The Right to Explanation:

Legal Challenges and Regulatory Solutions in the EU Digital Framework
The Future of Fairness in Automated Decision-Making

*Jacopo Dirutigliano and Davide Baldini**

I. Introduction

In today's data-driven society, Artificial Intelligence[1] (AI) has become pivotal in reshaping the technological and societal landscape. AI, including Machine Learning[2] (ML) leverage vast datasets and computational power to perform tasks previously requiring human intelligence, such as pattern recognition, decision-making, and predictive analytics. From optimising healthcare outcomes to enhancing public governance, the potential of AI to revolutionise processes across sectors is immense.

One of the most significant applications of AI is in Automated Decision-Making (ADM) systems, which enable decisions to be made without human intervention, also in critical areas such as credit scoring, recruitment, and criminal justice.

While these tools promise efficiency and consistency, they also introduce challenges related to transparency, accountability, and fairness. Concerns about algorithmic opacity, biases, and the potential for unforeseen consequences have led to widespread calls for regulatory oversight and ethical safeguards.

Since ADM systems often operate in ways that are not easily interpretable, individuals subjected to their decisions face significant cognitive obstacles that can hinder their ability to understand, challenge, or even anticipate outcomes. To enable decision-subjects to exercise their rights, protect their interests, and engage meaningfully with these systems, access to a clear and comprehensible explanation of how and why an automated decision was reached is crucial.

In this respect, the 'Right to Explanation' (RTE) has emerged as a legal and ethical response to these concerns, seeking to provide individuals affected by automated decisions with insights into how and why such decisions were made. The RTE has been the subject of extensive academic debate and legal development, particularly within the European Union's horizontally applicable frameworks which aim at regulating ADM, including the General Data Protection Regulation[3] (GDPR) and the AI Act[4] (AIA). While the GDPR provides individuals with the right to obtain 'meaningful information about the logic involved' in ADM, the AIA has introduced an explicit right to obtain explanations of high-risk AI decisions[5].

DOI: tbd

* Jacopo Dirutigliano, PhD in Law, Science and Technology. Davide Baldini, h.D. candidate at the Florence University (Florence, Italy) and Maastricht University (Maastricht, the Netherlands). For correspondence: please provide email.
Disclosure of Interests: The authors have no competing interests to declare that are relevant to the content of this article.
The overall work has been discussed and agreed by both the authors. Jacopo Dirutigliano has primarily drafted Sections II, IV, V, VI, and VII of the paper, while Davide Baldini has primarily drafted Section III of the paper. Both authors wrote Sections I, VI.2, and VIII and have read and agreed to the submitted version of the manuscript.

1 For this work, 'AI' will be intended as 'artificial intelligence' according to the definition of the AIA.

2 ML is intended as a subset of AI that involves training algorithms to identify patterns in data and make predictions or decisions without being explicitly programmed for specific tasks. ML is an umbrella term that involves Deep Learning, which is a subset of ML that uses neural networks with multiple layers to process complex datasets and identify patterns.

3 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*) [2016] OJ L 119.

4 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (*Artificial Intelligence Act*) [2024] OJ L 2024/1689.

5 The issue of the RTE has been widely debated in literature, with extensive analyses delving into its nuances under the GDPR. However, for the purposes of this work, we deal with the GDPR but refrain from addressing the contested existence of a true RTE within the GDPR. Instead, we note that arts 13, 14, and 15 GDPR primarily grant individuals rights to information about automated decision-making processes without establishing a substantive right to a detailed explanatio(n Due to limitation constraints, our analysis will centre on the practical implications and evolutions of RTE within the context of the AI Act, addressing the GDPR for specific considerations.

However, the effectiveness of the RTE remains contested. Scholars have pointed out its limitations in practical enforcement, particularly given the complexity of AI models and the challenges of translating algorithmic processes into human-understandable explanations[6].

This article explores the conceptual foundations, legal developments, and practical implications of the RTE in the context of ADM. It examines how the RTE interacts with broader principles of fairness, fundamental rights, and with EU regulatory frameworks governing AI. By doing so, it aims to assess whether the RTE sufficiently addresses the challenges posed by ADM, or whether further legal and technical innovations are required.

II. The Cognitive Barriers of AI: Opacity, Correlation-Based Logic, and Bias

As ADM systems become increasingly embedded in society, their ability to shape individual rights and opportunities has drawn intense scrutiny. While these systems offer efficiency and scalability, their inherent characteristics also create significant challenges for transparency and accountability. Three fundamental traits of AI – opacity, correlation-based logic, and bias – are particularly crucial in understanding these challenges. Those traits do not function in isolation; rather, they collectively create a deeper issue, that has been described by Bayamlioglu[7] as an informational asymmetry between decision-makers and decision-subjects.

Informational asymmetry arises when those affected by automated decisions lack the necessary knowledge or access to understand how a decision was made, why it was made, and whether it was fair. This imbalance is exacerbated by opacity, which may render AI systems inscrutable to both decision-subjects and even developers. The use of correlation-based logic means that AI does not establish causality but merely identifies patterns, often leading to decisions that seem arbitrary or inexplicable. Finally, bias, whether originating from flawed datasets or systemic prejudices embedded in the model, can distort decision-making, reinforcing inequalities and further complicating efforts to contest or understand an outcome.

Together, these three characteristics of AI create significant cognitive barriers, limiting individuals' ability to exercise autonomy, challenge unfavourable decisions, and protect their rights. Without a bridge for this asymmetry, subjects remain disadvantaged, unable to meaningfully engage with AI-driven decisions. This underscores the necessity of *explainability*, to provide individuals with useful insights into the reasoning behind automated decisions. In the following sections, we explore each of these characteristics in depth, analysing their implications for decision-subjects and for the RTE.

1. Opacity

Opacity is a defining feature of many AI systems. As Burrell describes, there are three kinds of opacity: 'intentional opacity,' where the workings of an AI system are withheld to protect intellectual property; 'illiterate opacity,' where the system is understandable only to those with specialised expertise in coding and computing; and 'intrinsic opacity,' where the system's complexity makes it inherently unintelligible to humans[8]. These forms of opacity represent a significant cognitive barrier both for decision-subjects and decision-makers, leaving them unable to understand the reasons that informed a decision shaped by a 'black-box'[9].

The opacity problem is particularly acute when decisions impact rights and freedoms, such as credit, criminal justice, or employment. For example, ADM

6 In this article, 'explanation' refers to the conclusive 'explanation' provided to the decision-subjects through RTE, which initially takes the form of an 'algorithmic explanation' extracted by the algorithm, and which is later translated into human terms.

7 Emre Bayamlioglu, 'Contesting Automated Decisions: A View of Transparency Implications' (2018) 4 European Data Protection Law Review 4. The concept of 'informational asymmetry' can be also traced back to: Bruno Lepri and others, 'The tyranny of data? The bright and dark sides of data-driven decision-making for social good' (2016) arXiv:1612.00323, and to George A Akerlof, 'The Market for 'Lemons': Quality Uncertainty and the Market Mechanism' (1970) 84(3) Quarterly Journal of Economics 488.

8 Jenna Burrell, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data and Society 1.

9 A term used to describe the lack of interpretability and transparency in complex AI models, where the process by which an input is transformed into an output is not easily understandable by humans. See: Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015).

systems may deny loans or flag individuals as high-risk without explanation, making challenges difficult. This lack of transparency erodes trust and undermines fairness, as affected individuals are left without the tools to hold decision-makers accountable.

The opacity of ADM has been recognised as a critical challenge, which has led to the establishment of subjective rights of access to the decision. The GDPR, for instance, requires the data controller to provide 'meaningful information about the logic involved' in fully automated decisions (Articles 13-15), while the AIA and the Platform Workers Directive[10] (PWD) are more explicit: the former establishing at Article 86 a full-fledged 'right to explanation' for high-risk AI systems (HRAIS) and the latter providing at Article 11 a detailed right to 'human review' which includes the provision of an explanation.

However, these measures face practical difficulties in operationalization, especially for highly complex or proprietary systems. Opacity creates a cognitive deficit, thus worsening the 'informational asymmetry' between decision-makers and decision-subjects.

2. Correlation-Based Logic

AI relies on correlation-based reasoning, as opposed to causal logic[11]. In particular, AI models establish statistical correlations that may provide only a first intuition of potential causal relationships within data, but do not confirm them[12]. Indeed, from several studies on the inference of causal relationships from data[13], it has emerged that correlations do not always correspond to causal relationships[14] and causality requires a wide frame of knowledge to assess that observed effects are indeed causal[15]. At its core, correlation-based decision logic involves identifying statistical relationships between variables to predict outcomes, rather than establishing causal connections. Put simply, a causal relationship between input and output 'may simply not exist, no matter how intuitive such relationships might look on the surface'[16]. Hence AI models may have difficulty establishing the causes of certain phenomena or distinguishing causal from non-causal relationships. For example, an algorithm might deny credit without considering the applicant's actual creditworthiness. Some of these correlations may make sense to a human, such as the applicant's neighbourhood or – more intuitively – purchasing habits, but most of them may appear random, such as the applicant's clickstream on a website, or even mundane and seemingly irrelevant characteristics such as 'dog ownership' (which has been found to correlate negatively with creditworthiness)[17].

