2021

Algorithmic Impact Assessments Under the GDPR: Producing Multi-Layered Explanations

Margot E. Kaminski
University of Colorado Law School

Gianclaudio Malgieri
Augmented Law Institute, EDHEC Business School

Algorithmic impact assessments under the GDPR: producing multi-layered explanations

Margot E. Kaminski and Gianclaudio Malgieri*

Key Points

- Policymakers, scholars, and commentators are increasingly concerned with the risks of using algorithms for profiling and automated decision-making.
- This article addresses how a Data Protection Impact Assessment (DPIA), applied as an algorithmic impact assessment (AIA), links the two faces of the General Data Protection Regulation (GDPR) approach to algorithmic accountability: individual rights and systemic governance.
- We propose that AIAs simultaneously provide systemic governance of algorithmic decision-making and serve as an important 'suitable safeguard' (Article 22) of individual rights.
- As a nexus between the GDPR's two approaches to algorithmic accountability, DPIAs have a heretofore unexplored link to individual transparency rights.
- Our examination of DPIAs suggests that the current focus on the right to explanation is far too narrow. We call, instead, for data controllers to consciously use the mandatory DPIA process to produce what we call 'multi-layered explanations' of algorithmic systems.
- This concept of multi-layered explanations not only more accurately describes what the GDPR is attempting to do, but also normatively fills potential gaps between the GDPR's two approaches to algorithmic accountability.

Introduction

To date, the discussion of the GDPR's regulation of algorithmic accountability has largely focused on whether there is an individual right to explanation of an algorithmic decision. Only more recently have legal scholars begun to focus on the GDPR's systemic accountability tools.[1]

Impact assessments have received particular attention, on both sides of the Atlantic, as a tool for algorithmic accountability. This article aims to address how a Data Protection Impact Assessment (DPIA) (Article 35) links the GDPR's two approaches to algorithmic accountability—individual rights and systemic governance—and potentially leads to more accountable and explainable algorithms. Examining the GDPR's approach to impact assessments suggests that the scholarship has been getting explanations wrong. Algorithmic explanations should not be understood as static statements but as a circular and multi-layered process. The literature has largely to date focused on what information goes to whom and when; we argue that the impact assessment process plays a crucial role in connecting internal company heuristics and risk mitigation to outward-facing rights and in forming the substance of several different kinds of explanations.

We begin by introducing the algorithmic accountability tools in the GDPR (section 'Algorithmic accountability in the GDPR'). In section 'Individual rights in the GDPR and the multi-layered explanation', we explore the individual rights of data subjects as regards

*Gianclaudio Malgieri, Augmented Law Institute, EDHEC Business School, Lille, France. E-mail: gianclaudio.malgieri@edhec.edu; gianclaudio.malgieri@gmail.com. The authors, in alphabetical order, have contributed equally to this work. This work was funded under Fullbright Schuman Program [to Kaminski], and EU Horizon 2020 Project PANELFIT, Grant Agreement n. 788039. No conflict of interest to declare.

1 This article originally builds on and further develops the research conducted in preparation for Margot E Kaminski and Gianclaudio Malgieri, 'Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR' in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* '20 (Barcelona, Spain: Association for Computing Machinery 2020), 68–79, https://doi.org/10.1145/3351095.3372875.

algorithmic decisions. In section 'Collaborative governance in the GDPR', we explain the GDPR's collaborative governance of algorithms. In section 'The DPIA as an algorithmic impact assessment', we explain the requirements of the DPIA under the GDPR. In section 'A model algorithmic impact assessment', we discuss the broader literature on Impact Assessments and how our interpretation of the GDPR's impact assessment requirement in fact leads to a better, more complex understanding of the GDPR's explanations of algorithmic decision-making than a focus on Article 22 alone. We close by calling for what we call a multi-layered approach to explanations, stemming from the impact assessment process.

As a methodological disclaimer: this article gives a relatively large amount of attention to the opinions and guidelines of the Article 29 Working Party, now the European Data Protection Board (EDPB), an advisory body consisting of national Data Protection Authorities. We are aware that these guidelines are not binding and cannot be considered to be the only possible interpretation of the GDPR. However, the GDPR (at Article 70) states that the EDPB is required to issue opinions, guidelines, and recommendations in order to ensure the consistent application of the GDPR. Accordingly, the interpretational activity of the EDPB is not only influential for commentators, but also for the activity of national Data Protection Authorities. Moreover, we understand such opinions and guidelines to be an essential component of the GDPR's regulatory approach, discussed further below. That approach—referred to in scholarship as 'collaborative governance' or 'new governance'—often entails broad or vague binding textual requirements, clarified over time in regulatory guidance or industry best practices or back-and-forth between regulated companies and regulators. To dismiss EDPB guidelines and opinions as non-binding 'soft law' is to overlook the central role this softer law serves in the design of the GDPR as a regulatory system.

Algorithmic accountability in the GDPR

The GDPR has significant implications for algorithmic decision-making. At first, the legal debate focused on whether the GDPR created an individual right to an explanation of an individual algorithmic decision.[2]

Subsequent legal analysis, however, began to focus instead on other accountability tools,[3] required either in the text of the GDPR, or in its Recitals, or in guidelines issued by the Article 29 Working Party (now the EDPB).[4]These tools include third-party auditing, the appointment of Data Protection Officers (DPOs)(Article 37), and the requirement of DPIAs (Article 35) under certain circumstances.

As one of us has argued at length elsewhere, the GDPR establishes a two-pronged approach to algorithmic accountability. It combines a series of individual rights (Articles 12–23) with a systemic governance regime overseen by regulators, targeted at more comprehensive oversight over the algorithm and the people around it (Articles 24–43 and throughout). These two systems interact and overlap. An individual right is often also a company's duty. But even if individuals (data subjects) fail to invoke their rights, companies (data controllers) have significant obligations—both procedural and substantive—under the GDPR.[5]

For example, in the algorithmic governance context, a data subject has a right to contest an individual algorithmic decision (Article 22), to receive notice of solely automated decision-making (ADM) (Article 13), and to request access to 'meaningful information about the logic involved' (Article 15). Should this fail to invoke any of these rights, however, the GDPR still puts in place significant obligations on data controllers using ADM, whether that decision-making involves a human or not.[6] The GDPR requires an array of systemic accountability tools, including third-party auditing, the appointment of DPOs (Article 37), and DPIAs (Article 35). These obligations arise from the text of the law, in

2 Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7(2) International Data Privacy Law 76; Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7(4) International Data Privacy Law 243; B Goodman and S Flaxman, 'EU Regulations on Algorithmic Decision-Making and a 'Right to Explanation' (2016) <http://arxiv.org/abs/1606.08813> accessed 30 June 2020; A Selbst and J Powles, 'Meaningful Information and the Right to Explanation' (2017) 7(4) International Data Privacy Law 233; Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond' (2019) International Journal of Law and Information Technology <https://doi.org/10.1093/ijlit/eay017> accessed 8 October 2020; Margot E Kaminski, 'The Right to Explanation, Explained' (2019) 34(1) Berkeley Technology Law Journal <https://papers.ssrn.com/abstract=3196985> accessed 8 October 2020.

3 Antoni Roig, 'Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)' (2018) 8(3) European Journal of Law and Technology <http://ejlt.org/article/view/570> accessed 8 October 2020; Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For' (2017) 16(1) Duke Law & Technology Review 18; Bryan Casey, Ashkon Farhangi and Roland Vogl, 'Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise' (2018) 34 (2019) Berkeley Technology Law Journal.

4 Article 29 Working Party Guidelines (n 4) 29.

5 Margot E Kaminski, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92(6) Southern California Law Review; Kaminski (n 2).

6 Edwards and Veale (n 3) 74–80.

accompanying Recitals, and in the Guidelines on Automated Individual Decision-making and Profiling ('Guidelines on ADM') released in October 2017 and revised in February 2018 by the Article 29 Working Party (now the EDPB).[7]

It is also crucial to understand the mode through which the GDPR governs. The GDPR largely governs—both in the sense of coming up with the substance of data controllers' duties, and in the sense of monitoring compliance—through an approach known in the legal literature as 'collaborative governance': the use of public–private partnerships.[8] This form of regulatory design has alternatively been referred to as 'new governance', 'co-governance', partial delegation to the private sector, and 'meta-regulation'. Rather than create strict top-down rules enforced by the government, the government works with both regulated industries and with third parties to come up with the substance of, and enforce, regulations. Importantly, collaborative governance is not equivalent to self-regulation; the government still has an important, even central, role to serve.

Because the GDPR effectively outsources many governance decisions to private companies, accountability takes on added significance. Accountability in the GDPR is not just about protecting individual rights. It is about ensuring that this process of co-governing with private parties receives appropriate input and oversight from the public, from civil society, and from both expert and affected third parties.[9]

With this background in mind, the next two sections of this article go into more detail on both the individual rights and systemic governance elements of the GDPR's approach to algorithmic accountability, before turning to the role of the DPIA in linking the two facets.

Individual rights in the GDPR and the multi-layered explanation

The GDPR gives individuals several important rights with respect to algorithmic decision-making. The GDPR contains both general data protection rights (such as notification rights, access rights, rectification rights, and the right to restrict processing) and rights specific to profiling (such as the right to object), which also apply to algorithmic decision-making.[10] On top of this, the GDPR establishes rights specific to algorithmic decision-making, which include: a right to be notified of solely ADM (Articles 13, 14); a right of both notification and access to meaningful information about the logic involved (Articles 13, 14, 15); a right to be informed of the significance of and envisaged effects of solely ADM (Articles 13, 14, 15); and a right not to be subject to solely automated decision making (Article 22), with safeguards and restraints for the limited cases in which ADM is permitted. Those safeguards include, but are not limited to, a right to contest a decision, to express one's point of view, and to human intervention (Article 22).

We do not intend to revisit the legal debate over these rights in detail here, but an overview may be useful. As mentioned, discussion of these individual rights has largely focused on whether or not—or really, how—solely ADM must be explained to individuals. As Selbst and Powles point out, it is disingenuous to say that there is no right to an explanation in the GDPR; the GDPR's text clearly requires companies to explain at least 'meaningful information about the logic involved' in ADM, in addition to its significance and envisioned effects (Articles 13, 14, 15).[11] What this information constitutes in practice, however, has been the subject of hot debate, including whether it is a system-wide (model-wide) explanation or specific to individual decisions, and what depth of explanation is required.[12]

The core debate has primarily focused on whether or not Article 22 creates an *ex post* right to explanation of an *individual* decision made by an automated system.[13] Our view, discussed at length by each of us elsewhere, is that it does.[14] Automated decisions with significant effects must be made 'legible' to individuals, in the sense that individuals must be able to understand enough about the decision-making process to be able to invoke their other rights under the GDPR, including the right to contest a decision.[15] Several of the Member States

7 Article 29 Working Party Guidelines (n 4). See Michael Veale and Lilian Edwards, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' (2018) 34(2) Computer Law & Security Review 398.

