AI and Ethics
https://doi.org/10.1007/s43681-025-00754-0

ORIGINAL RESEARCH

Non-discrimination law, the GDPR, the AI act and the - now withdrawn - AI liability directive proposal offering gateways to pre-trial knowledge of algorithmic discrimination

Ljupcho Grozdanovski[1,2]

Received: 2 December 2024 / Accepted: 7 May 2025

Summary
This article focuses on the evidence necessary to support claims of discrimination arising from AI-assisted recruitment. It addresses two main issues. First, given that discrimination may be subtly expressed by (possibly opaque) AI systems, this article examines the EU legal frameworks designed to facilitate access to explanations and evidence capable of revealing discriminatory bias in automated recruitment processes. Those provisions include the Equality Directives, the GDPR, the AI Act (AIA), and the now-withdrawn AI Liability Directive (AILD) proposal. In analysing those provisions, particular attention is paid to the types of information that may be sought: the logic behind an AI's output, the reasons a human decision-maker relied on that output, and the AI system's compliance with the AIA. Second, the article determines which among the various applicable provisions should be treated as *lex specialis*, that is, the specific rule that should be preferentially applied to obtain pre-trial knowledge of algorithmic discrimination. In this context, special emphasis is placed on Articles 22 GDPR and 86 AIA, both of which recognize a right to an explanation and are potentially applicable to automated recruitment systems, since those can be classified as both high-risk under Annex III of the AIA and involving personal data processing, under the GDPR. From the standpoint of a litigant's ability to satisfy the procedural requirements of both provisions, the article argues that Article 86 AIA may offer a more accessible pathway than Article 22 GDPR, both in terms of the scope of information provided and the conditions required for access. Nonetheless, neither provision guarantees automatic disclosure; access remains conditional and often subject to stringent procedural requirements. This selective, rather than automatic approach to transparency raises important questions about its implications for fundamental rights, particularly the right to access justice and effective remedies.

Keywords Artificial intelligence · Explanation · *Lex specialis* · GDPR · AI act · AI liability directive · Evidence · Effective judicial protection– Access to remedies

.

1 Introduction

In the emerging caselaw dealing with Artificial Intelligence (AI) systems, the *Loomis*[1] case has become unavoidable, in discussions of AI-related discrimination. Famously addressing the judiciary's use of the (potentially biased) COMPAS recidivism prediction system, the Wisconsin Supreme Court maintained that the system served merely as a tool, intended to assist, rather than replace, the court's reasoning. Importantly, the Loomis case does not involve actual or probable discrimination supported by concrete evidence. Rather, it

✉ Ljupcho Grozdanovski
lgrozdanovski@uliege.be

[1] Fund for Scientific Research, Brussels, Belgium

[2] University of Liège, Liège, Belgium

[1] Supreme Court of Wisconsin, 1 3 July 2016 (decided), *State of Wisconsin v. Eric L. Loomis*, 881 N.W. 2 d 749 (2016) 2016 WI 68.

Published online: 29 May 2025

addresses the risk of discrimination, based on the potential of the COMPAS algorithm to exhibit a bias.

Though certainly a legitimate concern, that potentiality raises the issue of whether it is possible to know that one has been the victim of discrimination. The variable association (i.e. the correlation between various types of data) underlying a specific output (recruitment, loan approval, university admission etc.) is often not visibly biased. For instance - and hypothetically speaking - an AI used for college admission might, on the surface, base its decision on the applicants' Grade Point Average (GPA), but also consider their places of residences. The result might be the rejection of applicants residing in racially mixed areas. However, it may not be easy for the rejected applicants to, practically speaking, acquire knowledge of those facts.

For a few years now, scholars have been triggered by the covert ways in which AI systems can (and sometimes do) violate fundamental rights.[2] As the use of those systems is enhanced and normalised, we see the emergence of real-life situations that illustrate the challenges they raise regarding the assumption that, as any harm, discrimination is a *prima facie* knowable occurrence.[3] Usually, when a person requests that harm be compensated, they are able to present a causal narrative, showcasing how event X led to event Y (and not to event Z).[4] To do so, they should have unhindered or, at least, reasonably open access to facts and knowledge of how those events form a - ideally, convincing - causal structure. We, indeed, tend to view situations of injustice (understood as a violation of a legal and/or ethical norm[5]) as fact-of-the-matter, verifiable events which - precisely because they are verifiable - justify the launching of procedural mechanisms meant to restore justice, not universally but individually, in a litigant's particular circumstances.

AI systems have come to upset the knowability-of-harm postulate. Recently, ChatGPT has been found to express a gender bias, by using gendered language when tasked with the drafting of reference letters for job applicants.[6] The expression of that bias was subtle, discernible to the trained eye of one sensitised to gendered language uses, but perhaps not to a person who might not view a trait like industriousness as inherently masculine. In a similar vein, van de Waerdt mentions a study of how Facebook likes were used to predict people's sexual orientation. The study found that "for unclear reasons,"[7] likes of all things Britney Spears were "moderately indicative of homosexuality.[8] Stereotypes aside, the association between music tastes and sexual orientation should, in principle, not be considered as an *indicium* (a probative element) of discrimination a claimant could confidently rely on before a court. Of course, this is a general observation and does not apply to all cases where Meta may have correlated music tastes with sexual orientation. For a case of discrimination to be made, a person must present, what Bex et al. call *justified arguments* that is, arguments that "survive the competition with their counterparts"[9] and prove persuasively, if not conclusively, that their Facebook likes triggered conduct or a practice that differentiated them from other users based on a protected characteristic.

The issue of (un)knowability and (un)provability of AI-related discrimination has also been raised in judicial instances following *Loomis*. We will mention two such

[2] For a usefully informative study on the AI-human rights connection, see, amongst others, McGregor, L., Murray, D., Ng, V.: International Human Rights Law as a Framework For Algorithmic Accountability. The Int'l & Comp. L. Quart'ly, 68/2, 309–343 (2019).

[3] In hist study on AI decision-making in criminal law and procedure, Chiao addressed the issue of biased AI output, upholding a view of AI systems as being merely 'transmission belts,' replicating various overtly or covertly biased *human* practices. Under this view, Chiao does not seem to suggest that AI biases are particularly problematic (in terms of detectability) since they are extensions of already existing and recorded practices. See Chiao V.: Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice. Int'l J. L. in Cont. 15, 126–139 (2019) at 127.

[4] Persuasively showcasing a fact's/event's causal power is, essentially, a concrete application of a criterion– well-known in liability circles– that is the criterion of necessity. When litigants give evidence that show how two events are casually linked (as opposed to merely correlated), they essentially show how they are linked under the criterion of *causal necessity*. As Ridley put it, causal explanations are contrastive because when these answer a why-question, they provide understanding on why "event p happened instead of some event q." See Michael Ridley M.: Explainable Artificial Intelligence (XAI): Adoption and Advocacy, Inf. Tech'y & Lib. 41/2, 1–17 (2022), at 4.

[5] For simplicity, we synonymize 'injustice' to 'legal injustice' that is, violation of the law. Of course, injustice and its counterpart (justice) are much more complex and broader than the narrow 'justice is whatever the law says' view. Let us however introduce a brief comment here. In any system based on the rule of law, legal provisions are meant to give expression to the principle - initially conceptualised in ethics - of equality or proportionality. The gist is that law should distribute rights and duties in a way that does not create or maintain asymmetries of entitlements, means and power. In this sense, Delaney connected injustice to systemic asymmetries of power, such as domination, exploitation and marginalisation both in the world and with respect to access to law. See Delaney D.: Legal Geography II: Discerning injustice. Pr. H. Geo'y, 40/2, 267–274 (2016), at 268. For justice as foundation and aspiration to most modern conceptions of the rule of law, see Tasioulas J.: The Cambridge Companion to the Philosophy of Law. Cambridge University Press (2020) at 117 seq.

[6] See Yin L., Alba D., Nicoletti L.: OpenAI's GPT is a recruiter's dream tool. Tests show there's racial bias. (2024). https://www.bloomberg.com/graphics/2024-openai-gpt-hiring-racial-discrimination/?embedded-checkout=true, accessed on 23 November 2024.

[7] van de Waerdt P. J.: Information asymmetries: recognizing the limits of the GDPR on the data-driven market CMLRev. 38, 1–18 (2020), at 6.

[8] *Ibid*.

[9] Bex F. J., van Koppen P. J., Prakken H., Verheil B.: A hybrid formal theory of arguments, stories and criminal evidence. A. I. L., 18, 123–152 (2010) at 130.

instances. The first is from a UK Court. In 2020, the Cardiff Court of Appeal (Civil Division)[10] gave a ruling on police authorities having used biometric identification cameras to perform indiscriminate facial image recognition, face detection, feature extraction, face comparison and matching at a public event.[11] The cameras focused on the arena's entrance and performed facial recognition based on three watchlists: people previously arrested, persons wanted on warrants and suspects. The UK courts were called to address the lawfulness of such an operation, regarding the right to privacy (Article 8 of the European Convention of Human Rights - ECHR) and the principle of equality/non-discrimination. To assess the probability that the system might discriminate, the Cardiff Court of Appeal referred to "scientific evidence that facial recognition software can be biased and create a greater risk of false identification in the case of people from black, Asian and other minority ethnic backgrounds, and also in the case of women."[12] Be that as it may, no concrete evidence of discrimination was presented before the Court, urging it to double down on the human control requirement, arguing that prior to any human intervention based on the system's output, "there must be two human beings, including at least one police officer, who have decided to act on the positive match."[13] In essence, the Cardiff Court of Appeal considered that the mere likelihood - not certainty, or even high probability - of harm, warranted enhanced human involvement in the assessment of the system's output.

We find a similar consideration in the second example from a Brazilian court.[14] ViaQuatro was a camera surveillance data processing system, located above several entrances ('digital doors') to the São Paulo metro. The system was able to recognise human faces and detect emotion, gender and age, triggering the possibility for systematic discrimination of trans and non-binary individuals. Since the Brazilian data protection law was not yet in force when the cameras were installed, the São Paulo civil court was called to evaluate the risk of discrimination on the grounds of consumer law. Said court did, indeed, find that there was violation of consumer protection rules, considering that the passengers were not given clear and precise information about how the product (i.e. the cameras) were being used.

The cited cases have two points in common. First, they deal with biometric identification systems. Second and more importantly, the element missing in both was convincing evidence of discrimination. Like *Loomis*, these rulings address the potential rather than the (proven) probability of the systems involved to give discriminatory outputs.

From a procedural point of view, and with the knowability-of-harm assumption in mind, we cannot help but wonder if the mentioned cases sketch out the features of evidentiary debates that may become more frequent in future litigation. Will courts seek evidence on the probability of harm AI systems are likely to cause, instead of harm they have actually caused? Perhaps this shift from evidence of 'what happened' to evidence of 'what might happen' is integral to the growing pains associated with the coming of age of AI-related litigation. As a collective, we may need to get used to the fact that, as AI technologies become more sophisticated, it will be less possible to know (with reasonable certainty) or discover (with reasonable feasibility) how they caused or contributed to causing harm. Judicial fact-finding[15] might, indeed, shift from being a process of *demonstration* (discovery and assessment of facts associated with an actual dispute) to being a process of *verification* (assessment of whether there is enough evidence to establish a *justiciable* dispute).

The challenge AI raises regarding the access to knowledge about facts has a significant impact on the right to access remedies, as guaranteed under Article 47 of the European Union's (EU) Charter of Fundamental Rights

[10] Court of Appeal (Civil Division), on appeal from the High Court of Justice - Queen's Bench division (Administrative Court), Cardiff District Registry, Case No: C1/2019/2670.

[11] *Id*., para. 9.

[12] *Id*., para. 164.

[13] *Id*., para. 184.

[14] 37a Vara Civel Do Foro Central Da Comarca E Capital Do Estado De São Paulo, 1,090,663 −42.2018.8.26.0100.

[15] The emphasis on *judicial* proceedings allows us to limit the scope of this article to fact-finding conducted within the context of trials and primarily incumbent on the parties. This is a necessary clarification because, in our further developments, we will not focus on pre-trial fact-finding procedures conducted by public authorities like, say, the European Commission in Competition law. In that connection, see for example, Hjelmeng E.: Competition law remedies: Striving for coherence or finding new ways?. CMLRev., 50 −4, 1007–1037 (2013). With that said, and regardless of whether a specific evidence-collection endeavor formally includes a pre-trial procedure (e.g. EU Competition law) or is subject to a court's assessment within the context of a trial (e.g. ordinary civil proceedings), the stages of what we call 'fact-finding' remain the same. Pattenden usefully summarised the function of fact-finding (its procedural framing notwithstanding) into two essential stages: empirical and non-empirical. The empirical stage (or fact-finding proper) is the discovery of the - as she calls them - actual facts, the non-empirical stage (classification) being where the judges decide whether "the semantic meaning of the text fits the actual facts." See Pattenden R.: Pre-verdict Judicial Fact-Finding in Criminal Trials with Juries. Oxford J.L.S. 29/1, 1–24, at 5 (2009). Considering that in this article, our focus is on proof of standing, required for a court to declare admissible specific legal proceedings, the attention we pay to 'legal classification' will be limited. Indeed, 'legal classification' (that is, the subsumption of facts to law) is relevant to the extent that we need to answer the question: which facts does the law require to be 'found' and proven so that a litigant's action can be admissible, enabling them to exercise their entitlements (procedural and substantive) before a court? This is essentially the question we will aim to answer in our analysis of Article 86 AIA (n. 32) in Sect. 3 of this article.