These inherent features of algorithmic reasoning prioritise predictive performance over interpretative insight[18] by relying on quantitative reasoning rather than qualitative, causal analysis[19]. This focus results in decisions whose underlying rationale remains obscure, hindering affected individuals from assessing its legality or fairness. Due to the lack of causal insights, algorithms offer statistical likelihoods without explaining the 'why' behind out-

10 Directive (EU) 2024/2831 of the European Parliament and of the Council of 23 October 2024 on improving working conditions in platform work [2024] OJ L 2024/2831.

11 Bernhard Schölkopf, 'Causality for Machine Learning' in H. Geffner, R. Dechter and J.Y. Halpern (eds), *Probabilistic* and Causal Inference (ACM,2022).

12 Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (Basic Books, Inc. 2018), ch 10.

13 Judea Pearl, *Causality: Models, Reasoning and Inference* (vol 64, Cambridge University Press 2000).

14 Correlation indicates a relationship between two variables, but it does not imply causatio(n A strong correlation may suggest causality, but other factors could explain the connection: (a) it may be due to chance, with no real underlying relationship; (b) a third hidden variable could influence both variables, creating a misleading association.
Correlation reveals patterns in data that move together, but it does not confirm one variable causes the other. A statistically significant correlation can exist without a causal link, often due to a shared external factor. Empirical research helps identify causal relationships, making it essential to distinguish correlation from causatio(n Techniques such as randomization, controlled experiments, and predictive models aid in establishing causality.

15 Pearl and Mackenzie (n 12).

16 Cary Coglianese and David Lehr, 'Regulating by Robot: Administrative Decision Making in the Machine-Learning Era' (2017) 105 Georgetown LJ 1157.

17 Sandra Wachter, 'The Theory of Artificial Immutability: Protecting Algorithmic Groups under Anti-Discrimination Law' (2022) 97 Tulane Law Review 149.

18 Galit Shmueli, 'To Explain or to Predict?' (2010) 25(3) Statistical Science 289–310; Ronan Hamon et al., 'Bridging the Gap Between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making' (2021) 1 17IEEE Computational Intelligence Magazine 80–81.

19 Daniel J. Solove and Hideyuki Matsumi, 'AI, Algorithms, and Awful Humans' (2024) 92 Fordham Law Review 1923.

comes[20]. Therefore, it is up to individuals to fill the gap between the information they receive and the inference they make[21], creating an epistemic obstacle.

The limitations of correlation-based reasoning also create challenges in explanation. Correlations often fail to provide explanatory power because they do not reveal the causes of specific events. Effective explanations typically require causal understanding, as people find explanations rooted in causality easier to grasp and more satisfactory[22] (see §6.1.2). The reliance on correlation-based reasoning further complicates the provision of detailed reasoning in contexts where explanations are required to justify decisions (eg, in public decision-making).

3. Bias

Algorithms, though seemingly objective, are influenced by subjective design choices and potentially flawed training data, making their decisions neither inherently accurate nor impartial. These so-called biases emerge at every stage of the algorithm's lifecycle – from the selection and weighting of features during design, to the quality and representativeness of the training data, and the fine-tuning of the algorithms[23]. For instance, a programmer's decision to omit certain criteria can inadvertently result in decisions that are incorrect or unfair. Similarly, datasets riddled with inaccuracies or incomplete information perpetuate biases, further compromising the integrity of algorithmic outcomes[24]. In summary, biases exist precisely because ADM processes are the result of a choice to consider or not a given criterion[25].

Biases, intended as the inclination of a system to favour certain criteria, perspectives, or outcomes over others, generate informational asymmetry by introducing partiality into ADM systems, thereby obscuring the logic and fairness of their outcomes. This partiality prevents an objective evaluation of how decisions are made, as algorithms prioritise certain data points or rules while neglecting others. As a result, individuals subjected to these systems often lack the knowledge of which factors influenced a decision, why those factors were deemed relevant, and how alternative outcomes were ruled out. This gap in understanding fundamentally hinders transparency, as users are effectively prevented from tracing or contesting the decision-making process.

4. Conclusions: Cognitive Obstacles in ADM and the Request for Explainability and Explanations

Bayamlioglu's analysis highlights that the three critical characteristics of AI can contribute to significant cognitive obstacles and create 'informational asymmetries' between decision-makers and subjects[26]. As a result, individuals face increased vulnerability to digital powers[27], encountering challenges in comprehending or contesting decisions that affect their rights, freedoms, or legitimate interests.

20 Having said that, it must be pointed out that there are a great many ML systems that have very different characteristics. ML comprises a variety of techniques, which range from traditional linear regression models over support vector machines and decision tree algorithms to different types of neural networks. As indicated by Wischmeyer '[t]he difficulty to establish a causal nexus ex post between a specific input and a specific output differs considerably between these techniques. While decision tree algorithms, which […] are employed, eg, in financial forecasting or in loan application programs, allow causal explanations once a tree has been built, this is not necessarily true for artificial neural networks […]. Here, the re-construction problem is serious, because even with complete information about the operations of a system, an ex-post analysis of a specific decision may not be able to establish a linear causal connection which is easily comprehensible for human minds. […] The difficulty to identify causal relations in neural networks […] it does not affect the possibility to collect information about the system and its operations. Incrementally seeking explanations for opaque phenomena by gathering data and by testing a series of hypotheses is the standard way to produce knowledge and to comprehend the data in science and in society'. See Thomas Wischmeyer, 'Artificial Intelligence and Transparency: Opening the Black Box' in *Regulating Artificial Intelligence* (Springer International Publishing 2020), 75–101.

21 Peter Lipton, *Inference to the Best Explanation* (Routledge 1991), 7; Daniel Kahnemann, *Thinking, Fast and Slow* (Farrar, Straus and Giroux 2011).

22 Tim Miller, ''But Why?' Understanding Explainable Artificial Intelligence' (2019) 25(3) XRDS: Crossroads, The ACM Magazine for Students 20–25.

23 Tea Mustać and Peter Hense, *AI Act Compact: Compliance, Management & Use Cases in Corporate Practice* (Fachmedien Recht und Wirtschaft 2024).

24 As shown by De Mulder and Valcke: 'the system is unable to produce predictions, or any other type of outcome, that rely on other variables than the given input variables. Consequently, if outcomes are produced that are influenced by values of input variables which are discriminatory, such as race, it is only because the system developer poorly performed the feature selection step'. See: Wim De Mulder and Peggy Valcke 'The Need for a Numeric Measure of Explainability' (2021) IEEE International Conference on Big Data, 15 December 2021, 2712–20.

25 Bayamlioglu (n 7), 433–46.

26 Ibid.

27 Stefano Rodotà, *Il Diritto Di Avere Diritti* (Editori Laterza 2012), 335.

The lack of helpful information deprives decision-subjects of the necessary tools to self-determine and contest potentially unlawful automated decisions, leaving them disadvantaged in critical areas of life. While transparency is a necessary step to redress this asymmetry, it is insufficient by itself, as merely providing access to data does not empower individuals who lack the means to interpret and act upon it. Structured and functional explanations, designed to be comprehensible and actionable, are essential to redress this imbalance[28]. Such measures could restore agency to individuals and rebalance power dynamics – eg, power to control information, to control one's own identity, to understand a decision –, particularly as society increasingly delegates decision-making to machines. Crafting a regulatory framework that ensures algorithmic explainability while fostering innovation is undoubtedly challenging, but it is imperative to protect fundamental rights and democratic values in the digital age. In this scenario, we contend that a RTE should form a pillar of such a framework.

In light of the foregoing analysis, it should be now apparent that ADMs produce multiple negative impacts on the various dimensions of fairness that underpin EU law in general, and EU digital regulation in particular, and which provide the theoretical foundations thereof. Before delving into the value of the RTE as a means to redress the black-box problem, it is necessary to examine the impact of ADMs vis-à-vis said dimensions of fairness. This analysis enables us to appreciate how the RTE should be interpreted as a corrective mechanism to mitigate these adverse effects. In other words, by unpacking the ways in which ADMs challenge fairness – whether through opacity, bias, or the reinforcement of existing inequalities – we can better understand the precise role that the RTE plays in restoring it.

III. The Right to Explanation and the Rationale of Fairness: Procedural and Substantive Dimensions

When viewed through the lenses of the *polysemic* fairness rationale which underpins the ever-increasing EU's digital regulatory framework[29], the RTE lays down a pivotal bridge between the ethical dimensions of fairness and their legal operationalisation in EU digital law. As will be discussed below, this right is deeply rooted in different fairness dimensions, aligning with broader justice traditions embedded in EU Law which frame fairness as both an aspirational value and a functional normative standard.

In the field of AI regulation, achieving fairness often becomes synonymous with preventing unfair automated biases and the occurrence of other substantive algorithmic harms [30]. Two of the most prominent EU legal instruments which aim at regulating AI – namely, the GDPR and the AI Act,[31] are no exception, as they clearly embed this understanding. Both instruments adopt a (albeit different[32]) risk-based and procedural approach, with the ambition to prevent, or mitigate, violations of fundamental rights and freedoms occasioned by AI systems, be it in the context of automated processing and profiling (in the case of the GDPR[33]), or the deployment of so-called high risk

28 Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (arXiv.org, 1 November 2017) https://arxiv.org/abs/1711.00399.

29 For an overview of the different dimensions of fairness in the EU corpus on data regulation and digital law, see *inter alia*: Giovanni De Gregorio and Pietro Dunn, 'The European risk-based approaches: Connecting constitutional dots in the digital age' (2022) 59 Common Market Law Review 2; Philipp Hacker, Johann Cordes and Janina Rochon, 'Regulating Gatekeeper AI and Data: Transparency, Access, and Fairness under the DMA, the GDPR, and beyond' (2022) Working Paper.

