Multi-layered Explanations from Algorithmic Impact Assessments in the GDPR

Margot E. Kaminski
Associate Professor of Law
University of Colorado
Boulder, CO USA

Gianclaudio Malgieri
Doctoral Researcher at LSTS
Vrije Universiteit Brussels
Brussels, Belgium

ABSTRACT

Impact assessments have received particular attention on both sides of the Atlantic as a tool for implementing algorithmic accountability. The aim of this paper is to address how Data Protection Impact Assessments (DPIAs) (Art. 35) in the European Union (EU)'s General Data Protection Regulation (GDPR) link the GDPR's two approaches to algorithmic accountability—individual rights and systemic governance—and potentially lead to more accountable and explainable algorithms. We argue that algorithmic explanation should not be understood as a static statement, but as a circular and multi-layered transparency process based on several layers (general information about an algorithm, group-based explanations, and legal justification of individual decisions taken). 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 kinds of explanations.

CCS CONCEPTS

Applied computing → Law, social and behavioral sciences; • Computing methodologies → Machine learning

KEYWORDS

Law, General Data Protection Regulation, Impact Assessments,

ACM Reference format:

Margot E. Kaminski, Gianclaudio Malgieri. 2019. Multi-layered Explanations from Algorithmic Impact Assessments in the GDPR. In *Proceedings of ACM FAT* Conference (FAT* 2020). ACM, Barcelona, Spain, 12 pages.*

1 Introduction

The discussion of the GDPR (General Data Protection Regulation) and 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.

FAT* '20, January 27–30, 2020, Barcelona, Spain

ACM ISBN 978-1-4503-6936-7/20/02.
https://doi.org/10.1145/3351095.3372875

Impact assessments have received particular attention on both sides of the Atlantic as a tool for implementing algorithmic accountability. The aim of this paper is to address how Data Protection Impact Assessments (DPIAs) (Art. 35) link the GDPR's two approaches to algorithmic accountability—individual rights and systemic governance—and potentially lead to more accountable and explainable algorithms. Indeed, we argue that algorithmic explanation should not be understood as a static statement, but as a circular and multi-layered transparency process based on several layers (general information about an algorithm, group-based explanations, and legal justification of individual decisions taken). We argue that the impact assessment process plays a crucial role both in connecting internal company heuristics and risk mitigation to outward-facing individual rights, and in forming the substance of several kinds of explanations.

Section 2 introduces the algorithmic accountability tools in the GDPR, while Section 3 explores the individual rights of data subjects. Section 4 develops the idea of collaborative governance of algorithms in the GDPR. The subsequent sections address the concept of an Algorithmic Impact Assessment (§5) and how the DPIA in the GDPR can serve as an AIA (§6). The final section (§7) explains how the combination of Algorithmic DPIA and individual rights can create a virtuous circle of multi-layered explanations.

2 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. [1,2,3,4,5,6] Subsequent legal analysis, however, began to focus instead on other accountability tools [7, 8, 9], either required in the text of the GDPR or recommended in interpretative guidelines. [10] These tools include third-party auditing, the appointment of Data Protection Officers (DPOs)(Art.37), and the requirement of Data Protection Impact Assessments (DPIAs)(Art.35).

As one of us has argued, the GDPR combines a series of individual rights (Arts. 12-23) with a systemic governance regime overseen by regulators (Arts. 24-43 & 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. [11, 12]

For example, a data subject has a right to contest an individual algorithmic decision (Art.22), to receive notice of solely automated decision-making (Art.13), and to request access to "meaningful information about the logic involved" (Art.15). Should she fail to invoke any of these rights, however, the GDPR still puts in place significant obligations on data controllers using automated decision-making, whether solely automated or not. [7] The GDPR requires an array of systemic accountability tools, including: third-party auditing, the appointment of Data Protection Officers (DPOs)(Art. 37), and Data Protection Impact Assessments (DPIAs)(Art. 35). These obligations arise from the text of the law, but also in 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. [10, 12]

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 with them—through an approach known in legal literature as "collaborative governance": the use of public-private partnerships. [13,14] Because the GDPR often effectively outsources governance decisions to single data controllers, 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 oversight from the public and both expert and interested third parties. [11]

With this background in mind, the next two sections of this paper 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 the Data Protection Impact Assessment (DPIA) plays in linking the two.

3 Individual Rights in the GDPR and the Multi-layered Explanation

The GDPR gives data subjects several important rights with respect to algorithmic decision-making. The GDPR contains both general data protection rights and rights specific to profiling that also apply to algorithmic decision-making. [15] On top of this, it establishes rights specific to algorithmic decision-making, which include: a right to be notified of solely automated decision-making (Arts. 13, 14); a right of both notification and access to meaningful information about the logic involved (Arts. 13, 14, 15); a right to be informed of the significance of and envisaged effects of solely automated decision-making (Arts. 13, 14, 15); and a right not to be subject to solely automated decision making (Art. 22), with safeguards for the limited cases in which automated decision-making 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 (Art. 22).