(ECFR).[16] The main reason is that the standard requirements of admissibility of proceedings cannot be easily satisfied when litigants are called to prove standing in connection with a suspected AI-related harm. The EU courts have stressed the critical difference between claims proper and mere assertions,[17] implying that the admissibility of actions launched on the grounds of EU law depend on the litigants' ability to meet pre-defined duties to give evidence. In most, if not all national and supranational procedures forming part of the EU's 'complete system of legal remedies,'[18] claimants are typically asked to prove that their legal situations have been affected (e.g. a right was violated, illegality was committed, harm was suffered…) by a specific and verifiable measure, conduct or practice. Generally, if the evidence of standing is deemed sufficient and satisfies a standard of proof (most frequently, preponderance of evidence[19]), proceedings may be declared admissible.

As the above-mentioned cases show, the trouble with actions whereby the debate on evidence is focused on discussing *potential of harm* (as opposed to materialised or highly probable harm) is that they are evidentially challenging because the facts are not within the (alleged) victim's immediate reach. Naturally, demonstrating the likelihood of harm is more difficult than proving that harm has actually occurred. This already daunting task is made even harder when a potentially opaque AI system limits access to the information needed for a claimant to launch judicial proceedings. In this context, and considering the above-mentioned case law, two questions can be raised. First, what pre-trial factual knowledge might lead claimants to suspect a probable risk of having suffered discrimination occasioned by an AI system? Second, does EU law in the fields of non-discrimination and AI open access to such knowledge, in view of ultimately enabling effective judicial redress?

For simplicity (and in line with part of the zeitgeist in AI scholarship[20]) we will focus on AI-related discrimination in the context of recruitment. Three reasons justify our choice. First, whether AI-assisted or not, recruitment is by nature an opaque process of selection. Proving discrimination in that context is already a difficult task, made even more challenging by the involvement of an AI system. In the following Sections, we will refer to already existing legal practices and case law in the EU, seeking to uncover if and how evidence can be made more accessible for job applicants intending to bring discrimination claims before courts. Clarifying these issues will help form informed opinions on how we should approach the evidence in disputes involving AI, not only in the field of access to labour, but across the high-risk sectors listed in Annex III. The reason for this is simple: the most pressing concern in those sectors (from credit scoring to automated judicial decision-making) is the risk that the systems deployed express unfair biases.

Second, the so-called Equality Directives[21] 'put into effect'[22] the principle of equal treatment[23] essentially in instances dealing with labour, particularly employed labour, and all things labour-adjacent (like access to vocational training or social advantages). In this context, said Directives prohibit discrimination regarding, namely, the access

[16] Article 47 ECFR (OJ C 326/2012, p. 391: "everyone whose rights and freedoms guaranteed by the law of the Union are violated has the right to an effective remedy before a tribunal in compliance with the conditions laid down in this Article (…)." See also Gutman K.: The Essence of the Fundamental Right to an Effective Remedy and to a Fair Trial in the Case-Law of the Court of Justice of the European Union: The Best Is Yet to Come. German L.J., 20/6, 884–903 (2009).

[17] The difference between a claim and a mere assertion is the degree to which each is supported by evidence. In the *British Airways* case for instance, the, at the time, Court of First Instance (CFI) was very critical toward "general assertions and assumptions unsupported by any concrete evidence." See joined cases T-371/94 and T-394/94 [1998] *British Airways et al.*., EU: T:1998:140, para. 70.

[18] We refer to case 294/83, *Les Verts v. European Parliament* [1986] EU: C:1986:166, para. 23.

[19] In national doctrines of procedural law, 'preponderance of evidence' is the standard of proof associated with civil proceedings, while 'beyond reasonable doubt' is the standard of proof applied in criminal proceedings. Standards of proof typically translate a 'certainty' or 'sufficiency' threshold applied to evidence. This is a text-book *summa divisio*, which we will not elaborate on in this article. May it be noted, however, that standards of proof have given way to an elaborate scholarship which has inquired, amongst other things, how procedural law should articulate various considerations (and difficulties) that the parties face when striving to achieve a predefined sufficiency threshold. These considerations include efficiency (informational costs), accessibility of evidence etc. These and other points have been addressed in expert scholarship and will not be further elaborated here. For a specialised and critical view on 'evidence thresholds,' see, for example, Stein, A.: Evidence, Probability and the Burden of Proof. Arizona L.R., 55, 557–603 (2009), at 580 seq.

[20] See *inter alia* Pan Y., Forese F., Liu N., Hu Y., Ye M.: The adoption of artificial intelligence in employee recruitment, The influence of contextual factors,. Int'l J. Human Ress. Management, 33/6, 1125–1147 (2022); Köchling A., Wehner M. C., Warkocz J.: Can I show my skills? Affective response to artificial intelligence in the recruitment process. Rev. Managerial Sc. 17/6, 2019–2138 (2023); Ligeiro N., Dias I., Moreira A., Recruitment and Selection Process Using Artificial Intelligence: How Do Candidates React?. Admin. Sci., 14/7, 155–172 (2024).

[21] Council Directive 2000/78 establishing a general framework for equal treatment in employment and occupation, OJ L 303/2000, p. 16; Council Directive 2000/43 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, OJ L 180/2000, p. 22 and Directive2006/54 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast), OJ L 204/2006, p. 23.

[22] 'Putting into effect the principle of equal treatment' is the rationale of each of the Equality Directives. See, for instance, Directive 2000/43 (n. 24), Art. 1.

[23] The constitutional provisions relative to the principles of equality and non-discrimination are 19 Treaty on the functioning of the European Union (TFEU) and Art. 21 ECFR (n. 19°.

to employment, self-employment,[24] vocational training,[25] dismissal and pay,[26] union membership,[27] social security and healthcare.[28] Mirroring those grounds of discrimination, Annex III of the AI Act (AIA)[29] labels as high-risk the sectors of vocational training and employment,[30] workers' management and access to self-employment.[31] Given the overlap between the Equality Directives and two high-risk sectors in the Annex III AIA, AI-assisted recruitment can, indeed, fall in the scope of application of both provisions. Additionally, bearing in mind the revised and recently enforced Product Liability Directive (PLD)[32] and the - now withdrawn[33] - AI Liability Directive (AILD) proposal,[34] the overlap with EU Discrimination Law is also visible in the field of evidence and procedure. Indeed, the evidence relating to discrimination in connection to automated recruitment could be regulated by the systems of evidence in the Equality Directives as well as the AILD, had the latter been enacted. Hence, the need to reflect on how those provisions may or may not complement each other in view of supporting the litigants' quest for evidence-worthy knowledge about the facts and effective judicial redress.

Third, and expanding on the overlap between the Equality Directives and the EU's substantive and procedural regulation of AI, we should stress that future AI-assisted recruitment cases will be adjudicated in a regulatory context where transparency (particularly *informational transparency*) is translated in a multitude of subjective entitlements (rights). Depending on how a given case is framed, claimants could rely on Article 22 of the General Data Protection Regulation (GDPR),[35] Article 86 AIA, Articles 8 PLD and 3 AILD (had it been enacted). These entitlements give specific expressions to the principle of explainability, evangelised as one of the pillars of the EU's AI regulation.[36] Frequently paired with transparency and human control and oversight,[37] explainability has been defined in many ways, which can be reduced to two key approaches: technical and functional. Technically, explainability is the aptitude of the system *to be explained*. It is seen as a characteristic of interpretable (ergo transparent)[38] AI, overseen by humans who - because of those systems' transparency - are able to explain the functionalities and rationale underlying given AI output. Functionally (from the vantage point of explanatory purpose), explainability is a *narrative aptitude* - that of the explainer to deliver an understanding of situations involving AI systems. We will not focus too much on the epistemic constraints that frame explanatory 'goodness'; these have been sufficiently analysed elsewhere.[39] Against the backdrop of explainability as a technical feature (of AI systems) and a rhetoric capacity (of explainers), our analysis will focus on the procedural conditions that frame the *access* to explanations, which could, in turn, facilitate the litigants' access to remedies.

[24] Directive 2000/43 (n. 24), Art. 3(1)(a),

[25] *Id*., (b).

[26] *Id*., (c).

[27] *Id*., (d).

[28] *Id*., (e). In the other Directives, the provisions equivalent to Article 3 Dir. 2000/43, *cit. supra* are Articles 3, Dir. 2000/78 (n. 24) and 7 to 9 Directive 20,006/54 (n. 24),.

[29] Regulation No 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act - AIA), OJ L 2024/1698.

[30] *Id*., Annex III, pt 3.

[31] *Id*., pt 4.

[32] Directive 2024/2853 on liability for defective products, OJ L 2024/2853.

[33] Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to Artificial Intelligence (AI Liability Directive - AILD) COM (2022) 496 final. Beyond the initial enthusiasm of the AILD proposal, it has recently experienced a halt for various reasons (such as lack of clarity in the design and content of the rules in the proposal). See Bertuzzi, B., "France leads charge in effort to torpedo EU's AI Liability Directive" 24 October 2024, available on: https://mlexmarketinsight.com/news/insight/france-leads-charge-in-effort-to-torpedo-eu-s-ai-liability-directive (last accessed on 30 November 2024).

[34] Directive 2024/2853 on liability for defective products (n. 35).

[35] Regulation No 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of data, JO L 119/2016, p. 1.

[36] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI (2019), available on: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (last accessed on 6 May 2025), at 13.

[37] *Id*. at 18: "(Transparency) is closely linked with the *principle of explicability* and encompasses transparency of elements relevant to an AI system: the data, the system and the business models." In the AIA, transparency is primarily associated with the 'appropriate' use of a high-risk AI system (see Art. 13(1) AIA (n. 32): "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 and use it appropriately"). However, the appropriate use includes the component of human oversight measures "including technical measures put in place to facilitate the interpretation of the outputs of the high-risk systems by the deployers." See AIA (n. 32), Art. 13(3)(d).

[38] Although explainability and interpretability are often viewed as equivalent, they have different meanings from an engineering/data science perspective. Interpretability pertains to the explanations on how and why a model gave a specific output. Explainability pertains to explanations relative to that output, delivered in a human-understandable form. See Bahalul Haque, A.; Najmul Islam A.; Mikalef P.: Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. Tech. F. & Soc. Ch. 186/1, 1–17 (2023), at 12.

[39] See Grozdanovski, L.: The Explanations One Needs for the Explanations One Gives. The Necessity of Explainable AI (XAI) for Causal Explanations of AI-related harm - Deconstructing the 'Refuge of Ignorance' in the EU's AI Liability Regulation. Int'l J. L. Tech'y & Ethics. 2, 155–262 (2024).

Building on these observations, the main objective of this article is to examine how transparency and explainability, as translated in subjective rights (the right(s) to an explanation, the right to request the disclosure of evidence), can open access to pre-trial knowledge about a possibly biased AI output, affecting an individual's personal situation. This is, presumably, the value this article would add to the existing doctrine on AI-related discrimination. Well-established scholarship has already delved into the variable association/bias interrelationship, calling it discrimination by proxy, the work of Xenidis[40] being a must-read in that connection. This article pursues a different goal. It will not focus on how AI-related discrimination might occur, but how it can be made known to the potential victims.

To that end, Sect. 2 will set the stage, by outlining the evidentiary requirements connected to the assumption of knowability. The constant case law of the Court of justice of the EU (CJEU) shows that, be it in direct or indirect discrimination cases, claimants are assumed to have elementary knowledge of being subjects to differential treatments, meaning that in most cases, they can reasonably meet the burden of proving standing required by the EU Equality Directives. Against the backdrop of Sect. 2, Sect. 3's point of departure is that AI systems often express biases in subtle, unknowable ways, making it difficult for claimants to even suspect unfair treatment. This Section will therefore explore if Articles 22 GDPR and 86 AIA can be relied upon to access such knowledge, finding that, as far as the accessibility of explanations is concerned, Article 86 AIA will likely be viewed as a *lex specialis* in high-risk Sectors. The downside (we contend) is that the explanations due under that provision will not be conducive to pre-trial knowledge of discrimination, but to cases of salient (*ergo* provable) discrimination.

Section 4 will inquire if the right to request disclosure of evidence in the AILD proposal could have opened a more effective access to empirical knowledge in comparison to that enabled by Article 86 AIA. Considering the withdrawal of the AILD proposal, it is difficult to prospect whether these provisions could have applied jointly (with Article 86 AIA being the explanatory antechamber to Article 3 AILD) or separately (where litigants may have preferred Article 3 AILD to Article 86 AIA, when seeking to uncover information about the defendant's level of compliance with the AIA). The impact on effective judicial protection of the possible hierarchies and interpretations of the rights discussed are analysed in the concluding remarks in Sect. 5.

Before we proceed, we must open a methodological parenthesis to clarify our operative concepts and key methodological choices. First, we should define evidence and explanations. 'Evidence' will be viewed holistically as both a *material demonstration* of an empirical truth with the support of specific items of evidence (documentary, testimonials etc.) and an *argumentative demonstration* a litigant offers in view of persuading a court that what they claim is true. As Bex et al. note, that reasoning is complex and multilayered because it departs from the "available evidence (which) constraints which facts can be proven, which in turn are the grounds for the legal consequences on the basis of the applicable legal rules."[41]

What we call the 'holistic understanding of evidence' is, in fact the oldest one: an *argumentum* as a narrative about facts is persuasive because of the orator's rhetoric skills, but also because of their ability to give *probatio* that is, items upon which their empirical assertions can stand.[42] In light of this view of evidence as an argument about reality, its connection to explanations is not difficult to discern. We have analysed explanations in the broader context of knowledge theory elsewhere.[43] In this article, may it be sufficient to define explanations as interpretations of experience[44] meant to yield understanding (as opposed to knowing proper) of that experience.

