BINARY GOVERNANCE: LESSONS FROM THE GDPR'S APPROACH TO ALGORITHMIC ACCOUNTABILITY

MARGOT E. KAMINSKI*

Algorithms are now used to make significant decisions about individuals, from credit determinations to hiring and firing. But they are largely unregulated under U.S. law. A quickly growing literature has split on how to address algorithmic decision-making, with individual rights and accountability to nonexpert stakeholders and to the public at the crux of the debate. In this Article, I make the case for why both individual rights and public- and stakeholder-facing accountability are not just goods in and of themselves but crucial components of effective governance. Only individual rights can fully address dignitary and justificatory concerns behind calls for regulating algorithmic decision-making. And without some form of public and stakeholder accountability, collaborative public-private approaches to systemic governance of algorithms will fail.

In this Article, I identify three categories of concern behind calls for regulating algorithmic decision-making: dignitary, justificatory, and instrumental. Dignitary concerns lead to proposals that we regulate

*. Associate Professor of Law, Colorado Law School; Faculty Privacy Director at Silicon Flatirons; Affiliated Fellow, Information Society Project at Yale Law School; Faculty Fellow, Center for Democracy and Technology. Many thanks to Jef Ausloos, Jack Balkin, Michael Birnhack, Frederik Zuiderveen Borgesius, Bryan H. Choi, Kiel Brennan-Marquez, Giovanni Comandé, Eldar Haber, Irene Kamara, Derek H. Kiernan-Johnson, Kate Klonick, Mark Lemley, Gianclaudio Maglieri, Christina Mulligan, W. Nicholson Price, Andrew Selbst, Alicia Solow-Niederman, and Michael Veale for reading and for detailed comments. Thanks to the Fulbright-Schuman program, Institute for Information Law ("IViR") at the University of Amsterdam, and Scuola Sant'Anna in Pisa for the time and resources for this project. Thanks to the faculty of Tel Aviv University, the Second Annual Junior Faculty Forum on the Intersection of Law and Science, Technology, Engineering, and Math (STEM) at the Northwestern Pritzker School of Law, and my own Colorado Law School faculty for excellent workshop opportunities. Extra thanks to Matthew Cushing, whose incredible support made this project possible, and to Mira Cushing for joy beyond words.

algorithms to protect human dignity and autonomy; justificatory concerns caution that we must assess the legitimacy of algorithmic reasoning; and instrumental concerns lead to calls for regulation to prevent consequent problems such as error and bias. No one regulatory approach can effectively address all three. I therefore propose a two-pronged approach to algorithmic governance: a system of individual due process rights combined with systemic regulation achieved through collaborative governance (the use of private-public partnerships). Only through this binary approach can we effectively address all three concerns raised by algorithmic decision-making, or decision-making by Artificial Intelligence ("AI").

The interplay between the two approaches will be complex. Sometimes the two systems will be complementary, and at other times, they will be in tension. The European Union's ("EU's") General Data Protection Regulation ("GDPR") is one such binary system. I explore the extensive collaborative governance aspects of the GDPR and how they interact with its individual rights regime. Understanding the GDPR in this way both illuminates its strengths and weaknesses and provides a model for how to construct a better governance regime for accountable algorithmic, or AI, decision-making. It shows, too, that in the absence of public and stakeholder accountability, individual rights can have a significant role to play in establishing the legitimacy of a collaborative regime.

TABLE OF CONTENTS

INTRODUCTION

In 2011, fifth-grade teacher Sarah Wysocki was fired by a machine.[1] Wysocki had been a star teacher, receiving rave reviews from both parents and her principal. But these positive reviews from humans were outweighed by a low score Wysocki received from an algorithm (a computer program), and the school district was required by policy to fire her, along with 205 other teachers.

When Wysocki and others questioned this outcome and asked how the scores were calculated, they were told that the algorithm was too complicated for them to understand. Wysocki suspected, however, that the model relied heavily on standardized test scores and that due to cheating, over half of her students started the school year with artificially inflated exam scores. When she tried to make her case to the school district, however, she was told that the decision was final.[2]

Algorithms—that is, computer programs, including AI—are now used to make a host of decisions about human beings that have significant impacts on human welfare and dignity.[3] Decisions about whether to extend credit, whether to hire or fire somebody, how to price discriminate, and what

1. CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION 11 (2016); Bill Turque, '*Creative . . . Motivating' and Fired*, WASH. POST (Mar. 6, 2012), https://www.washingtonpost.com/local/education/creative--motivating-and-fired/2012/02/04/gIQAwzZpvR_story.html.