8 See Jody Freeman, 'The Private Role in the Public Governance' (2000) 75 New York University Law Review 543; K Bamberger, 'Regulation as Delegation: Private Firms, Decisionmaking, and Accountability in the Administrative State' (2006) 56Duke Law Journal 377.

9 Kaminski (n 5) 28.

10 Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'?' (2018) 16(3) IEEE Security & Privacy 46.

11 Selbst and Powles (n 2).

12 Wachter, Mittelstadt, and Floridi (n 2) 78; Malgieri and Comandé (n 2) 244; Selbst and Powles (n 2) 240–41.

13 Brkan (n 2); Stefanie Hänold, 'Profiling and Automated Decision-Making: Legal Implications and Shortcomings' in Marcelo Corrales, Mark Fenwick and Nikolaus Forgó (eds), *Robotics, AI and the Future of Law, Perspectives in Law, Business and Innovation* (Singapore: Springer Singapore 2018), 123–53, See also Edwards and Veale (n 3) *passim*.

14 Malgieri and Comandé (n 2); Kaminski (n 2).

15 Malgieri and Comandé, ibid 250.

implementing Article 22(2)b of the GDPR have outlined the Article 22 explanation duties in greater detail.[16]

Existing discussions of the Article 22 right to explanation, however, largely obscure the more complex approach to algorithmic transparency taken by the GDPR as a whole. As we discuss below in section 'Towards a model AIA', our view is that the GDPR's transparency rights are best discussed together as a system. That is, the GDPR, is best understood as establishing a system of *multi-layered explanations*.[17] Individuals have a right to both a system-wide but detailed description of the logic of an algorithm (Articles 13, 14, 15), and more specific insights on individual decisions taken.[18] We discuss layers of explanations further in this section.

There have been legitimate concerns voiced in the legal literature both in Europe and in the USA about the capacity of individuals to both invoke their rights and execute oversight over algorithmic decision-making.[19] These range from concerns about access to justice to concerns about individual capacity and expertise. Consequently, most policy proposals call either for a dual regime, like the GDPR, that mixes individual rights with systemic forms of accountability;[20] or for foregoing individual accountability in favour of expert and external oversight.[21]

The latter approach—foregoing individual rights entirely—ignores the dignitary and legitimizing value of such rights.[22] Individual rights allow individuals to exhibit autonomy and exert control, and to protest or reject their objectification by profiling or decision-making machines.[23] Individualized explanations also serve to establish the legitimacy, or illegitimacy, of a decision-making system by subjecting its logics and performance to inspection and assessment as to whether they are socially acceptable or even illegal (what we and others call a 'justification' of algorithmic decisions).[24]

Rejecting individual rights, as we discuss below, also ignores the symbiosis between the GDPR's two regimes. Individual rights can play a crucial role in the GDPR's systemic collaborative governance. Understanding the GDPR's dual approach to algorithmic accountability has the potential to answer important questions in the literature about the value, in practice, of individual rights in algorithmic accountability.[25]

Collaborative governance in the GDPR

The other side of algorithmic governance in the GDPR is its systemic governance regime. This regime aims, primarily, to address instrumental goals: to prevent error, bias, and discrimination.[26] It focuses on assessing and mitigating system-wide risks, including before an algorithm is deployed. It is largely constituted through collaborative governance, or a cooperative public–private approach to regulation. We here illustrate two examples of how this works in the GDPR.

Article 22's suitable safeguards on ADM is one example of this in practice. The GDPR's text does not comprehensively dictate what companies using ADM must do to protect individual rights (Article 22). Instead, it lists examples of safeguards (contestation, expression, human intervention), but leaves it to both companies

16 Gianclaudio Malgieri, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations' (2019) Computer Law & Security Review 105327 <https://doi.org/10.1016/j.clsr.2019.05.002> accessed 8 October 2020. See in particular the cases of French and Hungarian laws that provide more explicit explanation of individual decisions taken (based on criteria and methods used in algorithmic processing).

17 Karthikeyan Natesan Ramamurthy and others, 'Model Agnostic Multilevel Explanations' (12 March 2020) <https://arxiv.org/abs/2003.06005v1> accessed 8 October 2020.

18 Article 29 Working Party Guidelines (n 4) 25: 'The controller should find simple ways to tell the data subject about the rationale behind, or the criteria relied on in reaching the decision. The GDPR requires the controller to provide meaningful information about the logic involved, not necessarily a complex explanation of the algorithms used or disclosure of the full algorithm. The information provided should, however, be sufficiently comprehensive for the data subject to understand the reasons for the decision'. See also at page 27: 'the controller should provide the data subject with general information (notably, on factors taken into account for the decision-making process, and on their respective 'weight' on an aggregate level) which is also useful for him or her to challenge the decision' and 'Recital 71 highlights that *in any case* suitable safeguards should also include: specific information to the data subject and the right (. . .) to obtain an explanation of the decision reached after such assessment and to challenge the decision'.

19 Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20(3) New Media & Society 973; M Hildebrandt, 'The Dawn of a Critical Transparency Right for the Profiling Era' in J Bus (ed), *Digital Enlightenment Yearbook* (2012th edn, Amsterdam: IOS Press 2012) 41–56; Edwards and Veale (n 3) 67; Bryce Goodman, 'A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection' (2016), 29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona. NIPS Foundation, 3–4; Joshua Kroll and others, 'Accountable Algorithms' (2017) 165(3) University of Pennsylvania Law Review 633; Deven R Desai and Joshua A. Kroll, 'Trust But Verify: A Guide to Algorithms and the Law' (2017) Harvard Journal of Law & Technology 27; Edwards and Veale (n 3) 65–67.

20 See generally Kaminski (n 5). See also Kate Crawford and Jason Schultz, 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms' (2014) 55(1) Boston College Law Review 93; Danielle Citron, 'Technological Due Process' (*Faculty Scholarship*, 30 April 2009) 1310; Danielle Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (*Faculty Scholarship*, 1 January 2014) 20, 26.

21 Kroll and others (n 19) 660–63; Desai and Kroll (n 19); Edwards and Veale (n 3) 76.

22 Lee A Bygrave, 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling' (2001) 17(1) Computer Law & Security Review 18.

23 Kaminski (n 5) 4; Hildebrandt (n 19) 47.

24 Kaminski, Ibid 15.

25 Ibid, *passim*.

26 Ibid 27.

and regulators to determine what additional safeguards are necessary. The accompanying Recital famously adds a right to individual explanation (Recital 71). A more detailed list of best practices can be found in several sources, including the interpretative guidelines of the Article 29 Working Party.[27] These include, but are not limited to regular quality assurance checks, algorithmic auditing, independent auditing, establishing data minimization and clear retention periods, using pseudonymization techniques, certification mechanisms, ethical review boards, and more.[28]

All these tools are attempts at establishing systemic accountability and oversight, in a comprehensive and ongoing manner. But the Guidelines make clear that what counts as adequate safeguards will be established through an ongoing conversation between companies and regulators, involving government guidelines and potentially involving industry-wide efforts to come up with codes of conduct or other forms of standards (Article 40).[29] The GDPR thus harnesses companies' efforts to help come up with both the substance and the method of regulation in this space.

The GDPR's approach to preventing bias and discrimination in algorithmic decision-making is another example of collaborative governance in action. Recital 71 tasks companies with preventing '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' in profiling and algorithmic decision-making. The GDPR does not lay out *how* to do this. Instead, the Guidelines suggest that companies check data sets for bias, regularly review the accuracy and relevance of decisions, deploy systems that audit algorithms, and use 'appropriate procedures and measures to prevent errors, inaccuracies or discrimination' on the basis of sensitive data such as race, religion, or health information, deployed on a cyclical basis.[30]

Again, the GDPR does not tell companies precisely what to do. It identifies the problem, provides suggestions of what regulators might consider adequate, and often tasks companies with cooperatively coming up with solutions. Such company-created solutions may then inform what regulators ultimately require.[31]

The DPIA as an algorithmic impact assessment

Within this dual system of algorithmic accountability—individual rights accompanied by extensive but collaborative governance of companies' behaviour—the DPIA plays a special role. We claim that as applied to algorithmic decision-making, the DPIA is best understood as a nexus between the GDPR's two approaches to algorithmic accountability. Understanding it in this way allows us to better understand what is or might be required, and to observe the tool's potential shortcomings as implemented in the GDPR.

The Guidelines on ADM interpret the GDPR to mandate DPIAs for all ADM with significant effects.[32] Article 35(3)(a) requires a DPIA for 'a systematic and extensive evaluation of personal aspects relating to natural persons which is based on *automated processing*... on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person' (emphasis added). As Casey, Farhangi, and Vogl have noted, 'demonstrating that a DPIA is not necessary will, in many instances, itself require a DPIA'.[33] We note, too, that at least one Member State, Slovenia, requires algorithmic impact assessments (AIAs) for ADM under Article 22(1) of the GDPR.[34]

In this section, we address the DPIA as an AIA. We identify what the purpose of the DPIA is in the GDPR, and what it must include. Understanding the DPIA's purpose in algorithmic governance both clarifies what the content should be and points to several shortcomings in the current conception of it. The GDPR's DPIA will serve, in the ADM context, as an AIA. It thus may prove to be an example for governments around the world considering using impact assessments as a tool to achieve algorithmic accountability.

What is required in a DPIA?

In this section, we discuss the GDPR's requirements for a generic DPIA, before turning in the next sections to requirements specific to algorithmic decision-making. The GDPR requires DPIAs only under certain circumstances. The GDPR describes a DPIA as 'an assessment of the impact of the envisaged processing operations on

27 Article 29 Working Party Guidelines (n 4) 31–34.

28 Ibid 32.

29 Ibid 32. On Certifications and algorithms see also Edwards and Veale (n 3) 50.

30 Article 29 Working Party Guidelines, ibid 28.

31 Roig (n 3) 2 ('the requirement of data protection impact assessment (DPIA)... could compile all the relevant safeguards for specific technologies and automatic processing and turn into a data generator for policy purposes').

32 Creating a categorical requirement that applies 'in the case of decision-making including profiling with legal or similarly significant effects that is not wholly automated, as well as solely automated decision-making defined in Article 22(1)'. Article 29 Working Party Guidelines (n 4).