Defined in such a way, it is apparent how explanations can be integral to evidence (as we understand it): both seek to generate an understanding of 'how' an event occurred, 'why' it occurred and 'why' it is relevant for the law, to a point where court proceedings are justified. We can even contend that legal evidence is a form of explanation since both are factual statements that can be 'true' or 'false,' depending on the empirical facts that support specific

[40] See, namely, Xenidis, R.: The polysemy of anti-discrimination law: the interpretation architecture of the framework employment Directive at the Court of Justice. CMLRev. 58/6, 1649–1696 (2021); Xenidis R.: Tuning EU equality law to algorithmic discrimination: Three pathways to resilience. Maast. J. Eur. Comp. L. 27/6, 736–758 (2020); Xenidis R.: Beyond bias: algorithmic machines, discrimination and the analogy trap. Trans. Leg. Th'y, 14/4, 378–412 (2023).

[41] Bex F., Verheij B.: Legal stories and the process of proof. AI & L. 21, 253–278 (2013) at 255.

[42] There is a tradition which some scholars have attributed to Cicero, which suggests that *argumentum* was used - namely by that philosopher - as being equivalent to *probatio*. Strictly speaking, the latter is a material demonstration (evidence is adduced to establish facts) whereas the former is a rhetoric demonstration (a narrative explaining how the facts established tie together). See Lévy J.P.: Les classifications des preuves dans l'histoire du droit. In: Perelman C., Foriers P. (eds.), La preuve en droit: études. pp. 27–58. Bruylant (1981), at 29.

[43] Grozdanovski L.: The Explanations One Needs for the Explanations One Gives. The Necessity of Explainable AI (XAI) for Causal Explanations of AI-related harm - Deconstructing the 'Refuge of Ignorance' in the EU's AI Liability Regulation (n. 42).

[44] Franz Lamberti W. F.: An overview of explainable and interpretable AI. In: Baratesh F., Freeman L. (eds.), AI Assurance. Towards Trustworthy, Explainable, Safe and Ethical AI Elsvier, pp. 55–123. Elsevier (2022), at 57.

empirical statements.[45] Both explanations and evidence are, indeed, arguments that aid "a particular apprehension"[46] of the facts and can be viewed as equivalent. At least for the purpose of this article.

The conceptual synonymy between evidence and explanations allows us to suggest a procedural one: the rights to explanations in the EU's digital legislation can be viewed as rights to access evidence. The tricky question is whether those rights can be relied upon pre-trial, for the purpose of verifying the presence of harm. This goes against traditional ways of interpreting the right to explainability, the exercise of which presupposes that there be something to be explained. We will comment more on this when we elaborate on the knowability of harm as a precondition to exercising the right(s) to an explanation.[47]

Second, we should clarify how we define discrimination in connection to equality. The fundamental issue in this article is whether all violations of the equality principle can be qualified as discriminations. The answer is, of course, 'no.' People are not treated equally for tax purposes simply because they do not fall in the same income brackets. Similarly, if a legislator wished to encourage access to labour for young, fresh-off-college people, they would provide advantages (e.g. easier conclusion of short-term contracts, lower income tax etc.) for below-25 college graduates.[48] Now, if it so happened that employers in the sectors dominated by young, entry-level workers primarily hired men, there would be unequal treatment qualifiable as gender discrimination.

A cursory overview of the CJEU's caselaw based on the Equality Directives and, more generally, on EU citizenship and the freedoms of movement, shows that, in most cases, non-discrimination and equality are treated interchangeably, although it would be more correct to consider the former as an expression of the latter. The *Schmelz* case[49] for instance, dealt with the differential application of a tax exemption to a German homeowner in Austria. The referring court asked the CJEU if the VAT Directive[50] was consistent with - at the time of the ruling - Articles 12 EC (non-discrimination), 43 EC, 49 EC (freedom of establishment) and "the general principle of equal treatment."[51] Both the CJEU[52] and Advocate General (AG) Kokott,[53] found that in the main proceedings, the principle of non-discrimination did not apply independently, since it was particularised in specific Treaty provisions, stressing that this reasoning applied "*mutatis mutandis* to the principle of equal treatment, which is recognised as a general principle of law."[54] What qualifies an instance of unequal treatment as 'discrimination' is that differential distribution of benefits is based on one or several of the so-called protected characteristics, which essentially pertain to a person's ontological (or identitarian) features like race, gender, sexual orientation, political beliefs etc.[55] Showing that these features are the basis for a difference of treatment is what litigants will be called to prove in cases involving AI systems on the grounds of EU Non-Discrimination law. We will explore further how the different provisions (Equality Directives, GDPR, AIA and the PLD/AILD) correlate in regulating the procedural conditions under which algorithmic discrimination can be argued and established.

Finally, a self-evident but necessary observation should be made. The cases of discrimination we will consider will be associated with high-risk systems, in particular those used in the field of labour. The high-risk systems are defined in Article 6 AIA as being either safety components of a product, covered by existing Union legislation[56] and required to undergo third party conformity assessment[57] or are systems used in the sectors listed in Annex III AIA.[58]

If one asks 'which fundamental rights in particular are being threatened by these systems?' the answer would be 'everything under the Sun,' as the AIA confirms (in a Recital!) by referring, not only to rights protected under EU law, but also under International instruments like the UN Convention on the Rights of the Child.[59] Non-discrimina-

[45] Bathaee, Y.: Artificial intelligence opinion liability. Berkley Tech'y L. J. 35/1, 113–170 (2020), at 120.

[46] Pell S. K., Soghoian C.: Can You See Me Now: Toward Reasonable Standards for Law Enforcement Access to Location Data That Congress Could Enact. Berkley Tech'y L. J., 117–196 (2012) at 156.

[47] See Sect. 3 of this article.

[48] This example is based on case C-143/16, *Abercrombie & Fitch Italia Srl* [2017] EU: C:2017:566.

[49] CJEU, 26 October 2010, *Schmelz*, case C-97/09, [2010] EU: C:2010:632.

[50] Council Directive 2006/112 on the common system of value added tax, OJ L 347/2006, p. 1.

[51] Case C-97/09, *Schmelz* [2010] EU: C:2010:632, para. 33.

[52] *Id*., para. 75.

[53] Case C-97/09 *Schmelz* (Opinion) [2010] EU: C:2010:354.

[54] *Id*., para. 75.

[55] Pursuant to Article 21 ECFR (n. 19), the protected characteristics are sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation.

[56] AIA (n. 32), Art. 6(1)(a). The legislation referred to in this provision is listed in Annex I.

[57] *Id*., Art. 6(1)(b).

[58] The sectors concerned are biometric identification and categorization of natural persons; management and operation of critical infrastructure; education and vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private services and public services and benefits; law enforcement; predictive policing and migration, asylum and border control management.

[59] AIA (n. 32), Rec. 48.

tion is, of course, included in those and will be - as stated several times - this article's point of focus. Methodological parenthesis closed. On with the analysis.

2 'If you see something, say something.' the assumption of knowability in EU Non-discrimination law

"In the law of evidence 'ought' implies 'can.'"[60] This is the foundation of most, if not all modern systems of evidence. When a legal provision asks for specific evidence to be given, it assumes that that evidence is within the litigant's reach. And indeed, who better to establish harm than the person having suffered it?

The accessibility of relevant evidence determines the success with which claimants can prove standing. The systems of evidence in the Equality Directives are not particularly demanding in that regard. They define the proof of standing based on a *prima facie* evidence model originating in the - now-repealed - Directive 97/80.[61] Under that model, the victim carries the initial burden of proving the plausibility of their claim. If a court finds the *indicia* adduced to be sufficient, discrimination can be presumed, shifting the burden of proof on the defendant. This lightening (but not removal) of the victims' *onus*, is meant to contribute to the substantive effectiveness of EU non-discrimination law: the effective detection and sanctioning of discrimination demands that the victims' *prima facie* evidence be feasible.[62] Be that as it may, accessing the evidence needed to prove standing can be more or less arduous, depending on whether a claimant brings an allegation of direct or indirect discrimination.

Instances classified as 'direct discrimination' often involve unequal treatment straightforwardly based on one or several protected characteristics.[63] Sometimes, the facts speak for themselves,[64] like in *Firma Feryn.*[65] An employer's overt reluctance to recruit non-nationals revealed, without a doubt, direct discrimination based on nationality. In other cases, the difference of treatment was less overt but still unmistakenly based on a protected characteristic. The *Hey*[66] case e.g., deals with a French bank's practice to award bonuses to married employees, but not to registered partners. Although the conditions to receive those bonuses were not expressly dependent on the employees' sexual orientation, the Court nevertheless found there was direct discrimination because, at the time the national proceedings were launched, only persons of different sexes could marry, depriving homosexual employees of the ability to be eligible for the benefit concerned.[67]

In contrast, indirect discrimination is evidentially more challenging because the facts do not often 'scream' unfair unequal treatment. This type of discrimination is 'indirect' because it pertains to a measure which, while appearing neutral, works to the disadvantage of certain individuals or groups in comparison to individuals or groups in similar situations (unless that difference in treatment is justified by objective factors unrelated to the protected grounds).[68]

To be plausible, the evidence supporting an indirect discrimination claim should be minimally fact-correspondent, meaning that there should be verifiable facts (*indicia*) able to show that a specific group has been disadvantaged in comparison to another group. In other words, the availability and nature of the *indicia* directly impacts a court's ability to identify the relevant groups to be compared, in view of uncovering a disadvantage. The comparability test includes a consequentialist reasoning, consisting in the comparison between the outcomes of situations involving individuals or groups that deserve to receive equal legal treatment.[69] To avoid 'overfitting' - a case of everyone being comparable

[60] Cheng E. K., Pardo M. S.: Accuracy, optimality and the preponderance standard. L. Prob'y & Risk, 14/3, 193–212 (2015), at 198.

[61] Council Directive 97/80 on the burden of proof in cases of discrimination based on sex, OJ L 14/1998, p. 6 repealed by Directive 2006/54 (n. 24).

[62] Rec. 31 of Directive 2000/78 (n. 24): "the rules on the burden of proof must be adapted when there is a *prima facie* case of discrimination and, *for the principle of equal treatment to be applied effectively*, the burden of proof must shift back to the respondent when evidence of such discrimination is brought" (emphasis added).

[63] This observation is supported by Articles 2(1) (direct discrimination based on race); Directive 2000/43 (n. 24); Art. 2(2)(a) (direct discrimination based on religion or belief, disability, age or sexual orientation as regards employment and occupation) Directive 2000/78 (n. 24); Art. 2(1)(a) (for a generic definition of direct discrimination) and Art. 2(2) (for the grounds of discrimination being harassment, instruction to discriminate against persons on the grounds of sex and less favourable treatment of a woman related to pregnancy or maternity leave), Directive 2006/54 (n. 24).

[64] We allude to the *res ipsa loquitur* principle, typically seen as marking an evidentiary threshold of self-evidence of the disputed facts. Once that threshold is reached, no further evidence is needed for a fact or a set of facts to be proven conclusively. Hart and Honoré argued that, in causal scenarios, the application of this principle presupposes that some part of the causal process is known, but what is lacking is evidence on its connection with defendant's act or omission. Presumably, the implication here is that, while the items of evidence establish certain facts, the litigant must still demonstrate causation (i.e. an explanation on how that evidence proves the link between a cause and harm). See Hart H.L. A., Honoré T.: Causation in the Law. Oxford University Press (1985), at 419.

[65] Case C-54/07, *Firma Feryn* [2008] EU: C:2008:397.

[66] Case C-267/12, *Hey* [2013] EU: C:2013:823.

[67] *Id*., para. 44.

[68] See *inter alia* case C-25/02, *Rinke* [2003] EU: C:2003:435, para. 33.

[69] Case C-311/97 (Opinion), *Royal Bank of Scotland* [1998] EU: C:1998:557, para. 48.

to everyone else - the CJEU confirmed, namely in *Römer*,[70] that comparability should be carried out, "not in a global and abstract manner, but in a specific and concrete manner in the light of the benefit concerned."[71]

Bearing in mind the previously discussed *Hey* case,[72] the similarity between same-sex registered partnerships and heterosexual marriages presents an interesting case, from the standpoint of comparability. We could argue that the two are different conjugal statuses. The opposite argument is also possible, since it all depends on the specific ways in which a national law assimilates marriages and registered partnerships. Before *Hay*, the CJEU addressed the comparability between marriages and registered partnerships in *Römer*,[73] dealing with a measure by which, same-sex registered partners received lower supplementary retirement pensions than married pensioners. The CJEU tasked the national courts with determining if the difference in treatment was (indirectly) discriminatory and could be justified. However, it gave an interesting guideline on comparability. There would be discrimination based on sexual orientation - the Court argued - if the national law in the main proceedings reserved the institution of marriage to heterosexual couples (which, for the Member State concerned, may be a legitimate thing to do) but *assimilated* registered partnerships to married couples, in terms of their rights and duties.[74] It follows that the *locus* of comparison lies in how a public measure distributes specific advantages amongst groups - e.g. married couples and registered partnerships - that, for all intents and purposes, are legally treated as equal or equivalent. The same can be said for private measures, as the headscarf jurisprudence confirms. An employer seeking to preserve an image of religious neutrality in front of their clients can have a legitimate motive to ask their employees to refrain from exhibiting religious beliefs.[75] An undertaking's terms of employment that prohibit workers from manifesting their religious or philosophical beliefs through clothing, are not directly discriminatory, *provided they are applied in a general and undifferentiated way*.[76]

Indirect discrimination can generally be presumed when there are *facts to confirm* that it is relevant to compare and explain the conditions under which specific groups receive a specific benefit. The cited cases allow us to make two important points, in relation to AI.