2. Recently, Houston-area teachers who were similarly fired based on the use of a secret computer algorithm successfully argued that their dismissals may have violated their procedural due process rights. Houston Federation of Teachers, Local 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1180 (S.D. Tex. 2017). This ruling, however, applies only to government actions and to dismissals during the term of a contract or a continuing contract. *Id.* at 1173–74. Thanks to Mark MacCarthy for identifying this case.

3. For more extended descriptions and definitions of decision-making algorithms, see Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 67–88 (2017) (discussing the actuarial algorithms used in recidivism risk assessment); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 658–62 (2017) (discussing machine-learning algorithms).

educational track to put a student on now can all involve algorithmic decision-making, or a heavy reliance on suggestions arrived at by algorithms. These decisions, when made by the private sector, go largely unregulated in the United States.[4] The types of algorithms used can present significant challenges to our legal system, as they are often secret, impossible to predict, hard to explain to nonexperts, and continuously changing over time.

Earlier calls for algorithmic accountability propose addressing algorithmic decision-making through both individual "due process" rights and an array of systemic accountability measures.[5] The systemic measures include public disclosure of source code, agency oversight, expert boards, and stakeholder input. More recent literature, however, moves away from these proposals, questioning both the value of individual rights and stakeholder input by nonexperts and the costs of public disclosure. Public-facing accountability and individual due process in particular have become

4. This is not to say that algorithmic decision-making goes entirely unregulated. There are a number of existing laws in the United States that were not written for, but may be applicable to, algorithmic decision-making. For example, the Department of Housing and Urban Development charged Facebook in 2019 with allegedly violating the Fair Housing Act, by using machine learning algorithms to discriminate in housing advertisements against users based on membership in protected classes. Facebook, Inc., FHEO No. 01-18-0323-8 (U.S. Dep't of Hous. & Urban Dev. Mar. 28, 2019), https://www.hud.gov/sites/dfiles/Main/documents/HUD_v_Facebook.pdf. As the Obama White House noted, however, the use of big data has "the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace." JOHN PODESTA ET AL., EXEC. OFFICE OF THE PRESIDENT, BIG DATA (2014), http://obama whitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf. For a more detailed analysis of the difficulties inherent in applying antidiscrimination laws, such as Title VII of the Civil Rights Act, to algorithmic decision-making, see Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 714–23 (2016) [hereinafter Barocas & Selbst, *Disparate Impact*]. *But see* James Grimmelmann & Daniel Westreich, *Incomprehensible Discrimination*, 7 CALIF. L. REV. ONLINE. 164, 170 (2017) (arguing that with judicial adaptations antidiscrimination law can address algorithmic decision-making). The Fair Credit Reporting Act ("FCRA"), too, contains due process-like and explanatory requirements for credit reporting agencies that apply to the use of algorithmic decision-making. Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 16–18 (2014) (discussing the coverage and limitations of the FCRA); Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1100–05 (2018) [hereinafter Selbst & Barocas, *Intuitive Appeal*] (discussing the FCRA and Equal Credit Opportunity Act). In the context of employment decisions where the employer is in the public sector, due process rights may also apply. *See supra* note 2 and accompanying text.

5. I join several scholars in calling for governance of algorithmic decision-making that incorporates both individual rights and systemic oversight. Danielle Citron and Frank Pasquale presciently propose a mix of individual rights with systemic governance, as do Jason Schultz and Kate Crawford. These scholars do not, however, characterize their proposed systemic regulation as collaborative governance nor discuss at length the interaction between the individual and systemic approaches. Citron & Pasquale, *supra* note 4, at 27–32; Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 124–28 (2014). And as other scholars in this space have since pushed back against the efficacy of individual rights, it is worth revisiting their value.

the straw men of algorithmic accountability: held up as ineffective or even harmful, in contrast to the specific tools a particular author or set of authors propose in their place.