30 See, *inter alia*: Tobias Baer, *Understand, Manage, and Prevent Algorithmic Bias. A Guide for Business Users and Data Scientists* (Apress 2019); Danielle Keats Citron and Frank Pasquale, 'The scored society: due process for automated predictions' (2014) 89 Washington Law Review Association 1; Melissa Hamilton, 'The sexist algorithm' (2019) 37 Behavioral Science & the Law, 145; Paul Hayes et al, 'Algorithms and values in justice and security' (2020) 35 AI & Society, 533; Anna Lauren Hoffmann, 'Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse' (2019) 22 Information, Communication & Society 7, 900; Joseph E. Stiglitz, 'Artificial Intelligence and Its Implications for Income Distribution and Unemployment' (*NBER Paper*, 2017) https://www.nber.org/papers/w24174 accessed 28 February 2025.

31 We contend that these two EU legislative instruments lay down the most generally applicable rules and principle that regulate AI, due to their broad scope (in the case of the GDPR, due to the broad notion of 'personal data'; in the case of the AIA, due to the broad notion of 'AI system'). Naturally, other EU legislative instruments, such as the DSA, DMA and PWD, also regulate AI; however, their scope is much narrower.

32 De Gregorio and Dunn (n 29).

33 This is expressly recognised by recital 71 GDPR, in the context of profiling and ADM: 'the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate […] that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect'.

AI systems[34]. Most GDPR and AI Act provisions ultimately seek to achieve the objective of mitigating risks of violating the fundamental rights protected under EU primary law - namely, the EU Charter of Fundamental Rights[35] (EUCFR) - with algorithmic discrimination being one of the most salient harms to be mitigated, as a topical example of harm commonly associated with the so-called 'high-risk' AI systems[36].

Nevertheless, at first sight the foundation of the RTE seemingly deviates from the underlying assumption of advancing the substantive fairness dimension that has taken roots in AI regulation[37]. In fact, providing to the decision-subject even the clearest, most comprehensive and truthful explanation of ADM does not, in itself, prevent unlawful discrimination, or other negative impacts on the decision-subjects' substantive rights.

However, the RTE can be said to meaningfully advance substantive fairness, albeit indirectly, in two significant ways. Firstly, the effectiveness of the RTE necessitates that AI systems be designed and developed with a certain degree of explainability. Without this requirement being incorporated within the very architecture of the ADM system, it would not be possible to achieve any explanation in the first place. Consequently, the requirement that ADM systems be explainable enhances the accountability of AI developers and deployers, reducing the possibility for inexplainable and unfair biases to arise. The RTE thus helps integrate normative 'fairness constraints' into ADM systems, ensuring that their design ultimately upholds individuals' rights.

Secondly, once explainability is achieved, the exercise of the RTE enables affected individuals to contest and challenge unfair ADM practices, thereby also advancing procedural fairness[38]. By equipping the decision-subject with meaningful information about how the decisions are taken, the RTE can be leveraged to uncover hidden impacts on fundamental rights and freedoms, empowering individuals to seek redress, hold decision-makers accountable, and demand corrective measures (when applicable, see §6.1 and 6.4). In this respect, the RTE functions as supporting the effective exercise of subjective rights, ensuring that substantive fairness is not only a regulatory aspiration.

Against this background, the requirements of AI explainability demanded by the RTE emphasise equitable outcomes by aiming to mitigate biases and ensure that algorithmic systems do not perpetuate discrimination or unduly impact other rights, drawing on fairness principles historically tied to proportionality and equality. Procedurally, it manifests as an entitlement to contestability, access to remedies, and accountability mechanisms, consistent with the EUCFR and which is finding expression in secondary law instruments like the GDPR, Digital Services Act[39] (DSA), PWD and AIA, albeit with different levels of intensity and scope. By embedding both dimensions of fairness into the operational structures of EU's AI regulation, the RTE becomes a normative tool for harmonising technological advancement with ethical and legal obligations, contributing to the scholarly understanding of fairness as a unifying, albeit contextually nuanced, principle.

The foundations of the RTE in the context of digital fairness must thus be understood as integral to the broader fairness rationale that underpins EU digital regulation. As an expression of fairness, the RTE anchors itself in cardinal EU law principles embedded such as equity, due process, proportionality, and non-discrimination.

This procedural, but ultimately substantive, entitlement also challenges the traditional fairness conceptions and encourages a pluralistic understanding that accommodates the complexities of modern digital ecosystems. While grounded in the EU's regulatory discourse, the RTE exemplifies how fairness can serve as both a guiding principle and an actionable frame-

34 While art 1 AIA generically refers to the objective of, *inter alia*, ensuring a high level of protection of health, safety, fundamental rights enshrined in the Charter', recital 27 expressly states that: 'Diversity, non-discrimination and fairness means that AI systems are developed and used in a way that includes diverse actors and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory impacts and unfair biases that are prohibited by Union or national law'.

35 Charter of Fundamental Rights of the European Union [2009] OJ C 202.

36 See, *inter alia*: Stefano De Luca and Marina Federico, 'Algorithmic discrimination under the AI Act and the GDPR' (European Parliamentary Research Service, 2025) https://www.europarl.europa.eu/thinktank/en/document/EPRS_ATA(2025)769509 accessed 28 February 2025.

37 Namely, fairness as the prevention of automated biases and other substantive algorithmic harms. See also (n 30).

38 In particular, the RTE clearly aligns with the conceptualisation of procedural fairness adopted by the HLEG's Guidelines for Trustworthy AI, which defines it at p 13 as 'the ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them. In order to do so (…) the decision-making processes should be explicable'.

39 Regulation 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (*Digital Services Act*) [2022] OJ L 277.

work, bringing clarity on how 'justice-as-fairness' should be operationalised in the age of AI and digital markets. Therefore, the RTE offers a critical lens for examining how the EU's digital regulation, and especially ADM regulation, seeks to balance innovation, accountability, and fundamental rights protection.

Having thus reconstrued the RTE's fairness rationale within the EU digital regulatory framework, the following sections explore how this right finds expression in the context of two key instruments in regulating ADMs, that is, the GDPR and AI Act.

IV. The Background of the Right to Explanation: The GDPR and the Right to Contest

The RTE finds its roots in constitutional values and principles of public law, such as due process and the duty to give reasons. These principles, which impose procedural obligations, are traditionally foreign to private law where the RTE also applies[40]. Yet, the transposition of public law principles into private law in the context of ADM can be understood through the lens of the analysis proposed by Celeste and Di Gregorio. They highlight that ADM systems have facilitated a shift where private entities increasingly assume roles traditionally held by state actors, exercising quasi-regulatory power through their control over vast datasets and sophisticated algorithms: this shift represents a structural transformation in governance, in which private actors dominate decision-making processes with profound implications for individual autonomy and collective interests. This shift, characterised by the increasing reliance on private actors for decision-making processes that significantly affect individuals[41], calls for a reconfiguration of legal safeguards. In this context, Celeste and Di Gregorio recognised that the GDPR aims to serve as a key instrument translating constitutional guarantees into private law[42], extending due process protections to interactions between private actors. To rebalance the power asymmetry between individuals and data controllers, the GDPR establishes key safeguards, including the right to contest automated decisions under Article 22(3), also out-of-court and directly before the data controller[43]. This provision arguably implies an implicit right to review, given that the right to contest would be rendered ineffective without an obligation for review (the delegation of such quasi-regulatory functions to private actors is not unprecedented, as exemplified by the *Google Spain* case, wherein private entities were entrusted with the responsibility of balancing fundamental rights). Accordingly, this provision seems to uphold the constitutional value of due process in algorithmic decision-making[44], ensuring that constitutional protections extend beyond state actions to dominant private entities, reinforcing the idea that constitutional constraints should apply, not only to States, but also to dominant private actors[45].

Against this background, and taking stock from the right to contest in the GDPR, many scholars[46] ar-

40 Notably, while public institutions are bound by constitutional requirements such as providing reasons for their decisions and ensuring effective remedies, private actors lack equivalent obligations, and this discrepancy raises concerns regarding procedural justice and the legitimacy of decisions.

41 Herwig Hofmann, 'Automated Decision-Making (ADM) in EU Public Law' (2023) SSRN Electronic Journal https://doi.org/10.2139/ssrn.4561116.

42 Edoardo Celeste and Giovanni De Gregorio, 'Digital Humanism: The Constitutional Message of the GDPR' (2022) 3(1) Global Privacy Law Review 4, 15.

43 While existing legal frameworks, such as art 47 EUCFR, already provide the right to challenge unlawful decisions, the GDPR expands these protections by explicitly establishing an out-of-court avenue for contestation. It is not uncommon for legislative acts to reiterate or restate these established rights existing under art 47 EUCFR or other treaty. Such reaffirmations may simply reference existing protections, but they may also introduce limitations or balancing measures in the sense of art 52(1) EUCFR. In this context, the GDPR's 'right to contest' should be understood both as a novel out-of-court remedy against data controllers and as a reaffirmation of the EU's broader commitment to protecting fundamental rights and freedoms in the digital sphere.