First, all the mentioned cases deal with discrimination either suffered, or likely to be suffered by specific natural or legal persons. This remark contrasts the observation made in the Introduction,[77] relative to AI-related discrimination. In the cases mentioned in that connection, the courts were called to examine *risks of discrimination* that were hypothesized but not strongly supported by evidence. This approach to AI-related discrimination allows us to, second, contend that, if AI challenges the knowability of discrimination, it is because AI systems obstruct the access to the knowledge thereof.

Indeed, a point that weaves through the cited case law is that, in none of the examples given were the victims unaware of the unfavourable treatment they received. Like many other systems of evidence in national law and EU law, those in the Equality Directives assume that reasonably vigilant claimants can notice (and establish) an unfavourable treatment, compared to that of individuals or groups in similar situations. This assumption applies even in cases where discrimination presents itself, not as a *fait accompli*, but as a threat. For instance, a religious institution might consider the adherence to its beliefs a genuine requirement for its employees.[78] Hiring a non-believer would present the risk of violating that institution's ethos; a risk that should be proven as sufficiently "probable and substantial"[79] - *dixit* the CJEU.

No doubt, because of the assumed knowability of discrimination, none of the Equality Directives include a *specific procedural right* based on which claimants can request access to relevant facts. The assumption applies even in discrimination cases where the claimants clearly experienced difficulties in adducing the *prima facie* evidence. For instance, in *Kelly*,[80] the claimant - a man - applied for admission in a master's degree course and was rejected. He contested the decision, arguing discrimination on the grounds of sex, contending that he was more qualified than the least qualified admitted female candidate. When the dispute reached the CJEU, the Court was called to determine if the academic establishment concerned in the national proceedings was under a duty to disclose information. Indeed, for the claimant to argue that he was disadvantaged compared to admitted female applicants, he would need to have access to data on those applicants' qualifications and performance. The evidentiary conundrum the CJEU faced was obvious. On the one hand, full access to those data would allow the claimant to substantiate his claim. On the other hand, the

[70] Case C-147/08, *Römer* [2011] EU: C:2011:286.

[71] *Id*., para. 42.

[72] Case C-267/12, *Hey* (n. 86).

[73] Case C-147/08, *Römer* (n. 93).

[74] *Id*., para. 52.

[75] Case C-157/15, *Achbita* [2017] EU: C:2017:203.

[76] Case 344/20, *L.F.* [2022] EU: C:2022:774.

[77] See *supra*, Sect. 1 ('Introduction').

[78] Art. 4(2), Directive 2000/78 (n. 24).

[79] Case C-414/16, *Egenberger* [2018] EU: C:2018:257, para. 67.

[80] Case C-104/10, *Kelly* [2011] EU: C:2011:506.

defendant was justified in not allowing full disclosure of said data.

In its ruling, the CJEU treaded carefully, arguing that Article 4(1) of Directive 97/80[81] did not create a specific entitlement for a claimant to request, from the defendant, disclosure of information needed to prove standing. Of course, the absence of such an entitlement presented the risk of diminishing the Directive's substantive effectiveness: what if the claimant's suspicions were correct and there was, indeed, discrimination based on gender? In such instances, it cannot be excluded - the CJEU pursued - that "a refusal of disclosure by the defendant, in the context of establishing such facts, could risk compromising the achievement of the objective pursued by that directive and thus depriving that provision in particular of its effectiveness."[82] However, this issue was left for the national courts to address.

The *Meister* case[83] is often cited as a follow-up to *Kelly*, because it also addresses the issue of the claimants' inability to access the evidence necessary to substantiate their claim. *Meister* deals with an allegation of discrimination based on nationality made by a rejected job applicant. This raised the question of whether a recruiter is obliged, under EU law, to allow the access to information relative to the recruitment process. Here again, the CJEU argued that EU law creates no specific duty for recruiters to disclose information on how a recruitment was performed, concluding that - like in *Kelly* - it was for the referring court to determine if and how the absence of such disclosure would impact the effectiveness of the EU non-discrimination law.

Against the backdrop of the assumption of knowability of discrimination, the trouble with AI is that observing reality to pick up on unfairness is not always easy. 'Does it matter?' one might ask. After all, discrimination is inherently consequentialist. One should merely look at an AI's output to determine if it specific individuals or groups were impacted. If only it were that simple.

First, without even a basic knowledge of the functioning of a given system, a litigant would not know *who to compare themselves to*. Consider a credit-scoring scenario where the basis for differential treatment is the loan applicants' place of residence. It would take a closer inspection of the variable association to reveal that that criterion might coincide with ethnic background, resulting in a system's favouring of loan applicants who live in primarily white areas. If the bank allowed a rejected loan applicant to have access to the loans approved by the system, that applicant would arguably not immediately pick up on the unfair treatment because they would, in essence, *not know where to look for the basis for comparison*. Indeed, referring to the shortlist alone and without additional information, how would a non-expert loan applicant figure out that, to uncover a disadvantage, they would need to compare themselves to applicants with ethnic backgrounds different from their own?

Second, gaining an appropriate understanding of AI-related discrimination is further challenged by the fact that AI output is, in principle, a component of a human decision. Even if a bank used a credit scoring system, the loan approval is ultimately incumbent on the bank.[84] With the AIA in mind, we can reasonably assume that the system's deployer possesses the necessary knowledge of what the system can do (and has done), as well as the data and instructions fed into it, making it possible to explain why applicant X and not applicant Y was approved for a loan. Of course, no reasonable business or public administration would ever admit to being biased or to using a biased system, which implies that the knowability (and provability) of the bias cannot be assumed. The European Commission highlighted this difficulty as early as the White Paper on AI.[85]

Ideally, in AI-related disputes, either positive law or a court's jurisprudence would establish a *procedural right* allowing claimants to request disclosure of relevant information, necessary to prove standing and enable the access to a remedy. And indeed, rights to explainability and evidence disclosure have been mushrooming across the recently enacted and proposed EU legislation. The question is, however, whether those rights can be exercised in a pre-trial (evidence-constituting) phase or whether they open access to evidence once disputes are brought before courts (evidence-adducing phase). The following Sections offer some insight on these points.

[81] Article 4(1) Directive 97/80 (n. 24) states that "Member States shall take such measures as are necessary, in accordance with their national judicial systems, to ensure that, when persons who consider themselves wronged because the principle of equal treatment has not been applied to them establish, before a court or other competent authority, facts from which it may be presumed that there has been direct or indirect discrimination, it shall be for the respondent to prove that there has been no breach of the principle of equal treatment."

[82] Case C-104/10, 1, *Kelly* (n. 114), para. 34.

[83] Case C-415/10, *Meister* [2012] EU: C:2012:8.

[84] This was one of the observations made by the referring court in case C-634/21, *Schufa* [2023] EU: C:2023:957 (Request for preliminary ruling), para. 25: "even though, at least from a purely hypothetical point of view, third-party controllers can make their own decision as to whether and how to enter into a contractual relationship with the data subject, because an individual decision taken with the involvement of human beings is in principle still possible at that stage of the decision-making process, that decision is in practice determined by the score transmitted by credit agencies to such a considerable extent that the score penetrates through the decision of the third-party controller, so to speak."

[85] White Paper, Artificial Intelligence - A European approach to excellence and trust, COM(2020) 65 final, at 13.

3 The AIA, a *lex specialis* to the GDPR?

While the AIA's application does not affect the GDPR's,[86] we can argue that the latter laid down the normative foundation upon which the former took shape. For instance, deployers of high-risk systems can use the information provided under Article 13 AIA[87] to carry out a data protection impact assessment (DPIA), under the GDPR.[88] Fundamental rights impact assessment (FRIA) performed under the AIA should "complement" the DPIAs.[89] When deployers of high-risk systems carry out bias detection and data correction, they must comply with the GDPR's requirements.[90] From the perspective of governance, it is possible that data protection supervisory authorities under the GDPR assume the role of market surveillance authorities, under the AIA…[91] Given the degree of complementarity between the AIA and the GDPR, we might think that - personal data cases aside - the AIA serves as a lex specialis in the domain of high-risk AI.

However, it can also be argued that the systematic application of the AIA in cases involving high-risk systems is not a given. It depends on the legal basis of a claim dealing with (possibly biased) automated recruitment. Explanations in such a case could be sought on the grounds of both the GDPR and the AIA. In such circumstances, Article 86(3) AIA states that it will apply "to the extent" that the right to an explanation "is not otherwise provided under Union law." An on-the-surface reading of this provision would suggest that in cases like automated recruitment (possibly governed by the GDPR and the AIA), Article 22 GDPR is to be viewed as a *lex specialis* in relation to Article 86 AIA.

In this article, we will take the litigant's perspective and argue that, what might nudge them to frame their claim as a 'data protection' or a 'high-risk AI' issue is, presumably, the accessibility and relevance of the explanations they would receive, under each of the instruments discussed. From that perspective, the designation of the explanatory *lex specialis* would be largely determined by the type of explanation sought (**Sub-Sect. 3.1.**). An overview of the content of the explanations given based on Articles 22 GDPR and 86 AIA, allows us to contend that latter provides a more convenient gateway to knowledge of the relevant facts. However - as we will argue - that gateway is not unconditional but conditioned on the litigants' ability to meet possibly arduous evidentiary requirements (**Sub-Sect. 3.2.**).

[86] AIA (n. 32), Art. 2(7).

[87] *Id*., Art. 13, 'Transparency and provision of information to deployers.'

[88] *Id*., Art. 26(9).

[89] *Id*., Art. 27(4).

[90] *Id*., Art. 10(5).

[91] *Id*., Art. 74(8).

3.1 The type of explanation sought - key in designating the *lex specialis ad explicandum*

When discrimination is associated with an AI system, proof of standing typically requires access to knowledge about that system's *causal power*[92] that is, information able to uncover the decisional process (variable association) having resulted in a discrimination occurring. The challenge is to determine if a *relevant* explanation (i.e. one that can actually be useful to the launching of proceedings) should include information about a system's inner workings, or a human agent's reasons to rely on that system's output. Article 22 GDPR opens a pathway to the former (**A**); Article 86 AIA yields explanations about the latter (**B**).

3.1.1 'GDPR explanations' yielding knowledge about the rationale of automated output

Before the AIA was even submitted as a legislative proposal, scholarship[93] had already reflected on a legal basis in EU law that could be seen as a gateway to knowledge of facts in cases of AI-related harm, like discrimination. In standard liability law and doctrine, when a litigant is called to prove standing, they are required to give evidence on two points: the harm suffered (or likely to be suffered[94]) and causality i.e. the 'necessary and sufficient'[95] cause of that harm. AI opacity has been interpreted as obstructing knowledge on

[92] *Causal conditionality* ties with the 'necessary and sufficient cause' principle in liability doctrine which asks us to perform a counterfactual reasoning that is *contextual* and *exclusionary*. It asks the question of whether harm (like discrimination) would have been suffered had the environment (or context) been different (e.g. an AI would not have been used at all). It is exclusionary because it allows us to identify correlational structures that are not causal, that is, do not causally explain the occurrence of harm. See, namely, Mackie, J.L.: Causes and Conditions. Am. Ph. Quart'y, pp. 245–264 (1965), at 250.

[93] See Grozdanovski, L.: In Search For Effectiveness and Fairness in Proving Algorithmic Discrimination in EU law. CMLRev., 58/1, 99–136 (2021).

[94] For an example of a case addressing the 'risk of harm' In the field of non-discrimination, case C-414/16 *Egenberger* (n. 113), para. 67: "as regards the 'justified' nature of the requirement, that term implies not only that compliance with the criteria in Article 4(2) of Directive 2000/78 (n. 24), can be reviewed by a national court, but also that *the church or organisation imposing the requirement is obliged to show*, *in the light of the factual circumstances of the case*, *that the supposed risk of causing harm to its ethos or to its right of autonomy is probable and substantial*, so that imposing such a requirement is indeed necessary" (emphasis added).

[95] Necessity and sufficiency are the traditional criteria against which causality is assessed. An event is causal to another event if the latter had not at all occurred, without the former. See, inter alia, Hart, H.L.A., Honoré, T., Causation in the Law, *cit. supra*, at 82 seq.; Moore M. S., Causation and Responsibility: An Essay in Law, Morals, and Metaphysics, Oxford University Press (2009), at 33 seq.

both points, possibly affecting the litigants' right to access remedies and courts.

To overcome this hurdle and before the AIA was enacted, part of scholarship suggested that, in cases of so-called algorithmic discrimination, Article 22 GDPR be interpreted as generating a procedural right to request evidence and explanations 'about the machine' that would allow litigants to meet the proof-of-standing requirements in the Equality Directives. With the AIA now in force, we could argue that a reading of Article 22 GDPR *cum* EU Non-Discrimination law no longer has priority, at least in cases where high-risk systems are involved. Indeed, Article 86 AIA now creates a duty to give explanations when such systems play a role in shaping human-driven decisional processes. That article thus fills a normative *lacuna* that the GDPR was called to fill, at a time when the EU's AI regulation was still in its inception.

Complications arise, however, in cases like AI-assisted recruitment, where a high-risk AI system is used to processes *personal* data. Such a case would fall in the scope of application of either Article 22 GDPR *and* Article 86 AIA, making the choice of a *lex specialis ad explicandum* (special law to an explanation) more challenging. Presumably, that choice would depend on the *type of explanation* a claimant would require, as well as on the defendant's quality as data controller, within the meaning of the GDPR[96] or provider or deployer, within the meaning of the AIA.

