; Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences' (2019) 267 *Artificial Intelligence* 1; Clément Henin and Daniel Le Métayer, 'Beyond Explainability: Justifiability and Contestability of Algorithmic Decision Systems' (2021) 37(4) *AI & Society* 1397-1410.

[33] Miller (n 32); Samuli Laato and others, 'How to Explain AI Systems to End Users: A Systematic Literature Review and Research Agenda' (2022) 32 *Internet Research* 1.

[34] de Laat (n 26); Edwards and Veale (n 26).

[35] Surden (n 2).

[36] McQuillan, 'Data Science as Machinic Neoplatonism' (n 16); McQuillan, 'Non-Fascist AI' (n 16).

[37] Reuben Binns, 'Algorithmic Accountability and Public Reason' (2018) 31 *Philosophy & Technology* 543.

[38] Dan L. Burk, 'Algorithmic Legal Metrics' (2021) 96 *The Notre Dame Law Review* 1147.

[39] Hildebrandt (n 16).

[40] Karen Yeung and Adam Harkens, 'How Do "Technical" Design Choices Made When Building Algorithmic Decision-Making Tools for Criminal Justice Authorities Create Constitutional Dangers? (Part I)' (2023) *2023 Public Law* 265.

In such a process, much nuance can be lost regarding the particularities of an individual.[41] Each of the abstraction choices reduces the complexity of the world, as well as that of the individual litigant who is being profiled, into abstract representations, whose accuracy in reflecting the individual in question, can be contested on a variety of grounds.

2.2.1 The grounds of contestation

In the previous paragraphs we saw that profiling systems do not offer ground truths about a litigant's behavior. Rather, their analyses offer merely an imperfect approximation of reality that is based on certain presumptions and abstractions. A litigant can therefore contest whether the analysis of the system is accurate or not. But what information about a profiling system could then be provided to a litigant, that would enable contestation? What are the (epistemic) grounds for contestation?

First of all, a litigant could question the parameters that the algorithm took into account. Are the factors that the system analyzed truly accurate representations of their identity? Do they present a fair picture of the litigant, or do they show irrelevant features that should not influence the verdict? Secondly, a litigant can contest the input data that the model relied upon. To this end, they might provide more accurate information about themselves. However, they may also share non-quantifiable information about themselves that the system could not have accounted for.[42] Thirdly, the pattern and correlations that an algorithm found between different parameters can be contested as well by a litigant. A *'reasonable hypothesis'* as to what patterns the algorithm found between data points should therefore be disclosed.[43] Perhaps the algorithm found a connection between a certain characteristic of the litigant, and a predicted type of behavior. However, a litigant might question whether this connection relies on *causation* or *correlation*. Lastly, it might be the case that the parameters that were looked at, are indeed relevant to profile the litigant, but the system could have attached too much or too little weight to them in its analysis. On all of these different grounds, a litigant can contest a profiling system.

2.2.2 What an explanation is (not) for

However, when implementing a right to an explanation, we should be vigilant to not put the burden of remedying broken, risky, and inaccurate profiling-based decision-making on a decision subject.[44] Apart from any explainability requirements, profiling systems must be safe, accurate, and well-audited before their use. A right to an explanation cannot be used as a form of '*transparency washing*' of bad judicial profiling systems.[45]

Having said that, procedural rights such as a right to an explanation still have their place in the regulatory toolkit for judicial profiling systems. Such rights enable a litigant to better engage with their trial, as litigants are now able to contest the profiling system in question. Safeguarding contestability and, in turn, the participation of a litigant should, therefore, be seen separately from the question of whether or not a profiling system is suitable in the first place.[46]

2.3 Obstacles to explainability

In the previous paragraphs, we saw that there are multiple grounds on which a litigant could contest a profiling system. However, how can we explain these grounds to a litigant? There are a number of technical and organizational obstacles that can obstruct the process of providing an explanation. In this section I first discuss three main concerns in this regard: technical opacity, integrity concerns, and external private interests. I then discuss two main explainability methods: model-agnostic methods and intrinsic methods. Subsequently, I also outline a number of other measures that judiciaries can take to ensure the contestability of judicial profiling systems.

[41] Hildebrandt (n 16); Burk (n 38); McQuillan, 'Data Science as Machinic Neoplatonism' (n 16).

[42] Hildebrandt (n 16); Burk (n 38).

[43] Yeung and Harkens (n 40).

[44] Edwards and Veale (n 26); Reuben Binns, 'Human Judgment in Algorithmic Loops: Individual Justice and Automated Decision-Making' (2022) 16 *Regulation & Governance* 197.

[45] Monika Zalnieriute, '"Transparency-Washing": The Corporate Agenda of Procedural Fetishism' (2021) 8(1) Critical Analysis of Law 39-53.

[46] See for a similar argument that has been made in regard to human oversight policies: Ben Green, 'The Flaws of Policies Requiring Human Oversight of Government Algorithms' (2022) 45 Computer Law & Security Review 105664, 1–12.

2.3.1 Technical opacity

There are a number of technical limits that hinder our ability to interpret how highly complex ML algorithms function. These can be seen as a form of '*epistemic opacity*'[47] or '*epistemic constraints*'[48] that limit our inquiries into the models created by ML algorithms.[49] Bayamlıoğlu delineates three main issues in this regard: the complexity of the model, its lack of human-interpretable patterns, and its adaptive and changeable nature.[50]

Firstly, after the model has been created, the rules that guide its functioning can be so incredibly numerous, that inspection of the system becomes impractical.[51] In part, this opacity is caused by the huge scale at which some algorithms currently operate.[52] This is especially the case with ML techniques such as deep-learning.[53] Secondly, algorithms reason through statistical relationships between parameters, which can rely on patterns that go beyond our comprehension. Such correlations can be non-intuitive for humans. For example, an algorithm could categorize the chance of recidivism as high because an inmate's name contains five letters. It will be difficult then for a human to ascertain, or guess, the pattern that the algorithm has found.[54] Lastly, some types of ML systems are adaptive in nature, changing their output continuously as the model improves itself. Its parameters may then change continuously for each case that it handles, making it hard to keep along for a human seeking an explanation of why the system came to a certain output.[55]

2.3.2 Integrity concerns

Besides these technical limitations, there also exists a number of organizational obstacles to disclosing an explanation. In this regard, *integrity concerns* play an especially difficult role for sensitive judicial profiling systems. Concerns have been raised that the integrity of a system can be harmed, if its heuristics are fully disclosed.[56] This is because people might '*game the system*'; manipulating it to their benefit.[57] Disclosing the criteria that a judicial profiling system relies upon, could lead to litigants abusing such information. Litigants might namely provide input data about themselves that is more likely to lead to a positive output. For example, if a profiling system relies on the answers to a questionnaire, a litigant might give specific answers so that the system gives them a favorable output. This could then harm the objectivity and accuracy of the system, which would necessitate that an explanation be withheld from the litigant.

Criminal profiling systems could, for example, rely on psychological questionnaires that rely on the input of the defendant. Moreover, profiling systems used in administrative judicial cases to help determine the eligibility for social benefits, might also rely on documents that a litigant would need to provide.[58] In both these cases, explaining the specific parameters that the system would take into account, could risk opening up the system to manipulation by the litigants.