33 Casey, Farhangi and Vogl (n 3) 176.

34 Predlog Zakona o varstvu osebnih podatkov – predlog za obravnavo – nujni postopek – Novo Gradivo ŠT. 2 <http://www.mp.gov.si/fileadmin/mp.gov.si/pageuploads/mp.gov.si/novice/2018/ZVOP-2_NG_2_apr.pdf> accessed 8 October 2020.

the protection of personal data' (Article 35). That assessment, per the text of the GDPR, must include: a description of the 'processing operations' (in this case, the algorithm) and the purpose of the processing; an assessment of the necessity of processing in relation to the purpose; an assessment of the risks to individual rights and freedoms; and importantly, the measures a company will use to address these risks and demonstrate GDPR compliance, including security measures (Article 35(7)) (Recitals 84, 90).

The DPIA must occur before a company implements a system. That is, a company must assess a system and propose risk-mitigation measures, before data processing takes place (Article 35(1)). But the GDPR also envisions iteration. For example, if the risk posed by a system changes, a company must assess whether it is complying with its own Impact Assessment (Article 35(11)). It should also under such circumstances review and possibly revise the DPIA itself.

The DPIA Guidelines suggest an even more dynamic view of DPIAs. They suggest that DPIAs should as a matter of good practice actually be continuous, 'updated throughout the lifecycle [of the] project', and that they should be re-assessed or revised at least every three years. 'Carrying out a DPIA is a continual process, not a one-time exercise', per the DPIA Guidelines.[35] This continual process involves assessing risk, deploying risk-mitigation measures, documenting their efficacy through monitoring, and feeding that information back into the risk assessment. The DPIA Guidelines interpret this process by running 'multiple times'.

The GDPR also lays out procedural requirements for the DPIA. Differing from most of the impact assessments imagined in the literature and discussed in section 'Proposals for AIAs', DPIAs do not involve a period of public comment or input. Many companies that are required to perform AIAs will likely have an internal but independent DPO in place (Article 38).[36] The GDPR requires consultation with this DPO, if a data controller has one.

In lieu of public or formal stakeholder consultation, the GDPR requires consultation 'where appropriate' with impacted individuals (Article 35(9)).[37] This puts in place one method for external input from impacted individuals rather than external experts or the public. The DPIA Guidelines suggest that this input could be, for example, in the form of surveys crafted by companies and sent to future customers. This would make external input less meaningful than, say, deep consultation with a board of representatives of civil society members or chosen community representatives, as envisioned in the literature on impact assessments reviewed in section 'Proposals for AIAs'.[38] The DPIA Guidelines explain that if companies do not seek these external views, they have an obligation to justify this decision.[39] In addition, if companies do seek these views and then disregard them, they must document why they have chosen to disregard external input.[40]

As for other forms of external oversight, the Guidelines recommend but do not require seeking advice from independent experts, ranging from lawyers and sociologists to data security experts.[41] The GDPR does not generally require most DPIAs to be overseen by a public authority (the Data Protection Authority). But if a risk assessment indicates that processing would result in *high risk* in the absence of measures taken by the controller to mitigate the risk, then a company must consult with the regulator before processing (Article 36). Thus a company effectively decides itself whether it should be subject to regulatory oversight, as part of the impact assessment process.

In the biggest departure from the impact assessment proposals discussed below, DPIAs are not legally required to be released to the public, even when finalized.[42] As the Guidelines explain, '[p]ublishing a DPIA is not a legal requirement of the GDPR ... [h]owever, data controllers should consider publishing their DPIA, or perhaps part of their DPIA.'[43] The Guidelines caution that it is a good practice to publish DPIAs, especially where members of the public are impacted. But companies need not publish the entire assessment; the

35 Article 29 Data Protection Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is 'likely to result in a high risk' for the purposes of Regulation 2016/679, 14.

36 Ibid 15. See also Article 29 Working Party Guidelines (n 4) 30: 'An additional accountability requirement is the designation of a DPO, where the profiling and/or the automated decision-making is a core activity of the controller and requires regular and systematic monitoring of data subjects on a large scale (Article 37(1)(b).'

37 In the original proposal of the Commission, consultation with data subjects was mandatory (art 33[4]). The Parliament's text argued that this 'represents a disproportionate burden on data controllers' (amendment 262). Accordingly, the approved art 35(9) requires consultation only 'where appropriate' and 'without prejudice to the protection of commercial or public interests or the security of the processing operations'. Reuben Binns, 'Data Protection Impact Assessments: A Meta-Regulatory Approach' (2017) 7(1) International Data Privacy Law 28.

38 Article 29 Guidelines on Data Protection Impact Assessment (n 35) 15. See also Dariusz Kloza and others, 'Data Protection Impact Assessments in the European Union: Complementing the New Legal Framework towards a More Robust Protection of Individuals' D.Pia.Lab Policy Brief No 1/2017, n.d., 4 <https://cris.vub.be/files/32009890/dpialab_pb2017_1_final.pdf> accessed 8 October 2020.

39 Article 29 Guidelines on Data Protection Impact Assessment (n 35) 15.

40 Ibid 15.

41 Ibid 15.

42 Ibid 18.

43 Ibid 17.

published DPIA 'could even consist of just a summary of the DPIA's main findings'.[44] As some scholars have remarked, there are cases in which full disclosure of the assessment results may be limited by the legitimate interests of the data controller, such as interests in the confidentiality of information, in security, and in competition.[45]

The GDPR and the DPIA Guidelines thus give an overview of, but little specific guidance on, what exactly a company must put in a DPIA report in the context of AIAs. Unlike the impact assessments proposed in the legal literature, the GDPR does not require public input or public disclosure, though the DPIA Guidelines suggest both as best practices. This has led one policy proposal to dismiss the GDPR's DPIAs as 'not shared with the public, and hav[ing] no built-in external researcher review or other individualized due process mechanisms'.[46] As we discuss below, this is not entirely correct, if one re-evaluates the role of the DPIA in the specific context of ADM.

What is the purpose of a DPIA, in the context of the GDPR's algorithmic governance?

Having discussed the requirements for a generic DPIA, we now turn to the specific application of the DPIA as an 'algorithmic impact assessment', aided both by our understanding of the GDPR's approach to algorithmic governance, and by the Guidelines on ADM. Thus far few commentators have linked the GDPR provisions on ADM to the DPIA process.[47] The DPIA as an AIA plays a particularly central role in the GDPR's approach to governing algorithmic decision-making. We posit that in the context of the GDPR's algorithmic governance regime, the DPIA should be understood as a nexus between the GPDR's two approaches to governing algorithmic decision-making. The DPIA links the GDPR's individual rights to its systemic governance of algorithms.

Understanding the DPIA in this way both clarifies its potential content and leads us to observations about how the DPIA as an AIA might be implemented and even improved. The DPIA is not a perfect AIA. As a tool in the GDPR's overall algorithmic governance regime, however, it has more potential than might initially meet the eye.

How understanding the DPIA's dual role helps clarify its content

The DPIA has two roles: as a tool in the GDPR's systemic (and collaborative) governance regime, and as an element of the GDPR's protection of individual rights. Understanding the DPIA in this way—as a connection between the two regulatory subsystems—lets us better understand how it is meant to function as an AIA, to the extent of further clarifying its content. It also leads us to some insights in the next section ('Towards a model AIA') about the layers of algorithmic explanations produced by, and to be released according to, the GDPR.

When understood as part of the GDPR's collaborative governance of algorithms,[48] the DPIA can be characterized as a form of monitored self-regulation. That is, the DPIA tasks companies with identifying problems and coming up with solutions, with internal oversight and some external input, under a threat of regulatory oversight but ordinarily minimal regulatory supervision. Binns has similarly identified the DPIA as 'meta-regulation'.[49]

Monitored self-regulation attempts to change both a company's decision-making processes and its decision-making heuristics.[50] Collaborative governance generally is centrally concerned with affecting management culture and creating meaningful changes within a company.[51] The DPIA, applied in the context of algorithmic decision-making, tasks companies with considering risks of unfairness, error, bias, and discrimination, and with coming up with concrete ways of mitigating those risks. This aims to affect firms' decisional heuristics by dictating, through the GDPR's text, the Recitals, and the Guidelines, what values a company must consider in building and overseeing algorithmic decision-making.

44 Ibid.

45 Alessandro Mantelero, 'AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment' (2018) 34(4) Computer Law & Security Review 766; Frank Vanclay and others, 'Social Impact Assessment: Guidance for Assessing and Managing the Social Impacts of Projects' (International Association for Impact Assessment, April 2015) <https://www.iaia.org/uploads/pdf/SIA_Guidance_Document_IAIA.pdf> accessed 8 October 2020; Simon Walker, *The Future of Human Rights Impact Assessments of Trade Agreements*, School of Human Rights Research Series, v. 35 (Antwerp; Portland: Intersentia 2009) 39–42.

46 Dillon Reisman and others, 'Algorithm Impact Assessment: A Practical Frameworks for Public Agency Accountability' (AI Now Institute, n.d.) 7 <https://ainowinstitute.org/aiareport2018.pdf> accessed 8 October 2020.

47 But see Casey, Farhangi and Vogl (n 3) 170–84, largely focused on practical compliance, discussing the Guidelines on Automated Decision-Making and DPIA.

48 Kaminski (n 5) 57.

49 Binns (n 37) 23, 29 has similarly described DPIAs as 'meta-regulation,' which he characterizes as a narrower subset of co-regulation, 'a means for the state to make corporations responsible for their own efforts to self-regulate'.

50 Bamberger (n 8) 435.

51 Alexander A Boni-Saenz, 'Public-Private Partnerships and Insurance Regulation' (2008) 121 Harvard Law Review 1375; Freeman (n 8); Bamberger (n 8).

The process of conducting the DPIA—taking input from impacted individuals, consulting with an independent DPO, consulting with a regulator where required, and involving both internal and external experts and stakeholders—is meant to change internal company processes.[52] Baking in a compliance culture can be valuable, even where public oversight and input is not sought.[53] The DPIA can also be understood in this context as a necessary precursor to government regulation, serving as a documentation requirement, and creating records that can later be sought and inspected by regulators under the GDPR's extensive information-forcing capabilities.[54]

The DPIA also, however, has an unexplored role in the GDPR's system of individual rights.

First, the DPIA can serve as a source of material for the much-discussed disclosures to individuals about algorithmic decision-making: the individual notification and access rights. Remember, data subjects have a right to receive 'meaningful information' about the 'logic involved, as well as the significance and the envisaged consequences' of ADM (Articles 13, 14, 15). A DPIA must contain, as mentioned above, 'a systematic description of the envisaged processing operations and the purposes of the processing. . .' (Article 35(7)). If companies must already internally describe ADM at a systematic level as part of the DPIA process, those internal descriptions could be disclosed to individuals, or at least serve as the basis for these disclosures, in addition to being released to the public in the form of summaries.