Regarding the types of explanations, the components of the *explananda* that is, 'that which ought to be explained,' under Article 22 GDPR have been somewhat shrouded in mystery.[97] Recital 71 GDPR gives a few hints by mentioning that suitable safeguards "should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of *the decision reached after* such assessment and to challenge the decision."[98] In outlining those 'suitable safeguards,' said Recital emphasizes the controller's compliance with the GDPR's lawfulness conditions, namely fairness and transparency,[99] accuracy of the data processed[100] and the prevention of risks of fundamental rights violations, in particular discrimination.[101] While Recital 71 GDPR essentially asks for explanations on the system's process having led to output that "produces legal effects concerning him or her or similarly significantly"[102] a data subject, it does not mention *how* that subject can practically access explanations. Article 22 GDPR, however, does.

Let us imagine an instance of automated recruitment where the *probandum* (the fact for which evidence is sought) is gender discrimination: a recruitment AI produces an all-male shortlist of applicants. Let us assume that a rejected female applicant would have agreed to her resumé being automatically processed, thus meeting the lawfulness (consent) requirements of Article 22(2) GDPR.[103] Let us also assume that our applicant had no access to the candidate shortlist, but she is generally suspicious of automated data processing. To verify if her 'hunch' is correct, she might consider bringing an action based on Directive 2000/78[104] which would, naturally, require her to prove standing, i.e. give evidence that could minimally but plausibly justify her hunch. To that end, she could request an explanation based on Article 22 GDPR and come to realise that a gatekeeper stands before the explanatory gateway: the personal data processing decision should be *fully automated* and legally or significantly affect her. We will elaborate on the 'legal or significant effect' requirement in connection with Article 86 AIA, which also makes the access to explanations conditional on that criterion. Regarding Article 22 GDPR, we will focus more on the somewhat sibylline requirement of 'full automation.' Two readings can be suggested in that regard.

[96] See GDPR (n. 38), Article 4(7).

[97] Bibal, A., Lognoul, M., de Streel, A., Frénay, B.: Legal requirements on explainability in machine learning. AI & L. 29, 149–169 (2021), at 152: "the type of explanation to be given by the processors of personal data is not clear."

[98] Emphasis added.

[99] GDPR (n. 38), Rec. 71: "in order to ensure *fair and transparent processing* in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed (…)" (emphasis added).

[100] *Ibid*: "(…) the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised (…)."

[101] *Ibid*: "(…) secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject, and prevent, 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 processing that results in measures having such an effect."

[102] *Ibid*.

[103] The principle, announced in Article 22(1) GDPR (n. 38) is that a data subject has the right not to have their personal data automatically processed. However, para. 2 mentions three exceptions to this principle: when such processing is necessary for the entering into or performance of a contract with the data controller (a); when it is authorised by the Union's or a Member State's law to which the controller is subject (b) and when the data subject has given their explicit consent. A recruitment scenario may be interpreted as a stage necessary for the entering into a contract within the meaning of Article 22(2) (b) GDPR (n. 38), although there is no court practice to confirm this. If, however, a job applicant gave their consent to their résumé being automatically processed, the exception listed in Article 33(2)(c) GDPR (n. 38) would, in any case, apply, warranting the delivery of the right to an explanation under the conditions mentioned in Article 22(3) GDPR (n. 38).

[104] Directive 2000/78 establishing a general framework for equal treatment in employment and occupation (n. 24).

First, full automation can be defined *instrumentally* as pertaining to output arrived at through automated means. In that sense, an AI-generated shortlist of candidates resulting from an automatically performed data processing would fit the bill. Second, a decision would be 'fully automated' if it was reached through automated means *and* was automatically relied upon by the controller. The claimant's burden would then be heavier because they would need to establish the defendant's automatic adherence to the output. Though procedurally more tedious, this second reading was suggested by the Article 29 Working Party (A29 WP) who interpreted 'full automation' as implying the total absence of human involvement in automated data processing. Their Guidelines state that "to qualify as human involvement, the controller must ensure that any oversight of the decision is meaningful rather than a token gesture;"[105] 'meaningful' oversight being performed by someone with the authority and competence to change the decision.[106]

Against this backdrop, a person suspecting discrimination and seeking an Article-22 explanation would need to prove that the human oversight was reduced to a 'token gesture,' revealing adherence to the system's output, without prior discernment of that output's desirability and/or accuracy. Of course, what can be considered a 'token gesture' is up for debate, the million-dollar question being "when does 'nominal' human involvement become no involvement?"[107] As Kaminski rightly stressed, it might mean that even the slightest human involvement, like rubber-stamping, would upset the fullness of the automation of a specific data processing.[108] The conceptual vagueness of the 'full automation' requirement in Article 22 GDPR has repercussions on the explainee's practical efforts to meet that requirement. They could, presumably, ask the recruiter to explain if they used a system's shortlist with no prior checks, although no reasonable recruiter would ever admit this…

For argument's sake, let us assume that a recruitment decision is uncontroversially automated (under one of the previously flagged readings), triggering the explanatory mechanism of Article 22 GDPR. The *content* of that explanation still remains an open issue. In the *Schufa* case,[109] the CJEU did not give a particularly enlightening guideline on what should be included in the *explanandum*. In the *Dun* case, AG De La Tour focused more on the explanation's *format* (meaningfulness, clarity etc.) rather than on the types of information that should be included in explanations given based on Article 22 GDPR.[110] In its ruling, the CJEU aligned with the AG's Opinion.[111]

The Explanatory Report to the Council of Europe's (CoE) Framework Convention includes a slightly clearer specification of Article 14 ('Remedies'), stating that "information-related measures should be context-appropriate, sufficiently clear and meaningful, and *critically provide a person concerned with an effective ability to use the information in question to exercise their rights in the proceedings* in respect to the relevant decisions *affecting their human rights*." In other words, the 'meaningfulness' of an explanation is not a function of formatting (clarity, etc.) but one of substantive relevance (i.e. utility of the information shared for the purpose of launching proceedings).

In scholarship, we may, of course, refer to an article that has, by now, entered the GDPR's doctrinal folklore *i.e.* the study penned by Wachter *et al.*'s[112] Distancing themselves from the formatting explanations should have to be 'meaningful,'[113] the authors - refreshingly - focused on what those explanations should be about, so that data subjects can receive information that is 'meaningful' because it provides the 'understanding why'[114] that is, the understanding of how harm occurred as the result of a system's, not a human's, casual power (*i.e.* ability to cause harm). They argued that Article-22 explanations ought to provide information on two points of the automated decisional process: *ex ante* and *ex post*. The former yields understanding of a system's functionalities and intended purpose. The latter is meant to reveal the rationale underlying the system's outcome.[115] The US Blueprint for an AI Bill of Rights[116] aligns with this

[105] Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purpose of Regulation 2016/679 (3 October 2017, last revised on 6 February 2018), at 21.

[106] *Ibid*.

[107] Edwards, L., Veale, M.: Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for. Duke L. & Tech'y Rev., 16, 18–84 (2017–2018), at 45.

[108] Kaminski, M. E.: The Right to Explanation, Explained. Berkley L. Tech'y L. 34/1, 189–218 (2019), at 197.

[109] Case C-634/21, *Schufa* (n. 119).

[110] Case C-203/22 (Opinion), *Dun* [2024] EU: C:2024:745, para. 66.

[111] Case C-203/22, *Dun* [2025] EU: C:2025:117, para. 66.

[112] Wachter, S., Floridi, L, Mittelstadt, B.: Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. Int'l D. Pr'y L. 7/2, 76–99 (2017).

[113] This has been noted in Casey, B., Farhangi, A., Vogl, R., Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise. Berkley Tech'y L. J., 34/1, 143–188 (2019), at 159.

[114] The understanding-why as a consequence of explainability is also a consequence of transparency, as mentioned in Edwards, L., Veale, M.: Slave to the Algorithm: Why a Right to an Explanation Is Probably Not the Remedy You Are Looking for (n. 142), at 41.

[115] For a comment on ad hoc and *post hoc* explanations under Article 22 GDPR, see Grozdanovski, L.: In Search For Effectiveness and Fairness in Proving Algorithmic Discrimination in EU law (n. 128), at 121 seq.

[116] Blueprint for an AI Bill of Rights. Making Automated Systems Work For The American People (2022), available at: https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/ (last accessed on 6 May 2025), at 40.

interpretation. 'Notice and Explanation,' within the meaning of this document, include a person's ability to know that an automated system had been used and understand how and why it had yielded outcomes impacting the person concerned. It follows that the substantive requirement for the *explanandum* (*i.e*. the object of the explanation) is the system's role in affecting a person's situation. The formal requirement is that the explanation be clear and meaningful (primarily, from the explainee's point of view).[117]

Following Wachter et al.'s interpretation, the following question can be raised: how would an explanation, as they define it, enhance the knowability of discrimination and usefully support the launching of proceedings under the Equality Directives? If we are optimistic, 'Wacherian' explanations could provide sufficient information to support a person's burden to prove standing. For example, in providing the *ex ante* explanation, the recruiter could highlight that the system was trained on historical data in a sector notorious for gendered recruitment. That information could encourage a rejected female applicant to gather statistical data (expertise) confirming the consistency of those practices. As the CJEU's case law shows,[118] statistical evidence can be enough for discrimination to be presumed. However, we do not yet have case law that can clarify the elements of proof a claimant could bring before a court, in cases where they would rely on Article 22 GDPR to meet the *prima facie* burden of proving standing, under the Equality Directives.

May it suffice to stress the following: if pursuant to Article 22 GDPR, an explainee can establish full automation if they prove two facts (use of automated means for data processing *and* automatic human reliance), we can legitimately ask ourselves if that burden is at all feasible for the reasons outlined above. It remains that, absent the information Article 22 GDPR promises to open access to, claimants would arguably not be able access the pre-trial knowledge they could rely on to prove standing under the Equality Directives. A few years ago, we could have interpreted this as an important gap in the system of effective judicial protection[119] but today there is hope, since we have the AIA to step in and save the day. Enter Article 86 AIA.

3.1.1.1 'AIA ***Explanation' yielding knowledge about human reliance on automated output.*** At first glance, Article 86 AIA offers an attractive alternative to Article 22 GDPR (access-to-explanation wise), because its applicability appears to be conditioned on less stringent requirements. First, Article 86-explanations are not due only in contexts of full automation, but whenever high-risk AI systems are deployed. Second and more importantly, the object of those explanations (*explanandum*) seems more realistic than that of explanations given under Article 22 GDPR. The deployer is required to provide "*clear and meaningful explanations* of the *role of the AI system in the decision-making procedure* and the *main elements of the decision taken*."[120]

Article 86 AIA does not mention - at least not explicitly - the sharing of information on the system's functionalities, or on the rationale underlying its output. It rather pertains to 'the role of the AI system in the decision-making procedure,' the implication being that a high-risk system was, in any case, an *auxiliary* to a decision primarily incumbent on the deployer. Article 86 AIA essentially invites a deployer to give *reasons* (about their reliance on an AI's output), yielding understanding about the rationale behind *their* decision and the degree to which the deployed system shaped it. To give such an explanation in a 'clear and meaningful way,' the deployer would presumably be called to explain their motives for using a specific system (in connection with its intended purpose), the output produced, the deployer's assessment of that output's trustworthiness/accuracy and the motives for their choice to rely on it (or not).[121]

Compared to Article 22 GDPR (in its Wachterian interpretation), Article 86 AIA appears to, indeed, create more realistic explanatory duties because the explanations it refers to ultimately have, as objects, *human decisions*, which presents a remarkable advantage from the standpoint of knowability. When a human agent is called to explain their (AI-assisted) conduct, the rationale underlying their decision is reasonably within their ability to communicate it. In contrast, under Article 22 GDPR - and, again, depending on how we interpret 'full automation' - the human agent would be called to primarily explain the rationale of an AI's decision-making process. If that process is opaque (*ergo* unknowable and inexplainable), the giving of such explanations would, of course, be more challenging. To refer back to the recruitment scenario: it would be one thing to ask a recruiter to explain why *they* considered a system's output to be correct and worth complying with (Article 86 AIA) and another thing to ask them to explain if and why the system gave a possibly biased output (perhaps, without the recruiter having intended it).

[117] *Ibid*.

[118] See the case law cited in Sect. 2 of this article.

[119] Grozdanovski, L.: In Search For Effectiveness and Fairness in Proving Algorithmic Discrimination in EU law. (n. 128).

[120] AIA (n. 32), Art. 86(1) (emphasis added).

[121] Explanation pertaining to an AI-assisted human decision seems to be a type of explanation due within the meaning US Blueprint for an AI Bill of Rights, which mentions that a person should know how and why an outcome impacting them was determined by an automated system, "including *when the automated system is not the sole input determining the outcome*" (emphasis added). See Blueprint for an AI Bill of Rights. Making Automated Systems Work For The American People (n. 151), at 40.

It follows that, to launch actions on the grounds of the Equality Directives, Article 86 AIA may act as a *lex specialis* in relation to Article 22 GDPR. All things considered, a victim of discrimination resulting from personal data processing *and* performed by a high-risk system may - even strategically - choose to rely on Article 86 AIA to gain access to the knowledge they may rely on to bring a discrimination claim before a court. A noteworthy limitation, however, is the quality of the defendant: under Article 86 AIA, they should be a system's *deployer*. In the Amazon automated recruitment case, dealing with an HR system ultimately shown to be biased against women, Amazon would fit the bill as it was both a provider and deployer of the recruitment system used.[122] In actions of discrimination involving only a provider as a defendant, Article 86 AIA would presumably not apply, leaving the alleged victim with the possibility to rely on Article 22 GDPR, provided they show that their personal data processing was 'fully automated' (whatever that means).

