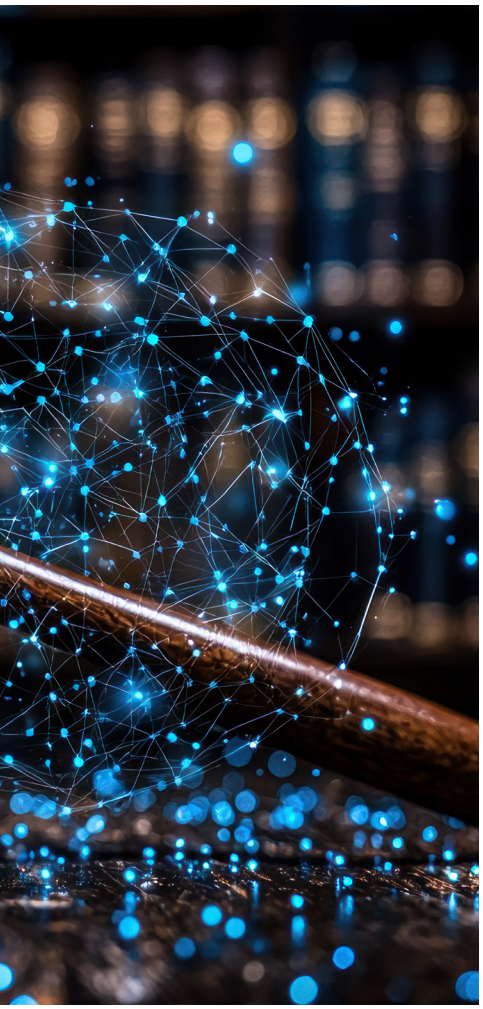
A Procedural Sedative: The GDPR's Right to an Explanation

COLUMN

AUTHOR: **LJUBIŠA METIKOŠ**

UNIVERSITY OF AMSTERDAM

What remedies do you have when AI errs, when it discriminates, or harms you in some other way? How can we hold organizations accountable when they cause people harm during the development, distribution, or use of AI? Arguably, the first step is understanding how the system in question works. To this end, the right to an explanation, provided in EU law under the GDPR and the AI Act, is one of the most important remedies individuals have to contest AI.



In recent years, the Court of Justice of the EU (CJEU) elaborated upon what the right to an explanation specifically entails. For example, in the landmark SCHUFA judgment, the CJEU expanded the scope of the right to an explanation to also include certain preparatory decisions.[1] This ensured that a glaring loophole in its scope would be sealed. However, in the recent Dun & Bradstreet judgment, the right to an explanation seems to shift away from being an effective accountability instrument to something akin to a procedural sedative. The danger here, I argue, is that explanations can (falsely) placate the concerns of decision-subjects with information that is not truly conducive to contestation.[2]

To understand these issues, it is helpful to first know how the right to an explanation has been legally constructed. Under the GDPR, this right has been cobbled together from a number of different provisions. Most important in this regard is art. 15(1)(h) GDPR, which grants data subjects a right to meaningful information about the logic involved in an automated decision process as defined by art. 22 GDPR.

The scope of this right can, subsequently, be found in art. 22(1) GDPR, which provides a ban on solely automated decision-making procedures, except in a few instances, such as when there is a contractual or legal ground for the use of the system in question, or if the data subject consents to its use. In these instances, certain safeguards must be implemented according to art. 22(3) GDPR, such as the right to contest the decision. This should also include a right to an explanation, according to Recital 71 GDPR.

Based on these provisions, the CJEU states that data subjects have a genuine right to an explanation about a specific automated decision, that should enable them to contest that decision.[3] However, there is a difference between nominally allowing a person to contest a decision, versus enabling them to actually hold organizations accountable and to remedy wrongdoing on their part. Arguably, Dun & Bradstreet actively hinders the latter in a number of ways.

First of all, the CJEU explicitly excludes information that would be deemed too technical for a lay person to understand. This would not only include expansive information such as the training dataset of the system, but also the *'mathematical formula'* underpinning the system.[4] The problem here, is that explanations will therefore be wholly generated and selected by data controllers. Of course, they have a vested incentive to not be held accountable for any wrongdoing on their part. Rather, they could wish to placate the concerns of data subjects, offering them a procedural sedative rather than any path toward true recourse.