We do not intend to revisit the legal debate over these rights in detail, but an overview may be useful. As mentioned, discussion of these individual rights has largely focused on whether or not—or really, how—solely automated decision-making must be explained to data subjects. As Selbst & 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 automated decision-making, in addition to its significance and envisioned effects (Arts. 13, 14, 15). [4] 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 is an explanation specific to individual decisions, and what depth of explanation is required. [1,2,4]

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. [5, 8, 16] Our view, explored at length elsewhere, is that it does. Such decisions must be made "legible" to data subjects, in the sense that data subjects must be able to understand enough about the decision-making process to be able to invoke their other rights under the GDPR.[2] Several of the Member States implementing Article 22(2)b of the GDPR, have outlined such Article 22 explanation duties in greater detail.[17]

As we discuss below, our view is that the GDPR's transparency rights are best discussed together as a system. The GDPR, that is, is best understood as establishing a system of *multi-layered explanations*. Data subjects have a right to both a system-wide but detailed description of the logic of an algorithm (Arts. 13, 14, 15), and more specific insights on individual decisions taken (Art. 22 and recital 71).[10] The level of granularity of explanations provided depends on several factors and is not clearly dictated in the GDPR.

It is important to remark that different data controllers have different accountability 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 organisational measures to ensure compliance with the GDPR. Accordingly, algorithmic decision-making involving bigger risks for data subjects (considering the consequences and implications of the decision or the predictive power of the data controller) should entail more safeguards.[18] As we will explain in Section 4, data controllers to a certain extent choose their own algorithmic accountability safeguards: some are required in the GDPR both at Article 22(3) and at recital 71 (algorithmic auditing, ensuring the right to context, to have a new decision and to a human in the loop,

but also the right to explanation), but these are not closed lists, and the Guidelines on ADM suggest additional techniques.

Accordingly, in case of more intrusive and riskier automated decision-making processes, the data controller should implement all possible safeguards, including the right to explanation of an individual decision. However, there have been legitimate concerns voiced in the legal literature about the capacity of data subjects to both invoke their rights and execute oversight over algorithmic decision-making.[19,20,21,22,8] 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 other more systemic forms of accountability [11,23,24,25]; or for foregoing individual accountability in favour of expert and external oversight. [22, 26, 8].

Foregoing individual accountability ignores the dignitary and legitimizing value of such rights.[27] Individual rights allow data subjects to exhibit autonomy and exert control, and to protest or reject their objectification by profiling or decision-making machines.[11, 20] 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.[11] 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 system of collaborative governance. 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 governance.[11]

4 Collaborative Governance in the GDPR

The other side of algorithmic governance in the GDPR is its systemic governance regime. It aims, largely, to address instrumental goals: preventing algorithmic error, bias, and discrimination.[11] This governance regime, as one of us discusses at length elsewhere, is largely constituted through collaborative governance, or public-private cooperation. We here illustrate two examples of how this works in the GDPR.

Article 22's suitable safeguards on automated decision-making are one example of this in practice. The GDPR's text does not comprehensively dictate what data controllers must do to protect individual rights when they use solely automated decision-making(Art.22). Instead, the GDPR lists examples of safeguards (contestation, expression, human intervention), and leaves to both data controllers and regulators to determine what additional safeguards are necessary. The accompanying Recital famously adds a right to individual explanation (Rec.71). The Guidelines, too, fill in this gap, proposing a list of best practices.[10] These include, but are not limited to: regular quality assurance checks, algorithmic auditing, independent auditing, establishing data minimisation and clear retention periods, using pseudonymisation techniques, certification mechanisms, ethical review boards, and more.[10] All of these are attempts at systemic accountability and oversight, and notably attempt to establish that oversight in a comprehensive and ongoing manner.

Another example is the GDPR's approach to preventing bias and discrimination in algorithmic decision-making. Recital 71 tasks data controllers 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, however, lay out *how* to do this. Instead, the Guidelines suggest that data controllers 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.[10]

The GDPR does not tell data controllers precisely what to do. It identifies the problem, provides suggestions of what regulators might consider adequate, but also tasks data controllers with cooperatively coming up with solutions. Such company-created solutions may then feed back into what regulators ultimately require. Roig [7] explains that "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".

5 Algorithmic Impact Assessments

Within this dual system of algorithmic accountability—individual rights accompanied by extensive but collaborative governance of data controllers' behaviour—the Data Protection Impact Assessment (DPIA) serves a special role. We claim here that 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 shortcomings as implemented in the GDPR. We turn here to both the general discussion that has arisen recently over Algorithmic Impact Assessments, and to the specific role of the DPIA in the GDPR,

We are not the first to focus on Algorithmic Impact Assessments, or impact assessments in closely related fields. [9,28-39] We are, however, the first to discuss Algorithmic Impact Assessments 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 system of data governance. Our GDPR-specific analysis, then, may have implications for proposals for algorithmic impact assessments

in other legal systems.[37] It suggests that impact assessments best serve a role in conversation with other accountability tools, as part of overarching governance design.[33] We suggest that impact assessments play a central role both as a source of, and mediator between, the multi-layered explanations indicated in the GDPR.

5.1. Proposals for Algorithmic Impact Assessments

Algorithmic Impact Assessments 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 U.S. federal law [39]. We here briefly discuss several important precursors to the AIA: Human Rights Impact Assessments (HRIA), Privacy Impact Assessments (PIA), Ethical Impact Assessments (EIA), and Surveillance Impact Assessments (SIA).

The Human Rights Impact Assessment (HRIA) process outlined by the United Nations [41] 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 [34]. Mantelero, for example, draws partially on the HRIA model, as does Katyal.[34, 33]. Wright and Friedewald [39] look to Ethical Impact Assessments (EIA), voluntary assessments that go beyond legal compliance to assess the ethical implications of new technologies, and involve consultation with a wide number of stakeholders, and publication of the assessment [42].