With that said, the attractiveness of Article 86 AIA as a more convenient gateway to explanations might be dimmed by some of the requirements that frame its application, as discussed in the following sub-Section.

3.2 Accessibility of explanations afforded under Article 86 AIA as *lex specialis ad explicandum*

3.2.1 The unknowability of legal/significant effects of AI-assisted decisions

Article 86 is featured in Section IV (Remedies) of Chapter IX ('Post-Market Monitoring, Information Sharing, Market Surveillance') AIA. It states, in para. 1, that any affected person subject to a decision taken by a deployer on the basis of the output from a high-risk system listed in Annex III (with the exception of systems listed under point 2 thereof[123]) 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 on the role of the AI system in the decision-making procedure and the main elements of the decision taken." This right - paragraph 2 pursues - will not apply in the presence of restrictions contained in Union or national law.

Article 86 AIA is, without a doubt, a praise-worthy provision. The praise is warranted by the fact that, absent contrary indication, its application is not confined to adjudication. The right to an explanation seems to be a *residual entitlement*, afforded to addressees of (high-risk) AI output, regardless of whether harm was suffered or any kind of proceeding launched. The Achilles' heel of Article 86 AIA is that it asks addressees to know *how* exactly they were affected by an AI-supported decision, in order to claim access to an explanation. As will be argued further,[124] that knowledge may not be within reach. Yet, based on the wording of the Article considered, explanations are warranted if three conditions are met: (1) there is a *human* decision based on an AI output (a point already discussed), (2) that decision produces a legal or a similarly significant effect on the addressee, (3) the significance of that effect is measured by the extent to which it impacts a person's health, safety and fundamental rights.

The requirement - sibylline and GDPR-inspired[125] - we will focus on is precisely the 'legal' or 'similarly significant effect,' especially regarding fundamental rights violations. Article 14 of the CoE Framework Convention[126] also includes the expression 'significantly affect human rights' which, according to the Explanatory Report,[127] is a "threshold requirement"[128] asking the Parties to examine if, given the applicable international and domestic human rights law, as well as the "relevant circumstances" in relation to a given AI system, the latter can have a 'significant effect' or 'significant impact' on human rights.[129]

Proving such impacts may, however, be trickier than proving the 'effects' on health and safety. The Arkansas

[122] Jeffrey Dustin, "Insight - Amazon scraps secret AI recruiting tool that showed bias against women" (11 October 2018), available on: https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/ (last accessed on 6 May 2025).

[123] Point 2 Annex III AIA cit. supra, concerns critical infrastructure: AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity.

[124] See *infra*, pt B.

[125] We allude here to the Article 29 Working Party's (A29 WP) Guidelines on Automated individual decision-making and profiling where it is stated that an automated decision should produce effects that "must be sufficiently great or important to be worthy of attention." Significant effects can be the result of say automatic refusal of an online credit application or e-recruitment practices, not involving human intervention. Such decisions should present the risk of significantly affecting "the circumstances, behavior or choices of the individuals concerned; have a prolonged or permanent impact on the data subject or at its most extreme, lead to the exclusion or discrimination of individuals." See Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purpose of Regulation 2016/679 (n. 140), at 21.

[126] CoE Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, CoE Treaty Series, No 225.

[127] Explanatory Report to the CoE Framework Convention on AI and Human Rights, Democracy and the Rule of Law (n. 161).

[128] *Id*., para. 101.

[129] *Ibid*.

Supreme Court's *Ledger Wood* ruling[130] shows why this would be the case. A Medicaid program used a system to issue health charts for low-income patients suffering severe health conditions. The system executed this task by collecting patients' health data, based on which it classified them into multiple groups, scheduling nurses' visits, prescribed treatments, etc. It is important to mention that the deployer gave specific instructions for nurses to fully (that is, automatically) comply with the system's classification and treatments. Needless to say, the consequences were catastrophic and importantly, tangible and verifiable. Many patients were left without food, care and medication, which led to a significant worsening of their conditions. The manifest, beyond-doubt nature of the harm was such that Arkansas Supreme Court qualified is as 'irreparable' that is, "harm that cannot be adequately compensated by money damages or redressed in a court of law."[131] If *Ledger Wood* was adjudicated on the grounds of EU law, the application of Article 86 AIA would presumably be uncontroversial both in- and out-of-court, as the facts of the case clearly include a human decision to entirely rely on AI output, the effects of which on people's health were clearly significant because they could be proven beyond the shadow of a doubt.

The effects of AI on fundamental rights can be, and have often been, much subtler. To use our recruitment example: access to labour is classified as a high-risk sector in Annex III AIA and has already given rise to much scholarly debate.[132] However, how would a rejected job applicant *even suspect* that their personal situation was 'significantly affected' by, say, a (discriminatory) recruitment system? The same question can be raised across the high-risk sectors in instances where high-risk AI is used to approve, or not the benefit from a specific right. Take systems used to determine admission to educational and vocational training institutions.[133] Is the refusal to gain such access a 'significant effect' on a person's rights? Admission - AI-assisted or not - in higher-education institutions has always been selective, invariably resulting in the exclusion of certain applicants. Unless these were aware of a University's long-standing preferential admission of, say, white students, they would have no realistic way of arguing that their right to non-preferential, merit-based treatment was 'significantly affected.' Similar observations can be made in connection with the access to essential public services (like unemployment benefits)[134] and certainly in migration, asylum and border control management, regarding, say, the examination of applications for asylum, visa and residence permits.[135]

Returning to the recruitment example: both under the Equality Directives[136] and - it turns out - under Article 86 AIA, if a job applicant felt they were discriminated against, they would be expected to have at least a founded (that is, fact-based) suspicion of that discrimination and be able to show that the treatment they received was unfair, compared to that of a group (men, nationals, entry-level workers…) in a similar situation. Depending on how we interpret Article 86 AIA (which we will discuss below), rather than being a gateway to such evidence, it, in fact, requires that that evidence be *already* in the litigant's possession when they make the claim for an explanation.

Suppose the following: a recruiter has conducted background checks on a job applicant. They accessed pictures taken at political rallies published on the applicant's Instagram account. Eventually, the applicant is no longer considered for the position. For all we know (as imaginary third-party observers of that scenario), the candidate's exclusion may have been based on job-relevant factors like the lack of specific skills or work experience. However, if the candidate in question was somehow aware of the background check (which, in reality, is not likely), they could of course, spin a narrative of discrimination based on their political beliefs. But - again - to do so, they would need to have the information that their social media were combed through.

Our point is the following: in cases of - shall we call it - unostentatious discrimination (which indirect discrimination is, by definition), the addressees of AI-assisted human decisions, as defined in Article 86 AIA interpreted literally, would need to somehow acquire knowledge of those decisions' effects before receiving the explanations that Article promises to deliver. The implication here is that the AI's effects should be characterised by a level of salience *ergo* verifiability.

In a way, this defeats the purpose of Article 86 AIA: what is the point of seeking an explanation about specific facts if the explainee *already* has knowledge of those facts? Take a scenario similar to the above-mentioned *Ledger Wood* case: when the harm is salient (e.g. deterioration of health) and

[130] Supreme Court of Arkansas, 9 November 2017 (Opinion Delivered - Appeal from the Pulaski County Circuit Court, N° 60 CV-17-442), *Arkansas Department of Human Services v. Bradley Ledger Wood et al.*, No. CV-17-183.

[131] *Id*., at 9.

[132] See Dustin, J.: Insight - Amazon scraps secret AI recruiting tool that showed bias against women. *Cit. supra*. See also Köchling, A., Wehner, M. C.: Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. B. Res. 13, 795–848 (2020), at 796.

[133] AIA (n. 32), Annex III, point 3(a).

[134] *Id*., point 5.

[135] *Id*., point 7, (c). In this regard, see Palmiotto, F.: When is a decision automated? A taxonomy for a fundamental rights analysis. German L. Rev. (2024), 25/2, p. 210–236.

[136] See our developments in Sect. 2 of this article on the *prima facie* evidence in discrimination cases.

causally linked with an AI, what additional (relevant) information could the deployer possibly provide in their Article-86 explanation? The victim of such harm would already have enough elements of fact to seek compensation, without receiving such an explanation. In a discrimination scenario, if the victim is already able to gather *indicia* showing they were discriminated against, they could launch an action directly based on the Equality Directives without bothering beforehand to seek an explanation based on Article 86 AIA. And why would they? If they met the *prima facie* burden in EU Non-Discrimination law, there would be no need - at least as far as proof of standing is concerned - to learn more about how a deployer used an AI system, as required by Article 86 AIA.

In the absence of consolidated administrative and court practice relative to Article 86 AIA, we can but prospect on how that provision should be read regarding the knowability of the 'significant effects' on a person's fundamental rights. We contend that two readings are possible in that connection.

3.2.2 *Two solutions to the unknowability of legal/significant effect of AI-assisted decisions*

3.2.2.1 First solution: the right to an explanation is unconditional, First, we could argue in favour of a synonymy between the deployment of a high-risk AI and its 'significant' effect on a person's rights: the *mere use* of a system would generate a presumption of significant impact.

In a recruitment scenario, the argument would roughly go as follows: (1) a claimant would consent to a recruiter using an AI system to short-list job applicants, (2) familiar with the type of harm (e.g. gender discrimination) typically associated with that system, the claimant could argue that the likelihood (i.e. the 'high' risk) of them suffering such harm is enough to meet the 'significant effect' requirement; (3) seeking to verify if the residual risk had indeed materialised in their case, the claimant would consider themselves entitled to an explanation, pursuant to Article 86 AIA. In other words, a person would rely on an AI's 'bad rep' to request an explanation, which, as an approach, is neither surprising, nor new, especially in litigation dealing with proof of risks. In that litigation, expertise and statistical data are frequently used as evidence to show, amongst other things, the *consistency* of a given threat: a system reputed to discriminate in the past is likely to discriminate in the future (…because, according to the available statistics, discrimination is how that system has the habit to 'mess up').

However intuitive, that reasoning is fallacious and - if we push the irony - symptomatic of intellectual idleness: to posit that 'what was true before, will hold true moving forward' is to assume the unchangeability of AI technologies. This, of course, is not epistemically tenable. Be that as it may, the AIA seems to be permissive to consistency-based reasoning simply because we still lack enough practice and empirical data that would allow us to confidently map out the *archetypal forms of harm* in reference to which, plausible consistency arguments could be made. For instance, Article 9(2) AIA asks providers to implement a risk-management system which will include the identification and analysis of "the *known and reasonably foreseeable risks*" that a high-risk system can pose to health, safety or fundamental rights.[137] That assessment should be coupled with "the estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose, and under conditions of reasonably foreseeable misuse.[138] In our view, these two assessments are the two sides of the consistency coin: only known harm can - because it is 'known' - inform a provider of what can be reasonably (as opposed to hypothetically) foreseeable harm and misuse. For instance, in cases where a recruitment system relied on a recruiter's historical data, it was known to discriminate based on gender (think Amazon[139]). Depending on the recruiter's past hiring trends, they could probably also discriminate based on ethnic background (reasonably foreseeable). In all imaginable scenarios, the *genus* of the materialised and hypothesised harms associated with a recruitment system is discrimination, as opposed to, say, a broken leg.

Consistency is also useful in the field of AI liability. Article 10(3) of the revised Product Liability Directive[140] states that the causal link between defectiveness and harm shall be presumed "when it has been established that the product is defective and that the damage suffered is of a kind typically consistent with the defect in question." Causation is not presumed *ex officio*, but when the claimant proves two relevant facts: defectiveness and harm 'typically consistent' with the established defect.

Following this idea of 'typical consistency,' a rejected job applicant could argue that the harm recruitment systems 'typically' cause is discrimination (based on gender, ethnic background etc.). In doing so, they would seek evidence of harm where it is reasonably accessible, relying on general knowledge about the AI as a stand-in for what should be

[137] Art. 9(2)(a) AIA (n. 32) (emphasis added).

[138] *Id*., Art. 9(2)(b).

[139] See Dustin, J.: Insight - Amazon scraps secret AI recruiting tool that showed bias against women. *Cit. supra*.

[140] Directive 2024/2853 (n. 35).

known in its specific context of use.[141] The *Pickett* case[142] for instance, dealt with a DNA-testing AI used to identify a person charged with committing a crime. Conclusive evidence of the system's accuracy was practically inaccessible, given that the reverse-engineering of its output was said to take 8,5 years.[143] As an imperfect but within-reach substitute for that evidence, the court turned to *general* expert opinion on the system's performance which confirmed that the system generally but not invariably gave accurate output.

Like the relevant EU law provisions, the CoE Framework Convention[144] also adheres to the idea of stability and consistency of the causal relations between specific biases and AI systems. The Explanatory Report to the Framework Convention[145] states, in connection to Article 10 ('Equality and Non-Discrimination'), that the data on biases documented in the past essentially draws a roadmap for uncovering biases moving forward. In other words, the available data show where *it is empirically relevant*[146] to look for (risks of) biases, while the Parties to the Framework Convention are encouraged to "adopt new or maintain existing measures aimed at overcoming structural and historical inequalities, to the extent permitted by domestic and international human rights obligations."[147]

The CoE Framework Convention places much faith in the compliance with the law: documented past occurrences of AI biases inform the (possible updates on) future compliance with fundamental rights obligations, assuming that more rigorous compliance would somehow reduce or prevent biases. A similar view is upheld in the definition of 'sensitive domains' in the US Bill of Rights in the era of AI.[148] Those are areas where activities conducted "can cause material harms, including significant adverse effects on human rights such as autonomy and dignity, as well as civil liberties and civil rights." Again, the past informs us where to look for rights violations moving forward. These are instances "that have historically been singled out as deserving of enhanced data protections or where such enhanced protections are reasonably expected by the public (and) include, but are not limited to, health, family planning and care, employment, education, criminal justice, and personal finance."[149]

This 'past harm informs future harm' approach to AI-related risks is understandable, debatable but will not be further deconstructed here. Our detour on consistency allows us to better support the above-mentioned presumption of *ex officio* explainability, under Article 86 AIA: a person could claim an explanation if they were affected by the output of a system with a longstanding association to a specific type of harm. In this context, it is tempting to interpret Article 86 AIA as generating an automatically applicable entitlement to an explanation. As a matter of common sense (and a basic fairness drive), the receiving of explanations should not be conditioned on stringent requirements of proving that the explainer's situation was significantly altered (whatever 'significantly' may mean). We might defend the view that said Article's fairness-upholding purpose is *verificationist* as opposed to *demonstrative*, as it seeks to open access to information that can allow a person to verify (not yet prove, strictly speaking) if a risk of harm had, indeed, materialised.