In this Article, I make the case for both individual rights and public- and stakeholder-facing accountability not just as goods in and of themselves but as crucial components of effective governance. I begin by examining the reasons behind calls for algorithmic accountability and find that this explains the split in the literature over the role of individual rights. Then, unlike earlier authors, I draw on regulatory design literature—specifically on collaborative governance, or governance through public-private partnerships—to explain why both public- and stakeholder-facing accountability are not luxuries but necessary tools.

My claims are both descriptive and normative. As a descriptive matter, I identify that many recent calls for systemic approaches to algorithmic accountability are in fact calls for collaborative governance. I also identify that the EU's GDPR represents an attempt to use collaborative governance towards algorithmic accountability, combined with a system of individual rights.

As a normative matter, I make two claims. First, governing algorithmic decision-making *should* include both individual rights and systemic approaches. Recent authors are correct that individual rights are not the best way to approach instrumental goals, but should not so readily dismiss dignitary and justificatory concerns about algorithmic decision-making. These concerns dictate a need for some form of individual process, at least when decisions have a significant effect on an individual. But at the same time, a systemic approach is necessary to address risks on a system-wide level and to target the various stages of development and training when algorithms (or really the humans who build them) are more amenable to regulation.

Second, if we are going to use the tools of collaborative governance to govern algorithmic decision-making—and there are good reasons, at least on paper, to do so—we have to be vigilant about designing a regulatory system that is legitimate, immunized from capture, and strong enough that it is different in kind from self-regulation. Collaborative governance is described, in brief, as a better way to govern fast-changing, risky systems with a high degree of technological complexity. There are reasons to think it might be a good fit for governing algorithmic decision-making. But effective collaborative governance requires both a regulator with real power (for example, the GDPR's famously stringent fines) and what Jody Freeman has

called "aggregate accountability":[6] accountability that runs through the system as a whole, and accounts not just to a regulator or group of technical experts but to representatives of affected stakeholders and to the public.

Otherwise, the regulatory system will be both less legitimate and less effective. It will be less legitimate because of the high risk of capture. And it will be less effective because it will fail to deploy those third-party resources—from external expertise, to increased enforcement power, to naming and shaming, to market feedback, to public input into policy—that make collaborative governance appealing to resource-strapped regulators in the first place.

If we fail to design the regulatory system with these inputs and outputs in place, then—as I argue in Part III is the case with the GDPR—we may find ourselves needing to rely on individual transparency rights to accomplish systemic accountability goals. This forms yet another argument for putting individual rights into place. That is, if the systemic side of governance involves only the government in dialogue with affected parties, then third-party stakeholders and members of the public will need to devise creative ways of using individualized disclosure about algorithmic decision-making to ensure the system of governance is not captured and remains legitimate.

This Article proceeds as follows: in Part I, I discuss at length the three distinct categories of reasons behind calls for regulating algorithmic decision-making. The first is a dignitary rationale, concerned with both personhood and autonomy in the face of complex and secret profiling and decision-making systems. The second is a justificatory rationale, concerned with ensuring that decisions are made based on socially and legally acceptable reasoning and are legitimized by acceptable process or oversight. The third is an instrumental rationale, advocating regulation to ensure that algorithmic decisions are not erroneous, faulty, biased, or outright discriminatory. These rationales often overlap, and regulating in the name of one will often help address another. But distinguishing them explains both why articles on algorithmic decision-making currently speak past each other and why a number of recently proposed regulatory solutions are incomplete.

If governments wish to address all three concerns, they will need to adopt a two-pronged, or binary, approach to algorithmic accountability. Part II describes this proposed system. The first prong, based on individual due process, addresses dignitary and justificatory concerns, but may be less

6. Jody Freeman, *The Private Role in Public Governance*, 75 N.Y.U. L. REV. 543, 664–67 (2000).

effective at producing a system-wide, instrumental impact. The second prong, a systemic regulatory approach, addresses some systemic justificatory concerns, but largely serves the instrumental goals of addressing error, unfairness, bias, and discrimination.