44 Celeste and De Gregorio (n 42), 19-20.

45 Ibid, 16.

46 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(2) International Data Privacy Law 91 https://doi.org/10.1093/idpl/ipx005. Also the Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' has clarified the provisions of the GDPR and specified that the data subject 'will only be able to challenge a decision […] if they fully understand how it has been made and on what basis'. Amongst many, Brkan argues that: 'the absence of concrete safeguards, in particular the right of data subject to contest the decision and the related right to be informed about the reasons for decision, is problematic and might even violate the fundamental rights standard from Article 47 of the EU Charter of Fundamental Rights which provides for a fundamental right to an effective remedy. If the data subject does not understand the reasons behind the decision (for example a decision to arrest her), she is also not in a position to bring an effective remedy against such a decision. Even if the decision is taken with a human intervention, the human would still need to provide the data subject with reasons, giving her an opportunity to effectively challenge the decision'. See Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond' (2019) 27(2) International Journal of Law and Information Technology 119.

gue that the existence of an RTE is necessary to ensure meaningful contestation. Aware that the right to contest the outcome of an ADM is arguably different when it comes to private law and public law[47], it can be argued that a RTE takes inspiration primarily from the right to a reasoned decision in administrative law, effectively constituting the private-law declination of such right. Similarly to the right to a reasoned decision in administrative law, the RTE can be used to address power asymmetries between the State and citizens.

In the next section we argue that, in addition to being a means to contest automated decisions, the RTE is also a mechanism to address the informational asymmetries generated by AI.

V. The Foundations of a General Right to an Explanation

As previously observed, before the AIA coming into force, scholarly debate on the RTE largely centred on whether it exists within the GDPR[48], with less focus on whether a general 'right to explanation' is necessary[49], and why automated decisions should be explained. As noted in literature, starting from the assumption that, except in cases provided for by law[50], a human decision does not entitle the decision-subject to an explanation, the question arises as to whether – if the decision is instead made by an algorithm – it is necessary to have a right to explanation[51]. In other words, what justifies the right to have an explanation when individuals are subject to an automated decision?[52]

With reference to the debate on the existence of the RTE in the GDPR, many have argued that 'to contest a decision, an explanation is necessary'. However, this approach risks implying that all decisions, even non-automated ones, require explanations. Additionally, it overlooks the unique risks associated with ADM. As previously argued, the reasoning behind ADM greatly differs from human reasoning: AI systems rely on complex statistical correlations and operate opaquely, often failing to account for real-world complexities. Unlike human decision-making, which is inherently qualitative and context-sensitive, ADM lacks the capacity to consider the nuanced, case-specific factors that inform human judgment, as algorithmic decisions may fail to capture the holistic reasoning and adaptability characteristic of human decision-making. These traits introduce risks of bias, opacity, and errors, necessitating safeguards like explanations to ensure contestability.

Explanations are thus a targeted mechanism to address the informational gaps and accountability challenges unique to ADM. In sum, explanations are

47 The forms of contestation and judicial review differ considerably. Details must be reviewed according to the applicable law of substance and procedure (national and EU). This is also relevant as to what standard and what kind of information is required (problematic are cases where the public delegates to private parties, eg, art 17(4) Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market, and the Case C-401/19 *Republic of Poland v European Parliament and Council of the European Union* [2022] ECLI:EU:C:2022:297. See: Hofmann (n 41). In this scenario, the approach of providing new rights is a legislative approach, where some considerations are transposed from administrative law into contractual relations. In public and administrative law, the right to a fair hearing prior to a decision which could affect an individual comes with the right to access documents concerning that case and the right to a subsequent reasoned decision in order to assess the possibilities of review. This does not generally exist in private law relations.

48 As Fink and Finck noted, 'it is paradoxical that discussions around the explainability of AI have focused almost exclusively on data protection law, neglecting not only obligations in administrative law, but also other areas of EU law where similar obligations exist, such as public procurement law, consumer protection law, and financial regulation. The acknowledgement that explanation obligations already exist in other areas of EU law is important more generally, especially in the context of claims that EU data protection law should 'introduce' explanation requirements. See: Melanie Fink and Michèle Finck, 'Reasoned A(I)dministration: Explanation Requirements in EU Law and the Automation of Public Administration' (2022) 47(3) European Law Review 376.

49 As Bobbio argued, rights are the outcome of social conflict. Rights do not arise all at once but gradually, they emerge when they must or can. Particularly, rights arise when the increase in man's power over man, which inevitably follows technological progress and creates new threats to the individual's freedom. The definition of individuals' rights requires an approach focused on their needs, which refers to the condition of those who, in different contexts, mostly suffer from inequalities, such as power asymmetries, subordination relationships, and discrimination. See Norberto Bobbio, *L'età Dei Diritti* (Einaudi Torino 1989).

50 For instance, concerning the duty to give reason for public decisions, whether or not a public authority relies on AI does not in principle affect their obligation towards those affected by the decision to provide reasons for such decisions. Also, there are other areas of EU law where similar obligations exist, such as public procurement law, consumer protection law, and financial regulation.

51 Jacopo Dirutigliano, 'Trasparenza e Spiegabilità degli algoritmi' in Ugo Pagallo and Massimo Durante (eds), *La politica dei dati: Il governo delle nuove tecnologie tra diritto, economia e società* (Mimesis 2022), 281.

52 In trying to answer this question about a general RTE, we will not take into account those scenarios where a right to a justified decision exists already (eg, the right to a reasoned decision).

not universally required for all decisions but are critical for decisions stemming from ADM due to the distinct cognitive challenges they generate. Extending this requirement to all decisions risks diluting its purpose and misapplying a solution designed for the specific shortcomings of ADM. Consequently, central to the inquiry is the recognition that the debate on the RTE focuses on 'how' and 'by what means' an ADM decision is made, regardless of its content. Specifically, it matters whether the decision is taken by a human or an ADM system, even if the conclusion reached is the same. The AI-generated informational asymmetry alone, however, cannot justify a general RTE for all automated decisions. In light of our analysis above, it is necessary to define the specific conditions in which said right should apply, considering the interests of all stakeholders involved. We contend that those conditions are as follows:

1. A decision significantly impacts the decision-subjects or produces legal effects on them (similarly to the ratio of Article 22(1) GDPR). While everyday decisions delegated to ADM systems, such as Google Maps suggesting routes or Spotify recommending songs, typically do not affect an individual's legal sphere, others – like credit scoring – can profoundly impact one's life. If the RTE is to safeguard individuals from risks of violation to their rights and interests, a 'significance'-based approach becomes essential. Generating explanations requires effort, and reducing unnecessary efforts by providers of ADMs may allow resources to be redirected to other priorities. In other words, the utility of explanations must be balanced against the cost of generating them[53]. A regulatory pressure to use interpretable models could inadvertently drive developers away from deploying state-of-the-art systems, potentially stifling innovation in AI applications. Thus, not all decisions require explanation – only those with significant consequences for the decision-subjects. Especially, what renders a decision problematic and thus worthy of an explanation is the impact such a decision may have on the decision-subject's life without knowing how and why it has been taken.
2. The second condition requires that the algorithm has the potential to cause an impact which is negative and may harm the decision-subject, thereby justifying the need for RTE. Algorithm's lack of explainability does not inherently mean they will produce unfair outcomes. In some cases, the algorithm is merely supportive to human decision-making, rendering the absence of or an incorrect output equivalent to not employing the AI tool at all. For instance, Corti, a black-box algorithm, identifies signs of cardiac arrest during phone calls[54]. While this involves high-risk scenarios affecting individuals' health, the AI only assists the operator. If the algorithm fails to detect cardiac arrest, the outcome is no different than if the AI were not present[55]. In such cases, prioritising accuracy over explainability may be advisable, as the system's ability to correctly identify emergencies takes precedence over its interpretability. In other words, a explanation is required only when the effect is adverse, whereas if the output is exclusively beneficial, it is not necessary.
3. The RTE should be limited in scenarios where explanations would lead to the disclosure of information likely to affect sensitive public interests. For instance, algorithms used in planning classified military operations, selecting taxpayers for audits, or conducting customs or counter-terrorism activities might justifiably remain undisclosed[56]. Such secrecy is justified when the interest in protecting higher-order objectives, such as national security or counter-terrorism efforts, outweighs the individual's right to explanation.
4. A RTE exists only if there is a clearly identifiable decision-maker who can be held accountable for providing rule-based normative and causal explanations. This accountability ensures that individuals can engage in informed self-advocacy by contesting unfair decisions, understanding the rules

53 Finale Doshi-Velez et al, 'Accountability of AI Under the Law: The Role of Explanation' (2017) arXiv:1711.01134 https://arxiv.org/abs/1711.01134.

54 James Vincent, 'AI That Detects Cardiac Arrests during Emergency Calls Will Be Tested across Europe This Summer' *The Verge* (25 April 2018) https://www.theverge.com/2018/4/25/17278994/ai-cardiac-arrest-corti-emergency-call-response accessed 16 June 2025.

55 Scott Robbins, 'A Misdirected Principle with a Catch: Explicability for AI' (2019) 29(4) Minds and Machines 509 https://doi.org/10.1007/s11023-019-09509-3 accessed 24 April 2025.

56 Maja Brkan, 'AI-Supported Decision-Making under the General Data Protection Regulation' in *Proceedings of the International Conference on Artificial Intelligence and Law* (2017) 7 https://doi.org/10.1145/3086512.3086513 accessed 24 April 2025.

governing their lives, and adjusting their behaviour accordingly. Without an accountable entity, the RTE would be ineffective since there would be no one to demand justification from[57].