The downside of this interpretation would be the cultivation of a perpetual state of techno-pessimism,[150] and the opening of the floodgates of explanation requests. Presumably, there is no AI under the sun with a clean (error-free) track record, suggesting that there will always be a risk to refer to, as a reason to claim an explanation. To avoid a possible inflation of Article-86 explanations, we could argue that the 'significant effect' criterion should be viewed not as a presumption but as a burden. This is the second reading of the Article under discussion.

3.2.2.2 Second solution: the right to an explanation is conditional on proof, Considering our previous arguments, we might advise a more restrictive reading of Article 86 AIA by suggesting some preliminary checks that competent authorities would perform with the view of distinguishing cases where explanations are justifiably warranted from cases where they are not. In procedural terms, this would mean that for the right to an explanation to be exercised, the 'significant effect' on the explainee's fundamental rights should

[141] For a comment on the evidence adduced in the *Pickett* case, see Grozdanovski, L.: The Explanations One Needs for the Explanations One Gives. The Necessity of Explainable AI (XAI) for Causal Explanations of AI-related harm - Deconstructing the 'Refuge of Ignorance' in the EU's AI Liability Regulation, (n. 42), at 204 seq.

[142] Superior Court of New Jersey (Appellate Court), 2 February 2021, *State of New Jersey v. Corey Pickett*, N° A-4207-19 T4.

[143] *Id*., at 17.

[144] CoE Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (n. 161).

[145] Explanatory Report to the CoE Framework Convention on AI and Human Rights, Democracy and the Rule of Law (n. 162).

[146] The Explanatory Report states that Article 10 of the CoE Framework Convention asks that "the required approach under this Article should not stop at simply requiring that a person not be treated 'without objective and reasonable justification' based on one or more protected characteristics that they possess in relevant matters of a protected sector." See Explanatory Report to the CoE Framework Convention on AI and Human Rights, Democracy and the Rule of Law (n. 162), para. 77.

[147] *Ibid*.

[148] Blueprint for a n AI Bill of Rights. Making Automated Systems Work For The American People (n. 151).

[149] *Id*. at 11.

[150] For an outline of techno-pessimism, see, namely, Turner, J.: Robot Rules. Regulating Artificial Intelligence. Palgrave (2019), at 32–33.

be provable and proven. There are, of course, advantages and disadvantages to this.

The advantage is the alignment with expert views expressed in relation to Article 22 GDPR (automated decision-making). The A29 WP Guidelines mentioned that automated decisions should have the potential to "significantly affect the circumstances, behaviour or choices of the individual concerned; have a prolonged or permanent impact on the data subject or at its most extreme, lead to the exclusion or discrimination of individuals."[151] Assuming that these Guidelines should inspire the interpretation and application of Article 86 AIA, we could argue that an AI output's effect on a person's situation is 'significant' when that person's rights have *plausibly*, *likely* or - in a perfect world - *uncontroversially* been affected. This approach to Article 86 AIA also aligns with the systems of evidence in the so-called Equality Directives under which - as previously mentioned[152] - discrimination is presumed not because a claimant said so, but because they can present initial evidence in support of their claim.[153]

Comparability - as already outlined in connection with indirect discrimination[154] - can play an important role in picking up hints on tangible, as opposed to imagined discriminations. Take the (in)famous Dutch case of - as Amnesty International put it - xenophobic automated profiling. Receivers of childcare benefits who were non-Dutch nationals were flagged, by an AI, as presenting a higher risk of having fraudulently benefited from such benefits. As a result, they were asked to reimburse the benefits 'unlawfully' received which represented exorbitant amounts. If an addressee of that decision wished to prove its legal or similarly significant impact, they would need to: (1) possess knowledge of the characteristics of other addressees in the same situation; (2) identify unequal treatment between beneficiaries (in this case, Dutch and Non-Dutch); (3) argue that, based on a comparison between the two groups, it is likely that the system relied on ethnic background to distinguish between the two. In other words, like in any other discrimination case, the claimant should - again - harbour a *founded suspicion* that a seemingly neutral measure or practice yielded discriminatory outcomes.

In this context, the maxim 'what you don't know won't hurt you' seems fitting to prospect the future application of Article 86 AIA. Based on its wording and the connections with provisions like the Equality Directives, it is tempting to assert that the right to an explanation will apply *selectively* instead of *generally*. The disadvantage with that reading is that the *onus* carried by the affected persons would, of course, be heavier since the legal/significant effect will need to be proven rather than presumed. Fair enough. The word of caution is, however, that - as already mentioned - discrimination can occur without any external signs to hint at its existence. Let us return to the recruitment scenario.

When an AI-supported recruitment process unfolds behind closed doors - as most recruitments do - which *indicia* could signal a bias? Let us recall that in the previously discussed *Kelly* and *Meister* cases,[155] the CJEU mentioned that recruiters are under no obligation to disclose information on the modalities of how they performed a given recruitment. In other words, the Court confirmed that job applicants had no legal means (subjective rights) to force recruiters to disclose any information that might be used against them, in a discrimination claim. Seeking not to add insult to injury, the CJEU nevertheless advised frustrated job applicants to consider all recruitment circumstances,[156] including the recruiter's attitude as a possible sign of bias. This is understandable advice but challenging to heed in automated contexts where an applicant may not have an effective way of monitoring what the system has done.

Consider the following: for an executive-level position, a company decides to set up two interview rounds. A woman applies for the job after a professional hiatus of several years, dedicated to raising her children. Suppose she makes the first-round cut. In the second interview, the recruiter might insist on knowing more about why a break from work was necessary. To connect their insistence to a bias against women may be an overstretch: after all, uninterrupted career of both men and women is usually perceived as a sign of ambition and value in the labour market. Be that as it may, our job applicant might have statistics to back her up: as has been often shown, women are, more so than men, likely to accept part-time jobs or even stop working when they

[151] Article 29 Working Party, Guidelines on Automated individual decision-making and profiling for the purpose of Regulation 2016/679 (n. 140).

[152] See *supra*, Sect. 2.

[153] See Art. 10(1), Directive 2000/78 (n. 24); Art. 8(1), Directive 2000/43 (n. 24); Art. 19(1), Directive 2006/54 (n. 24).

[154] See Sect. 2 of this article.

[155] Case C-104/10, *Kelly* (n. 114); case C-415/10, *Meister* (n. 118).

[156] The CJEU stated that, when applying EU non-discrimination law, national courts must, "in particular, take account of all the circumstances of the main proceedings, in order to determine whether there is sufficient evidence for a finding that the facts from which it may be presumed that there has been (direct or indirect) discrimination have been established." See case C-415/10, *Meister* (n. 118), para. 42.

have young children.[157] The CJEU[158] and the ECtHR[159] have admitted that statistics can be given as evidence of, say, residual biased practices. In that context, and bearing in mind the CJEU's ruling in *Meister*, if our job applicant wished to make a case of discrimination, theoretically, she could. It would be a long shot, but *indicia* would nevertheless be at her disposal, that is, the recruiter's insistence on her job hiatus and statistics showing that female workers who leave the job market to raise children have a harder time finding employment.

Suppose that, in the same scenario, instead of a first round of interviews, they tasked an AI system with the creation of a shortlist of candidates to be interviewed in the second round. Our job applicant would not be on that list and would have no clue as to 'why' because her exclusion would not follow from direct interaction with a human recruiter, whose attitude she could monitor. For our applicant to argue discrimination, she would need to access the list of shortlisted or rejected candidates, identify the characteristics they share (e.g. rejected applicants are women, mothers, of a certain age etc.) and only then, possibly, make an argument of discrimination. If we posit - as we do in this sub-Section - that the right to an explanation in Article 86 AIA is conditional on the proof of 'significant effect,' our job applicant would be left without the possibility of gaining knowledge about if and how they were affected by the system.

This prove-it-if-you-can reading of the 'significant effect' requirement in Article 86 AIA might seem unfair. There is something frustratingly formal in conditioning the benefit from subjective rights to procedural requirements that are practically difficult to meet. If, to receive an explanation, one should *prove* that a high-risk system has 'significantly' affected them, then in truth, the AIA asks us to turn a blind eye on fundamental rights violations that are likely to go unnoticed. In other words, 'what you don't know won't hurt you'… and won't give you the incentive to know if you were actually hurt.

We could argue this is already the state of the world: every day, we may be targeted by a multitude of biases that luckily go unnoticed. Be that as it may, suppose we wanted to receive explanations *in situations that matter to us* (like why we got excluded from a job or social benefits?). For those, Article 86 AIA includes a gatekeeper, in the form of an *onus probandi*. Does this suggest that the application of said Article would be the exception rather than the rule, leaving many individuals without the empirical information required to seek judicial redress? The answer is not a categorical 'no.' The application of Article 86 AIA will - again - likely be selective and justified by the *degree of salience* of specific human rights violations.

The bottom line is this: in cases of discrimination dealing with personal data processing and involving a high-risk system, Article 86 can be seen as a *lex specialis* to Article 22 GDPR. However, like the Article 22 GDPR, the application of Article 86 AIA is conditional, concerning instances where the harm suffered is, to some extent, salient and therefore, knowable.

The *lex specialis* story does not end there. The now withdrawn AILD proposal and the enforced PLD raise critical issues in terms of designating the adequate substantive and procedural law that ought to apply in cases of AI-related discrimination. Substantively, the main issue is whether the harm that said provisions seek to regulate includes discrimination. Procedurally, if AI-related discrimination is indeed harm they cover, could the PLD and the AILD (had it been enforced) act as *lege specialia* in relation to the Equality Directives? If the answer is 'yes', litigants would be more inclined to launch proceedings on the grounds of those instruments rather than on EU Non-Discrimination law, for the reasons discussed in the following Section.

4 The lex specialis to the lex specialis? The right to disclosure of evidence in the - now withdrawn - AILD proposal.

The revised and recently enforced PLD will be excluded from our analysis for two reasons. First, algorithmic discrimination, as this article's point of focus, does not correspond to the PLD's notion of damage, defined in Article 4(6) as material loss resulting from death or personal injury, including medically recognized harm to psychological

[157] For examples of scholarship, see *inter alia* Baker, L.: Sex discrimination against part-time workers: the 'Buggs' issues for women. Feminist L. Stud. VI/2, 257–271 (1998); Kjeldstad, R., Nymoen, E. H.: Part-time work and gender: Worker versus job explanations. Int'l Lab. Rev., 151-1/2, 85–107 (2012); Blazquez M. C., Carcedo, J. M., Women's part-time jobs: 'Flexirisky' employment in five European countries. Int'l Lab. Rev., 153/2, 269–292 (2014); Weber, G., Williams C.: Mothers in 'good' and 'bad' part-time jobs: different problems, same results. Gender & Soc'y. 22/12, 752–777 (2008).

[158] See case C-385/11, *Moreno* [2012] EU: C:2012:746 dealing with Spanish legislation by virtue of which part-timers retired at an older age and received lower pension than full timers. Given that about 80% of the part-time workers in Spain were - at the time of the ruling - women, the Court sanctioned the Spanish legislation on the grounds of discrimination based on gender.

[159] ECtHR, *Hoogendijk v. Netherlands*, App. No. 58,641/00 [2005], at 21: "the Court considers that where an applicant is able to show, on the basis of undisputed official statistics, the existence of a prima facie indication that a specific rule - although formulated in a neutral manner - in fact affects a clearly higher percentage of women than men, it is for the respondent Government to show that this is the result of objective factors unrelated to any discrimination on grounds of sex. If the onus of demonstrating that a difference in impact for men and women is not in practice discriminatory does not shift to the respondent Government, it will be in practice extremely difficult for applicants to prove indirect discrimination."

health.[160] Said notion also includes harm to, or destruction of, any property[161] other than the defective product itself,[162] a product damaged by a defective component[163] or property used exclusively for professional purposes.[164] Finally, damage can ensue from the loss or corruption of data not used exclusively for professional purposes.[165] Clearly, discrimination (as a stand-alone harm) does not fit the PLD's definition of damage. What is more, the PLD applies, second, in the context of *non-professional use* of defective products (including software). Presumably, in the high-risk sectors listed in Annex III AIA (like access to labour, education, biometric identification in the field of migration etc.), a harm like discrimination will likely arise in a professional use context. For these reasons, the PLD will not be further discussed.