However, the most direct precursor for the notion of the AIA, especially in the context of the GDPR, is the Privacy Impact Assessment [29,34,30,39,35]. As Clarke explained in 2009, PIAs originated in the 1990s around the world, with multiple regulators issuing guidance in the early 2000s [30]. While Privacy Impact Assessments as conducted in the United States have been widely decried as toothless,[28,30,32] elsewhere they are more substantial [30]. It is interesting to note that in at least several European countries, the PIA may have originated in part in the system of "prior checking" under earlier data protection regimes, which was effectively a system of licensing prior to data processing [29,30,43]. 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.

Clarke characterizes the characteristics of the ideal PIA 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 mere legal compliance; being oriented towards surfacing solutions, not just problems; emphasizing process over product; and requiring engagement from executives and managers.[30] In 2012, Wright and Raab [35, 44] proposed the concept of a Surveillance Impact Assessment (SIA), 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."[35] But this is the ideal. Mantelero observes that in practice, DPIAs 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.[34]

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

Selbst proposes the use of an Algorithmic Impact Statement, modelled after the Environmental Impact Statement (EIS) (established in the United States in 1969 in the National Environmental Policy Act (NEPA)), with some modifications. His AIS 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, the AIS would, like an EIS, require "(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."[38] Second, a police department would have to "devote substantial treatment to each alternative" including "the alternative of no action."[38] And finally, police must include proposed mitigation measures in the AIS.[38] To 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.[38]

Katyal incorporates elements of Selbst's proposal into her suggestion of a Human Impact Statement in Algorithmic Decision-making.[33] She recommends as a backstop a substantive, rather than purely procedural, commitment to algorithmic accountability and antidiscrimination.[33] And she adds that companies should also (1) identify potentially impacted populations and determine their status-based categories; (2) identify the effect of uncertainty or error on those groups; and (3) study whether the decision will have an adverse impact on a particular subpopulation.[33] The single biggest difference, however, between Katyal's and Selbst's proposals is that Katyal recommends the HIA in AI as a voluntary measure undertaken by private industry, rather than required by law.

The AI Now Institute,[37] issued a report also building on Selbst's proposal.[38] The report's authors call for a pre-procurement Algorithmic Impact Assessment, before public agencies commit to the use of an automated decision-making

system. Like Selbst's proposal, the AI Now proposal is limited to the public sector. Like Selbst's proposal, it would be mandatory rather than voluntary. Unlike Selbst's proposal, it goes beyond the policing context.[37] The proposed AIA requires agency disclosure and a public comment period, including both meaningful access for researchers and auditors once systems are deployed and individual due process for those affected by the system's decisions. The AIA is envisioned as being renewed every two years [37].

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.[34] A hybrid between a Human Rights Impact Assessment and a Privacy Impact Assessment, the HRSEIA suggests that businesses voluntarily take into account ethical and social impact, in addition to human rights.[34] Mantelero emphasizes the role of such an impact assessment in addressing the collective dimensions of data harms. At its core, the HRSEIA has three features: it is participatory, it is transparent, and it is circular in nature.[34] Practically, it consists of a self-assessment questionnaire, sometimes leading to evaluation by an ad hoc committee of experts.[34] Stakeholder engagement is encouraged but not required.[34] Similarly, public disclosure is encouraged.[34] Mantelero explains that while this proposal is "[i]n line with the declared intent of the GDPR," he does not understand the HRSEIA to be required by the GDPR.[34] Several other commentators have recently discussed the DPIA and the role it plays in the context of algorithmic accountability more generally.[8,9]

Most of these proposals for Impact Assessments centrally depend on release of information to the public.[37,38] 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 source code.[45,46,38] Some suggest a tiered release of information, with summaries released to the public and detailed or sensitive information released only to regulators or experts.[34] Thus more recent proposals also call for expert input and oversight by suggesting 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 use.[47,37]

6 The Data Protection Impact Assessment (DPIA) as an Algorithmic Impact Assessment

A version of an Algorithmic Impact Assessment might be derived from the GDPR's Data Protection Impact Assessment (DPIA). Article 35(3)(a) requires a DPIA in case of (inter alia): "a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person." Interpreting this provision, the Article 29 Working Party Guidelines on DPIAs ("DPIA Guidelines") mandate DPIAs for any automated decision-making, 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)".[9] Moreover, as Casey, Farhangi, & Vogl have noted, "demonstrating that a DPIA is not necessary will, in many instances, itself require a DPIA".[9] We note, too, that at least one Member State (see the Slovenian Data Protection Act that implements the GDPR) requires algorithmic impact assessments as a specific safeguard in case of automated decision-making under Article 22(1) of the GDPR.[48]

In this section, we address what the purpose of the DPIA is in the GDPR, and what it must include. Understanding the DPIA's purpose clarifies what the content should be, and points to several shortcomings in the current conception of the DPIA.

6.1. What is Required in a DPIA?

The GDPR describes a DPIA as "an assessment of the impact of the envisaged processing operations on the protection of personal data" (Art.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 (Art.35(7))(Rec. 84, 90).

The GDPR's version of a DPIA must take place before a company implements a system. That is, a company must assess a system, and propose risk-mitigation measures, before data processing takes place (Art.35(1)). But the GDPR also envisions iteration. If the risk posed by a system changes, a company must assess whether it is complying with its own Impact Assessment (Art.35(11)). It should also review 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 3 years. "Carrying out a DPIA is a continual process, not a one-time exercise," state the DPIA Guidelines.[49] This continual process involves assessing risk, deploying risk-mitigation measures, documenting their efficacy through monitoring, and feeding that information back into the risk assessment and ongoing process. The DPIA Guidelines envision this process as running "multiple times."