5. Lastly and as previously noted, the presence of 'informational asymmetry'. In this respect, it should be emphasised that not all ADM involves opacity. For instance, simpler algorithms (which might not fall under the AI Act's definition of AI) execute processes fully designed by programmers, leaving no decisions to chance or to the machine itself. Consequently, these systems do not exhibit the same levels of opacity as those associated with more complex algorithms. However, it remains essential to ensure that the programming phase considers all relevant real-world factors, with human oversight ensuring the system adequately reflects human reasoning and decision-making.

As argued in the next section, Article 86 AIA seems to consider, to a certain degree, the aforementioned conditions. Hence, we analyse how and to what extent the RTE established by the AIA acts as a remedy against ADM opacity.

VI. The Right to an Explanation in the AIA

The AIA has introduced a novel and explicit RTE, setting an important precedent in AI governance and distinguishing itself from the GDPR, which does not provide for such a right, at least explicitly[58]. This provision underscores the growing importance of 'algorithmic transparency' in the governance of AI systems, addressing gaps in existing legal frameworks, including areas outside the GDPR's scope. Notably, the RTE was introduced during the trialogue[59], representing a significant shift in the Act's normative framework and diverging from the AIA's product-safety-oriented approach.

The AIA's RTE provides as follows:

1. Any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system listed in Annex III, with the exception of systems listed under point 2 thereof, and which produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken.

2. Paragraph 1 shall not apply to the use of AI systems for which exceptions from, or restrictions to, the obligation under that paragraph follow from Union or national law in compliance with Union law.

3. This Article shall apply only to the extent that the right referred to in paragraph 1 is not otherwise provided for under Union law.

Despite several shortcomings (analysed in the following paragraphs), Article 86 seems to include the conditions that justify the obligation to explain an automated decision based on the need to protect decision-subjects' self-determination and ability to exercise their rights. In particular, the AIA's RTE:

1. applies only in case of decisions that significantly impact the decision-subjects or produce legal effects (first condition);

57 Reference is made to the work of Vredenburgh, where it has been argued that 'Individuals also have an interest in being able to hold decision-makers to account for mistakes or unfairness. This interest is in an interest in living under systems of rules that are predictably and fairly applied. Being able to hold decision-makers TO account for mistakes is necessary to engage in robust forward-looking exercise of agency: it is rational for agents to engage in temporally extended planning only if they are reasonably confident that they can reliably correct mistakes and there is not systemic unfairness that would curtail their plans. […] the inability to engage in informed self-advocacy generates weighty complaints in hierarchical and non-voluntary institutions, as it undermines their fairness and legitimacy'. See Kate Vredenburgh, 'The Right to Explanation' (2022) 30(2) Journal of Political Philosophy 209.

58 See Bryce Goodman and Seth Flaxman, 'European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation'' (2017) 38(3) AI Magazine 50 https://doi.org/10.1609/aimag.v38i3.2741; Wachter, Mittelstadt and Floridi (n 46), 76; Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7(4) International Data Privacy Law 4 https://doi.org/10.1093/idpl/ipx019; Andrew D Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7(4) International Data Privacy Law 233 https://doi.org/10.1093/idpl/ipx022; Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a Right to Explanation Is Probably Not the Remedy You Are Looking For' (2017) SSRN Electronic Journal 1 https://doi.org/10.2139/ssrn.2972855; Brkan, 'Do Algorithms Rule the World?' (n 46)

59 Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on 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))(1), art 68(c).

2. applies when the AI affects a person in a way that they consider having an adverse impact (second condition);
3. shall not apply to the use of AI systems for which exceptions from, or restrictions to, the obligation under paragraph 1 follow from Union or national law in compliance with Union law (third condition);
4. requires deployers of AI systems to provide those affected with 'meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken' (fourth condition); and
5. applies, both to fully automated and semi-automated ADM, when using AI systems that, pursuant to Article 3, are 'designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment' and thus can create 'informational asymmetries' (fifth condition).

While Article 86 represents a crucial step toward algorithmic transparency and individual empowerment, its scope and formulation raise several concerns regarding its effectiveness and practical enforceability. We now turn to examine in detail these concerns.

1. Critiques of the Right to an Explanation in the AIA

Despite the commendable intent by the EU legislator in introducing an RTE in the AIA, we contend that that right falls short in providing an effective mechanism to address the risks deriving from ADM.

a. The Role of the Deployer

The AIA assigns the responsibility for ensuring the RTE to the deployer. However, as Demková correctly noted[60], this exclusive focus is problematic since it neglects the indispensable role of the AI providers in delivering the relevant information. In Recital 93 AIA, the EU Legislator justified this decision by alleging that deployers are 'best placed to understand how the high-risk AI system will be used concretely' and can thereby identify risks unforeseen during development. Yet, as observed, such an assumption overlooks scenarios in which deployers lack the technical insight into the system's inner workings that the provider possesses[61]. Even Article 13 AIA – which provides that high-risk AI systems 'shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output' – is arguably insufficient to address the issue: while technical standards are currently being developed by European standardisation organisations, led by CEN and CENELEC following the EU Commission's request[62], as of today there are no clear standards for what constitutes a meaningful explanation. Moreover, despite the standardisation process currently under development, many disagree on the promise that EU standards will solve the issue, given the inherent shortcomings of standardisation in addressing individual rights issues. Consequently, achieving a truly meaningful explanation requires that deployers secure, through contractual agreements, access to the provider's critical information which may be necessary to effectively provide a RTE. This requirement, however, is fraught with challenges, as deployers are often in a weaker bargaining position relative to providers, which undermine their ability to obtain the necessary contractual rights. This shortcoming calls for a reassessment of the lifecycle approach to AI governance, one that recognises the integrated and interdependent obligations of both deployers and providers in protecting fundamental rights. The absence of a clear obligation for the provider to collaborate with the deployer in ensuring this right ultimately undermines the effectiveness of the RTE, risking its reduction to a merely formal safeguard rather than an effective mechanism of accountability.

60 As noted by Demková, '[b]y omitting any mention of the AI provider in the right to explanation, the above formulation ignores the reality that AI deployers […] cannot provide a meaningful explanation of the AI's influence on their decisions. […] Expecting deployers to provide a meaningful explanation without involving the AI provider is both unrealistic and flawed. While deployers may understand the context of their decision-making better than the provider, the provider holds key information about the system's inner workings, which is crucial to understanding the final decision. […] Without recognising the integrated nature of obligations between the deployer and the provider, the right to explanation becomes the weakest link of the AI responsibility chain'. See Simona Demková, 'The AI Act's Right to Explanation: A Plea for an Integrated Remedy' *MediaLaws* (31 October 2024) https://www.medialaws.eu/the-ai-acts-right-to-explanation-a-plea-for-an-integrated-remedy/ accessed 24 April 2025.

61 Ibid.

62 Josep Soler Garrido et al, *Harmonised Standards for the European AI Act* (European Commission, 24 October 2024) https://publications.jrc.ec.europa.eu/repository/handle/JRC139430 accessed 24 April 2025.

b. Lack of (Sufficient) Guidance on the Explanation Content

The AIA's RTE fails in clarifying what should be included in an explanation, a significant shortcoming that limits its practical effectiveness. The provision only mentions '*the role of the AI system in the decision-making procedure and the main elements of the decision*' but fails to define the scope of the required explanation. This lack of guidance is bound to make compliance with the RTE inconsistent. This regulatory uncertainty is likely to persist for an extended period, as Article 96 AIA does not include Article 86 among the provisions requiring specific guidance from the Commission or clarification through technical standards, nor does Article 40 address this issue[63]. Therefore, we can reasonably expect a waiting period of several years before any relevant administrative and court practice provide effective guidance.

Without clear criteria on what makes an explanation meaningful, fundamental questions – such as what should be disclosed and how detailed an explanation must be – remain unresolved. Scientific theories of explanation—which argue that different contexts demand distinct types of information—alongside strands in philosophy of science, cognitive psychology, and legal theory, provide valuable insights into how explanations should be structured. For instance, causal models[64] emphasise the necessity of linking decisions to underlying causes rather than mere correlations. Their structural causal model posits that an explanation must consider the explainee's epistemic state[65] by eliminating possibilities and pinpointing a precise causal factor, thereby enabling a robust understanding of the decision-making process. This aligns with legal and ethical imperatives, as explanations that merely describe statistical associations without causal reasoning risk failing to provide the explainee with the ability to contest the decision. Furthermore, theories of explanation[66] highlight the contextual nature of what constitutes a 'good' explanation.

Different legal and philosophical perspectives demand different degrees of transparency, with explanations varying based on the intended function – whether ensuring legal compliance, fostering user trust, or supporting contestability.

Given these considerations, explanations under the RTE should be both pragmatically structured and causally grounded, ensuring that individuals can meaningfully understand and challenge decisions. For instance, an adequate explanation could include some crucial elements such as (i) the input data that significantly influenced the decision, (ii) the role of key data features in shaping the outcome, (iii) the normativity of the decisional process, and (iv) a causal narrative that contextualises the decision within a broader framework of reasoning. This last element is particularly critical, as explanations that lack causality may fail to provide the necessary justificatory force required for accountability. Additionally, a pragmatic approach to explanation should consider the explainee's epistemic state[67], background knowledge, and legal context to ensure that the information provided is both accessible and actionable.

By integrating these principles, the RTE can move beyond a mere procedural requirement and serve its intended purposes: empowering individuals to understand, assess, and contest AI-driven decisions in a meaningful way.

---

63 Still, indirect guidance on the explainability requirement may come from the technical standards applicable to the 'Human Oversight' obligation laid down by art 14 AIA, especially from the requirements that the HRAIS be developed in a way that it is possible 'to properly understand [its] relevant capacities and limitations' and to 'to correctly interpret the high-risk AI system's output'.