This argument has in fact already been raised against explanation requests. In a case before the Court of Appeals of Amsterdam, the ride-sharing company Uber argued that their fraud-risk detection system could be manipulated by their drivers, if an explanation was granted about its main parameters.[59]

[47] Florian Eyert, Florian Irgmaier and Lena Ulbricht, 'Extending the Framework of Algorithmic Regulation. The Uber Case' (2022) 16 *Regulation & Governance* 23.

[48] Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in Thomas Wischmeyer and Timo Rademacher (eds), *Regulating Artificial Intelligence* (Springer International Publishing 2020) 75-101.

[49] See Bayamlıoğlu's (n 17), for an in-depth discussion of these technical limitations.

[50] Bayamlıoğlu (n 17).

[51] Andrew D. Selbst and Solon Barocas, 'The Intuitive Appeal of Explainable Machines' (2018) 87 *Fordham Law Review* 1085, 1094.

[52] Jenna Burrell, 'How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms' (2016) 3(1) *Big Data & Society* 1.

[53] Rahul Iyer and others, 'Transparency and Explanation in Deep Reinforcement Learning Neural Networks' (arXiv, 10 September 2018) <http://arxiv.org/abs/1809.06061> accessed 7 October 2021.

[54] Selbst and Barocas (n 51) 1129.

[55] Karen Yeung, 'Algorithmic Regulation: A Critical Interrogation' (2018) 12 *Regulation & Governance* 505.

[56] Wischmeyer (n 48).

[57] Jane Bambauer and Tal Zarsky, 'The Algorithm Game' (2018) 94 *The Notre Dame Law Review* 1.

[58] Stankovich (n 12).

[59] Uber BV v Personenvervoer Nederland BV [2023] Court of Appeal of Amsterdam, ECLI:NL:GHAMS:2023:793, para 3.28.

2.3.3 External private interests

Another obstacle pertains to the fact that judiciaries can rely on models developed by private parties.[60] These developers are often unwilling to share their proprietary software as they fear that this might disclose their trade secrets and therefore harm their business interests.[61] This was for example the case with COMPAS.[62] Here, the proprietary nature of the system, made it difficult for anyone besides Northpointe, its developer, to assess how the system functioned. Moreover, in the past, companies have also invoked their intellectual property rights and trade secrets rights, to block the exercise of a right to an explanation.[63]

Burrell describes these aforementioned obstacles to explainability as '*intentional forms*' of opacity.[64] While it might not necessarily be the case that judiciaries would purposefully cause the aforementioned obstacles to occur, there does remain a severe risk that the careless development of profiling systems might lead to situations where technical opacity, integrity concerns or external private interests hinder or block an explanation.[65] In the following section, I lay down a number of approaches that can help overcome the aforementioned obstacles to the right to an explanation.

2.4 Explaining profiling systems

As we saw in the previous paragraphs, providing an explanation of a profiling system can be hindered by a number of technical and organizational obstacles. There are two main types of explainability methods that can be distinguished in computer science literature, which could help resolve the technical opacity of ML models. These are *intrinsic*, and *model-agnostic* methods. In addition, there are also a number of organizational measures that judiciaries can take during the development of the system, to overcome issues related to integrity concerns and external private interests.

2.4.1 Model-agnostic methods

Within the field of interpretable ML, there exist two broad approaches to providing explanations: *intrinsic* explanations and *model-agnostic* explanations.[66] Intrinsic explanations aim to showcase directly how an algorithm functions by developing a ML model that is directly interpretable. However, what if one wishes to use a ML method that is highly complex, and therefore not directly interpretable? In such an instance, a model-agnostic method can be applied to such an opaque model.

Model-agnostic methods, in short, can be applied externally to an opaque model to deduce information about its inner heuristics. A variety of different methods have been developed in this regard. One can use '*counterfactuals*', to test and see what kind of input would lead to a certain output.[67] Methods like LIME use '*surrogate models*' that mimic the opaque model while remaining directly interpretable. These models try to recreate the reasoning of the algorithm by relying on the same input and output of the opaque model. Moreover, as they do not need to understand the inner workings of complex models, they also do not need to access proprietary information, circumventing the trade secrets and interests of external private developers.[68]

Such *post-hoc* methods have in the past been applied to the aforementioned COMPAS system. Unfortunately, this has only shown limited success. Post-hoc explanations of profiling systems have been noted to not provide accurate depictions of these systems, and can sometimes be insufficient to reliably explain complex

[60] Paul B. de Laat, 'Algorithmic Decision-Making Employing Profiling: Will Trade Secrecy Protection Render the Right to Explanation Toothless?' (2022) 24 *Ethics and Information Technology* 17; Grad and Koprowska (n 13); Bayamlıoğlu (n 16).

[61] Bart Custers and Anne-Sophie Heijne, 'The Right of Access in Automated Decision-Making: The Scope of Article 15(1)(h) GDPR in Theory and Practice' (2022) 46 *Computer Law & Security Review* 105727.

[62] Cynthia Rudin, Caroline Wang, and Beau Coker, 'The Age of Secrecy and Unfairness in Recidivism Prediction' (2020) 2 Harvard Data Science Review <https://doi.org/10.1162/99608f92.6ed64b30> accessed 11 March 2024; Grad and Koprowska (n 13).

[63] Andrew Mitchell and Lingxi Tang, 'AI Regulation and the Protection of Source Code' (2023) 31 *International Journal of Law and Information Technology* 123; de Laat (n 26); Custers and Heijne (n 61).

[64] Burrell (n 52).

[65] de Laat (n 26).

[66] Molnar (n 19).

[67] Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2017) 31 *Harvard Journal of Law & Technology* (Harvard JOLT) 841.

[68] Molnar (n 19).

models such as COMPAS.[69] For example, the most famous model-agnostic analysis of the COMPAS system has arguably been created by ProPublica. By using this method, researchers argued that COMPAS contained racial bias.[70] However, this finding has been criticized by others in computer science literature, which shows that model-agnostic methods are not a fool-proof method for explainability.[71]

2.4.2 Intrinsic methods

In lieu of model-agnostic methods, we can also rely on intrinsic explainability methods. These methods aim to make the system directly interpretable to an observer of the system. However, systems such as COMPAS rely on ML methods that are too complex to be directly interpretable. To therefore be able to rely on intrinsic methods, we could for example use simpler algorithms that can easily be interpreted.[72] In essence, this would entail developing a system that is contestable *by design*.[73] That is to say, one would develop a profiling system with the goal of making it explainable, so that it can be contested by a litigant.

In this regard, judiciaries could simply refuse to use complex and opaque ML models in the first place. Still, in some instances, opaque ML techniques can provide us with high quality information and pattern-recognition that is not always possible with intrinsically interpretable methods.[74] Potentially, the accuracy of profiling systems could therefore be higher if we use an opaque method.[75] However, this is not the case in every instance. Studies have shown that simpler and more interpretable algorithms can provide the same level of accuracy as opaque algorithms, such as in the case of COMPAS.[76] Various authors have therefore argued that simple intrinsically interpretable recidivism prediction algorithms can produce as accurate analyses as the opaque algorithms used to create COMPAS.[77]

2.4.3 Technical measures

To address the issues related to technical opacity, integrity concerns, and external private interest, it could be warranted that developers take certain technical measures. Such measures can be understood to include all measures undertaken regarding the design and creation of the system. Such measures reflect the idea that judicial profiling systems should be made contestable *by design*.[78]

These measures could include relying on intrinsically explainable models from the start. At the very least, such should be done for profiling systems that try to predict the risk for recidivism, as some have argued that there is no loss in accuracy when intrinsically explainable models are used for such systems.[79] Regarding integrity concerns and conflicts that can arise from external private interests, developers could also potentially choose to design the system to not rely on input data that can easily be manipulated or rely on private developers who would not hinder the disclosure of an explanation.