The GDPR also lays out procedural requirements for the DPIA. Differing from the Impact Assessments imagined in the literature, DPIAs do not involve a period of public comment or

input. They do require consultation with an internal but independent Data Protection Officer, if a company has one. Many data controllers that are required to perform impact assessments will have Data Protection Officers in place (Art.38).[49,10]

In lieu of public or formal stakeholder consultation, the GDPR requires consultation "where appropriate" with impacted data subjects (Art.35(9)). In the original proposal of the Commission, consultation with data subjects was mandatory (Article 33[4]). The Parliament's text argued that this 'represents a disproportionate burden on data controllers' (amendment 262). Accordingly, the approved Article 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''. [29] This puts in place one method for external input, from impacted data subjects rather than external experts or the public.

The DPIA Guidelines envision that this input could be, for example, in the form of surveys crafted by data controllers and sent to future customers, which would make it less meaningful than, say, deep consultation with a board of representative of civil society members or chosen community representatives, envisioned in the literature.[49,50] The DPIA Guidelines explain that if data controllers do not seek these external views, they have an obligation to justify this decision. [49] In addition, if data controllers do seek these views and then disregard them, they must document why they have chosen to disregard this input. [49]

As for other forms of external oversight, the DPIA Guidelines recommend, but do not require, seeking advice from independent experts, ranging from lawyers and sociologists to data security experts.[49] 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 (Art.36).

In the biggest departure from the Impact Assessment proposals discussed above, DPIAs are not legally required to be released to the public, even when finalized.[49] 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."[49] The Guidelines caution that it is a good practice to publish DPIAs, especially where members of the public are impacted. But data controllers need not publish the entire assessment; the published DPIA "could even consist of just a summary of the DPIA's main findings."[49] There are of course cases in which full disclosure of the assessment results may be limited by the legitimate interests of the data controller, such as confidentiality, security and competition.[34,51,52]

The GDPR text and DPIA Guidelines thus give an overview, but little specific guidance on what exactly a company must put in a DPIA report. Unlike the Impact Assessments proposed in the legal literature, they do not require public input or public disclosure, though they suggest both as best practices. This has led one 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".[37] As we discuss below, this is not entirely correct.

6.2. What is the Purpose of a DPIA, in the context of the GDPR's Algorithmic Governance?

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 system of individual rights to systemic governance.

Understanding the DPIA in this way both clarifies its potential content, and leads us to observations about how implementing the DPIA as an Algorithmic Impact Assessment might be improved. DPIAs are not the perfect Algorithmic Impact Assessment. As a tool in the GDPR's overall algorithmic governance regime, however, they have more potential than might initially meet the eye.

6.3. How Understanding the DPIA's Dual Role Helps Clarify Its Content

The DPIA has two roles: as a tool in the GDPR's systemic collaborative governance regime, and as an element of the GDPR's approach to protecting and enabling individual rights. Understanding the DPIA in this way—as a connection between the two systems—lets us better understand how it is meant to function as an Algorithmic Impact Assessment, even to the extent of further clarifying its content.

When understood as part of the GDPR's collaborative governance system,[11] the DPIA is a form of monitored self-regulation.[29] Collaborative governance is centrally concerned with affecting management culture and creating meaningful changes within a company.[53,13,14] Monitored self-regulation attempts to change both company decision-making processes and decision-making heuristics.[14] The DPIA, in the context of algorithmic decision-making, tasks data controllers with considering risks of unfairness, error, bias, and discrimination, and with coming up with concrete ways of mitigating those risks. This affects firms' decisional heuristics by dictating, through Recitals and the Guidelines, what values a company must consider, and what harms it must prevent.

The process of conducting the DPIA—taking input from impacted data subjects, consulting with an independent Data Protection Officer, consulting with a regulator where required, and involving both internal and external experts—is meant to change internal company processes.[29] As others have noted, baking in a compliance culture can be valuable, even where public oversight and input is not sought.[14,33] The DPIA can also be understood in this context as a documentation or even

a reporting requirement, creating records that can later be sought and inspected by regulators.[49]

The DPIA also, however, has an unexplored role in the GDPR system of individual rights. First, the DPIA can serve as a source of material for the much-discussed disclosures to data subjects 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 automated decision-making (Arts.13, 14, 15). A DPIA must contain, as mentioned above, "a systematic description of the envisaged processing operations and the purposes of the processing…"(Art.35(7)). If data controllers are already internally describing automated decision-making at a systematic level as part of the DPIA process, those internal descriptions could be disclosed to data subjects, or at least serve as the basis for these disclosures. In addition, they might be 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 data subjects, and data subjects have a right in the context of automated decision-making to be informed of the "significance and envisaged consequences" of such decision-making (Arts.35, 13, 14, 15). Again, as a company conducting automated decision-making 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.

Second, despite other commentators' dismissal of DPIAs as failing to put in place individual due process,[37] the Guidelines explicitly envision the DPIA as an essential part of establishing suitable measures to safeguard individual rights, including a version of individual due process. The GDPR requires data controllers that fall under the exceptions to its ban on solely automated decision-making to still implement suitable measures to protect individual rights (Art.22). The Guidelines counsel that data controllers should use DPIAs to "identify what measures they will introduce to address the… risks involved".[10] As discussed, suggested measures include a number of individual rights: informing data subjects 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. [10] This list effectively imports the various individual rights laid out in the GDPR that are restricted to *solely* automated decision-making, into DPIAs that the Guidelines say go beyond the solely automated context (that is, algorithmic decisions that involve a human decision-maker).