In contrast, the AILD proposal, had it not been withdrawn, could have applied to AI-related discriminatory recruitment. A first element in favor of this reading is the fact that the AILD applied to harm associated with high-risk systems,[166] such as those used in recruitment practices.[167] Discrimination in the access to labour could be captured by the AILD, if shown to stem from fault that is, the non-compliance with a duty of care laid down in Union or national law directly intended to protect against the damage occurred.[168] Though the agent at fault could be a provider, for the sake of simplicity, let us focus on recruiters acting as deployers. By virtue of the AILD proposal, those would be at fault if discriminatory recruitment had resulted from their failure to meet their duty to comply with the accreditation certificates delivered by notification bodies under Article 29 AIA.

If the recruiter in our hypothetical scenario was a public authority, or a private one providing public services, they would be required to perform a fundamental rights impact assessment (FRIA), pursuant to Article 27 AIA. It is worth noting that the AILD proposal did not include, under the 'fault' heading, a provider's failure to conduct, or conduct properly, such an assessment. Nevertheless, we could imagine a scenario of, say, a public administration using a recruitment system that would be required to carry out a FRIA, by evaluating the modalities of the system's deployment in connection to its intended use,[169] the description of the timeframe and frequency of use,[170] the categories of natural persons and groups likely to be affected[171] and the specific risks of harm likely to impact those groups, bearing in mind the information available in connection to reasonably foreseeable uses and misuses generating risks of fundamental rights violations.[172] Failure to show (i.e. prove) care on any of these points could qualify as non-compliance with the duties the AIA places on deployers, reinforcing the *indicia* that a biased recruitment had, indeed, resulted from the deployer's fault, as defined in the AILD. proposal.

Had that proposal not been withdrawn, a claimant intending to launch proceedings could have chosen to base their claim on it instead of EU Non-Discrimination law, for an obvious reason: Article 3 AILD granted the right to request disclosure of evidence whereas the Equality Directives - as already mentioned - do not include such a right. That said, much like the receiving of explanations under Article 86 AIA, the disclosure of evidence in the AILD was not unconditional. Claimants - and this would be a disadvantage - carried the initial burden of proving two facts: first, that they had sought the disclosure of information from the defendant *prior to asking a court to order it*;[173] second, that, based on the evidence possibly disclosed (either spontaneously by the deployer or following a court order) or the deployer's refusal to grant disclosure, there are sufficient *indicia* to support a presumption of fault.[174]

Although claimants acting on the grounds of the AILD were not altogether discharged from proving standing, their burden did appear to be more bearable than its counterparts in the previously discussed legislation. As already mentioned, we continue to lack clarity on the requirements regarding the evidence of 'fully automated data processing' (under Article 22 GDPR) or of 'significant effects' on a person's situation (Article 86 AIA). In comparison to those, in

[160] PLD (n. 35), Art. 4(6)(a).

[161] *Id*., Art. 4(6)(b).

[162] *Id*., Art. 4(6)(b)(i).

[163] *Id*., Art. 4(6)(b)(ii).

[164] *Id*., Art. 4(6)(b)(iii).

[165] *Id*., Art. 4(6)(c).

[166] Article 1 ('Subject matter and scope') mentions that the ALD shall lay down common rules on the disclosure of evidence on "*high-risk artificial intelligence (AI) systems* to enable a claimant to substantiate a non-contractual fault-based civil law claim for damages" (AILD (n. 36) Art. 1(1)(a)) (emphasis added).

[167] AIA, *cit. supra*, Annex III pt 4 'Employment, workers' management and access to self-employment.'

[168] AILD (n. 36), Art. 4(2).

[169] AIA (n. 32), Art. 27(1)(a).

[170] *Id*., Art. 27(1)(b).

[171] *Id*., Art. 27(1)(c).

[172] *Id*., Art. 27(1)(d). This provision refers to Article 13 AIA ('Transparency and Provision of Information of Deployers') which mentions that, for the purpose of deployment, high-risk systems will be accompanied by instructions enabling "concise, complete, correct and clear" use (Art. 13(2) AIA (n. 32)). Those instructions should outline the characteristics, capabilities and limitations of performance of the high-risk system, including its level of accuracy, metrics against which it was measured and any "known or foreseeable circumstance, related to the use of the high-risk AI system in accordance with its intended purpose or under conditions of reasonably foreseeable misuse which may lead to risks to the health and safety of fundamental rights" (Art. 13(3)(b)(iii) AIA (n. 32)).

[173] AILD (n. 36), Art. 3(2).

[174] *Id*., Art. 4(5).

the AILD proposal, disclosure requests seemed to be based a *founded suspicion* that a high-risk system was involved in the suffering of harm.[175] However, both the claimant's suspicion and the deployer's attitude (to grant disclosure or not) were required - it seemed - to meet a certain level of fact-correspondence given that, when a court was asked to order disclosure, the claimant was required to "present facts and evidence sufficient to support the plausibility of a claim for damages."[176]

In requiring that the Article 3 machinery be put into motion by a founded and fact-based suspicion, the system of evidence in the AILD was akin, in terms of its design, to its counterparts in the Equality Directives. Under those, proof of standing is admissible when a claimant adduces *indicia* of discrimination, suggesting - especially in indirect discrimination cases - a level of sufficiency and relevance of the available facts. The AILD followed a similar logic in its framing of the right to request disclosure of evidence. However, the curious question we could have raised, had the AILD not been withdrawn, was how Article 86 AIA (right to an explanation) and Article 3 AILD (right to request disclosure of evidence) could correlate to one another. To answer this question, our assumption - from a claimant's point of view - would be to rely on the provision that leads to the shortest and least procedurally taxing route to a judicial remedy. Based on that assumption, the interrelationship between the two discussed provisions can take two forms: *complementarity* and *strategic exclusion*.

In a world where both the AIA and the AILD are binding, we could imagine that Article 86 AIA and 3 AILD would apply jointly. Article 86 AIA would apply first for two reasons, mentioned earlier: first, because it does not require judicial proceedings for explanations to be given. Second, because contrary to the Article 22-explanations (GDPR), the Article-86 ones do not pertain to the rationale of a system's output, but to that of the human decision to rely on that output (or not).

The complementary application of Article 86 AIA and Article 3 AILD would go as follows: a person having unsuccessfully applied for a job at, say, a local administration would have the nagging suspicion of being the victim of a discriminatory automated selection. She would rely on Article 86 AIA to receive an explanation regarding the extent to which the recruiter adhered to the system's output in deciding to exclude her from the recruitment process. If the recruiter believed the system's output to be skill-based and accurate, our claimant would be encouraged to request information on how the recruiter applied the system's use instructions and - since we are talking about a public administration - how they performed a FRIA. If the recruiter refused to disclose that information, our claimant would be justified to rely on the AILD and request a court order of disclosure of evidence, on the basis of Article 3 AILD.

There is something organic about this unfolding of the stages between a person's initial suspicion of harm and the access to the evidence able to justify that suspicion. Article 86 AIA would appear as the explanatory antechamber to Article 3 AILD, providing claimants with initial knowledge (of 'what happened') based on which they could decide on whether or not to launch proceedings, on the grounds of AILD. Because the latter was withdrawn, claimants can still launch proceedings on the grounds of EU Non-Discrimination law, if the explanation given under Article 86 AIA provides enough *indicia* to substantiate those proceedings.

What could have disrupted this organic unfolding of a person's access to knowledge relative to a possibly harmful high-risk AI is the burden of proving the legal or significant effect on a person's situation, required under Article 86 AIA.[177] As previously argued, that burden may not be easy to satisfy, except in cases where an AI's effect on a person is salient, justifying, without much controversy, the explanation based on said Article. In this scenario, the AIA and the AILD would not have been as congruent as we could have imagined, because their application may not have been an issue of chronological succession but one of selection: if both instruments were binding, litigants would have been able to choose between the explanation in the AIA and that in the AILD, depending on the type of information sought. This is the second hypothesis we previously called strategic selection.

Save in cases of beyond-doubt harm, both provisions would apply in an 'ambience of suspicion' where a person would suspect, without being certain, that they were affected by an AI system included in some ultimately human decisional process (like recruitment). Since the application of both Article 86 AIA and Article 3 AILD is conditional on meeting the burdens of proof outlined earlier, a person would be confronted with an *electa una via* (choose a path) situation. If they wished to receive a statement of reasons explaining (e.g. if and why a recruiter relied on a system), they would take the Article-86 route. If, however, they wished to verify the recruiter's compliance with the AIA - or similar national legislation - then they would have taken the AILD route, if the EU legislature had gone ahead with the enactment of that Directive proposal.

[175] *Id*., Art. 3(1).

[176] *Ibid*.

[177] See Sect. 3(B) of this article.

5 Conclusion

While our ears are still ringing from the echo of the vivacious cries for more transparency, explainability and fairness in the field of AI, the legislative execution somewhat falls short for the reasons detailed in this article which we can sublimate in two important points. This observation is warranted by one fundamental question: given the stringency of the conditions under which litigants can access explanations under the provisions discussed, what happens with the 'essence'[178] of the right to a fair trial, in particular the right to access courts?

First, though a 'well done' is certainly due to the EU legislature for including rights to explanation/disclosure in several instruments (the GDPR, the AIA, the PLD and the AILD proposal), what seems to be missing is a vision of how those instruments will interact with one another in practice. Bearing in mind that the GDPR is the 'mother' of virtually all subsequent data processing regulation, the kinship between the AIA, on the one hand and the PLD/AILD proposal, on the other hand, would *prima facie* mean that the rights contained in those instruments do not pertain to separate explanations but to separate stages of a person's quest of explanations. As we mentioned, it would have been possible to view Article 86 AIA as the explanatory 'antechamber' to the right to request disclosure of evidence, as enshrined in Article 3 of the withdrawn AILD. And perhaps, if this Directive was enacted, institutional and court practice would have evolved in that direction.

Second, it is also likely that Articles 22 GDPR, 86 AIA, 9 PLD and 3 AILD give way to different types of explanations, depending on whether those are requested in-court or out-of-court. We can make the pluralist argument that different persons in different contexts involving AI systems will require different types of knowledge about those systems. Fair enough. Except that a rigid, sectoral approach seems ill-adapted to the instruments discussed because of their obvious normative kinship. The PLD and the AILD proposal considered together were intended as procedural extensions of the AIA. We could argue that the right to request disclosure of evidence in the PLD/AILD is, in truth, a specific procedural expression of the right to an explanation in the AIA. Similarly, whatever the A29 WP has written *à propos* the right to a human explanation in Article 22 GDPR and however the CJEU[179] and AG De La Tour may have interpreted it,[180] a map was drawn out for how to approach explainability (and, more generally, access to empirical knowledge) across the legislation on new technologies where a right to an explanation/explainability is recognised.

Choosing between an integrity-driven or a pluralist reading of the instruments discussed depends on whether we view them as sharing the same, normative understanding of explainability as being a right to access understanding (i.e. a form of knowledge)[181] about an AI system and the impact it had, or not, in a given situation. Most of us would be presumably agree with this. But what is the point of acquiring such understanding? The obvious answer is 'fairness.' Whether applied out-of-court or in-court, the rights conducive to explanations are all specific gateways to empirical knowledge that a person might rely on to determine if they have enough evidence to actually build a case.

In the realm of judicial protection, integrity-driven and pluralist readings are also possible. Let us begin by the pluralist approach. A traditional EU procedural lawyer would probably swear by the availability of remedies and defend a 'the-more-the-merrier' take on the effective judicial protection principle. The gist is the following: judicial protection is effective if there are multiple remedies which, though framed by different procedural conditions, support the same type of judicial review (e.g. legality). Litigants would then have a choice between several options and launch the proceedings for which they could meet the procedural requirements. We can see why it is tempting to transpose this reasoning to the benefit from rights. If we look at the legislation discussed in this article, it seems that the legislature was pressed to include a variant of the right to explainability everywhere, in the hope of 'covering all bases' regarding the explanatory needs litigants might have down the line. Praise-worthy, indeed. The dangerous germ there is *ineffectiveness*: what is the point of multiplying the procedural pathways to explanations if litigants are likely unable to prove standing in many cases? The cases referred to here are those of *prima facie* undetectable harm like, possibly, algorithmic discrimination. In those, does the 'acceptance of risks' the AIA ties to high-risk systems mean that there might be entire classes of harms that, for all we know, have never happened? In itself, this is not alarming. What is alarming is that the law seems to offer several limited procedural means to inquire if harm had been suffered. What does this tell us about the EU's AI regulation?

The Human rights rhetoric aside, the AIA and its procedural progeny (the PLD and even the AILD proposal) are market-supporting regulations. If the trust/excellence

[178] See, in this regard, the insightful study of Hallinan, D.: The Essence of Data Protection: Essence as a Normative Pivot. Eur. J. L.& Tech'y. 12/3, 1–24 (2021).

[179] See case C-634/21, *Schufa* (n. 119).

[180] Case C-203/22 (Opinion), *Dun* (n. 145).

[181] Grozdanovski, L.: The Explanations One Needs for the Explanations One Gives. The Necessity of Explainable AI (XAI) for Causal Explanations of AI-related harm - Deconstructing the 'Refuge of Ignorance' in the EU's AI Liability Regulation, (n. 42), at 180 seq.

components of the EU's regulatory ecosystem[182] on AI were balanced, excellence would most certainly outweigh trust. Indeed, excellence is sought through concrete strategies on competitiveness and innovation[183] while trust turns out to be a matter of faith: because high-risk AI providers are tasked with several demanding requirements which particularise noble principles like transparency, human oversight etc., we should trust that when AI systems reach us, end-users, they meet those requirements. Otherwise, they would not be made available at all in the market. The trouble with this trust-as-faith interpretation is that - like in faith - one is encouraged to refrain from questioning it too much.

Author contributions L.G. - sole author.

[182] European Commission, White Paper. Artificial Intelligence - A European approach to excellence (n. 120).

[183] *Id*., at 5 seq.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.