[69] Caroline Wang and others, 'In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction' (arXiv, 9 May 2020) <http://arxiv.org/abs/2005.04176> accessed 18 November 2023; Cynthia Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1(5) *Nature Machine Intelligence* 206.

[70] Julia Angwin and others, 'Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks' (ProPublica, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=k0PuAmvq_Xy63TS9ofcxNn6J431eO1RK> accessed 20 February 2024.

[71] Rudin, Wang and Coker (n 62); Wang and others (n 69).

[72] Molnar (n 19).

[73] Almada (n 20).

[74] Yeung and Harkens (n 40).

[75] Molnar (n 19); Bayamlıoğlu (n 16).

[76] Julia Dressel and Hany Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4 *Science Advances*eaao5580 <https://www.science.org/doi/10.1126/sciadv.aao5580> accessed 8 March 2024.

[77] Nienke Tollenaar and Peter G. M. van der Heijden, 'Which Method Predicts Recidivism Best?: A Comparison of Statistical, Machine Learning and Data Mining Predictive Models' (2013) 176 *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 565; Wang and others (n 69); Rudin (n 69); Rudin, Wang and Coker (n 62).

[78] Almada (n 20).

[79] Rudin (n 69).

However, such technical measures must not be overly specified by a right to an explanation. Intrinsically explainable models are not always preferable over the use of model-agnostic methods. Intrinsically explainable models can, for example, be too overly complex for humans to interpret directly. This can lead to a situation of '*information overload*'.[80]

A right to an explanation must therefore not be understood to prescribe specific development measures that must be implemented in *every* situation. Rather, a more nuanced and generalizable approach is needed, which can be amended to the specifics of every situation. Thus, a right to an explanation must require the use of intrinsic models to be implemented *by design*¸ only insofar a model-agnostic method proves insufficient to explain a (particular) profiling system.

2.4.4 Organizational measures

However, if we wish to achieve contestability *by design*¸ we cannot rely only on the technical measures. Organizational measures can also be necessary. Such measures include non-technical actions that can be undertaken by judiciaries, to promote the contestability of the profiling system. Profiling systems do not function in isolation after all. Rather, they are embedded in an institutional context where technology, power structures, and human organizations intersect.[81]

Organizational measures could for example include the requirement to log and register certain actions. This would include either decisions undertaken by the developers during the design and training process of the system, or those made by its human operators while using the system. This could also include documentation on why certain choices were made in designing the system, such as why the system focuses on certain parameters. Moreover, this could also include disclosing the input data that was chosen by the operator of the profiling system.[82] An '*institutional setup*' might also be created, so that the organization that uses the algorithm, is ready to perform the necessary supervision tasks to provide the explanation in question.[83] Lastly, judiciaries must be wary of not putting unnecessary '*social, financial, and organizational*' burdens on the litigants who wish to receive the explanation.[84] This might for example occur when extensive or complicated bureaucratic procedures are put in place, through which an explanation must be requested.[85]

Moreover, judiciaries could also collaborate more closely with the developers of the algorithm, to ensure that the profiling system can be explained to a litigant.[86] Such a collaboration could help ensure that important values such as contestability are safeguarded in the design of the system. In turn, this could also prevent judiciaries from relying on a profiling system that cannot be explained without harming the trade secrets of an externally involved private party or one that is prone to manipulation if it is explained. Still, increased collaboration between developers and legal professionals could perhaps require too much from, sometimes already overworked, judges and legal clerks.[87]

Alternatively, judiciaries can include clear transparency agreements with an external private developer in a procurement contract.[88] Through this, they can assure that the developer will not hinder the disclosure of the inner workings of the algorithm based on their business interests. Secondly, such a contract could also ensure that the developer takes adequate measures to prevent integrity concerns from arising as well. The use of a standard set of contracts that is commonly used by multiple judiciaries, could moreover strengthen the negotiation power of such public bodies vis-à-vis private developers.

[80] Andrew Bell and others, 'It's Just Not That Simple: An Empirical Study of the Accuracy-Explainability Trade-off in Machine Learning for Public Policy' (2022) ACM Conference on Fairness, Accountability, and Transparency <https://dl.acm.org/doi/10.1145/3531146.3533090> accessed 12 April 2024.

[81] Waldron (n 14); Bayamlıoğlu (n 16).

[82] Selbst and Barocas (n 51).

[83] Bayamlıoğlu (n 16); Tal Z. Zarsky, 'Transparent Predictions' (2013) 2013 *University of Illinois Law Review* 1503.

[84] Aviva de Groot, *Care to Explain?: A Critical Epistemic in/Justice Based Analysis of Legal Explanation Obligations and Ideals for 'AI'-Infused Times* (PhD thesis, Tilburg University 2023) <https://www.tilburguniversity.edu/current/events/phd-defense-a-groot> accessed 9 April 2024.

[85] de Groot (n 84) 170.

[86] Yeung and Harkens (n 40).

[87] Hildebrandt (n 16); Yeung and Harkens (n 40).

[88] Lavi M. Ben Dor and Cary Coglianese, 'Procurement as AI Governance' (2021) 2 *IEEE Transactions on Technology and Society* 192.

3. The right to a fair trial

In the previous paragraphs, I argued that, to effectively safeguard contestation, a litigant must receive an explanation. An explanation cannot always be assured *post-hoc,* after an opaque model has already been developed. Therefore, a right to an explanation must require that contestability is safeguarded *by design,* through different technical and organizational measures. In the next few paragraphs, I will analyze the case law of the European Court of Human Rights (ECtHR) on art. 6 ECHR to see if it contains a right to an explanation for judicial profiling systems that can safeguard contestation *by design*.[89]

3.1 The adversarial principle

In legal scholarship, academics have argued that algorithmic opacity could be in breach with the right to a fair trial, as it would not be in line with the *adversarial principle*.[90] This principle has been formulated as '*the opportunity for the parties to a criminal or civil trial to have knowledge of and comment on all evidence adduced or observations filed*' before the court.[91] Under the scope of the adversarial principle fall not only documents filed by one of the litigating parties, but also the observations of various other persons, such as members of a national legal service, representatives of the administration, lower courts, or even the court hearing the case itself.[92]

Therefore, as all evidence and observations in the hands of the courts need to be disclosed to the litigant during their trial, we can state that the same holds true for documentation about the judicial profiling system that a judge might have access to. This could occur for example if the judge knows on what grounds a person was profiled, what weight has been assigned to certain parameters, or how these parameters correlate. In other words, if the court understands on what grounds a profiling system assessed a litigant, it should disclose this information to that litigant as well.