64 As elaborated by Hume, Lewis, Gärdenfors, Pearl and Halpern, and Hitchcock. See David Hume, *An Enquiry Concerning Human Understanding* (1748) s VII; David Lewis, 'Causation' (1973) 70(17) The Journal of Philosophy 556 https://doi.org/10.2307/2025310; David Lewis, 'Causal Explanation' in David Lewis (ed), *Philosophical Papers* (vol 2, OUP 1986); Peter Gärdenfors, *Knowledge in Flux: Modeling the Dynamics of Epistemic States* (MIT Press 1988); Joseph Y Halpern and Judea Pearl, 'Causes and Explanations: A Structural-Model Approach. Part I: Causes' (2005) 56(4) British Journal for the Philosophy of Science 843 https://doi.org/10.1093/bjps/axi147; Christopher Hitchcock, 'Causal Models' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Metaphysics Research Lab, Stanford University Spring 2022) https://plato.stanford.edu/archives/spr2022/entries/causal-models/ accessed 24 April 2025.

65 Pearl and Halpern borrow the following example from Gärdenfors: in explaining why Mr Johansson developed lung cancer, an individual who already knows that asbestos causes cancer would not need an explanation of the causal model but rather confirmation that Mr. Johansson worked in asbestos production. Conversely, someone unfamiliar with the causal model (asbestos cause cancer) would require that explanation instead. This distinction illustrates that explanations must be tailored to what information the explainee lacks to refine their epistemic state.

66 James Woodward and Lauren Ross, 'Scientific Explanation' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Summer 2021 edn, Metaphysics Research Lab, Stanford University) para 2.4 https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation accessed 24 April 2025..

67 Halpern and Pearl (n 64), 843; Gärdenfors (n 64).

c. The Lack of Effective Enforcement Mechanisms

The lack of effective enforcement mechanisms for the RTE significantly undermines its potential impact and usefulness for affected individuals[68]. While the inclusion of the RTE in the AIA represents an important step toward algorithmic accountability, this provision remains largely symbolic without meaningful enforcement mechanisms, rendering the right ineffective in practice.

First, the AIA introduces the RTE without providing adequate legal avenues for affected individuals to seek redress when they believe the right has been violated. The absence of a structured framework for judicial remedies within the AIA exacerbates this issue, leaving individuals without the means to enforce their rights effectively[69]. Without robust enforcement, affected individuals are left to seek remedies elsewhere, often relying on fragmented recourse under overlapping frameworks, such as sector-specific laws. This creates a burden for individuals, especially in navigating the jurisdictional complexity of digital regulations.

Second, the enforceability of the RTE is further compromised by the lack of clear and measurable standards to evaluate compliance. Regulatory bodies face significant challenges in assessing whether outputs provided by AI systems meet the criteria for transparency and comprehensibility. Without predefined metrics or interpretability benchmarks (called for by De Mulder and Valcke[70]), compliance becomes subjective, creating uncertainty for both regulators and organisations deploying high-risk AI systems. This ambiguity may result in a regulatory vacuum, where the enforcement of the RTE is inconsistent or ineffective.

Third, the lack of robust enforcement mechanisms disincentivises the development of inherently transparent AI systems. As AI systems grow more complex to maximise accuracy leading to a trade-off between transparency and efficacy, organisations may prioritise efficiency or performance over explainability if the penalties for non-compliance are negligible or inadequately enforced. This dynamic may perpetuate the deployment and reliance on opaque systems, undermining the goal of promoting trust and accountability in AI.

2. Article 86(3): The Relationship Between the AIA's Right to an Explanation and Other Similar Entitlements

In light of the 'residuality clause' under paragraph 3 of Article 86, the RTE applies only in cases where individuals are not already granted the same right under other Union law provisions, such as the PWD or the DSA. Unlike the GDPR, which focuses primarily on granting individuals the right to information about the logic involved by an ADM (as per Articles 13–15), the AIA goes beyond mere 'meaningful information rights' by establishing a more substantive requirement for explanations.

When the GDPR applies – which is bound to happen quite often, given that AI systems included in Annex III almost always entail the processing of per-

68 Demková (n 60).

69 This is also reflected in the fact that, unlike the GDPR, the AIA does not impose an obligation to inform decision subjects about their subjective rights, including the RTE, leaving them unaware of their entitlement to an explanation.

70 De Mulder and Valcke insist for the need for a numeric measure of interpretability to make progress in the debate between proponents and opponents of the use of ML for high-stakes decisions. As they say, 'requiring a ML model to be explainable, without having any standard that acts as a baseline for explainability, is very similar to requiring citizens to drive slowly in a certain street without imposing any specific speed limit. The lack of a numeric measure, combined with the many disparate meanings of explainability, jeopardizes an objective decision whether a given ML system is explainable'. See De Mulder and Valcke (n 24). In this context, it is worth noting that the scholarship on explainable AI has evolved from early enthusiasm to a more critical position. Initial post-hoc explanation methods, used to justify decisions made by complex models, are now recognized as potentially misleading and insufficient for true transparency. These surrogate explanations can fail to meet accountability standards. As a result, scholars emphasize that terms like 'explainability' and 'interpretability' are often conflated, though they signify different objectives: interpretability relates to understanding a model's internal logic, while explainability pertains to communicating model outputs. The AIA reflects this conceptual vagueness by opting for the term 'interpretability' (eg, in art 13 AIA) and leaving the more contentious 'explainability' largely undefined, treating it as a usability concern rather than a rigorous technical or legal requirement. This lack of definitional clarity poses challenges for implementing effective legal safeguards or standards. Scholars such as Simon, von Luxburg, and Spiecker argue that effective explainability must be tailored to the cognitive and informational needs of users. Art 86 AIA, which references 'meaningful explanation', requires careful legal interpretation to bridge the gap between technical feasibility and normative expectations. As a result, the focus is shifting toward context-sensitive standards to make explainability a reliable safeguard under the AIA. See Judith Simon, Indra Spiecker gen Döhmann and Ulrike von Luxburg, *Generative KI – jenseits von Euphorie und einfachen Lösungen* (2024) Diskussion Nr 34, Nationale Akademie der Wissenschaften Leopoldina https://doi.org/10.26164/leopoldina_03_01226.

sonal data – we contend that Article 86 AIA is designed to complement, rather than replace, the rights provided by the GDPR, namely the rights to obtain human intervention, to express the point of view and to contest the decision under Article 22(3), and the rights to meaningful information under Articles 13-15 (assuming, without conceding, that the GDPR even provides a right to explanation under Article 22(3), as its provisions are widely interpreted to only guarantee transparency rights rather than explanatory obligations). There may be instances where a data subject's transparency rights under the GDPR (eg, the right to access under Article 15(1)[71]) are then complemented by the AIA's RTE, offering a more comprehensive framework for understanding ADMs.

The rights under Article 22(3) GDPR and the AIA's RTE differ in several ways. Firstly, the scope of the AIA's RTE is broader for some instances, narrower for others. Article 86 AIA has a broader scope than Article 22 GDPR, because it applies to decisions made using AI systems even when these decisions involve non-personal data. Additionally, the Article 86 AIA encompasses not only decisions made 'solely' by automated means but also decisions which involve some degree of meaningful human oversight, and are thus only partially automated, thereby covering semi-automated processes that fall outside the exceptions provided by Article 22(2) GDPR.

Secondly, the AIA's RTE is narrower in scope when it comes to its targeted approach to the specific systems to which it applies. In particular, it applies only to fully and semi-automated systems that are classified as 'high-risk' under the AIA's risk-based framework; more specifically, it applies to only those high-risk systems listed within Annex III of the AIA.

Accordingly, given the differences between the cited provisions, we argue that there is no scenario where the RTE under the AIA has to be considered overlapping and, therefore, replaced or entirely subsumed by GDPR transparency rights due to the former's residual nature. The two frameworks seem, instead, to operate *in tandem*, with the AIA filling critical gaps that GDPR does not address, and vice-versa.

In this context, the recent ruling by the Court of Justice of the European Union (CJEU) *Dun & Bradstreet Austria*[72], which derives a 'right to explanation' from Article 15(1)(h), raises a number of concerns. The ruling seems to take into account that the GDPR's RTE is one of the 'suitable measures to safeguard the data subject's rights' set out in Article 22(3) and Recital 71, which is applicable only to solely automated decisions which are legitimate, that is, when the exceptions under lit. a) and c) of Article 22(2) applies. However, by recognising a RTE in Article 15 and extending it to situations where decisions are not 'solely automated' (ie, into paragraph 1), the Court risks almost entirely defusing the scope of application of the RTE expressly provided by Article 86 AIA. In principle, due to the residuality clause laid down in Article 86(3), the AIA's RTE is bound to disapply every time the decision-subject can invoke another RTE under EU or Member State Law. If Article 15(1)(h) GDPR indeed provides a full-fledged RTE, and taking into account that high-risk AI systems are almost always bound to involve the processing of personal data due to their nature (thus triggering GDPR applicability), it follows that the GDPR's RTE would prevail over the AIA's RTE in almost any instance of high-risk AI systems being used for ADM, with the only notable exception to this being the case where the ADM is not fully automated due to meaningful human oversight measures having been implemented by the data controller/deployer[73]. Although seeking to provide a systematic and expansive interpretation of the RTE is understandable, this reading risks straying from the literal text of the GDPR. This approach may conflate the established understanding of the 'right to explanation' and thus necessitates further clarification.