This is especially problematic when we consider that certain types of explanations can be unreliable as a basis to contest an AI system. Depending on the AI model that a data controller uses, it can be difficult to reliably and faithfully produce an explanation for why a specific output has been generated. Rudin argues that we should not even use the term 'explanation' but rather 'approximations' when we try to explain certain black-box models.[5]

1 CJEU, 'SCHUFA', 7th of December 2023, ECLI:EU-:C:2023:957.

2 See also my blog on this matter: 'Dun & Bradstreet: A Pyrrhic Victory for the Contestation of AI under the GDPR' for 'The Law, Ethics & Policy of AI Blog' www.law.kuleuven.be/ai-summer-school/blogpost/Blogposts/dun-bradstreet-a-pyrrhic-victory-for-the-contestation-of-ai-under-the-gdpr

3 CJEU, 'Dun & Bradstreet', 27th of February, ECLI:EU-:C:2025:117, para 55.

4 CJEU, 'Dun & Bradstreet', 27th of February, ECLI:EU-:C:2025:117, para 59-61.

5 Cynthia Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead', Nature Machine Intelligence 1, nr. 5, May 2019, p. 206-15, doi.org/10.1038/s42256-019-0048-x

> THE GDPR GIVES US A RIGHT TO AN EXPLANATION — NOT A RIGHT TO AN ACCURATE EXPLANATION.

To illustrate, one of the issues plaguing explainable AI is that interpreting these models can sometimes result in multiple contradictory explanations of the same model, a phenomenon titled the 'Rashomon effect'. Babic and & Cohen argue, based on these and many other explainability issues, that opaque, black-box, systems are not reliably interpretable.[6] Rather, they, and scholars like Rudin, and myself[7], argue in favor of the use of simpler, 'intrinsically' interpretable models in high-stakes situations which can be reliably and faithfully explained. Unfortunately, neither the GDPR nor the CJEU truly address these issues. They provide a right to an explanation, not a right to an accurate explanation. With the exclusion of 'technical' information, moreover, independent assessment of these systems and the explanations that data controller provide about them, could become near impossible.

Another issue that arises in Dun & Bradstreet, is how the CJEU tries to accommodate the concerns of data controllers about the potential of explanations to harm their trade secrets and other rights and interests. The CJEU states that the requested explanation can first be sent to a supervisory authority or judge, who will assess if any rights will be disproportionately infringed upon.[8] However, these institutions are already overwhelmed with all the other responsibilities they have under the GDPR. Consequently, this balancing exercise can be an easy route for data controllers to delay, delay, delay.

In short, while we have a right to understand how important automated decisions are being made, we have no assurance that these explanations are reliable, nor can we independently assess these systems ourselves.

Still, we could argue that explanations should serve to help instill trust in AI, and that this is a valuable pursuit in its own right regardless of the issue of contestability. Indeed, empirical research on procedural justice shows that people can value procedures, transparency, and participation, regardless of whether they themselves have any impact on the final outcome. However, this view can be a 'double-edged sword'; leading to decision-making that leaves people subject to exploitation and manipulation.[9] Caring, sensible, and agreeable explanations may do more harm than good, if true resistance against automated decision-making remains impossible.

These issues are not a reason to despair, however, and wholly throw away the right to an explanation. Rather, it forces us to reevaluate explainability requirements. Potentially, this could still happen with the right to an explanation under art. 86 of the AI Act, which has not yet been interpreted by the CJEU. That said, this would require a radical different legal interpretation of explainability requirements in the years to come. In this regard, the AI Act's right to an explanation could do much more than the GDPR. It could be a requirement to rely on intrinsically interpretable models[10], to provide as much open access as possible to critical outsiders, and to ensure that true contestation remains possible.

6 Boris Babic and I. Glenn Cohen, 'The Algorithmic Explainability "Bait and Switch"', Minnesota Law Review, nr. 108, 15th of August 2023, p. 857-909.

7 Ljubisa Metikos, 'Explaining and Contesting Judicial Profiling Systems', Technology and Regulation 2024, 13th of September 2024, p. 188-208, https://doi.org/10.26116/techreg.2024.017.

8 CJEU, 'Dun & Bradstreet', 27th of February 2025, ECLI:EU:C:2025:117, para 76.

9 Robert J. MacCoun, 'Voice, Control, and Belonging: The Double-Edged Sword of Procedural Fairness', Annual Review of Law and Social Science, Volume 1, 5th of May 2005, escholarship.org/uc/item/011185w5.

10 Arno Cuypers, 'The right to explanation in the AI Act: a right to interpretable models?', CiTiP blog, KU Leuven, 12th of March 2024, www.law.kuleuven.be/citip/blog/the-right-to-explanation-in-the-ai-act-a-right-to-interpretable-models/

About the Author:

Ljubiša Metikoš is a PhD Researcher at the University of Amsterdam. He is affiliated with the Institute for Information Law, the Paul Scholten Centre for Jurisprudence, and the RPA Human(e)AI. His research focuses on the use of AI by judges and litigants' right to an explanation about these systems.

To this end, he is interested in legal-philosophical theories on procedural justice, computer science discussions on explainable AI, and legal debates on the GDPR and the AI Act.