In Part II, I identify that the systemic governance prong of the regulatory system will involve collaborative governance, or "new governance"—that is, public-private partnerships in governance.[7] This claim is, again, both descriptive and normative. It is descriptive in that a number of calls for algorithmic accountability in fact already deploy the tools of collaborative governance without realizing they are doing so: auditing, whistle-blower protection, risk assessments, and more.[8] It is normative in that collaborative governance—which is not to be confused with self-regulation or deregulation—has several advantages in these circumstances over purely top-down, command-and-control regulation. Collaborative governance is generally described as better suited for regulating highly complex systems that create hard-to-calculate risks, change too quickly for traditional regulatory approaches, and involve technical and industry

7. *See, e.g.*, *id.* at 592–664; Jody Freeman, *Collaborative Governance in the Administrative State*, 45 UCLA L. REV. 1, 21–33 (1997); Orly Lobel, *New Governance as Regulatory Governance*, *in* THE OXFORD HANDBOOK OF GOVERNANCE 65, 66–67 (David Levi-Faur ed., 2012) [hereinafter Lobel, *New Governance*]; Orly Lobel, *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, 89 MINN. L. REV. 342, 371–76 (2004) [hereinafter Lobel, *The Renew Deal*].

8. Only a handful of scholars have explicitly considered using collaborative governance, in sector-specific contexts, to govern algorithmic decision-making or AI. *See* Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473, 529–31 (2016) ("We advocate a collaborative-dynamic regulation . . . ."); W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421, 465–71 (2017) (discussing collaborative governance of black box, medical algorithms). Sonya Katyal has called for the greater involvement of the private sector in algorithmic accountability, pointing to both self-regulation and whistleblower protection, but does not situate this approach in the literature of collaborative governance. Sonya K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 61 (2019) ("[W]e are looking in the wrong place if we look to the state alone to address issues of algorithmic accountability. . . . [I]t makes sense to explore opportunities for greater endogeneity in addressing civil rights concerns, particularly given the information asymmetry between the industries that design AI and the larger public."). Michael Guihot and others have similarly outlined "this decentering of regulation and . . . examples of peer or self-regulation that has begun to proliferate in the vacuum of government control," discussing responsive regulation but noting that in the case of AI no backdrop regulatory framework is currently in place. Michael Guihot et al., *Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence*, 20 VAND. J. ENT. & TECH. L. 385, 427 (2017). They suggest that a "[r]eally [*r*]*eally* [r]esponsive [r]isk-[b]ased [r]egulation" might be appropriate, ultimately calling for a mix of self-regulation and risk regulation in the form of soft-law, regulatory "nudges." *Id.* at 441, 445. Alicia Solow-Niederman has joined this conversation about possible regulatory toolkits and regulatory design, in a forthcoming piece. Alicia Solow-Niederman, *Administering Artificial Intelligence*, 93 S. CAL. L. REV. (forthcoming May 2020).

expertise that regulators and legislators are unlikely to have.[9] Collaborative governance, at least in theory, should be well suited to improving not just algorithms but the complex human systems around them.[10] It should be better suited to addressing solutions towards design and training phases, rather than just addressing the running model.[11]

Identifying the use of collaborative governance in establishing algorithmic accountability lets us (1) take a thorough and system-wide approach to accountability rather than deploy partial solutions; (2) engage with existing governance literature and examples rather than attempt to reinvent the regulatory wheel; (3) expand the existing algorithmic accountability toolkit; and (4) recognize that the accountability problem is in fact multilayered. Because regulators will be delegating rulemaking of sorts to private parties, we need not just transparency and oversight over the algorithm, but second-order transparency and oversight over that rulemaking and compliance process.[12]

By characterizing algorithmic accountability as collaborative governance, this Article identifies a second-order accountability problem that largely has gone ignored. Calls for transparency of algorithmic decision-making cannot so easily be dismissed by those who focus on instrumental, rather than dignitary or justificatory, concerns. Transparency and deeper forms of accountability play an essential role not just in protecting human dignity but in establishing a legitimate and well-functioning system of collaborative governance.

Part II closes by discussing the interaction between the two parts of this proposed binary system: individual process and systemic regulation involving collaborative governance. The interaction between these two prongs will be complex. At times they will be complementary, and at other times their goals and approaches will conflict.