Similarly, a DPIA must include an assessment of 'the risks to the rights and freedoms' of individuals, and individuals have a right in the context of ADM to be informed of the 'significance and envisaged consequences' of such decision-making (Articles 35, 13, 14, 15). Again, as a company conducting ADM must conduct a DPIA, it should consider how the information it produces in that process might also feed into or even satisfy the individual rights requirements under the GDPR.

Secondly, the DPIA as envisioned by the Guidelines on ADM can push companies to establish protections for individual rights as part of the risk-mitigation process. Despite other commentators' dismissal of DPIAs as failing to put in place individual due process,[55] the DPIA is an essential aspect of establishing suitable measures to safeguard individual rights, including per the Guidelines on ADM a version of individual due process. The GDPR requires companies using solely ADM, under the exceptions to its ban on such practices, to implement suitable measures to protect individual rights (Article 22). Data controllers should use DPIAs to 'identify what measures they will introduce to address the. . . risks involved'.[56] The Guidelines on ADM suggest that measures include not just the use of audits or other forms of systemic accountability, but also a number of recognizable individual rights: informing individuals about the logic involved, explaining the significance and envisaged consequences of algorithmic decision-making, providing a way to contest a decision, and providing a way to express one's point of view.[57] The Guidelines on ADM counsel companies to import these various individual rights laid out in the GDPR's Article 22 as a form of risk management throughout the DPIA process. They suggest implementing these individual rights as part of a risk-management strategy even for algorithmic decision-making that is not 'solely automated', but that more significantly involves a human decision-maker.

In other words, we might interpret the GDPR provisions on the DPIA as serving as a form of commitment-making to protecting, or even enabling, individual algorithmic due process rights. By characterizing these individual rights as *risk-mitigation measures*, the Guidelines on ADM both provide a substantive backstop as to what must be included in a DPIA, and task companies with *constituting*—through the process of performing a DPIA—what these individual rights will look like in practice. Thus the DPIA serves as a means of expanding company commitments, changing company decision-making heuristics to include an assessment of individual due process rights. It simultaneously serves as a collaborative governance mechanism used to involve companies in constituting the substance, in practice, of individual due process rights. [58]

Finally, the DPIA has a role in linking the GPDR's system of collaborative governance to its individual rights regime through the imposition of systemic accountability measures such as audits or external reviews. Remember, the general DPIA Guidelines only suggest, and do not mandate, consultation with external experts. In the context of algorithmic decision-making, however, external expert involvement and oversight is more

52 See also Binns (n 37) 23.

53 Bamberger (n 8) 467. See also Sonia K. Katyal, 'Private Accountability in the Age of Artificial Intelligence' (2019) 66 UCLA Law Review 140.

54 Article 29 Guidelines on Data Protection Impact Assessment (n 35) 20. See also the investigatory powers of Data Protection Authority at art 58(1) GDPR. See also Selbst and Barocas's call for documentation requirements in 'The Intuitive Appeal of Explainable Machines' (2018) 87 Fordham Law Review 1085.

55 Reisman and others (n 46) 10.

56 Article 29 Working Party Guidelines (n 4) 30.

57 Ibid 30.

58 Kaminski (n 5) 18.

necessary and can be understood as an essential risk-mitigation measure for algorithmic decision-making.[59]

The reasoning goes as follows. Recital 71 requires, in the context of algorithmic decision-making, the use of '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. . .and that prevents, inter alia, discriminatory effects' (Recital 71). Malgieri and Comandé observe that this requirement effectively expands the GDPR's 'suitable safeguards' requirement from the series of individual due-process-like protections enumerated in the GDPR's text, to a far broader set of systemic accountability measures, including third-party auditing (Article 22).[60]

The Guidelines on ADM's list of best practices for suitable safeguards over algorithmic decision-making includes recommendations that companies use both internal and external audits and external review boards, supporting this interpretation.[61] This means that in the context of algorithmic decision-making, a company running through the cyclical DPIA process discussed above will likely incorporate external oversight and input at the risk mitigation stage, bringing external input into the cycle despite the fact that it is not a formal procedural requirement for DPIAs in general.

Conceptually, the implications of this are even broader. By characterizing third-party and expert oversight as a form of 'suitable safeguard' or 'suitable measure' to protect *individual rights* in the face of ADM, the recommendation in the Guidelines on ADM links individual rights protection with collaborative governance techniques. Companies are tasked with coming up with ways to prevent error, bias, discrimination, and other harms to individual rights, and external oversight is imposed over how they choose to address these problems. That external oversight itself is also conceptualized as a crucial aspect of individual rights in the GDPR, standing in for individuals to ensure that they are not subjected to an unfair, arbitrary, discriminatory, or erroneous system.

A simpler way to say this is that expert oversight in the DPIA process serves two, or even three, roles: it watches the companies as they come up with ways of addressing problems with algorithmic decision-making, and it reassures individuals that their dignity and other rights are being respected by a fair system.[62] It also provides legitimation or justification. As the mechanism through which this external oversight is implemented, the DPIA thus connects the two approaches to algorithmic governance in the GDPR.

Shortcomings of the DPIA

The biggest shortcoming of the DPIA is that it does not include a mechanism for mandatory disclosure to the public.[63] Public disclosure, as discussed in section 'Elements of a model AIA', is understood by many to be an essential element of impact assessments as a policy tool.[64] Public-facing disclosure enables public feedback, both in the form of market feedback (enabling individuals to avoid companies with bad policies) and in the form of regulatory feedback over the longer term (enabling individuals to elect representatives who will put in place laws that will prevent bad company behaviour). By failing to mandate public disclosure, the GPDR's DPIA fails to trigger both of these mechanisms, which are essential components of a functioning collaborative governance regime.

This failure could be drastic. The GDPR puts a lot of faith in the behaviour of companies and in the capacity of regulators. As discussed, the GDPR often tasks companies with coming up with the substance of (i) how individual rights will be implemented and (ii) how to address unfairness, biases, and discrimination-related concerns about algorithms. In the absence of public oversight, how can we be sure that this hybrid system of individual rights and collaborative governance is working towards the public good?

One possible answer is to use heavy regulatory oversight. But the GDPR's enforcers have not, historically, been well-resourced in relation to the companies they regulate. Tasking regulators with extensive monitoring also forgoes some of the touted benefits of governing through public–private partnerships, including lowered costs and incorporating external third-party expertise. By failing to require the public disclosure of impact assessments, the GDPR fails to activate necessary third parties in its governance regime, such as civil society actors or civic-minded experts who might not be recruited for auditing purposes. The DPIA also

59 See Article 29 Working Party Guidelines (n 4) 32.

60 Malgieri and Comandé (n 2) 248.

61 Article 29 Working Party Guidelines (n 4) 28, 32. See Casey, Farhangi and Vogl (n 3) 170–80, emphasizing the centrality of algorithmic audits.

62 See Kaminski (n 5) 28; Binns (n 37) 32 discusses a similar notion in the regulatory theory literature, Gilad's concept of regulatory tiers. Gilad, 'It Runs in the Family: Meta-regulation and Its Siblings' (2010) 4(4) Regulation & Governance485, 497.

63 Kloza and others (n 38) 3; Michael Veale, Reuben Binns and Jef Ausloos, 'When Data Protection by Design and Data Subject Rights Clash' (2018) 8(2) International Data Privacy Law 118.

64 Reisman and others (n 46) 13. See also Andrew D Selbst, 'Disparate Impact in Big Data Policing' (2017) 52 Georgia Law Review 119. See also A. Michael Froomkin, 'Regulating Mass Surveillance as Privacy Pollution: Learning from Environmental Impact Statements' (2015) 2015 *University* of *Illinois Law Review* 1790.

potentially fails to involve real stakeholder input, if companies follow the Guidelines and consult with stakeholders using only simplistic surveys.[65]

Individual notification and access rights could do some of the necessary accountability work for the GDPR's attempts at collaborative governance. This is somewhat more convincing. If companies indeed link their DPIA content to what they disclose to individuals (for example, disclosing the systemic description of processing uncovered during a DPIA to individuals as the 'logic involved' in a decision-making system), then it is likely that these disclosures will make their way to other third parties, who may be able to provide the expertise and oversight over company self-governance. For example, an individual who feels she has been discriminated against might disclose the information she has received about a system's decisions to a civil society group, which could in turn help publicize the story and the information, triggering market mechanisms or regulatory feedback from the public or oversight by external experts. This is, however, a more attenuated way of getting at the same outcome as public disclosure, and risks failing entirely if companies significantly disaggregate the DPIA process from individual disclosure rights.

A model AIA: towards multi-layered explanations

From examining the GDPR's DPIA mechanism generally to discussing the DPIA as an AIA, we now turn to imagining a more ideal DPIA.

We close this article with a call for more work on establishing a model AIA that could serve as a basis for what we call *multi-layered explanations* of algorithmic decision-making. This will involve interdisciplinary efforts: technologists to assess what risk-mitigation and accountability measures could be implemented, and lawyers and ethicists to think through how to better involve constituents and define problems. It will also involve a deeper exploration of how to link the material created during the DPIA process to the individual disclosures required under the GDPR.

We are not the first to focus on AIAs, or impact assessments, in closely related fields.[66] We are, however, the first to discuss AIAs not in isolation, but as a central component, among many components, of the GDPR's two-prong approach to algorithmic accountability. This changes the nature of the conversation. Instead of examining impact assessments in isolation from other accountability tools, it situates them within an overarching governance system.

Our GDPR-specific analysis, then, may have implications for proposals for AIAs in other legal systems.[67] It suggests that impact assessments best serve a role in conversation with other accountability tools, as part of overarching regulatory design.[68] And it suggests that impact assessments play a central role both as a source of and mediator between the multi-layered individual explanations we believe are indicated in the GDPR.

Proposals for AIAs

We begin with an overview of the discussion that has arisen recently over AIAs. AIAs have received a good deal of attention on both sides of the Atlantic as possible tools to address problems of algorithmic discrimination, bias, and unfairness—including in at least one proposed US federal law.[69] We here briefly discuss several important precursors to the AIA: Environmental Impact Statements (EISs), Human Rights Impact Assessments (HRIAs), Privacy Impact Assessments (PIAs), Ethical Impact Assessments (EIAs), and Surveillance Impact Assessments (SIAs). It is important to clarify that, apart from the EIS, many of the below impact assessment models are voluntary. That is, they are not required by law in any legal system. In this section, we discuss how these different elaborations, most taking inspiration from the EIA under US law, have led to more recent proposals for AIAs.