In other words, the GDPR (or really, the interpreting Guidelines) envisions DPIAs, in the context of algorithmic decision-making, as serving as a form of commitment-making to protecting, or even enabling, individual due process rights. By characterizing these individual rights as *risk mitigation measures*, it both provides a substantive backstop as to what must be included in a DPIA, and tasks data controllers with *constituting*—through the process of performing a DPIA—what these individual rights will look like in practice. Thus the DPIA serves as a collaborative governance mechanism used to constitute the substance of individual due process rights. [11] It also serves as a means of expanding company commitments, changing company decision-making heuristics to include an assessment of individual due process rights.

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 review. 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 like a requirement. The use of external experts is framed as a necessary risk-mitigation measure for algorithmic decision-making, functionally changing the required content of a DPIA in that context.

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" (Rec. 71). Malgieri and Comandé observe that this requirement effectively expands the "suitable safeguards" to protect individual rights from the harms of algorithmic decision-making from the series of individual due-process-like protections enumerated in the GDPR text, to a far broader set of systemic accountability measures, including third-party auditing (Art.22).[2]

The Guidelines' list of best practices for suitable safeguards over algorithmic decision-making supports this interpretation, including recommendations that data controllers use both audits and external review boards.[10,9] When a company that deploys algorithmic decision-making conducts its DPIA, it will refer to the Guidelines' list of best practices in establishing risk-mitigation measures that are already regulator-approved. This means that in practice, 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 "suitable safeguard" or "suitable measure" to protect individual rights, the Guidelines link individual rights protection with collaborative governance techniques. Data controllers 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 simultaneously conceptualized as a crucial aspect of individual rights in the GDPR, standing in for data subjects 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 roles: it watches the data controllers as they come up with ways of addressing problems with algorithmic decision-making, and it reassures data subjects that their dignity and individual rights are being respected by a fair system.[11,29,54] As the mechanism through which this external oversight is implemented, the DPIA thus connects the two approaches to algorithmic governance in the GDPR.

6.4. Shortcomings of the DPIA

The biggest shortcoming of the DPIA is that it does not include a mechanism for mandatory disclosure to the public.[50,55] Public disclosure is understood by many to be an essential element of Impact Assessments as a policy tool of collaborative governance.[31,37,38] Public-facing disclosure enables public feedback, both in the form of market feedback and regulatory feedback over the longer term. By failing to mandate public disclosure, the GPDR's DPIA fails to trigger both of these mechanisms.

This failure could be drastic. The GDPR puts a lot of faith in the behaviour of data controllers. As discussed, it often tasks data controllers with coming up with the substance of (a) how individual rights will be implemented and (b) how to address unfairness, biases, and discrimination concerns about algorithms. The counterbalance is regulatory oversight of DPAs (Data Protection Agencies). But the GDPR's enforcers have not, historically, been well-resourced compared to the data controllers they regulate.

Tasking regulators with extensive monitoring also forgoes some of the touted benefits of governing through public-private partnerships, including lowered costs. By failing to require 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. Similarly, the DPIA fails to involve serious stakeholder input, unless data controllers understand the Guidelines' emphasis on expert boards and third-party audits to include impacted stakeholders and to be mandatory.[29,9]

As an alternative, individual notification and access rights could do some work. If data controllers indeed link their DPIA content to what they disclose to data subjects (for example, disclosing the "logic involved" in a decision-making system), then it is likely that these disclosures will make their way to other third parties, including civil society and the press. This is a more attenuated way of getting at the same outcome as direct public disclosure, however, and risks failing if data controllers significantly disaggregate the DPIA process from individual disclosure rights.

6.5. Lessons for Calls for Algorithmic Impact Assessments Generally

Our GDPR-specific analysis has implications for proposals for algorithmic impact assessments generally. Our research into the GDPR's version of AIAs suggests that the proposals discussed above have 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.[8,11,15] 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).[33] We found only one author—Binns—discussing DPIAs generally who identified the GDPR's version of impact assessments as a kind of collaborative governance with the private sector (or what he identifies as "meta-regulation"). [29] But this led Binns to critique the GDPR's version of the DPIA for inadequate public disclosure and stakeholder involvement, not to look to how it connects to the broader system of both collaborative governance tools and individual rights in the GDPR.

Second, as part of a larger system of governance, there are unexplored connections between the GDPR's potential AIA and its underlying substantive individual rights and substantive principles. It is true that many of those rights and principles are articulated in broad, sometimes aspirational, terms.[34] The GDPR version of the AIA has a substantive backstop. The oddity is the GDPR's circularity: the DPIA helps not just to *implement* but to *constitute* those individual rights.

Third, impact assessments can serve as a connection between collaborative governance and individual rights (only one other proposal, to our knowledge, suggests using Impact Assessments to establish something resembling individual rights—a system of "enhanced due process mechanisms for affected individuals"[37]). The information a company creates during the Impact Assessment process can feed into what it provides to data subjects 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.

Fourth, because the DPIA 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 as "multi-layered explanations" contemplated in the GDPR. Several of the above scholars, including both Mantelero and Wright & Raab, emphasize the often collective dimensions of surveillance and data processing.[34,35] The GDPR's system of individual rights threatens by itself to miss the impact of surveillance or in this case, automated decision-making, on groups, locations, and society at large.[34,56]

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% of those subjected to stop and frisk by the NYPD were Black or Latino men.[37] The DPIA with its systemic approach to risk assessment and risk mitigation requires data controllers to analyze how the system impacts not just individuals but groups. We believe that systemic and group-based explanations uncovered during a DPIA can and should be communicated to outside stakeholders.