9. *See* Kenneth A. Bamberger, *Regulation as Delegation*: *Private Firms, Decisionmaking, and Accountability in the Administrative State*, 56 DUKE L.J. 377, 385–92 (2006).

10. Mike Ananny & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC'Y 973, 983–84 (2018) (discussing algorithmic "assemblages" of humans and machines); Lehr & Ohm, *supra* note 3, at 657 ("Because playing with the data occurs earlier in time and entails much more human involvement than the running model, this phase provides more opportunities and behavioral levers for policy prescriptions.").

11. Lehr & Ohm, *supra* note 3, at 655–58.

12. IAN AYRES & JOHN BRAITHWAITE, RESPONSIVE REGULATION 57–60 (1992); Ian Ayres & John Braithwaite, *Tripartism: Regulatory Capture and Empowerment*, 16 LAW & SOC. INQUIRY 435, 491 n.137 (1991); *see also* Margot E. Kaminski, Regulating AI Risk Through the GDPR 31–32 (June 24, 2019) (unpublished manuscript) (on file with author).

By way of illustration, Part III of this Article turns to Europe's GDPR, which takes a complex, binary approach to regulating algorithmic decision-making.[13] This Article identifies that the GDPR's systemic approach to governing algorithms relies heavily on collaborative governance. It is the first to discuss the GDPR's approach to algorithmic accountability as a binary system that illustrates how an individual rights approach might interact with its collaborative governance aspects. Understanding the GDPR in this way both illuminates its strengths and weaknesses and provides a model for how to construct a better regulatory regime for algorithmic decision-making.

I. WHY REGULATE ALGORITHMIC DECISION-MAKING?

A growing body of literature calls for regulating algorithmic decision-making.[14] This Part identifies three categories of concerns behind these calls: dignitary (which includes autonomy), justificatory, and instrumental.[15] These categories can overlap, but they also lead to divergent regulatory solutions. To understand current divides in the literature, it is necessary to understand the three goals motivating calls for algorithmic accountability.

13. Council Regulation 2016/679, 2016 O.J. (L 119) 1 (EU).

14. *See, e.g.*, Ananny & Crawford, *supra* note 10, at 984–85; Lee A. Bygrave, *Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, 17 COMPUTER L. & SECURITY REP. 17, 21–22 (2001); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1301–13 (2008); Citron & Pasquale, *supra* note 4, at 18–28; Crawford & Schultz, *supra* note 5, at 109–10; Deven R. Desai & Joshua A. Kroll, *Trust but Verify*: *A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 5–6 (2017); Guihot et al., *supra* note 8, at 427; Mireille Hildebrandt, *The Dawn of a Critical Transparency Right for the Profiling Era*, *in* DIGITAL ENLIGHTENMENT YEARBOOK 2012, at 41, 49–54 (Jacques Bus et al. eds., 2012); Katyal, *supra* note 8, at 107–08; Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189, 197–202 (2017); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 696–99 (2017); Isak Mendoza & Lee A. Bygrave, *The Right Not To Be Subject to Automated Decisions Based on Profiling*, *in* EU INTERNET LAW 77, 96–97 (Tatiani-Eleni Synodinou et al. eds., 2017); Price, *supra* note 8, at 432–37; Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 408–09 (2014); Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. TECH. 353, 373–76 (2016); Paul Schwartz, *Data Processing and Government Administration: The Failure of the American Legal Response to the Computer*, 43 HASTINGS L.J. 1321, 1374–86 (1992); Solow-Niederman, *supra* note 8; Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1, 64–78 (2005); Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 105–11 (2017); Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1530–53.

15. Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1117–19 (identifying the three overlapping reasons for calls for transparency in the literature as (1) "autonomy, dignity and personhood"; (2) an "instrumental value [to] . . . educat[e] the subjects of automated decisions about how to achieve different results"; and (3) "the idea that explaining the model will allow people to debate whether the model's rules are justifiable"); *see also* Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking for*, 16 DUKE L. & TECH. REV. 18, 27–43 (2017) (identifying regulatory reasons as "[d]iscrimination and [u]nfairness," "[i]nformation[] [p]rivacy," and "[o]pacity and [t]ransparency").