In summary, the AIA introduces a robust and targeted RTE, creating a necessary and forward-looking layer of accountability for high-risk AI systems, complementing and expanding existing GDPR provisions. This dual-layered approach reflects a recognition of the unique challenges posed by AI and the

71 For example, an individual – as data subject – may leverage art 15 GDPR to require the data controller (the deployer, in AIA's terms) to confirm whether an automated decision pursuant to art 22 GDPR has taken place. If this is the case, the same individual – as affected person – may then leverage the AIA's RTE to obtain an explanation of such decision, without prejudice to their right to obtain meaningful information about the logic involved, under the GDPR.

72 Case C-203/22 *CK v Dun & Bradstreet Austria GmbH* [2025] ECLI:EU:C:2025:117.

73 In compliance with the obligation on the deployer under art 26(2) AIA: 'Deployers shall assign human oversight to natural persons who have the necessary competence, training and authority, as well as the necessary support'.

need for nuanced governance frameworks to address them effectively.

3. (Follows) Considerations on the Relationship Between the Right to a Reasoned Decision and the AIA's Right to an Explanation

The introduction of the AIA's RTE will have a significant impact not only in the private law domain but also the public sector. Building on the earlier analysis of Article 86(3), this paragraph investigates how the right to a reasoned decision intersects with the RTE, particularly considering the residual nature of the RTE, which only applies where no equivalent right exists under Union law.

a. The Duty to Give Reasons in EU Administrative Law

The duty to give reasons, which entails disclosing the rationale of a public authority's decision, has the function of allowing courts to exercise their powers to review the legality of the decision so as to ensure judicial control over administrative power, as well as giving affected individuals the opportunity to challenge the decisions and protect their rights (receiving enough information to be able to determine whether the decision is well-founded). In this regard, the CJEU often invokes Article 47 EUCFR, the right to an effective remedy, to support the requirement of reason-giving[74]. As explained by Demková and Hofmann[75], the right to given reasons, read in conjunction with Article 47 EUCFR, is both a procedural and a substantive guarantee which allows the plaintiffs to decide on the potential success of their claim and thus to effectively prepare defence. Accordingly, the reasoning must enable a person 'to defend his or her rights in the best possible conditions and to decide, with full knowledge of the relevant facts, whether there is any point in applying to the court with jurisdiction'[76].

In providing the reasoning followed to reach a decision, a decision-making authority is required to explain the facts and legal considerations that had a decisive importance in the context of their decisions[77]. However, the reasoning outline does not need to be exhaustive, including all points of fact and law[78], since it has to be assessed considering its context and 'all the legal rules governing the matter in question'[79].

According to the CJEU, the duty to state reasons varies according to the type of act, the nature[80], the interests of the individuals affected, its specific context[81], and the legal rules[82] governing the matter in question. For instance, for acts of general application, statements of reasons must provide legal justification[83], explanations of the circumstances having led to the adoption of that act and the objectives the latter is intended to achieve[84].

A more stringent duty to state reasons is required for individual acts because the addressees of those must be able to assess the lawfulness of the ways in which such acts affect them.

Another element affecting the statement of reasons is the discretion of the body adopting the act[85]. In the case of unilateral administrative acts, the reasoning should be detailed enough to enable the understanding of the reasons and to enable addressees the exercise of their right to defence, as well as the exercise of judicial review, and may not be limited to an indication of the factors analysed[86]. According to

74 As noted by Fink and Finck, 'on some occasions, it even deduced the right to a reasoned decision directly from art. 47 CFR itself, with or without additionally relying on arts 296 TFEU or 41(2)(c) CFR'.

75 Simona Demková and Herwig CH Hofmann, 'General Principles of Procedural Justice' in Katja S Ziegler et al (eds), *Research Handbook on General Principles in EU Law* (Edward Elgar Publishing 2022) 209 https://doi.org/10.4337/9781784712389.00020.

76 Case C-225/19 *R.N.N.S. v Minister van Buitenlandse Zaken* [2020] ECLI:EU:C:2020:951, para 43.

77 Case T-665/20 *Ryanair DAC v European Commission* [2021] ECLI:EU:T:2021:344.

78 Case C-63/12 *European Commission v Council of the European Union* [2013] ECLI:EU:C:2013:488.

79 Case T-122/15 *Landeskreditbank Baden-Wurttemberg - Forderbank v ECB* [2017] EU:T:2017:337, [124]–[125]; Case C-417/11 P *Council v Bamba* [2012] EU:C:2012:718; [2018] 1 C.M.L.R. 7, [54].

80 Case 5/67 *W. Beus GmbH & Co. v Hauptzollamt München* [1968] ECLI:EU:C:1968:13.

81 Case C-367/95 P *Commission v Sytraval and Brink's France* [1998] ECR I-1719.

82 Case T-220/00 *Cheil Jedang Corp. v Commission of the European Communities* [2003] ECR II-2473.

83 Case 158/80 *Rewe-Handelsgesellschaft Nord mbH and Rewe-Markt Steffen v Hauptzollamt Kiel* [1981] ECLI:EU:C:1981:163.

84 Case C-221/09 *AJD Tuna Ltd v Direttur tal-Agrikoltura u s-Sajd and Avukat Generali* [2011] ECR I-1655.

85 Case C-12/03 P *Commission of the European Communities v Tetra Laval BV* [2005] ECR I-987.

86 Case C-413/06 P *Bertelsmann AG and Sony Corporation of America v Independent Music Publishers and Labels Association (Impala)* [2007] ECLI:EU:C:2007:790.

the case law, in case of discretionary acts it is necessary to indicate objective and predetermined criteria on which such acts can be adopted[87].

Also, according to the CJEU, when the measure is based on established case law, measures may be reasoned in a summarised manner[88] (as long as the judicial review is possible). Conversely, in case of exceptional measures, the reasoning should be more rigorous.

In a nutshell, legal reasoning must be disclosed clearly and concisely, including the applied legal provisions, relevant facts, and key considerations such as legal interpretations, discretionary choices, or other influential factors. When decisions significantly impact the recipient, or deviate from precedent, more thorough reasoning is required and subject to stricter judicial review.

a. The Relationship Between the Right to a Reasoned Decision and the Right to an Explanation

As noted in Section IV, the RTE seems to draw inspiration from the right to a reasoned decision. Indeed, the two rights include procedural and substantive requirements intended to ensure fairness, legitimacy, and accountability in cases of administrative and private-sector decisions. Procedurally, they aim to provide individuals with reasons behind the decision, allowing them to potentially challenge it, assert the legality of the decision, defend their rights in the best possible conditions and – more generally – decide how to behave with full knowledge of the relevant facts. Substantively, they both aim at preventing arbitrary exercises of power, and addressing power imbalances between decision-makers and decision-subjects to ensure decisions are justifiable and acceptable (ie, procedurally and substantively fair).

Hence, the RTE can arguably be considered a declination of the right to a reasoned decision, which can additionally apply in decision taken by private actors. That said, there are fundamental differences between the right to a reasoned decision and the RTE.

Firstly, the obligation of a public authority to provide a reasoned decision applies regardless of whether AI is used while the AIA's RTE is applicable only to some AI-based decisions (eg when decisions have an adverse and significant impact), and has, therefore, a more limited scope of application.

Secondly (and most importantly), while the reasoning must demonstrate that the measure was adopted *secundum legem*, it is sufficient for the explanation to clarify that the decision was not taken in violation of the law (*contra legem*). Indeed, private entities—following the principle of freedom of contract—may legitimately make decisions as long as these do not violate imperative norms or specific legal provisions. For instance, a bank may lawfully refuse to grant a loan to an individual on the basis of certain behaviours (eg, a gambling addiction), provided the decision is not *contra legem* and notwithstanding that individual's right to be informed of the reasons for the refusal. In summary, whereas the right to a reasoned decision must be delivered by public authorities and provide information concerning the subsumption of algorithmic parameters to rules of positive law or principles of law, the explanation does not have to provide such information because, according to the RTE, there is no requirement to demonstrate that the decision adhered to the principles of administrative law and procedure.

In dealing with two similar rights and considering our earlier discussion, one problem is understanding if the right to a reasoned decision 'absorbs' the RTE. Considering the already discussed content requirements of those rights (see §6.1.2 and §6.4.1 for comparison), it is arguable that the right to a reasoned decision does not fully absorb the RTE. In fact, the right to a reasoned decision seems to have more stringent requirements in terms of the object of the information.

Furthermore, the AIA adds further content requirements for the explanation ('the role of the AI system in the decision-making procedure and the main elements of the decision taken'). In doing so, it is arguable that AIA's approach does do not fully tackle the problem of identifying the elements essential to enable the decision-subjects' understanding of the decision allowing them to challenge it. The AIA's RTE poorly identifies the information which may help public authorities to comply with its reason-giving duties when deferring a recommendation provided by an AI system – even if the reason-giving re-

87 Case 144/82 *Detti v Court of Justice of the European Communities* [1983] ECLI:EU:C:1983:211.

88 Case C-521/09 P *Elf Aquitaine SA v European Commission* [2011] ECLI:EU:C:2011:620.

quirements are more demanding[89]. For instance, it has been suggested that a higher explanation requirement should be applied when decision-making is automated to counter-act the risk of 'automation bias'[90], since 'preventing automation bias might necessitate additional safeguards, including, for instance, requiring the public authority that relies on AI to communicate how other available information or alternative outcomes were considered in reaching a decision'[91].