65 See Binns (n 37) 33; Casey, Farhangi and Vogl (n 3) 180.

66 See eg Kenneth A Bamberger and Deirdre K. Mulligan, 'PIA Requirements and Privacy Decision-Making in US Government Agencies' in D Wright and P De Hert (eds), *Privacy Impact Assessment*, Law, Governance and Technology Series (Dordrecht: Springer 2012) 225; Reuben Binns, 'Data Protection Impact Assessments: A Meta-regulatory Approach' (2017) 7 International *Data Privacy* Law 22; Casey, Farhangi and Vogl (n 3); Roger Clarke, 'Privacy Impact Assessment: Its Origins and Development' (2009) 25 Computer Law & Security Review 123; Froomkin (n 64) 1713; Chris Jay Hoofnagle, 'Assessing the Federal Trade Commission's Privacy Assessments' (2016) 14(2) IEEE Security & Privacy 58; Katyal (n 53) 54, 112; Mantelero (n 45) 754; David Wright and Charles D. Raab, 'Constructing a Surveillance Impact Assessment' (2012) 28 Computer Law & Security Review 613; Marc L Roark, 'Human Impact Statements' (2015) 54 *Washburn* Law *Journal* 649; Reisman and others (n 46); Selbst (n 64) 169; David Wright and Michael Friedewald, 'Integrating Privacy and Ethical Impact Assessments' (2013) 40 Science and Public Policy 755.

67 See eg more generally Reisman and others (n 46).

68 Katyal (n 53) 117 suggests this by emphasizing the concurrent need for whistleblower protection. But not an overarching governance system.

69 Wyden, Clarke, and Booker's Algorithmic Accountability Act. See <https://www.wyden.senate.gov/news/press-releases/wyden-booker-clarke-introduce-bill-requiring-companies-to-target-bias-in-corporate-algorithms-> accessed 8 October 2020.

The inspiration for many US-based impact assessment proposals is the EIS, established in the USA in 1969 by the National Environmental Policy Act (NEPA).[70] NEPA's impact statement requirement applies when a federal agency proposes to take a 'major Federal action significantly affecting the quality of the human environment'.[71] As a threshold matter, agencies assess coverage and query whether any 'Categorical Exclusions' apply.[72] If no exclusion applies, the agency as a first step performs an Environmental Assessment (EA), a public document that must 'provide sufficient evidence and analysis for determining whether to prepare an environmental impact statement'.[73] Then the agency either issues a Finding of No Significant Impact (FONSI), or goes on to prepare an EIS. Before a project can go forward, the full EIS must be prepared and must contain a detailed statement on the environmental impact of a project, including any adverse effects which cannot be avoided, and alternatives.[74] The EIS is subject to comment by the public and other government agencies,[75] and individuals can sue if an EIS is incomplete or inadequate, thereby delaying the project.[76]

A number of US commentators have used the EIS as a model for impact assessments in other contexts. Froomkin, for example, touts the EIS as an effective alternative to command-and-control regulation, and a model for his proposed Privacy Impact Notice.[77] According to Froomkin, the EIS is a good regulatory model for data privacy notices because it (i) pushes agencies to 'consider ... issues in the early design phase of their projects'[78] and (ii) informs the public and solicits public feedback.[79] Selbst, who similarly bases his call for Algorithmic Impact Statements (AISs) on the EIS model, agrees that an impact assessment is an 'action-forcing' regulation that 'push[es] decision-makers to do their homework and engage with the public'.[80] Selbst describes the EIS model not as an alternative to substantive regulation, but as a necessary precursor to it.[81] Froomkin, too, notes that public transparency can 'ignite a regulatory dynamic by collecting information about the privacy costs of previously unregulated activities that should, in the end, lead to significant results'.[82]Thus, these two scholars envision impact statements as having positive consequences both for the particular project at issue and for forward movement in a larger policy debate.

For some, however, the EIS model fails to go far enough. The EIS process is static in nature, taking place only prior to the commencement of a project.[83] It is procedural, rather than substantive; it does not set substantive requirements, nor prohibits anybody from doing anything.[84] And while the EIS process requires public transparency and input, it does not require ongoing monitoring for compliance.[85]

Other proposals for impact assessments thus draw on additional sources as models, some of which in turn also trace their origins to the EIS.[86] Mantelero, for example, draws partially on the model of HRIAs.[87] Katyal, too, references the HRIA process.[88] The voluntary HRIA process outlined by the United Nations[89] is a comparatively time- and resource-intensive process conducted on a business by third-party assessors, who collect data and interview stakeholders, experts, and management.[90] Wright and Friedewald look to EIAs as a model for PIAs. EIAs, similar to HRIAs, are voluntary assessments that go beyond legal compliance to assess the ethical implications of new technologies, and

70 (2012) 42 USC s 4332(C). See Reisman and others (n 46) 7 ('The EIS process combines a focus on core values with a means for the public, outside experts, and policymakers to consider complex social and technical questions'); Selbst (n 64) 168 ('before adopting predictive policing technology, police should be required to create "algorithmic impact statements" (AISs), modeled on the environmental impact statements (EISs) of the National Environmental Policy Act (NEPA)') See also Froomkin (n 64) 1749; See Roark (n 66) 18.

71 42 USC s 4332(2)(C) (1969); 40 CFR s 1508.18 (defining 'Major Federal Action'). This includes 'projects and programs entirely or partly financed, assisted, conducted, regulated, or approved by federal agencies'. Thus NEPA can apply to the behaviour of private actors, where they are funded or permitted by a federal agency. In fact, private actors applying for federal permits often participate in the EIS/EA process. See Froomkin (n 64) 1751, fn 205.

72 See Froomkin (n 64) 1750.

73 40 CFR s 1508.9(1), (3)(b) (1969).

74 42 USC s 4332(C) (2012); 40 CFR s 1502.14.

75 See Selbst (n 64) 178 (describing the two notice-and-comment periods, one to define the scope of the EIS and the second on the draft).

76 Froomkin (n 64) 1751.

77 Ibid 1755.

78 Ibid 1756

79 Ibid 1746

80 Selbst (n 64) 169.

81 ibid 168 ('It is hard to say in the abstract what stronger regulatory solutions may be required, or how big a problem the technology poses in reality, until more information about the technology's implementation is created.').

82 Froomkin (n 64) 1747.

83 Selbst (n 64) 172.

84 See Katyal (n 53) 115.

85 Selbst (n 64) 188; Clarke (n 66) 125 (describing an EIS as 'insufficiently auditable').

86 See Clarke, ibid 125. See also Mantelero (n 45) 757, which describes HRIA's as having their roots in the EIS.

87 Mantelero, ibid 762. His HRSEIA is a (voluntary) hybrid. Lighter touch than HRIA, but takes into account ethical, social, human rights (grounded in human rights law).

88 Katyal (n 53) 112.

89 United Nations, Human Rights Council, Office Of The High Comm'r, *Guiding Principles on Business and Human Rights* (2011) 23–26 <https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf> accessed 8 October 2020.

90 Mantelero (n 45) 764.

involve consultation with a wide number of stakeholders and publication of the assessment.[91]

However, the most direct precursor for the GDPR's version of the AIA is the PIA.[92] As Clarke explains, PIAs originated in the 1990s around the world, with multiple regulators issuing guidance in the early 2000s.[93] While PIAs as conducted in the USA have been widely decried as toothless,[94] elsewhere they are considerably more substantial.[95] Clarke identifies the EIS as a 'progenitor' of the PIA, but goes on to name a number of important differences.[96] In several European countries, for example, the PIA may have originated as part of the system of 'prior checking' under earlier data protection regimes, which was effectively a system of government registration or licensing of data processing systems, prior to processing.[97] In order to receive a license from a national authority, a company had to assess whether it was in compliance with national data protection law. This differs vastly from the EIS, which has no substantive underpinnings and does not serve as the basis for a licensing regime.

Clarke outlines the characteristics of an ideal PIA. He describes the assessment as being performed on a project rather than an organization; being anticipatory in nature rather than retrospective; being broad in scope with respect to individual, group, community, and other 'dimensions' of privacy; taking into account the perspectives of affected segments of the population; being broader than legal compliance; being oriented towards surfacing solutions, not just problems; emphasizing process over product; and requiring engagement from executives and managers.[98] In 2012, Wright and Raab proposed the concept of a Surveillance Impact Assessment,[99] wider in scope than a PIA but consisting of a similar 'process of engaging stakeholders in order to identify the impacts on privacy and other values of a new project, technology, service or other initiative in order to take remedial action to minimise, avoid or overcome the risks'.[100]

But these proposals articulate the ideal. Mantelero observes that in practice, DPIAs even in the European context have tended to focus on data quality and data security, leaving out broader social and legal impact despite aspirational language to the contrary.[101]

We now turn to recent proposals for AIAs, which draw to varying degrees on these precursors. We find both common threads and significant differences in the proposals. We also find a significant gap in this literature that our perspective on the GDPR helps to fill.

Selbst proposes the use of an AIS, modelled after the EIS with some modifications. His proposal would apply narrowly to police departments looking to acquire and use predictive policing technologies. An AIS would, in Selbst's proposal, be performed prior to using such technology. First, this Statement would, like an EIS, require policy departments to 'rigorously explore and objectively evaluate all reasonable alternatives', including by having third-party vendors '(1) explain the various design choices, (2) measure the resulting efficacy using the best available audit methods, and (3) evaluate the resulting disparate impact for the various systems and configurations'.[102] Secondly, a police department would have to 'devote substantial treatment to each alternative'.[103] It would be required to 'include the alternative of no action'.[104] It would be required to identify a preferred alternative among the various algorithm design choices disclosed.[105] And finally, police would have to include proposed mitigation measures in the AIS.[106] To

91 Wright and Friedewald (n 66) 755–66; See previously on this point David Wright and Emilio Mordini, 'Privacy and Ethical Impact Assessment', in Wright and De Hert (eds) (n 66) 397–418.

92 See Binns (n 66) 23; Clarke (n 66); Wright and Friedewald (n 66) 757–58. Wright and Raab (n 66) 755.

93 Clarke (n 66).

94 See eg Ibid 128; Hoofnagle (n 66) 64; Bamberger and Mulligan (n 66) 250.

95 Clarke (n 66) *passim*.

96 Ibid 125.

97 Binns (n 66) 24; Clarke, ibid 125. See also G Le Grand, and E Barrau, 'Prior Checking, A Forerunner to Privacy Impact Assessments' in Wright and De Hert (eds) (n 66) 97–115.