More importantly, to design a regulatory solution that works, we have to articulate what we want regulation to accomplish.

Calls for algorithmic decision-making often start from the premise that it replaces something worse: decision-making by humans. Human decision-making can be deeply, terribly flawed. Human decision makers can be outright discriminatory; can hold deep-seated biases about race, gender, or class; and can exhibit a host of cognitive biases that invisibly influence outcomes.[16] The example of "redlining" in which banks denied (and in some cases, continue to deny) housing loans to African Americans, Latinos, and other minorities exemplifies outright discrimination by human decision makers.[17] A 2016 study of interview practices by elite law firms evidenced bias against candidates on the basis of both social class and gender.[18] And thanks largely to Daniel Kahneman and Amos Tversky, the legal literature has become aware of the host of cognitive biases that human decision makers have.[19] Human decisions are influenced, for example, by anchoring—people tend to rely heavily on one piece of information, often the first piece of information acquired, rather than equally weight relevant inputs.[20] Humans are subject to confirmation bias—the tendency to read all information in a way that confirms already-held opinions.[21] These and other cognitive biases suggest that human decision makers, even when not being overtly discriminatory, are not impartial or even particularly accurate.

It is thus tempting to believe that machines will be better. But even complex algorithms are simplifications of reality, and these simplifications involve human choices along a number of axes.[22] As Cathy O'Neil has

16. *See, e.g.*, Amos Tversky & Daniel Kahneman, *Judgment Under Certainty: Heuristics and Biases*, *in* JUDGMENT UNDER UNCERTAINTY 3, 3–8 (Daniel Kahneman et al. eds, 1982); Jon D. Hanson & Douglas A. Kysar, *Taking Behavioralism Seriously: The Problem of Market Manipulation*, 74 N.Y.U. L. REV. 630, 643–45 (1999); Tamara R. Piety, *"Merchants of Discontent": An Exploration of the Psychology of Advertising, Addiction, and the Implications for Commercial Speech*, 25 SEATTLE U. L. REV. 377, 402–03 (2001) (describing cognitive biases).

17. JOHN PODESTA ET AL., *supra* note 4, at 53 ("'Redlining,' in which banks quite literally drew—and in cases continue to draw—boundaries around neighborhoods where they would not loan money, existed for decades as a potent tool of discrimination against African-Americans, Latinos, Asians, and Jews.").

18. Lauren A. Rivera & András Tilcsik, *Class Advantage, Commitment Penalty: The Gendered Effect of Social Class Signals in an Elite Labor Market*, 81 AM. SOC. REV. 1097, 1108–11 (2016).

19. For further discussion on this topic, see generally JUDGMENT UNDER UNCERTAINTY, *supra* note 16.

20. Hanson & Kysar, *supra* note 16, at 667–69; Piety, *supra* note 16, at 403.

21. Piety, *supra* note 16, at 402; *see also* Hanson & Kysar, *supra* note 16, at 647–50.

22. O'NEIL, *supra* note 1, at 20–21; Eaglin, *supra* note 3, at 63 (describing the "normative judgments embedded in actuarial risk assessment tools' construction"); Katyal, *supra* note 8, at 67 ("Since algorithmic models reflect the design choices of the humans who built them, they carry the biases of the

written, algorithms are not neutral: "Models are opinions embedded in mathematics."[23] Turning to algorithmic decision-making risks cloaking the very things we find problematic in human decision-making under a veneer of technical impartiality. And where human decision-making can often be contested, algorithmic decision-making, as we saw in the example of teacher Wysocki above, is often taken at face value and left unchallenged and unchallengeable.[24]

Algorithms are, effectively, mathematical models of the real world.[25] Statisticians, or "data scientists," construct algorithms to take in data and find correlations or make predictions. Humans actively design algorithms in a number of ways: they pick an algorithm's objectives, decide what the input will be, decide whether to use proxies, decide how to weight the data, decide how to clean the data, choose what type of algorithm to use, decide how to "validate" the algorithm (check that it is working), and determine how reliable the decisions need to be—and, once the model is running, decide whether and how to confirm in practice that it is producing correct results. Increasingly sophisticated algorithms can create their own rules and produce "intuitions" humans do not have.[26] But for any kind of algorithm, human choices and assumptions go into its construction, training, and oversight—or lack thereof.