In a nutshell, the AIA's RTE fails in defining the extent to which public authorities are permitted to base their decisions on AI-based recommendations, as well as in establishing the necessary requirements for transparency and interpretability of these systems. Instead, the resolution to these complexities will likely emerge from evolving case-law, which will help define the appropriate balance between transparency and explanation requirements. In the meantime, what can be observed is that when public authorities make a decision in any of the sectors specified in Annex III—whether entirely or partially automated using an AI system—it must be accompanied by a mandatory statement of reasons that also includes the 'clear and meaningful explanation' required by Article 86 AIA. This requirement is likely to require that any HRAIS used in these sectors should have an inherent degree of explainability necessary to produce reasoning and an explanation that complies with the EU administrative law[92].

In light of these brief considerations, the introduction by the AIA of an explanation duty addressed to private and public actors will be problematic[93].

4. Consequences of Failing to Explain

One of the major concerns surrounding AIA's RTE is the consequences of failing to provide an adequate explanation. In this regard, we argue that the absence of an explanation does not necessarily render the underlying decision unlawful but only affects its transparency and contestability. This is not surprising, given that the AI Act adheres more with the normative rationale of product safety than with that of fundamental rights protection. The AIA mainly imposes procedural obligations to ensure that AI-generated outputs meet standards of accuracy and fairness, without establishing substantive requirements for those outputs. In other words, the AIA regulates development and deployment of AI systems, and not the legality of AI-based decisions. This echoes the fact that the AIA does not provide specific remedies for decisions made by AI systems: rather, the assessment of legal compliance and the provision of redress mechanisms follow from sector-specific – frequently, national - regulations.

Differently from public law (where the absence of reasoning constitutes a violation of an 'essential procedural requirement'[94] potentially leading to the annulment of the measure irrespective of its substantive legality) this consequence is particularly relevant in private-sector decision-making, where contractual freedom allows businesses to make determinations—such as denying credit or adjusting insurance premiums—without necessarily providing an explanation unless, mandated by sector-specific laws (such as consumer protection, competition law, and insurance regulations, which may impose additional explanation duties). While an explanation may be required *ex post* to ensure fairness and accountability, the absence of an explanation does not inherently invalidate a decision. In the private arena, even in absence of an explanation under Article 86 AIA, the legitimacy of the AI-generated decision is left to be assessed under existing (labour, financial, etc) law frameworks, with substantive considerations that may fall within national laws, as substantive issues – such as compliance with non-discrimination laws – often fall within the jurisdiction of national legal frameworks, unless otherwise addressed by EU regulations.

However, where fundamental rights are at stake – effective judicial protection as per Article 47 EUCFR

89 For a thorough and interesting analysis, see the work of Fink and Finck (n 48) 376.

90 Fink and Finck (n 48), Sharon Alon-Barkat and M Busuioc, 'Human-AI Interactions in Public Sector Decision-making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice' (2022) Journal of Public Administration Research and Theory https://doi.org/10.1093/jopart/muac019.

91 Fink and Finck (n 48) 387.

92 Fink and Finck (n 48) 386-387.

93 As also noted by Hofmann, '[m]ixing public and private obligations is problematic since each have different legal obligations as to their procedures. Arguably, the use of AI in public decision-making should better be integrated into general EU administrative procedures act and address specific effects of ADM on decision-making and rule-making procedures'. See Hofmann (n 41).

94 C-521/09 P *Elf Aquitaine SA*, [146]; C-367/95 P *Commission v Sytraval and Brink's France*, [67]; Case C-17/99 *France v Commission* [2001] EU:C:2001:178, [35].

for instance – the absence of an explanation could undermine the fairness of the decision-making process. The decision-maker who provides inadequate or no explanation runs the risk of violating of the essence of the right to a fair trial protected under Article 47(1) EUCFR. As Brkan argues (in that case, referring to a right to explanation under the GDPR), the absence of concrete safeguards – particularly the right to contest the decision and the related right to be informed about the reasons for decision – might, indeed, violate Article 47 EUCFR[95].

VII. A Tailored Right to an Explanation

The growing integration of ADM systems across various sectors, ranging from criminal justice to administrative governance and commercial applications, necessitates enhanced scrutiny of the mechanisms used to provide adequate explanations to affected individuals. Indeed, many are the critiques that have been made to the RTE.

First, the potential for explanations to become performative rather than substantive. AI systems can generate 'plausible' explanations that may apparently seem to satisfy regulatory demands, without reflecting the system's true decision-making processes. This phenomenon reduces transparency to a mere formality by offering a '*façade*' of accountability.

Second, the RTE is neither not infallible in addressing the challenges posed by AI systems. Explanations, while empowering in theory, might fail to deliver their intended effects in practice. Research[96] has demonstrated that providing excessive explanations can overwhelm individuals with overly complex information, impairing their ability to critically assess decisions. Moreover, as De Mulder and Valke[97] highlight, explanations risk being inaccurate, especially given the opacity of many AI systems, where the true decision-making logic remains elusive. In such cases, explanations may inadvertently serve as tools of obfuscation, offering plausible but misleading justifications that conceal biases or unjust outcomes. Also, incomplete explanations–those that fail to fully capture the causal or logical reasoning behind a decision–can render the right to contest decision ineffective.

Lastly, there is the overlooked issue of cognitive accessibility. Even if explanations are technically accurate and provided as required, they might not be comprehensible to non-experts. This is compounded by disparities in digital literacy across the population, effectively limiting the ability of affected individuals to understand decisions, thereby eroding the practical value of the RTE. These challenges call, not just for legal innovation but also interdisciplinary approaches to ensure that the RTE can be meaningfully implemented in diverse, real-world contexts.

Accordingly, it is necessary to examine the challenges inherent to formulating a uniform RTE that is applicable across all legal domains, arguing instead for a tailored approach that reflects the specific regulatory, procedural, and substantive nuances of each sector.

ADM processes are governed by rules and transparency policies that are inherently tailored to the specific sector in which they are deployed, reflecting the multifaceted roles that legal systems play both nationally and globally[98]. As illustrated by Doshi-Velez et al, for instance, in legal contexts emphasising individual responsibility and restitution–such as criminal law and personal injury liability–individualised explanations are often pivotal to achieving just outcomes[99]. Conversely, in areas where the law is more concerned with social welfare, such as certain anti-discrimination statutes or administrative rule-making, there is typically a greater reliance on empirical evidence and procedural guarantees[100]. These sector-specific mandates broadly indicate the type and extent of information that must be communicated to decision-subjects, thereby challenging the effectiveness of instituting a uniform, horizontally applicable RTE across all areas of law. Instead, it becomes evident that legislators must tailor RTE provisions to the particular legal sectors where ADM systems significantly impact individual rights – a perspective supported by the flexibility embedded in Article 86(3) AIA. In essence, transparency requirements and the corresponding informational disclosures should be calibrated in accordance with the underlying purpose for which the right is invoked – whether that entails justifying an admin-

95 Brkan (n 58) 119. In such context, Brkan goes further arguing that '[e]ven if the decision is taken with a human intervention, the human would still need to provide the data subject with reasons, giving her an opportunity to effectively challenge the decision'.

96 Forough Poursabzi-Sangdeh et al, 'Manipulating and Measuring Model Interpretability' (2021) arXiv:1802.07810.

97 De Mulder and Valcke (n 24).

98 Doshi-Velez et (n 53), para 4.B.

99 Ibid.

100 Ibid.

istrative decision, explaining a disciplinary measure in employment law, or elucidating the rationale behind the manoeuvre of a self-driving car. Ultimately, establishing effective standards for explanation necessitates a preliminary and precise understanding of the goal for which the right is exercised.

VIII. Conclusions

The right to an explanation represents a crucial yet deeply ambivalent instrument in the evolving landscape of AI governance. Its inclusion in AIA signals an institutional commitment to transparency and accountability in ADM. Nonetheless, its current formulation risks fostering an illusion of control rather than granting substantive protection.

On the positive side, the RTE neither inherently impedes innovation in machine learning nor diminishes predictive performance. Instead, it aspires to introduce post-hoc justificatory duties and oversight mechanisms designed to temper the harms associated with AI systems. Yet the obligation to explain algorithmic decisions does not necessarily empower the individuals affected by them; rather, it threatens to burden lay users with opaque technical rationales. As Edwards and Veale observe, the interpretive burden thus 'remains disproportionately placed on end-users,' who typically lack the technical literacy required to contest algorithmic outcomes meaningfully[101]. This asymmetry not only weakens the protective function of the RTE but also risks entrenching a 'compliance theatre' in which AI deployers generate legally acceptable but ultimately meaningless explanations.

More troubling is the RTE's potential to become a symbolic concession rather than a substantive right. The absence of robust enforcement mechanisms renders it susceptible to being ritualised – a procedural checkbox rather than a tool for contestation. As Demková suggests, only an integrated, purposive approach – one that aligns the RTE with the AI lifecycle and guarantees under Articles 41 and 47 EUCFR – can render it practically effective, especially in public-sector applications.

If the RTE is to function as more than a symbolic gesture, it must be embedded within a broader regulatory architecture that includes procedural safeguards, algorithmic contestability, and meaningful human oversight. Without these complementary mechanisms, the RTE remains a formal right with limited transformative capacity – a legal artifact that acknowledges the problem of opacity but with minimal real-world impact on mitigating power imbalances and fairness concerns in ADM.

101 Edwards and Veale (n 58).