However, what if no information about the internal reasoning of the system is in the hands of the court? What if a judge only has access to the output of the system, which does not disclose the grounds on which it has been generated? In the civil case of *Yvon v. France*, the ECtHR stated that the adversarial principle does not require a party to '*transmit to its opponent documents which (…) have not been presented to the court either*'.[93] As a consequence of this, the disclosure of an explanation based on the adversarial principle would depend on whether or not the *judge, the opposing party*, or the *public prosecutor* are in possession of information about the grounds on which the litigant was profiled. In the case that the judge only has access to the system's *output*, it is only this *output* that needs to be divulged. If a judge is merely presented with, for example, a recidivism risk-score, it is only this risk-score that would need to be divulged.

This seems to be the case too, not only in civil and administrative trials, but also in criminal cases, as the disclosure of evidence is also similarly mandated by the adversarial principle in these instances.[94] There is for example no violation of the adversarial principle if the evidence in question does not form a part of the prosecution case, or was never put before the court or any other party.[95]

[89] While in this paper I do not engage directly with the EU Charter of Fundamental Rights (Charter), art. 47 Charter does provide similar protection to the right to fair trial as laid down in art. 6 ECHR. As a result of art. 52 (3) Charter, the meaning and scope of art. 6 ECHR also applies to art. 47 Charter. Therefore, the analysis provided in this paper can be applied to a certain extent to art. 47 Charter as well.

[90] Moreira (n 5); Ulenaers (n 5).

[91] Kress v France (2001) App no 39594/98 (ECtHR [GC], 7 June 2001) para 74; Lobo Machado v Portugal (1996) App no 15764/89 (ECtHR [GC], 20 February 1996) para 31; McMichael v the United Kingdom (1995) App no 16424/90 (ECtHR, 24 February 1995) para 80; Ruiz-Mateos v Spain (1993) App no 12952/87 (ECtHR, 23 June 1993) para 63; Vermeulen v Belgium (1996) App no 19075/91 (ECtHR [GC], 20 February 1996) para 33.

[92] *Köksoy v Turkey* (2020) App no 31885/10 (ECtHR, 19 May 2020) paras 34–35.

[93] *Yvon v France* (2003) App no 44962/98 (ECtHR, 24 April 2003) para 38.

[94] *Rowe and Davis v the United Kingdom* (2000) App no 28901/95 (ECtHR [GC], 16 February 2000) para 176.

[95] *Jasper v the United Kingdom* (2000) App no 27052/95 (ECtHR [GC], 16 February 2000) para 55.

3.2 The right to a reasoned judgment

The right to a fair trial also includes the obligation for judges to give sufficient reasons for their verdicts after the trial has concluded.[96] The proper reasoning of a verdict serves the important function of showing parties that they have truly been heard, which improves the acceptance of the verdict by the litigants.[97] Although the right to a fair trial does not contain a right to appeal, it does require that a judge '*must give such reasons as to enable the parties to make effective use of any existing right of appeal*'.[98] Consequently, the reasons that the judge formulates in their verdict, have to enable the litigant to be able to properly contest that verdict before a higher court.

One could argue that, if the accuracy rate of a system is on average high enough, it does not matter whether or not we know on what grounds it was generated.[99] If the output of a commonly used profiling system is used by a judge in their verdict, they might state that the output of the system is most likely accurate, even though we do not know on what grounds it was generated.[100]

However, such a verdict would not be based on actual reasons, but on the probability that the analysis of the system, and in turn also the verdict, is correct. Different scholars have argued that this goes against the very essence of reason-giving in legal procedures.[101] Providing reasons for a verdict entails that a judge must justify why the output is indeed accurate and not merely state that it is very likely to be accurate. To this end, the judge should disclose the grounds that the system relied upon. Litigants are then, as a result of such an explanation, better able to contest the output of the system in any appeal they might wish to file, as they can dispute whether these grounds are indeed justified.

Arguably, while ECtHR case law on the right to a reasoned judgment has always concerned human reason-giving, it can be applied to the novel situation where a ML-based system is generating important pieces of evidence. The ECtHR has stated that the manner in which a verdict should be justified, and what kinds of reasons need to be given, depend on the nature of the decision, and can only be ascertained in light of the specific circumstances of the case.[102] Considering the design measures we discussed in paragraph 1.3, these specific circumstances could for example be the availability of alternative design methods for the algorithm that can make the system more explainable and, therefore, contestable. Based on this, we can argue that a right to an explanation is mandated for profiling systems by the right on a reasoned judgment, as a litigant must be able to make '*effective use of any existing right to appeal*'. Or in other words, a litigant should be able to meaningfully contest a verdict that has been influenced by the system.[103] A litigant cannot properly do this, as we explained earlier, if the profiling system's criteria are not explained.

However, the *timing* of the right to a reasoned judgment does provide litigants with a fundamental obstacle when they wish to contest a profiling system. The right to a reasoned judgment namely applies to the *final verdict*. It does not require that the profiling system should be explained *during the trial,* unlike the adversarial principle. This makes it only effectively possible for the litigant to contest a verdict that has been influenced by the profiling system during an appeals procedure.

3.3 The limitations of the right to a fair trial

The right to a reasoned judgment could potentially provide litigants with an explanation in appeals procedures. However, what if it was simply impossible to interpret the algorithm, and therefore provide an explanation? What if the system was too technically complex? Would the use of such an opaque and inscrutable algorithm be in breach of art. 6 ECHR?

[96] *H v Belgium* (1987) App no 8950/80 (ECtHR, 30 November 1987) para 53.

[97] *Magnin c France* (dec) [2012] App no 26219/08 (ECtHR, 10 May 2012) para 29.

[98] *Hirvisaari v Finland* [2001] App no 49684/99 (ECtHR, 27 September 2001) para 30.

[99] Yeung and Harkens (n 40).

[100] Yeung and Harkens (n 40).

[101] Hildebrandt (n 16); Yeung and Harkens (n 40).

[102] *Hiro Balani v Spain* (1994) 19 EHRR 566, App no 18064/91 (ECtHR, 9 December 1994) para 27; *Ruiz Torija v Spain* (1994) 19 EHRR 553, App no 18390/91 (ECtHR, 9 December 1994) para 29.

[103] *Hirvisaari v Finland* (n 98) para 30.

The right to a fair trial does not generally lay down rules on whether evidence or observations are admissible. The weighing of evidence is a matter for the national court to decide on.[104] Moreover, as the right to a fair trial is limited to proceedings '*before a tribunal*', it is generally not applicable to acts undertaken outside of that scope.[105] Consequently, art. 6 ECHR cannot prescribe design requirements for the development of profiling systems, as the system is already developed before any trial has even started.

Art. 6 ECHR therefore cannot require a *contestability by design* requirement. Art. 6 ECHR can only offer a *post-hoc* approach to explainability. The right to a fair trial is namely a provision that is primarily focused on procedural safeguards. Its case law is therefore not well-suited, nor aimed at, regulating the development and design of ML-based systems. A more technology-specific approach is therefore needed to provide litigants with an effective right to an explanation.