In some applications—for example, in those used regularly in baseball—algorithms work extraordinarily well.[27] The clearer and more mathematical the objective ("Show me the player who is most likely to hit a home run") the more detailed and direct the data (measurements of actual performance rather than proxies for it), the more transparent the inputs and code (so that they can be double-checked by others), the more easily verifiable the outcomes (home runs hit or not), and the more likely it is that an algorithm tracks what you want it to track and produces "fair" or

observer or instrument."); Lehr & Ohm, *supra* note 3, at 705 n.187 ("[A]t the machine-learning stages we consider here, not only do analysts have to consider technical methods for achieving fairness, but they have to wrestle with highly normative questions of what kind of fairness matters most in a given context."); *see also* Danah Boyd & Kate Crawford, *Critical Questions for Big Data*, 15 INFO., COMM. & SOC'Y 662, 667 (2012) (observing that the process of "'data cleaning' . . . is inherently subjective").

23. O'NEIL, *supra* note 1, at 21.

24. Citron, *supra* note 14, at 1271–72 (describing the tendency to defer to machine decisions).

25. *But see* O'NEIL, *supra* note 1, at 19; Eaglin, *supra* note 3, at 91; Lehr & Ohm, *supra* note 3, at 684.

26. Steven Strogatz, *One Giant Step for a Chess-Playing Machine*, N.Y. TIMES (Dec. 26, 2018), https://www.nytimes.com/2018/12/26/science/chess-artificial-intelligence.html ("Most unnerving was that AlphaZero seemed to express insight.").

27. O'NEIL, *supra* note 1, at 17 ("Baseball is an ideal home for predictive mathematical modeling.").

"accurate" outputs.[28] But when an objective is hard to measure or articulate in mathematical terms ("Show me the best teacher"); the data is limited, muddy, and filled with proxies (standardized test scores or teacher evaluations); the algorithm itself is hidden (whether by technology, policy, or law); the results are hard to test or produce a skewing feedback loop (fired teachers cannot redeem themselves and, thus, prove that the model was wrong); then the risk that an algorithm will produce bad outputs goes up.[29]

Thus the dominant rationale for regulating algorithmic decision-making is an instrumental (or consequentialist) rationale. We should regulate algorithms, this reasoning goes, to prevent the consequences of baked-in bias and discrimination and other kinds of error. Algorithmic decision-making can be erroneous, based on incorrect facts or derived from incorrect inferences.[30] Algorithmic decision-making can be biased, reflecting biased decisions made by programmers or historic discrimination baked into the data sets on which algorithms are trained.[31] Algorithmic decision-making can be intentionally discriminatory, hiding discriminatory motives behind proxy rationales.[32] Or, algorithms can work perfectly well but be used for bad objectives, such as targeting individuals for exploitative payday loans.[33] Machine learning systems can also crash or function in extraordinarily out-of-the-box ways compared to human decision-making.

The instrumental rationale for regulating algorithmic decision-making counsels that regulation should try to correct these problems, often by using systemic accountability mechanisms, such as ex ante technical requirements, audits, or oversight boards, to do so.[34] Accountability for individual

28. *Id*.

29. *Id*. at 4–11.

30. *See* Citron & Pasquale, *supra* note 4, at 8; Crawford & Schultz, *supra* note 5, at 104; Zarsky, *supra* note 14, at 1506.

31. *See* Barocas & Selbst, *Disparate Impact*, *supra* note 4, at 677–94; Citron, *supra* note 14, at 1262; Citron & Pasquale, *supra* note 4, at 13–14; *see also* Joy Buolamwini, *How I'm Fighting Bias in Algorithms*, TED (Nov. 2016), https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms?language=en.