98 Clarke (n 66) 124–25.

99 See, previously, from the same authors: Charles Raab and David Wright, 'Surveillance: Extending the Limits of Privacy Impact Assessment' in Wright and De Hert (eds) (n 66) 363–83.

100 See Wright and Raab (n 66) 615, describing the 16 steps as follows:
Determine whether a PIA (or SIA) is necessary (threshold analysis).
Identify the PIA (or SIA) team and set the team's terms of reference, resources and time frame.
Prepare a PIA (or SIA) plan.
Determine the budget for the PIA (or SIA).
Describe the proposed project to be assessed.
Identify stakeholders.
Analyse the information flows and other impacts.
Consult with stakeholders.
Determine whether the project complies with legislation.
Identify risks and possible solutions.
Formulate recommendations.
Prepare and publish the report, eg on the organisation's website.
Implement the recommendations.
Ensure a third-party review and/or audit of the PIA (or SIA).
Update the PIA (or SIA) if there are changes in the project.
Embed privacy awareness throughout the organisation and ensure accountability.

101 Mantelero (n 45) 761.

102 Selbst (n 64) 173.

103 ibid.

104 ibid 176.

105 ibid.

106 ibid 177.

address various concerns about the EIS model, Selbst emphasizes the importance of public disclosure and comment, and judicial oversight with not just procedural but substantive bite.[107]

Katyal incorporates elements of Selbst's proposal into her suggestion of a Human Impact Statement in Algorithmic Decision-making.[108] She recommends as a backstop a substantive, rather than purely procedural, commitment to algorithmic accountability and anti-discrimination.[109] She adds that companies should also (i) identify potentially impacted populations and determine their status-based categories; (ii) identify the effect of uncertainty or error on those groups; and (iii) study whether the decision will have an adverse impact on a particular subpopulation.[110] Unlike Selbst, Katyal recommends the HIS as a voluntary measure undertaken by private industry, rather than required by law.

The AI Now Institute, a research institute housed at New York University,[111] issued a report that appears to build on Selbst's proposal.[112] The authors of the report call for a pre-procurement AIA before any public agency—not just the police—commits to the use of an ADM system.[113] Like Selbst's proposal, the AI Now proposal is limited to covering the public sector. Like Selbst's proposal, it would be mandatory rather than voluntary. Unlike Selbst's proposal, it goes beyond the policing context.[114]

At first glance, the AI Now proposal looks similar to an EIS in a number of ways. Like an EIS, the AIA must be done prior to implementing a project. Like an EIS, the proposed model requires agency disclosure and a public comment period. Unlike an EIS, however, the proposed model is envisioned as being renewed every two years.[115] Unlike an EIS, which does not create a system for ongoing monitoring, a substantial portion of AI Now's proposal is dedicated to ongoing processes to be established by the AIA, including both meaningful access for researchers and auditors once systems are deployed,[116] and individual due process for those affected by the system's decisions.[117]

Finally, we return to the context of the GDPR. Mantelero discusses the idea of a Human Rights, Social and Ethical Impact Assessment (HRSEIA) in the AI context.[118] A hybrid between a HRIA and a PIA, the HRSEIA suggests that businesses voluntarily take into account ethical and social impact, in addition to human rights.[119] Mantelero emphasizes the role of such an impact assessment in addressing the collective dimensions of data harms, that is, the impact of surveillance or processing on groups or locations.

At its core, Mantelero's HRSEIA has three features: it is participatory, it is transparent, and it is circular in nature.[120] Practically, it consists of a self-assessment questionnaire, sometimes leading to evaluation by an ad hoc committee of experts.[121] Stakeholder engagement is encouraged but not required.[122] Similarly, public disclosure is encouraged.[123] Mantelero explains that while this proposal is '[i]n line with the declared intent of the GDPR', he does not understand the GDPR to require a HRSEIA.[124] Several other commentators have recently discussed the DPIA and the role it plays in the context of algorithmic accountability more generally.[125]

Notably, many or even most of the above proposals for impact assessments centrally emphasize the release of information to the public.[126] This is necessary both to obtain external input into how a system is developed, trained, or monitored, and to gain public legitimacy and acceptance for the use of a system. The kind of information released to the public can be more in the nature of a summary or an overview; it is not necessarily the source code.[127] Some suggest a tiered release of information, with summaries released to the public and detailed or sensitive information released only to regulators or experts.[128] Thus, more recent proposals call for expert input and oversight as a central component of the impact assessment process—that companies (or government agencies) use impact assessments to come up with, and stick to, a plan for third-party expert oversight over a system's development and eventual ongoing use.[129]

107 ibid 178.
108 Katyal (n 53) 115.
109 ibid.
110 ibid 116.
111 Resiman and others (n 46) cit.
112 ibid.
113 Ibid 8.
114 ibid 6.
115 ibid 10.
116 ibid 18
117 ibid 16.
118 Mantelero (n 45) *passim*.
119 ibid 762.
120 ibid 759.
121 ibid 758.
122 ibid 769.
123 ibid.
124 ibid 762.
125 Casey, Farhangi and Vogl (n 3) 170; Edwards and Veale (n 3) 77–80.
126 Selbst (n 64) 118; Reisman and others (n 46) 4.
127 Council of Europe, 'Guidelines on the Protection of Individuals with Regard to the Processing of Personal Data in a World of Big Data' (Strasbourg, 23 January 2017) 4; Selbst (n 64) 190; Kristian Lum and William Isaac, 'To Predict and Serve?' (2016) 13(5) Significance 14–19.
128 Mantelero (n 45) 766.
129 Christian Sandvig and others, 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms' (2014), <http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20–

Lessons for calls for AIAs generally

Our GDPR-specific analysis has implications for proposals for AIAs generally. Our research into the GDPR's version of an AIA suggests that the proposals discussed above have largely missed several important observations.

First, AIAs are not best understood as a stand-alone mechanism. In the context of the GDPR, they are one part of a much larger system of governance.[130] Only one author among the above—Katyal—considers how impact assessments interact with other tools in the regulatory toolkit (discussing the concurrent need for whistleblower protection and exemptions from trade secrecy law).[131] In the context of the GDPR, both Edwards and Veale and Casey and others point to the DPIA's role in algorithmic accountability, but do not discuss at length its relationship to other accountability tools in the GDPR.[132] Our analysis suggests that impact assessments are just one tool in a larger regulatory ecosystem, and may work best when they are not deployed alone and are instead understood as entwined with other regulatory tools such as individual rights.

Secondly, impact assessments can serve as a connection between collaborative governance and individual rights.[133] The information a company creates during the impact assessment process can feed into what it provides to individuals and to the public at large. The procedures an impact assessment puts in place can serve not just to prevent error, bias, and discrimination, but also to legitimize a system or even respect an individual's dignity within it. This dual role is exemplified by the GDPR's DPIA. In the GDPR context, we found one author, Binns, who identified that the GDPR's version of impact assessments is a kind of collaborative governance with the private sector (or what he identifies as 'meta-regulation').[134] Binns, however, did not examine how the DPIA connects to the broader system of both collaborative governance tools and individual rights in the GDPR.

Thirdly, as part of a larger system of governance, there are unexplored connections between the GDPR's DPIA and its underlying substantive individual rights and substantive principles. It is true that many of the GDPR's individual rights and principles about algorithmic decision-making are articulated in broad, sometimes aspirational, terms.[135] Unlike an EIS, the GDPR's version of the AIA has a substantive backstop in, for example, Recital 71's admonishment that a data controller should minimize the risk of error and prevent discriminatory effects. The oddity is the GDPR's circularity: the AIA helps not just to *implement* but to *constitute* both these substantive backstops and the GDPR's individual rights. Thus, there is a substantive backstop to company self-regulation through impact assessments—but it is a moving target, in part given meaning by affected companies themselves.

Finally, because the AIA links individual and systemic governance, we understand the GDPR's version of the AIA to be both the potential source of and the mediator between what we refer to below as 'multi-layered explanations' contemplated in the GDPR. Several of the above scholars, including both Mantelero and Wright and Raab, emphasize the often collective dimensions of surveillance and data processing.[136] The GDPR's system of individual rights threatens by itself to miss the impact of surveillance, or in this case, ADM, on groups, locations, and society at large.[137] A recent AI Now report provides an illustrative example of the problem: providing an individualized explanation for a single 'stop and frisk' incident in New York City would have failed to reveal that over 80 per cent of those subjected to stop and frisk by the NYPD were Black or Latino men.[138] But the Impact Assessment with its systemic approach to risk assessment and risk mitigation requires data controllers to analyse how the system impacts not just individuals but groups. We believe that systemic and group-based explanations uncovered during an AIA can and should be communicated to outside stakeholders, and that a case can be made that such release is required under the GDPR.

%20Sandvig%20–%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf- accessed 8 October 2020; Reisman and others (n 46) 18–20.

130 Edwards and Veale (n 3) 77–80 understand this, as they discuss the DPIA in the context of many other rights in the GDPR. See also Kaminski (n 5) 69.

131 Katyal (n 53) 117.

132 Edwards and Veale (n 3) 77–80; Casey, Farhangi and Vogl (n 3) 170.

133 Only one proposal, to our knowledge, suggests using Impact Assessments to establish something resembling individual rights—a system of 'enhanced due process mechanisms for affected individuals'. Reisman and others (n 46).

134 Binns (n 66) 29 describing DPIAs as 'enforced risk-assessment, and compliance with self-imposed, stakeholder-influenced policies. . . as an instance of 'meta-regulation'.

135 Mantelero (n 45) 765 (discussing how 'Data protection laws adopt general principles. . . and general clauses. . . which are used to introduce non-legal social values into the legal framework').

136 Mantelero (n 45) 762–63; Wright and Raab (n 66) 615.

137 There is a growing field of scholarship devoted to 'collective data protection.' See Mantelero (n 45) 757, fn 21; Linnet Taylor, Luciano Floridi and Bart van der Sloot (eds), *Group Privacy: New Challenges of Data Technologies* (Cham, Springer International Publishing, 2017).

138 Reisman and others (n 46) 19.

Towards a model AIA: producing multi-layered explanations

We close with a framework for a model AIA under the GDPR. More research is clearly needed in both the technological and policy space to implement this in practice. But we hope to start that conversation here. Our envisioned model AIA is informed both by our understanding of the GDPR and by our overview of impact assessment proposals above. Like previous AIA proposals, we emphasize stakeholder input, expert oversight, and public disclosure as essential elements of an effective impact assessment. Unlike previous AIA proposals, we deploy our understanding of the GDPR to suggest how the DPIA process informs and connects into the GDPR's system of individual rights, through disclosures that we refer to as *multi-layered explanations*. Thus, our conclusions inform not only the ongoing policy discussion of impact assessments, but the ongoing debate about individual algorithmic transparency rights in the GDPR.