Moreover, various public interests can also restrict the disclosure of information under the right to a fair trial.[106] Both disclosure of materials during trials as well as in the final verdict, can be limited by such interests.[107] However, such non-disclosure must be necessary and proportional, and not harm the essence of the right to a fair trial.[108] On the other hand, if the non-disclosure would have no impact on the trial, the evidence and observations in question would also not need to be disclosed.[109]

In the past, such concerns have hindered explanation requests of profiling systems. In the Netherlands, there was a case in which an individual wished to understand how a psychological evaluation system assessed their application for a gun permit. In this instance, the government agency using the system feared that people might play into the criteria that the system relied upon by altering their responses. Because of this, the request for an explanation about the system in question was denied by the judge, who ruled that art. 6 ECHR was not breached by the non-disclosure of the explanation.[110]

Consequently, integrity concerns might prove to be a valid exception to the right to a fair trial, and therefore also the right to an explanation. However, the right to a fair trial cannot put forward design and development requirements for profiling systems to prevent such situations from arising in the first place. Art. 6 ECHR, with its ex post approach, is therefore, on its own, insufficient to safeguard contestation.

4. EU Data Protection Law

In the previous paragraphs, we saw that the right to a fair trial is not well-suited to effectively safeguard contestation through a right to an explanation. Its procedural focus entails that it cannot require development requirements for profiling systems. However, another foundation for the right to an explanation can potentially be found in Data Protection Law. The debate on this right has mostly revolved around the General Data Protection Regulation (GDPR).[111] In this paper, I will also discuss another important legal framework in this regard: the Law Enforcement Directive (LED),[112] and ascertain whether Data Protection Law can effectively safeguard a contestation enabling the right to an explanation.

[104] *García Ruiz v Spain* (1999) 31 EHRR 589, App no 30544/96 (ECtHR [GC], 21 January 1999) para 28.

[105] *Mantovanelli v France* (1997) 24 EHRR 370, App no 21497/93 (ECtHR, 18 March 1997) para 33.

[106] *Regner v the Czech Republic* (2017) 66 EHRR 8, App no 35289/11 (ECtHR, 19 September 2017) para 39.

[107] *Regner v the Czech Republic* (n 106) para 158.

[108] *Corneschi v Romania* (2022) App no 21609/16 (ECtHR, 12 May 2022) para 97; *Regner v. the Czech Republic* (n 106) paras 147–149.

[109] *Jasper v. the United Kingdom* (n 95) paras 54–55; *M v the Netherlands* App no 2156/10 (ECtHR, 25 July 2017) para 69.

[110] *Koninklijke Nederlandse Jagers Vereniging te Amersfoort v de Staat der Nederlanden* [2020] ECLI:NL:RBDHA:2020:1013, District Court The Hague, C-09-585239-KGZA 19-1221, para 4.33.

[111] Regulation (EU) 2016/679 (GDPR) OJL119/1 (n 21).

[112] Directive (EU) 2016/680 (LED) OJL119/89 (n 22).

4.1 The General Data Protection Regulation

In the GDPR we can find two regulatory tools that could potentially provide litigants with a right to an explanation. On the one hand, there is the right not to be subjected to solely automated decision-making, and its related set of safeguards such as the right to contest, present in art. 22 GDPR. On the other hand, there is the right to receive meaningful information about the logic involved in automated decision-making, present in art. 13, 14, and 15 GDPR.

4.1.1 Art. 22 GDPR and Recital 71

The term '*right to an explanation*' is not mentioned explicitly in any of the articles of the GDPR. To understand how we can deduce this right from the GDPR, we first need to explore the ban on fully automated individual decision-making, present in art. 22 GDPR. Art. 22 (1) GDPR contains 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*'. The ban of art. 22 (1) GDPR does not apply according to art. 22 (2) GDPR if the decision is necessary for the fulfillment of a contract (a), is authorized by EU or national law which '*also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests*' (b), or is based on explicit consent (c). Art. 22 (3) GDPR formulates a set of safeguards to protect the data subject's rights, freedoms, and legitimate interests, which need to be implemented when processing occurs under subparagraph (a) or (c). These include a set of active rights for the data subject that contain '*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*.'

The term '*right to an explanation*' can be found in Recital 71, which states that the following safeguards must be implemented by the data controller when a data subject is subjected to solely automated decision-making: '*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*'. There is disagreement among legal scholars, however, as to whether we can deduce from this a *direct* right to an explanation, as Recitals are not directly binding legal rules, but rather aid in the interpretation of provisions such as art. 22 GDPR.[113]

Alternatively, some scholars have argued that a right to an explanation is *indirectly* mandated by art. 22 GDPR, as it is explicitly mentioned that data controllers must safeguard the '*right to contest*', in any case.[114] As contestation is contingent on explanations, a right to an explanation is indirectly mandated through this provision. A basis for this view can be found in the EDPB guidelines on art. 22, which states 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*'.[115]

Assuming that there *is* a right to an explanation that can be deduced from a combined reading of art. 22 GDPR and Recital 71, there still remain some hindrances to applying these provisions. The scope of art. 22 GDPR is mainly limited to '*solely automated decision-making that produces legal or significant effects*'. The phrase '*legal or significant effects*' has sparked debate over its precise meaning.[116] In any case, for the purposes of litigant contestation, this provision seems to be clearly applicable, as algorithms used in the administration of justice directly affect the legally binding verdict of a judge.

However, an obstacle exists for the applicability of this provision in the field of adjudication, because of the phrase '*solely automated decision-making*'. This phrase entails that art. 22 GDPR applies to cases where decisions do not involve any human intervention. However, some have discussed whether the mere presence of a 'token' human with no decision-making power would be enough to satisfy this requirement.[117] In this regard, the European Data Protection Board has commented in their guidelines that human

[113] Edwards and Veale (n 26); Kaminski (n 26); Malgieri (n 26); Selbst and Powles (n 26); Wachter, Mittelstadt and Floridi (n 26).

[114] Bayamlıoğlu (n 16); Sarra (n 17).

[115] Article 29 Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (EDPB Guidelines, WP251rev.01, 2017) <https://ec.europa.eu/newsroom/article29/items/612053> accessed 18 February 2024.

[116] Michael Veale and Lilian Edwards, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' (2018) 34 *Computer Law & Security Review* 398.

[117] Riikka Koulu, 'Proceduralizing Control and Discretion: Human Oversight in Artificial Intelligence Policy' (2020) 27 *Maastricht Journal of European and Comparative Law* 720; Veale and Edwards (n 26).

intervention needs to be '*meaningful*'.[118] What '*meaningful*' entails also remains quite vague, although academic scholarship has formulated some requirements for how meaningful human intervention should be understood.[119]

However, this focus on solely automated decision-making can be problematic, as profiling systems typically support a judge, and do not fully automate the decision-making process. In this regard, the recent SCHUFA ruling of the CJEU clarifies how art. 22 GDPR applies to profiling systems. In the SCHUFA case, the output of the profiling system in question held such preponderant weight that a decision-maker simply never differed from the conclusions made by the system. According to the CJEU, decision-making can be seen as solely automated if the human decision-maker '*draws strongly*' on the information produced by the system.[120]

It is therefore important to assess the role that the profiling system takes in court proceedings. It is unclear, however, whether the preponderant weight assigned to the output of the system in the SCHUFA case will arise in the context of adjudication. This is because a judge is typically free in their assessment of the case before them. During the implementation of COMPAS in the U.S.A, it was for example emphasized that the risk-recidivism score created by the system could always be disregarded by the judge.[121] Still, in practice, concerns related to '*automation bias*', might lead to a situation in which a judge does not exercise proper human oversight over the system.[122]

In any case, we find that in this instance, the GDPR only provides litigants the ability to contest the system if there is no proper human oversight. However, the mere fact that a judge oversees, or even contests, the profiling system in question, does not negate the need or desire of a litigant to contest the system for themselves. While it can be important that the judge contests and criticizes the system, it remains vital that outside voices are also heard, as litigants can provide fresh and new perspectives on the system and its output.[123] Therefore, the fact that the profiling system does not play a preponderant role in the decision-making process, does not negate the need for a right to an explanation. It is therefore questionable to what extent art. 22 GDPR can safeguard an effective right to an explanation for litigants.