7 Comparing DPIA content with Algorithmic Accountability requirements under the GDPR

Thus far few commentators have linked the Guidelines on Automated Decision-Making to the DPIA process.[9] Here, we connect the GDPR's text to these Guidelines to show how the required content of DPIAs in fact easily serves as the basis for disclosures controllers must make to data subjects.

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 4. the measures envisaged to address the risks".

The Guidelines on DPIAs clarify that a *systematic description* of processing should include: the nature, scope, context and purposes of the processing, (categories of) personal data, recipients and storage; a functional description of the processing operation and the assets on which personal data rely.[49]

If we compare algorithm accountability requirements with DPIA content, there are multiple interesting similarities. In particular, 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 automated decision-making and how the algorithmic profiling is built. Analogously, the controllers' duty to assess necessity and proportionality of the processing operations in the DPIA is similar to the algorithmic transparency duty to explain the pertinency of personal data used and the relevance of the profiling [10]. The controllers' duty to assess the data processing risks and the impacts on individuals is similar to the transparency duty to explain the impact of the profiling use in automated decision-making [10]. Lastly, the controller's duty to find safeguards of individual rights in case of automated decision-making (under Article 22(3) and (4) GDPR) is similar to the duty to find and describe measures envisaged to address the risks in DPIA. In other words, in the case of automated decision-making, the DPIA steps might correspond to transparency duties (as interpreted by Article 29 Working Party [10], in Annex I).

7.2. Towards Multi-layered Explanations from Algorithmic DPIA

The DPIA process, combined with the GDPR's extensive individual transparency rights and Guidelines on suitable safeguards of individual rights, suggests what we call "multi-layered explanations" for automated decision-making.

Edwards & Veale have similarly suggested that data subjects should be given what they call "model-centric" and "subject-centric" explanations.[8] 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 a decision), the characteristics of similarly classified individuals, and the confidence a system has in an outcome. We understand these as really being three layers: individual explanations, group explanations, and systemic explanations. And unlike Edwards & Veale, we have more optimism that these multi-layered explanations can be found either in the text or subtext of the GDPR.

As a first layer, there should be the right to individual explanation, at least in riskier data processing operations (as required at Article 24 GDPR and explained supra in Section 3). This is an individual right to an explanation of an individual decision. Even this individualized form of disclosure can be informed by information revealed during a DPIA, as discussed further below.

As a second layer, 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 (see Article 22). That is: for these series of individual rights to be meaningful, data subjects 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. This group-based explanation, which we argue can be at least implied from if not required by the rights to contestation etc., could be created based on information on bias and discrimination uncovered during a DPIA.

As a third layer of explanation, the GDPR requires disclosure of "meaningful information about the logic involved" in automated decision-making on a systemic level (Arts. 13 and 14, [4]). The DPIA must include an assessment of necessity, proportionality, risks and safeguards. A DPIA is not required to be made public, but its public disclosure is highly recommended, at least in the form of meaningful summaries.[49] If we compare DPIA requirements to the safeguards in automated decision-making, as we do below in Table 1, it seems reasonable and efficient to suggest that a relevant summary of the Algorithmic DPIA could be used as "meaningful information" about the overall system of decision-making that must be provided to data subjects.

Further levels of details could be tailored from this systemic layer. Indeed, the description of algorithmic processing,

including rationales and criteria, might well be specified on both a *group* level and on *individual* level: the description of data categories and their pertinence, of the profile-building procedure and its relevance and use, of effects and safeguards can be based on groups of similar individuals affected by that automated decision or even on a single specific data subject who is making an explanation request.

Combining these two different tasks (the algorithmic DPIA and the duty to disclose meaningful information about algorithmic decisions) in coordinated actions would be highly beneficial for data controllers [2]. Combining these tasks would benefit data controllers because:

1. if conducted with an adequate degree of rigor, it could help controllers comply with both their transparency duties imposed by Article 13-15 (and 22) of the GDPR and with DPIA duties (Article 35 GDPR), optimizing efforts that would otherwise be spent on two different tasks;
2. publicly disclosing (at least some parts) of the DPIA as a basis for explaining automated decisions is considered a best practice,[2,50] in line with the data protection by design principle (Article 25 GDPR)[55];
3. disclosing information about algorithmic data processing to data subjects and collecting their reactions (through, e.g., the right to contest, to have a new decision, to have human involvement, etc., see recital 71)[7] 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 (see also recital 71)[61];
4. merging an algorithmic DPIA and multi-layered explanation might serve as 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[62];
5. developing an algorithmic DPIA and explanation safeguards in parallel (intrinsically related to the right to contest a decision, right to human-in-the-loop, etc.) might be the best way to enrich transparency with accountability safeguards [20] and overcome the "transparency fallacy" through a virtuous cycle of algorithmic auditing and continuous detection/mitigation of unfair effects. [8]

The idea of merging (at least partially) algorithmic accountability duties and DPIAs also seems useful considering the most advanced literature on explanation. Effective explanation is not only the result of an analysis, but also a two-stage process: both cognitive and social. [64] Multi-layered and multi-step explanations would be a continuous *process*, not merely a static product.[63,64]

Some scholars have remarked that what is needed is not merely an explanation, but a legal justification of automated decisions taken.[65] Connecting an algorithmic DPIA to individual transparency rights might address this: the information duties about the *pertinence* and *relevance* of decision-making processing required by the Guidelines may reflect the duty to assess the *proportionality and necessity* of data processing required at Article 35(7)(d) GDPR. If the data controller must prove the legal proportionality and necessity of automated decisions taken, they would be creating not merely an explanation, but a justification of both data used and profiling mechanisms.[10]

8 Conclusion

There is a growing literature suggesting that Algorithmic Impact Assessments are a crucial tool in establishing algorithmic accountability. This paper addresses that tool as implemented in the GDPR. We find that the GDPR's version of Impact Assessments 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 Algorithmic Impact Assessments beyond the GDPR. In particular, moving from individual transparency rights and governance accountability duties in the field of automated decision-making, we suggest a model of Multi-layered Explanations drawn from Algorithmic Impact Assessments. 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 this field, 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 efficient Algorithmic Impact Assessment that can be re-used as GDPR-complying explanations and disclosures.