Elements of a model AIA

A model AIA process should do at least the following. It should contemplate the involvement of civil society as a form of underused oversight. It should better involve and engage impacted individuals, not just through surveys but through representative boards, before an algorithm is deployed. It should contemplate requiring companies, or regulators, to help fund the involvement of both of the above and provide technical expertise or the resources for obtaining technical expertise. It should involve not just external technical experts, but external experts in law and ethics to help define, or at least frame discussions of, what we mean by terms like 'discrimination' or 'bias'.[139]

A model AIA process should also deliberately widen the lens from algorithms as a technology in isolation, to algorithms as systems embedded in human systems—both those who design the technology and those who use it.[140] There is a growing awareness that addressing problems of unfairness or bias in the technology in the abstract will be inadequate for mitigating these problems when an algorithm is implemented in practice. The risks come from the technology by itself, and from the humans who embed their values into the technology during its construction and training. Additionally, risks arise from how the humans using the algorithm are trained and constrained, or not constrained, in their use of it.[141] A model AIA should thus be truly continuous: a process that produces outputs or reports, but also includes ongoing assessment and performance evaluation, especially for those algorithms that change quickly over time and are deployed in multiple contexts.

Substantively, a model AIA could take advantage of the fact that it is conducted on a system-wide level to search for, and mitigate, social harms that go beyond impacted individuals.[142] For example, a model AIA could be used to root out discrimination not just against particular individuals but against marginalized communities, identifying discrimination patterns that would be impossible to find through individual disclosures alone.[143] A model AIA could explicitly require an assessment of performance metrics, on a system-wide and ongoing basis, and require disclosure of these metrics to external experts.[144] This would not stretch the purpose of the DPIA—at least one application of which explicitly focuses on collective surveillance in the context of monitoring public spaces—and would fill an existing gap in the GDPR's current algorithmic accountability and disclosure regime.[145]

As to substantive risk-mitigation measures, different data controllers may have different duties. Article 24(1) of the GDPR states that taking into account the nature, scope, context, and risks of data processing, the controller shall implement appropriate technical and organizational measures to ensure compliance with the GDPR. Accordingly, algorithmic decision-making involving bigger risks for data subjects should entail more safeguards. As discussed, data controllers to a certain extent choose their own algorithmic accountability safeguards:

139 For example, the COMPAS recidivism risk assessment algorithm led to a significant public discussion over different ways of defining discrimination. Julia Angwin and others, 'Machine Bias' (*Propublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 8 October 2020 (describing leading risk assessment tools forsentencing and corrections developed by Northpointe); Sam Corbett-Davies others, 'A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not that Clear' *The Washington Post* (17 October 2016) <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-thanpropublicas> accessed 8 October 2020.

140 See Andrew D Selbst and others, 'Fairness and Abstraction in Sociotechnical Systems' in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19 (New York, NY, USA: ACM, 2019) 59–68 <https://doi.org/10.1145/3287560.3287598>accessed 8 October 2020.

141 Ibid 61 on COMPAS Case.

142 Mantelero (n 45) *passim*; L Edwards, D McAuley and L Diver, 'From Privacy Impact Assessment to Social Impact Assessment' in *2016 IEEE Security and Privacy Workshops (SPW)* (2016) 53–57 <https://doi.org/10.1109/SPW.2016.19> accessed 8 October 2020; Wright and Raab (n 66) 613–26; Raab and Wright (n 99).

143 Reisman and others (n 46) 18; see also Pauline Kim, 'Auditing Algorithms for Discrimination' (2017) 166(1) University of Pennsylvania Law Review Online 196.

144 Edwards and Veale (n 3) 80.

145 Reisman and others (n 46) 8. Edwards and Veale (n 3) 80.

some are required in the GDPR both at Article 22(3) and at Recital 71 (algorithmic auditing, the rights to contest, to have a new decision, to a human in the loop, and to explanation), but these are not closed lists, and the Guidelines on ADM suggest additional techniques. In the case of more intrusive and riskier ADM processes, the data controller should and likely will implement all possible safeguards, including a right to explanation of an individual decision.

To make the Impact Assessment process meaningful, Data Protection Authorities must be willing to spot check and enforce against captured versions of it. While the GDPR does not require regulatory involvement in all DPIAs, DPAs could use the GDPR's broad information-forcing powers to inspect particular companies and check for compliance. This spot-checking might work not just to monitor and improve the efficacy of the process, but to identify substantive problems with algorithmic decision-making. DPAs might over time use what they have learned to establish more concrete best practices or support the establishment of sector-specific codes of conducts around algorithmic fairness, as suggested in the Guidelines on ADM. Several implementing Member States have already put in place substantive backstops around algorithmic decision-making, prohibiting decision-making based on particular factors, or that is discriminatory or biased. Slovenia, as mentioned, couples this substantive prohibition against discrimination with a required impact assessment process. This dual approach of linking impact assessments to substantive prohibitions may help to tether internal company risk mitigation measures to the public good.

Comparing DPIA requirements with algorithmic accountability requirements under the GDPR

The more granular but nonetheless central insight arising from our research is this: the DPIA process can and should inform the substance of the GDPR's individual algorithmic accountability rights. That is, when we take a close look at the substance a company or public agency is required to produce during a DPIA, it maps surprisingly well on to the disclosure requirements of the GDPR.

Article 35(7) GDPR requires that a DPIA should contain:

1. a systematic description of the envisaged processing operations and the purposes of the processing, (. . .)
2. an assessment of the necessity and proportionality of the processing operations in relation to the purposes
3. an assessment of the risks to the rights and freedoms of data subjects (. . .); and the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.

This required content of a DPIA largely maps onto the GDPR's individual algorithmic transparency rights. For example, where Article 35(7) requires a DPIA to contain a 'systematic description of the envisaged processing operations', this could be used as the foundation for the GDPR's disclosure requirement that individuals must be informed of 'meaningful information about the logic involved' in algorithmic processing. The Guidelines on DPIAs interpret the *systematic description* to include: the nature, scope, context, and purposes of the processing; categories of personal data, recipients, and storage; and a functional description of the processing operation and the assets on which personal data rely.[146] Any or all of this systematic description of processing produced during a DPIA could feed into what is disclosed to individuals regarding 'meaningful information about the logic involved'.

If we compare the GDPR's transparency requirements for ADM with this and other substantive requirements for DPIAs, the similarities are striking (see Table 1).

The Guidelines on ADM only make these parallels more apparent. The Guidelines on ADM once again echo the substantive requirements of a DPIA (Table 2).

That is, there are numerous similarities between information that a data controller is required to produce during a DPIA and what a data controller is required to release to individuals as part of the GDPR's algorithmic transparency duties. The data controller's duty to systematically describe the processing operations in a DPIA is similar to the algorithmic transparency duty to clarify the categories of personal data used in ADM and how algorithmic profiling is built. The controller's duty to assess the necessity and proportionality of the processing operations in a DPIA is similar to the algorithmic transparency duty to explain the pertinency of personal data used and the relevance of the profiling. The controller's duty to assess the data processing risks and impacts on individuals is similar to the transparency duty to explain the impact of the profiling use in ADM. Lastly, the controller's duty to establish safeguards of individual rights in the case of ADM is similar to the duty to find and describe measures envisaged to address the risks in DPIA.

146 Article 29 Guidelines on Data Protection Impact Assessment (n 35) 22.

Table 1. Comparison between DPIA duties and GDPR algorithmic accountability duties under the GDPR

Data Protection Impact Assessment requirements (Article 35(7))	Transparency rights about Automated Decision-Making (Articles 13–15, 22)
Systematic description of data processing	Meaningful information about the logic involved (Articles 13(2)(f); 14(2)(g); 15(1)(h))
Assessing necessity and proportionality of data processing	Meaningful information about the significance (Articles 13(2)(f); 14(2)(g); 15(1)(h))
Assessing risks to fundamental rights and freedoms	Meaningful information about the envisaged effects of the algorithm (Articles 13(2)(f); 14(2)(g); 15(1)(h))
Mitigating those risks through appropriate measures	Appropriate safeguards, including the right to contestation, individual explanation, human intervention (Article 22(3) and Recital 71)

In other words, in the case of ADM, the outputs of the DPIA process described in the GDPR appear to correspond to algorithmic transparency duties in the GDPR (as interpreted by Article 29 Working Party (now the EDPB)).

Towards multi-layered explanations from an algorithmic DPIA

Our perspective on the DPIA as linking systemic governance to individual rights thus has implications for the GDPR's overall approach to algorithmic explanations. The DPIA process in our view suggests what we call '*multi-layered explanations*' for ADM. These explanations will likely be crafted as part of the DPIA process and should be released either directly to the public or to affected individuals.

We are not the first to observe that there are multiple layers of explanations of algorithmic decision-making required in the GDPR.[147] These stem from the GDPR's two types of individual transparency requirements, articulated in Articles 13, 14, and 15 on individual notice and access, and its algorithm-specific provisions in Article 22. Edwards and Veale in particular have suggested that individuals subject to algorithmic decision-making should be provided both of what they call 'model-centric' and 'subject-centric' explanations.[148]Model-centric explanations, they suggest, should include: the family of model, input data, performance metrics, and how a model was tested. Subject-centric explanations should include counterfactuals (that is, what changes would change the outcome of an individual decision), the characteristics of similarly classified individuals, and the confidence a system has in an outcome.[149]

With our perspective on the GDPR's two-pronged approach to algorithmic accountability, and our emphasis on the role of the DPIA, we understand there to be more layers: individual explanations, group explanations, and systemic explanations, both internal and external. And unlike Edwards and Veale, we have more optimism that these multi-layer explanations can be grounded either in the text or subtext of the GDPR.

Looking at the GDPR through the lens of individual rights reveals the by-now-familiar two layers of explanations: a right to an explanation of the model, and a right to an individual explanation of an individual decision. The GDPR requires disclosure to individuals of 'meaningful information about the logic involved' in ADM on a systemic level.[150] It also establishes, we believe, the right to individual explanation of an individual decision.[151]

147 Wachter, Mittelstadt, and Floridi (n 2) 78; Selbst and Powles (n 2) 241; Edwards and Veale (n 3) 52ff. See also European Commission's High Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Thrustworthy AI*, 15 <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines> accessed 8 October 2020('The degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate').

148 Edwards and Veale (n 3) 22. They helpfully add to the conversation about the kinds of explanations that could be provided: (i) *model-centric explanations* that disclose, for example, the family of model, input data, performance metrics, and how the model was tested; and (ii) *subject-centric explanations* that disclose, for example, not just counterfactuals (what would I have to do differently to change the decision?) but the characteristics of others similarly classified, and the confidence the system has in a particular individual outcome.