4.1.2 Art. 13, 14, and 15 GDPR

We can find another basis for the right to an explanation in art. 13, 14 and 15 GDPR. Art. 13 and 14 GDPR, respectively, set forth a notification obligation for data controllers to inform data subjects when data is collected from them directly or indirectly. Moreover, art. 15 (1) (h) provides a right of access to information that needs to be actively exercised by data subjects themselves. These three provisions oblige data processors to provide a data subject with information on '*the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject*'.

There is extensive debate as to whether or not this merely requires an ex-ante explanation of how the system *generally* works, or whether they require an ex-post right to an explanation of the final decision.[124] Without going into too much detail on this debate, the current case law on art. 15 (1) (h) GDPR seems to have sided with the latter interpretation.[125]

[118] Article 29 Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling' (n 115).

[119] Ben Wagner, 'Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems' (2019) 11 *Policy & Internet* 104.

[120] SCHUFA Holding AG v. X [2023] CJEU C-634/21, ECLI:EU:C:2023:957, para 73.

[121] Angwin and others (n 70).

[122] Crootof (n 8); Koulu (n 117); Wagner (n 119).

[123] Waldron (n 14).

[124] Bayamlıoğlu (n 16); Kaminski (n 26); Kaminski and Urban (n 24); Sarra (n 17); Selbst and Powles (n 26); Wachter, Mittelstadt and Floridi (n 26).

[125] Datenschutzbehörde, 'Rights of the Data Subject' (2024) <https://www.data-protection-authority.gv.at/data-protection-in-austria/rights-of-the-data-subject.html> accessed 7 April 2024; Uber BV v Drivers Union [2021] District Court of Amsterdam, ECLI:NL:RBAMS:2021:1020.

The question arises however, whether the right to an explanation under art. 13, 14, and 15 GDPR, is limited to the same scope of '*solely automated decision-making*', as art. 22 GDPR and Recital 71. This is because it is unclear whether the phrase '*at least in those cases*' refers to '*automated decision-making including profiling*' in general, or specifically to the part that describes '*Article 22 (1) and (4) GDPR*'.

The District Court of Amsterdam promotes a strict reading of art. 15 GDPR that limits it to cases that fall under the scope of '*solely automated decision-making*' present in art. 22 GDPR.[126] On the other hand, the Austrian Data Protection Authority, the *Datenschutzbehörde,* published a ruling that contains a broader interpretation of art. 15 GDPR. It argues that this provision includes *any* form of automated data processing, and that it is not only applicable to solely automated decision-making.[127] The debate on the scope of art. 13, 14 and 15 GDPR therefore remains unresolved.

Besides the potentially limited scope of these provisions, concerns have also been raised concerning whether trade secret rights could leave the right to an explanation '*toothless*'.[128] Recital 63 states that data access rights '*should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software*'. Various authors have argued that this similarly applies to the right to an explanation under the GDPR.[129]

The question therefore rises to what extent intellectual property rights and trade secrets can curtail the right to an explanation.[130] Full disclosure of the source code or training data of the system, might be opposed by developers who do not want to give their competitors any advantages. However, both the District Court of Amsterdam, and the Austrian Data Protection Authority '*Datenschutzbehörde*', argue that the right to an explanation does not necessitate the full disclosure of the entire system. Rather, concrete and specific criteria based on which a certain output was generated needs to be disclosed. This would avoid harming the rights and interests of any external private partner involved.[131]

Integrity concerns could also lead to similar issues in the provision of a right to an explanation, as several member states provide exceptions to the data access rights of art. 13, 14, and 15 GDPR, when national security might be at risk. Such provisions can, for example, be found in the national data protection laws of the Netherlands, Germany, and Belgium.[132]

Be that as it may, as we discussed in paragraph 1.3, such a situation could be avoided altogether through the implementation of technical and organizational measures in the design process of the system. In this regard, it is interesting to note that the GDPR could necessitate that profiling systems should be made contestable '*by-design*'.[133] Art. 25 GDPR specifically requires that data controllers '*implement appropriate technical and organizational measures*', for the goal of meeting the requirements of the GDPR. Based on this, the right to an explanation could be seen as both an ex post procedural right, as well as a set of substantive requirements that can influence the development of the profiling system in question.

[126] Ljubiša Metikoš, 'Leg het me nog één keer uit: het recht op een uitleg na Uber en Ola. Annotatie bij Hof Amsterdam, 4 April 2023' (2023) *Privacy & Informatie* 3; Uber BV [2021] ECLI:NL:RBAMS:2021:1020.

[127] Datenschutzbehörde (n 125).

[128] de Laat (n 26).

[129] de Laat (n 26); Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2(2) *Columbia Business Law Review* 494; Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 *International Data Privacy Law* 243.

[130] Wachter and Mittelstadt (n 129); de Laat (n 26); Bayamlıoğlu (n 16).

[131] Uber BV v Personenvervoer Nederland BV [2023] Court of Appeal of Amsterdam, ECLI:NL:GHAMS:2023:793, para 3.28.

[132] Uitvoeringswet Algemene verordening gegevensbescherming (UAVG), art 41(1); Bundesdatenschutzgesetz (BDSG), art 29(1); Wet betreffende de bescherming van natuurlijke personen met betrekking tot de verwerking van persoonsgegevens, art 37(2)–(4).

[133] Almada (n 20); Bayamlıoğlu (n 16); Sarra (n 17).

4.2 The Law Enforcement Directive

The GDPR covers most processing of personal data undertaken by judiciaries. Nevertheless, art. 1 (2) (d) GDPR does state that the GDPR does not apply to data processing intended for the '*purposes of the prevention, investigation, detection or prosecution of criminal offenses or the execution of criminal penalties*'. While certain profiling systems have been used in civil and administrative procedures,[134] the most notable profiling systems have been those that have predicted the risk for recidivism in criminal trials.[135]

For these types of systems, the EU legislator has created a separate set of provisions in the LED.[136] This directive, therefore, regulates criminal data processing undertaken by judicial bodies. It remains quite unclear, nevertheless, what the interplay is between the GDPR and the LED, as some Member States place some tasks relating to criminal prosecution under the tasks of administrative bodies.[137]

Moreover, the LED also contains a prohibition on automated decision-making similar to the one present in art. 22 GDPR. Art. 11 (1) LED states that a decision based solely on automated processing, which produces an adverse legal effect concerning the data subject or significantly affects him or her, is to be prohibited, unless it has a legal basis in Member State or Union law. If such a basis exists, suitable safeguards must be implemented to protect the rights and freedoms of the data subject. However, the safeguards mentioned here are limited to simply requiring '*human intervention*'. Still, similar to Recital 71 GDPR, Recital 38 LED does provide a list of additional safeguards that should also be included, such as the ability '*to obtain an explanation of the decision reached after such assessment or to challenge the decision*'. This again mimics the safeguards of the '*right to contest'* and the '*right to an explanation'* present in the GDPR.