ACKNOWLEDGMENTS

We contributed equally to each paragraph. We are grateful to the fruitful comments received by FAT reviewers and CPDP reviewers. We are also grateful to the comments received by colleagues and other experts, in particular the LSTS team. Errors are our own.
The open access was financed by PANELFIT Project, European Union's H2020 research and innovation programme under grant agreement No 788039.

REFERENCES

[1] 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', *International Data Privacy Law* 7, no. 2 (1 May 2017): 76–99, https://doi.org/10.1093/idpl/ipx005.
[2] Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation', *International Data Privacy Law* 7, no. 4 (1 November 2017): 243–65, https://doi.org/10.1093/idpl/ipx019.

[3] Brice Goodman and S. Flaxman, 2016 EU Regulations on Algorithmic Decision-Making and a "right to Explanation," http://arxiv.org/abs/1606.08813, accessed 30 June 2018.
[4] Andrew Selbst and Julie Powles; 'Meaningful information and the right to explanation', *International Data Privacy Law* 7, no. 4 (1 November 2017:, 233–242, https://doi.org/10.1093/idpl/ipx022 .
[5] Maja Brkan, 'Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond', *International Journal of Law and Information Technology*, accessed 24 April 2019, https://doi.org/10.1093/ijlit/eay017;
[6] Margot E. Kaminski, 'The Right to Explanation, Explained', *Berkeley Technology Law Journal*, 34, no. 1 (2019), https://papers.ssrn.com/abstract=3196985.
[7] Antoni Roig, 'Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)', *European Journal of Law and Technology* 8, no. 3 (21 January 2018), http://ejlt.org/article/view/570;
[8] Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For', *Duke Law & Technology Review* 16, no. 1 (4 December 2017): 18–84;
[9] 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', *Berkeley Technology Law Journal* 34, no. 2019 (19 February 2018), https://papers.ssrn.com/abstract=3143325.
[10] Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, Adopted on 3 October 2017, As last Revised and Adopted on 6 February 2018, WP251rev.01, 29.
[11] Margot E. Kaminski, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability', *Southern California Law Review* 92, no. 6 (2019): 4, https://papers.ssrn.com/abstract=3351404.
[12] Michael Veale and Lilian Edwards, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling', *Computer Law & Security Review* 34, no. 2 (1 April 2018): 398–404, https://doi.org/10.1016/j.clsr.2017.12.002.
[13] Jody Freeman, 'The Private Role in the Public Governance', *New York University Law Review* 75 (2000): 543;
[14] K. Bamberger, 'Regulation as Delegation: Private Firms, Decision-making, and Accountability in the Administrative State', *Duke Law Journal*, 1 January 2006, 377.
[15] Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?', *IEEE Security & Privacy* 16, no. 3 (2018): 46–54.
[16] 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, https://doi.org/10.1007/978-981-13-2874-9_6.
[17] Gianclaudio Malgieri, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations', *Computer Law & Security Review*, 9 July 2019, 105327, https://doi.org/10.1016/j.clsr.2019.05.002 See in particular the cases of French and Hungarian laws, that provides more explicit explanation of individual decisions taken (based on criteria and methods used in algorithmic processing).
[18] Raphael Gellert, 'We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences between the Rights-Based and the Risk-Based Approaches to Data Protection', *European Data Protection Law Review* (EDPL) 2 (2016): 481.
[19] Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability', *New Media & Society* 20, no. 3 (1 March 2018): 973–89, https://doi.org/10.1177/1461444816676645.
[20] Mireille Hildebrandt, 'The Dawn of a Critical Transparency Right for the Profiling Era', 56, 2012, https://repository.ubn.ru.nl/handle/2066/94126.
[21] Bryce Goodman, 'A Step Towards Accountable Algorithms? : Algorithmic Discrimination and the European Union General Data Protection', 2016, 3–4.
[22] Joshua Kroll et al., 'Accountable Algorithms', *University of Pennsylvania Law Review* 165, no. 3 (1 January 2017): 633.
[23] Kate Crawford and Jason Schultz, 'Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms', *Boston College Law Review* 55, no. 1 (29 January 2014): 93.
[24] Danielle Citron, 'Technological Due Process', *Washington University Law Review* 85, 1249 (2008): 1310 https://digitalcommons.law.umaryland.edu/fac_pubs/1012.
[25] Danielle Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions', *Washington Law Review* 89, 1 (2014): 20, 26, https://digitalcommons.law.umaryland.edu/fac_pubs/1431.
[26] Deven R. Desai and Joshua A. Kroll, 'Trust But Verify: A Guide to Algorithms and the Law, '*Harvard Journal of Law and Technology* 31, 1.
[27] Lee A Bygrave, 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling', *Computer Law & Security Review* 17, no. 1 (1 January 2001): 18, https://doi.org/10.1016/S0267-3649(01)00104-2.
[28] Kenneth A. Bamberger & Deirdre K. Mulligan, 'PIA Requirements and Privacy Decision-Making in US Government Agencies', in D. Wright & P. De Hert (eds.), Privacy Impact Assessment (2012) at 225.
[29] Reuben Binns, Data protection impact assessments: a meta-regulatory approach, *International Data Privacy Law* 7, 22 (2017).
[30] Roger Clarke, Privacy impact assessment: Its origins and development. *Computer Law and Security Review* 25, 123 (2009).
[31] A. Michael Froomkin, Regulating Mass Surveillance as Privacy Pollution: Learning from Environmental Impact Statements, *University of Illinois Law Review* 1713 (2015).
[32] Chris Jay Hoofnagle, Assessing the Federal Trade Commission's Privacy Assessments, *IEEE Security U& Privacy* 14(2), 58 (2016).
[33] Sonia K. Katyal, 'Private Accountability in the Age of Artificial Intelligence', 66 *UCLA Law Review* 66 54, 112 (2019).
[34] Alessandro Mantelero, 'AI and Big Data: A blueprint for a human rights, social and ethical impact assessment', *Computer Law an Security Review* 34, 754 (2018).
[35] David Wright & Charles D. Raab, Constructing a surveillance impact assessment, Computer Law and Security Review 28, 613 (2012).
[36] Marc L. Roark, 'Human Impact Statements', 54 Washburn L.J. 649 (2015).
[37] Dillon Reisman et al., 'Algorithm Impact Assessment: A Practical Frameworks for Public Agency Accountability', AI Now Institute, n.d., https://ainowinstitute.org/aiareport2018.pdf
[38] Andrew D. Selbst, 'Disparate Impact in Big Data Policing', *Georgia Law Review* 52, 109 (2017), 169.
[39] David Wright & Michael Friedewald, Integrating Privacy and Ethical Impact Assessments, *Science and Public Policy* 40, 755 (2013).
[40] 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-.
[41] United Nations, Human Rights Council, Office of the High Comm'r, Guiding Principles On Business And Human Rights 23–26 (2011), https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf
[42] David Wright and Emilio Mordini, 'Privacy and Ethical Impact Assessment', in Privacy Impact Assessment, ed. David Wright and Paul De Hert, Law, Governance and Technology Series (Dordrecht: Springer Netherlands, 2012), 397–418, https://doi.org/10.1007/978-94-007-2543-0_19.
[43] Le Grand, G., & Barrau, E. (2012). Prior Checking, A Forerunner to Privacy Impact Assessments. (pp. 97–115).
[44] Charles Raab and David Wright, 'Surveillance: Extending the Limits of Privacy Impact Assessment. In: Wright David, De Hert Paul, Editors. Privacy Impact Assessment. Dordrecht', in Privacy Impact Assessment, ed. David Wright and Paul De Hert (Dordrecht: Springer, 2012), 363–83.
[45] 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).
[46] Kristian Lum and William Isaac, 'To Predict and Serve?', Significance 13, no. 5 (2016): 14–19, https://doi.org/10.1111/j.1740-9713.2016.00960.x.
[47] Christian Sandvig et al., 'Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms', 2014
[48] 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
[49] 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
[50] Dariusz Kloza et al., '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
[51] Frank Vanclay et al., '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;