149 See also Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' (2018) Harvard Journal of Law & Technology, *ArXiv:1711.00399 [Cs]*, 1 November 2017, <http://arxiv.org/abs/1711.00399> accessed 8 October 2020.

150 Arts 13 and 14 GDPR. See also Selbst and Powles (n 2) 241–42, discussing how this blends individualized and systemic explanations.

151 Kaminski (n 2) 199; Malgieri and Comandé (n 2) 246.

Table 2:. Comparison between DPIA duties and the WP29 Guidelines on GDPR algorithmic accountability duties

Content of DPIA (Article35(7)) GDPR	GDPR algorithmic accountability disclosure duties Articles 13–15, 22 (Guidelines on ADM, 30)
1. A *systematic description* of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller.	Describing: a. *The categories of data* used c. *How any profile* used in the automated decision-making process is *built*, including any statistics used in the analysis;
2. An *assessment of the necessity and proportionality* of the processing operations in relation to the purposes.	b. *Why* the categories of data are *pertinent* d. *Why* this profile is *relevant* to the automated decision-making process;
3. An *assessment of the risks to the rights and freedoms* of data subjects referred to in paragraph 1.	e. *How it is used for a decision* concerning the data subject: • ie which kinds of legal or similarly significant effects under Article 22(1
4. The *measures envisaged to address the risks*, including safeguards, security measures, and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.	*Which safeguards are adopted in compliance with Article 22(3) and (4)*: • eg Contestation, human involvement, making representation, explanation, algorithm audit, etc.

Looking through the lens of understanding the DPIA as a nexus between systemic governance and individual rights, however, reveals something more. The DPIA process entails a whole web of explanations: to internal oversight bodies ranging from the DPO to internal auditors, to external third parties such as auditors and expert boards, and as part of the overall assessment process.[152] These explanations are of differing degrees of breadth, depth, and technological complexity. But they establish a complex system of information flows, beyond the individual transparency requirements of the GDPR. These information flows will often require intermediation—that is, explanation—not just disclosure of existing information.[153] As discussed above, these various disclosures and explanations likely will include not just systemic and individual analysis, but group-level analysis of how an algorithm might impact particular classes of individuals, or particular locations. Thus, the DPIA process may address some of the concerns some scholars have about the DPIA's focus on individual rights, to the exclusion of groups.

Whether these explanations will go beyond the doors of companies is a different question. As discussed, a DPIA is not required to be made public, but its public disclosure is highly recommended, at least in the form of meaningful summaries.[154] We believe that analysis of how algorithms impact particular groups or places should be included in these public disclosures. This will help drive policy conversations in the way anticipated by most calls for public disclosure of impact assessments. It will also go some of the way to addressing concerns about a lack of stakeholder involvement and regulatory oversight over the impact assessment process, though we also counsel that companies aware of impacts on particular places or groups should seek out impacted individuals at an earlier stage of the process.

Moreover, there may be an argument for the disclosure of group- or location-based explanations to individuals as part of the GDPR's individual transparency rights regime. That is, even if DPIAs are not required to be made public, and even if companies decide not to disclose to the public what they discover about the impact of algorithmic decision-making on particular groups, they may nonetheless have to do so to impacted individuals under Article 22.

152 Margot E Kaminski, 'Understanding Transparency in Algorithmic Accountability' in Woodrow Barfield (ed), *Cambridge Handbook of the Law of Algorithms* (CUP 2020, forthcoming).

153 Frank Pasquale, 'Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries' (2010) 104(1) Northwestern University Law Review 105–74.

154 Article 29 Guidelines on Data Protection Impact Assessment (n 35) 17.

We understand the GDPR to suggest a connection between required DPIA analysis of systemic risks of unfairness and discrimination, and the individual rights to contestation, to express one's view, and to human intervention.[155] That is, for a person to be able to effectively invoke her right to contest an algorithmic decision, she may need to know whether she is being treated similarly to or differently from other similarly situated individuals. For the GDPR's series of individual rights to be meaningful, individuals need to know not just information about a particular stand-alone decision, but information about the algorithm's treatment of groups, and tendency towards bias and discrimination.

This group-based explanation, which we argue can be at least implied from—if not required by—the rights to contestation or to challenge the decision, could be created based on information on affinity or group profiling uncovered during a DPIA. This should not be hard to implement. In technological terms, multi-layered explanations based on general (or global), group, and individual (or local) explanation are already a reality.[156]

Finally, some scholars have remarked that what is needed is not merely an explanation, but a legal *justification* of automated decisions taken.[157] The full concept of justification is not easy to address in the data protection framework and is beyond the scope of this article. For the limited scope of this article, however, justifying a decision means not merely explaining the logic and reasoning behind it, but also explaining why it is a correct, lawful, and fair decision, ie that the decision is based on proportional and necessary data processing, using pertinent categories of data and relevant profiling mechanisms.

Again, connecting the DPIA to transparency requirements may clarify what this could mean. Language about the DPIA process suggests that in addition to technical explanations of a model, data controllers should produce *justificatory* explanations of a system during a DPIA. Under the DPIA process, data controllers must prove the legal proportionality and necessity of the data processing, and thus the legal necessity and proportionality of eventual automated decisions taken (Article 35(7)(d)). This may constitute a form of justification for data use and profiling mechanisms. Similarly, the Guidelines on ADM recommend that data controllers (in order to comply with Articles 13–15) explain the pertinence of categories of data used and the relevance of the profiling mechanism.[158] Assessing whether the data used are pertinent and the profile is relevant for a decision, as well as assessing the necessity and proportionality of the data processing in an ADM system, seems to constitute a type of justification of automated decision systems. The purpose of such assessment is not just transparency about the technology and its processes, but an explanation about the lawfulness, fairness, and legitimacy of certain decisions.[159]

Combining the algorithmic DPIA process and the duty to disclose information about algorithmic decisions in coordinated actions would be beneficial not just for individuals but for data controllers.[160] Combining these tasks could benefit data controllers because:

1. they could optimize efforts that would otherwise be spent on two different tasks (the DPIA and disclosure requirements) by taking compliance with DPIA duties (Article 35) and feeding them into transparency duties as imposed by Article 13–15 (and 22) of the GDPR;
2. publicly disclosing (at least some parts) of the DPIA as a basis for explaining automated decisions is considered a best practice recommended in the DPIA framework,[161] in line with the data protection by design principle (Article 25 GDPR);[162]
3. disclosing information about algorithmic data processing to data subjects and collecting their reactions (through, eg the right to contest, to have a new decision, to have human involvement, etc.)[163] could be considered compliant with the duty to seek the view of impacted data subjects (Article 35(9) GDPR), in the continuous cycle of the DPIA framework;[164]

155 Art 22 GDPR.

156 Ramamurthy and others (n 17).

157 Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2019(2) Columbia Business Law Review <https://papers.ssrn.com/abstract=3248829> accessed 8 October 2020; Kaminski (n 5), 'Binary Governance', 12–17.

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

159 About the difference between explanation and justification see Mireille Hildebrandt, *Law for Computer Scientists and Other Folk* (Oxford, New York: OUP 2020) 301.

160 On the list of positive externalities for data controller if they disclose a 'legibility' test on algorithms, see Malgieri and Comandé (n 2) 259–60.

161 See Article 29 Guidelines on Data Protection Impact Assessment (n 35) 17. See also Kloza and others (n 38) 2.

162 Veale, Binns, and Ausloos (n 63) 117–18.

163 See Art 22(3) and Recital 71. See also Roig (n 3).

164 See on the importance of continuous engagement of involved subject in PIA: Roger Clarke, 'An Evaluation of Privacy Impact Assessment Guidance Documents' (2011) 1(2) International Data Privacy Law 112.

4. in general terms, the dynamic merging of an algorithmic DPIA with multi-layered explanations might be a 'suitable safeguard' to protect fundamental rights and freedoms of individuals both under Article 22(3) and under Article 35(7)(d) of the GDPR;[165]
5. developing an algorithmic DPIA and explanation safeguards in parallel (intrinsically related to the right to contest a decision, right to a human-in-the-loop, etc.) might be the best way to enrich transparency with accountability safeguards[166] and overcome the 'transparency fallacy' through a virtuous cycle of algorithmic auditing and continuous detection/mitigation of unfair effects.[167]

The idea of at least partially merging algorithmic accountability duties with the DPIA process also seems useful considering the most advanced literature on explanations. As discussed above, a multi-layered and multi-step explanation would be a continuous *process*, not merely a product.[168]

Conclusion

There is a growing literature suggesting that AIAs are a crucial tool in establishing algorithmic accountability. This paper addresses that tool as it is implemented in the GDPR. We find that the GDPR's version of an AIA serves as a central connection between its two approaches to regulating algorithms: individual rights and systemic governance. That framing allowed us to identify both value in and shortcomings of the GDPR's Impact Assessment regime as applied to algorithmic governance.

This analysis, we hope, will have value for other discussions of AIAs beyond the GDPR. In particular, moving from individual transparency rights and governance accountability duties in the field of ADM, we suggest a model of multi-layered explanations drawn from an impact assessment process. Since there are several layers of algorithmic explanation required by the GDPR, we recommend that data controllers disclose a relevant summary of a system, produced in the DPIA process, as a first layer of algorithmic explanation, to be followed by group explanations and more granular, individualized explanations. More research is needed, in particular about how different layers of explanations—systemic explanations, group explanations, and individual explanations—can interact each other, and how technical tools can help in developing an AIA that might be re-used towards GDPR-complying explanations and disclosures.

doi:10.1093/idpl/ipaa020
Advance Access Publication 6 December 2020

165 About the link between 'risks to rights and freedoms' and impacts on individuals, see Niels van Dijk, Raphaël Gellert and Kjetil Rommetveit, 'A Risk to a Right? Beyond Data Protection Risk Assessments' (2016) 32(2) Computer Law & Security Review 304; See also Katerina Demetzou, 'Data Protection Impact Assessment: A Tool for Accountability and the Unclarified Concept of 'High Risk' in the General Data Protection Regulation' (2019) Computer Law & Security Review 105342 <https://doi.org/10.1016/j.clsr.2019.105342>.

166 Hildebrandt (n 19) *passim*.

167 Edwards and Veale (n 3) 65.

168 Tania Lombrozo, 'The Structure and Function of Explanations' (2006) 10(10) Trends in Cognitive Sciences 464–70. See also Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences' (2019) 267 Artificial Intelligence 6, 273 who explains that explanation has two processes: cognitive process and social process.