Moreover, this right can also be seen as both a procedural right, as well as a set of development requirements. Similar to art. 25 GDPR, art. 20 LED obliges that appropriate technical and organizational measures are implemented to fulfill the requirements of this directive. To this end, the right to an explanation could be seen as requiring that criminal profiling systems should be designed in a human-comprehensible manner, so that the decision based on its output may be challenged.

However, the right to contest is not explicitly present in the text of art. 11 LED itself. It only mentions the right to human intervention. Unlike with the GDPR, we cannot, therefore, argue that the safeguards present in art. 11 LED indirectly requires a right to an explanation, as the right to contest is not present in art. 11 LED at all.

But even if we assume that a right to an explanation can be read into art. 11 LED solely by relying on Recital 38 LED, there still remains the issue of scope. Art. 11 LED is after all also limited to cases of solely automated decision-making and would therefore not aid litigants facing profiling systems, whose output a judge can effectively ignore. If we assume that the analysis of art. 22 GDPR in the SCHUFA case could be applied to the LED *mutatis mutandis*, litigants faced with criminal profiling systems would only have a right to an explanation if the system would have a '*preponderant'* influence.[138]

One might then ask whether there is an alternative way to ground the right to an explanation in the information rights of data subjects, like in art. 13, 14 and 15 GDPR. Unfortunately, this is not possible. In art. 13, and 14 LED various information rights for data subjects are formulated. Neither article mentions any requirement to divulge '*meaningful information about the logic involved',* however. As such, the right to an explanation is even more difficult to deduce for judicial profiling systems from the LED than it is from the GDPR.

[134] Stankovich (n 12).

[135] Rudin, Wang and Coker (n 62).

[136] Directive (EU) 2016/680 (LED) OJL119/89 (n 22).

[137] Mireille M Caruana, 'The Reform of the EU Data Protection Framework in the Context of the Police and Criminal Justice Sector: Harmonisation, Scope, Oversight and Enforcement' (2019) 33 *International Review of Law, Computers & Technology* 249.

[138] SCHUFA (n 120).

In conclusion, we can see that EU Data Protection Law, even under the broadest interpretations, is inadequate to safeguard contestation in the context of algorithm-*assisted* adjudication. To a certain extent, both the GDPR and the LED can offer individuals procedural rights such as the right to an explanation.[139] However, they are not well equipped to regulate the use of technology in the administration of justice. Still, the AIA could provide a more sector-specific approach that can better cover the use of judicial profiling systems.

5. The AI-Act

The AIA is a new EU regulation that aims to regulate the development, deployment, and use of AI-systems.[140] It does this through a selective risk-based approach, providing more stringent obligations for AI-systems that present a higher risk to public health, safety, and fundamental rights. The AIA sets out to regulate these high-risk AI-systems through a variety of regulatory instruments that aim to increase and strengthen robustness, human oversight, and most importantly, transparency. In the following section, I discuss how the AIA applies to the judicial context and how it can safeguard a right to an explanation. I also discuss its shortcomings in light of the lessons we learned from the right to a fair trial and EU Data Protection Law.

5.1 Judicial high-risk AI

The scope of the AIA encompasses a wide array of different judicial systems. In this regard, Recital 28 AIA states that AI-systems can pose a danger for the rule of law and fundamental rights. Judicial AI-systems are especially impactful, in this regard, according to Recital 61 AIA, '*considering their potentially significant impact on democracy, rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial*.'

The AIA puts forward additional requirements on systems that have a high-risk of harming health, safety, and fundamental rights. According to art. 6 (2) and Annex III (8) (a) AIA, all AI-systems that intend '*to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts*' are to be classified as belonging to this high-risk tier. As a result of this classification, various obligations are put on the producers, providers, and users of judicial AI.

Here we see a clear difference in regard to the scope of art. 22 GDPR and art. 11 LED. The AIA is not limited to solely automated decision-making. Rather, *assistance* systems used in the administration of justice are now also regulated under this new regime. This fits much better with the current use of profiling systems and other types of assistance technologies in the judiciary.

However, art. 6 (3) AIA exempts AI-systems from this regime, if they do not pose a significant risk to health, safety, or fundamental rights by not materially influencing the outcome of a decision-making process. To specify the scope of this provision, art. 6 (3) AIA also lists a number of instances when this is the case. Nevertheless, this subparagraph ends by stating that all profiling systems of natural persons are considered to be high-risk AI. All judicial profiling systems are therefore considered to be high-risk under the AIA.

Besides regulating high-risk AI-systems, the AIA also lists a number of AI-systems that show '*unacceptable*' levels of risk. Amongst these, art. 5 (1) (d) prohibits AI systems that make '*risk assessments of natural persons in order to assess or predict the risk of a natural person to commit a criminal offense, based solely on the profiling of a natural person*.' At a first glance, this might mean that the use of criminal profiling systems such as COMPAS, which predict recidivism, will be prohibited. However, the addition of the phrase '*solely*' brings forth a similar issue to the one we noted in our discussion of the right to an explanation in EU Data Protection Law.

[139] Sarra (n 17).

[140] European Parliament, 'Legislative resolution of 13 March 2024 (Artificial Intelligence Act)' (n 23).

In practice, a judge will most likely have the final say over a verdict, as they can always overrule the profiling system. Recital 61 even prescribes this, by stating that AI can support judges, but should never replace their decision-making power. Perhaps future case law, similar to the SCHUFA ruling,[141] will interpret the phrase '*solely*' as referring to profiling evidence that has a preponderant influence on the final verdict. This might also be the case if automation bias were to occur.[142] However, if we interpret the phrasing '*solely*' more narrowly, we can argue that profiling systems that predict the likelihood of criminal offense, in cases where a judge decides whether or not to release a prisoner earlier, are only prohibited if judges are forced to automatically make a certain decision. In all other instances then, profiling systems would be allowed.

5.2 A new right to an explanation

The AIA provides litigants with a new right to an explanation in art. 86 (1) AIA. Here, it says that this right encompasses any: '*affected person subject to a decision which is taken by the deployer on the basis of the output from an high-risk AI system which produces legal effects*'. Based on this text it is very likely that high-risk judicial systems would fall under this provision. Such systems can be considered to produce legal effects, as they can contribute to a legally binding verdict.

A litigant would have the right to receive information '*on the role of the AI system in the decision-making procedure, the main parameters of the decision taken and the related input data*'. The wording of this provision is, arguably, much clearer than that of the GDPR and the LED. It provides specific categories of the kind of information that should be disclosed. The *'main parameters'* would require that a litigant be informed about the main criteria that the profiling system took into account. Such grounds can be contested, as we saw in paragraph 1.1.2, for being unsuitable to profile the specific litigant in question. Moreover, by knowing what '*input data*' the algorithm relied upon, the litigant can, in certain instances, also contest the information used about themselves for being inaccurate.

However, this provision misses any mention of the need to disclose the weight of different input variables. While the parameters and input data might be accurate, a litigant might also wish to contest the relative importance of certain data points. In addition, there is also no mention of disclosing the *relationship* between different parameters. Moreover, the pattern and correlation that an algorithm finds between different parameters can be contested as well by a litigant. As was discussed in paragraph 1.1.2, a *'reasonable hypothesis'* as to the reasoning of the algorithm in this regard should be disclosed as well.[143] While a specific parameter might be justified to profile a litigant, such a parameter could be connected by the profiling system to other characteristics of the individual that should not be taken into account, such as their race or ethnicity.