[52] 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.
[53] Alexander A. Boni-Saenz, 'Public-Private Partnerships and Insurance Regulation', *Harvard Law Review* 121 (2008): 1375.
[54] Gilad, It runs in the family,: Meta-regulation and its siblings. Regulation & Governance, 4(4)485, 497 (2010).
[55] Michael Veale, Reuben Binns, and Jef Ausloos, 'When Data Protection by Design and Data Subject Rights Clash', *International Data Privacy Law* 8, no. 2 (1 May 2018): 118, https://doi.org/10.1093/idpl/ipy002.
[56] Linnet Taylor, Luciano Floridi, Bart van der Sloot (eds), Group Privacy: New Challenges of Data Technologies (Springer International Publishing, 2017).
[57] Julia Angwin et al., Machine Bias, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (describing leading risk assessment tools for sentencing and corrections developed by Northpointe);
[58] Sam Corbett-Davies et al., A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not that Clear., WASH. POST, (Oct. 17, 2016) https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-thanpropublicas.
[59] 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.
[60] Pauline Kim, 'Auditing Algorithms for Discrimination', *University of Pennsylvania Law Review Online* 166, no. 1 (1 January 2017): 196, https://scholarship.law.upenn.edu/penn_law_review_online/vol166/iss1/10.
[61] Roger Clarke, 'An Evaluation of Privacy Impact Assessment Guidance Documents', *International Data Privacy Law* 1, no. 2 (1 May 2011): 112, https://doi.org/10.1093/idpl/ipr002.
[62] Niels van Dijk, Raphaël Gellert, and Kjetil Rommetveit, 'A Risk to a Right? Beyond Data Protection Risk Assessments', *Computer Law & Security Review* 32, no. 2 (1 April 2016): 304, https://doi.org/10.1016/j.clsr.2015.12.017.
[63] T. Lombrozo, The structure and function of explanations, *Trends in Cognitive Sciences* 10 (10) , (2006) 464–470.
[64] Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences', *Artificial Intelligence* 267 (1 February 2019): 6, https://doi.org/10.1016/j.artint.2018.07.007.
[65] Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI', *Columbia Business Law Review*, 2019(2).