The AIA also makes references to the understandability of the explanation. In Recital 171 AIA, it is stated that an explanation should be '*clear and meaningful and should provide a basis on which the affected persons are able to exercise their rights*', which could necessitate that the explanation be made understandable. This understandability requirement is an important step in providing effective explanations that enable contestation. However, the Recital leaves aside the need for more individualized explanations that take into consideration those who are less technically literate.

For example, on its own, full disclosure of complex technical information to a layperson will most likely not result in an understanding of the profiling system that would enable them to contest its output. Granting *too* much information might create a situation of information-overload. This would then actually increase the opacity of the system.[144]

[141] SCHUFA (n 120).

[142] Crootof (n 8); Wu (n 6).

[143] Yeung and Harkens (n 40).

[144] Cynthia Stohl, Michael Stohl and Paul M. Leonardi, 'Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age' (2016) 10 *International Journal of Communication* 15.

Ostensibly, those who lack either the technical skill, time, or other resources, to process such information, can be ignored when difficult-to-understand technological systems mediate parts of decision-making processes.[145] It is therefore important that the explanation is understandable from the perspective of the individual receiving it. To this end, explanations can for example be more personalized, and be adapted to the specific litigant's level of technical literacy as well as their capability to understand the explanation.[146] In this regard, the AIA lacks sufficient attention to those litigants who are less technologically literate, or who because of personal circumstances, temporarily are less able to function in complex technologically mediated decision-making processes.[147] Still, the fact that it does mention the need to provide legible and understandable explanations is commendable, as such an explicit requirement is not discussed in the provisions of the GDPR or the LED.

5.3 Technical and organizational measures

Art. 86 AIA does not mention any technical and organizational measures that can make judicial AI-systems be explainable and contestable *by design*. The AIA does not explicitly mention the need for such measures, to specifically safeguard the right to an explanation. Nevertheless, art. 13 (1) AIA does require that high-risk AI-systems '*be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable deployers to interpret the system's output and use it appropriately'*. Arguably, this provision would have as a consequence that, if model-agnostic methods cannot provide a sufficient explanation of the profiling system, one should rely on an intrinsically explainable system, as we discussed in paragraph 1.3.2. There is, however, no explicit reference to art. 86 AIA in art. 13 AIA, nor is there any mention to develop the system to be explainable to an individual affected by the high-risk AI-system. Art. 13 AIA only applies to explanations given to a judge.

Nevertheless, Recital 27 does state that transparency is understood to mean '*that AI systems are developed and used in a way that allows appropriate traceability and explainability'*, which includes duly informing '*affected persons about their rights*.' The ability to explain the system to an affected person could therefore be read into the requirement of art. 13 AIA that the system is developed to be '*transparent'*. Nevertheless, the wording of Recital 27 is not a completely perfect fit for such an argument, as it does not explicitly mention giving an affected person an explanation of the system and its output. Rather, Recital 27 only states that an affected person needs to be informed about what rights they have.

However, the judge will in any case still understand on what grounds a system profiled a litigant. This is because, as we just discussed, art. 13 AIA requires that the judge be able to understand the system's output. Moreover, art. 14 AIA also prescribes that the system be designed in such a way that it can be overseen by the natural persons who use it. Consequently, a judge should be able to understand the profiling system to perform human oversight tasks. Therefore, a litigant might still invoke their rights under the right to a fair trial, such as the right to adversarial proceedings or the right to a reasoned judgment, to request such information. In those circumstances, a litigant might therefore still have an indirect right to an explanation.

However, there is no direct requirement for the system to be designed to be explainable to the litigant. Issues related to integrity concerns or external private interests might therefore still arise. This problem is exacerbated by art. 86 (2) AIA. Here, a very broad and vaguely defined exception to the right to an explanation is formulated. It states that EU or national law may restrict the disclosure of an explanation. In paragraph 1.2.3, we saw that the rights of private external parties could hinder the disclosure of an explanation. In paragraph 2.3. we also saw that the right to a fair trial contains an exemption for integrity concerns that could arise as a result of the disclosure of an explanation. This, in turn, could provide a gap in the exercise of the right to an explanation, as there is no absolute requirement to develop effectively explainable and contestable systems. As a result of this, the right to an explanation can in theory be restricted in any way that a national legislator deems fit.

[145] Sofia Ranchordas, 'Empathy in the Digital Administrative State Automating the Administrative State' (2021) 71 *Duke Law Journal* 1341.

[146] Laato and others (n 33); Miller (n 32); Burrell (n 52).

[147] Ranchordas (n 145).

5.4 Procurement contracts

The AIA, in short, puts forward technical explainability requirements. It lacks, however, requirements that would prevent issues concerning the system's integrity and external private interests from arising. In this regard, more in-depth involvement from the judiciaries using profiling systems may resolve some of these issues. As we saw in paragraph 1.3.3, by being more involved in the design process of the systems that they employ, judiciaries may push for more contestable design choices.[148] Nevertheless, as we also discussed in paragraph 1.3.3. we should not overestimate the capabilities of judges and legal clerks in this regard.

Alternatively, by prescribing contestability requirements in procurement contracts, we could provide the necessary guidance to the external private developers of profiling systems.[149] In this regard, the use of standard procurement contracts might help strengthen the position of judiciaries to create and push for the development of contestable profiling systems.[150] The European Commission has recently provided a draft for such a document, specifically for public authorities that wish to use high-risk AI-systems that have been developed by external private parties.[151] Art. 6 of this standard contract mentions the requirement that both technical and organizational measures be taken to ensure that an explanation can be granted. This, in turn, could help promote the development of systems that are not only explainable to judges, but also litigants.

6. Conclusion

While ML profiling systems offer the potential to enhance judicial quality, their opacity poses challenges to litigants' ability to contest their output. Specifically, technical opacity, integrity concerns, and external private interests, form fundamental obstacles to the disclosure of an explanation and, subsequently, the exercise of contestation. The proposed solution, a right to an explanation, extends beyond being a mere procedural right. Rather, it requires for contestability measures to be integrated into the design of these systems from the outset.

The right to a fair trial, being primarily concerned with procedural safeguards, cannot effectively protect litigants in this regard. Similarly, the wording of the right to an explanation in Data Protection Law is currently quite vague and potentially ineffective in fully providing full protection. The AIA takes an important step in the right direction. However, some glaring loopholes still remain regarding the development obligations for these systems. Integrity concerns and the rights and interests of external private parties might still become fundamental obstacles to the exercise of the right to an explanation. However, judiciaries could step up to the task of safeguarding contestation and put forward additional transparency requirements on the developers of the profiling systems that they use.

In short, there is a need to shift the development of profiling systems towards explainability. Consequently, contestation can be better safeguarded in an era when judicial profiling systems are used.

[148] Hildebrandt (n 16); Yeung and Harkens (n 40).

[149] Dor and Coglianese (n 88).

[150] Dor and Coglianese (n 88).

[151] Jeroen Naves, 'Proposal for Standard Contractual Clauses for the Procurement of Artificial Intelligence (AI) by Public Organisations' (European Commission, 2023) <https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/proposal-standard-contractual-clauses-procurement-artificial> accessed 8 September 2023.