32. Barocas & Selbst, *Disparate Impact*, *supra* note 4, at 691–92.

33. O'NEIL, *supra* note 1, at 82–83.

34. *See, e.g.*, Desai & Kroll, *supra* note 14, at 39 ("[I]f an algorithm was not 'designed with future evaluation and accountability in mind,' no amount of software testing—even aided by total transparency—will always work to elucidate any particular question." (citation omitted)); Edwards & Veale, *supra* note 15, at 76, 82 (referring to a "structural approach" to accountability; calling for a "focus *a priori* on the creation of better algorithms, as well as creative ways for individuals to be assured about algorithmic governance"; and advocating for "creating better systems, with less, opacity, clearer audit trails, well and holistically trained designers, and input from concerned publics[, which] seems eminently more appealing than" individual transparency (footnote omitted)); *see also* Kroll et al, *supra* note 14, at 659.

decisions may have some place in this approach (for example, when experts such as lawyers or physicians might need to overrule recommendations made by algorithms),[35] but scholars driven primarily by instrumental concerns tend to discount the value of individualized transparency or process and instead emphasize a systemic regulatory approach, for reasons discussed further in Part II.

The other two rationales for regulating algorithmic decision-making, however, suggest that systemic oversight is not enough. Both dignitary and justificatory reasoning point towards including individual rights. Sometimes vague in the abstract and thus often discounted, dignitary and justificatory rationales suggest that fired teachers deserve a chance to understand and contest the data, the reasoning, and even the objectives behind their firing. It is not just that the system as a whole needs to be refined or corrected. The individual subject to decisions by the system is intuitively owed some form of process; the big questions are why, when, and what.

The dignitary argument—which for U.S. readers skeptical of dignity includes what are often characterized as autonomy concerns—posits that an individual human being should be respected as a whole, free person. Being subjected to algorithmic decision-making threatens individuals' personhood by objectifying them.[36] Objectification defeats autonomy: the freedom to make choices, be offered opportunities, or otherwise move freely through the world.[37] Objectification can take a variety of forms ranging from directly denying autonomy to treating somebody as fungible (that is, exchangeable with someone else).[38] Dignitary critiques of algorithmic decision-making reflect this range.

It may help to pin down some more specific versions of the dignitary

35. Daniel N. Kluttz & Deirdre K. Mulligan, Automated Decision Support Technologies and the Legal Profession 37 (June 22, 2019) (unpublished manuscript) (on file with author) ("Technical design should seek to put professionals and decision support systems in conversation, not position professionals as passive recipients of system wisdom who must rely on out-of-system mechanisms to challenge them. For these reasons, calls for explainability . . . should be replaced by governance approaches that promote contestable systems.").

36. IMMANUEL KANT, GROUNDWORK FOR THE METAPHYSICS OF MORALS 36 (James W. Ellington trans., Hackett Publ'g Co. 3d ed. 1993) (1785) ("Act in such a way that you treat humanity . . . always at the same time as an end and never simply as a means."); James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151, 1180–89 (2004) (describing the German philosophical origins of privacy law).

37. Zarsky, *supra* note 14, at 1545–50.

38. Martha C. Nussbaum, *Objectification*, 24 PHIL. & PUB. AFF. 249, 256–57 (1995) (arguing that there are seven forms of objectification: instrumentalizing to achieve a further purpose; denying autonomy; treating as inert, as fungible, as violable, as owned by another person; and denying subjectivity).

criticism.[39] The first, largely European, criticism of algorithmic decision-making is that allowing a decision about humans to be made by a machine inherently treats humans as objects, showing deep, inherent disrespect for peoples' humanity.[40] This particular version of a dignitary concern has been embraced by European legislators, while U.S. lawmakers and many U.S. persons have rejected it.

A second type of dignitary concern appeals, however, across cultural divides. Automatically making decisions based on what categories an individual falls into—that is, what correlations can be shown between an individual and others—can fail to treat that individual as an individual.[41] If algorithmic decision-making does not allow individuals to proclaim their individuality ("I may look like these other people, but I am not in fact like them"), then it violates their dignity and objectifies individuals as their traits, rather than treating an individual as a whole person.[42] Both decisional discretion and individual process rights are, under this reasoning, necessary not just to prevent error but to adequately recognize and respect individuality.[43]

This version of a dignitary concern resonates across the Atlantic with the legal tradition of equity. Long concerned with the unfairness of formalistic application of legal rules, the principles and practices of equity allow otherwise unfair decisions to be adapted to individual circumstances.[44] Decisions in equity can permit courts to look beyond factors the law ordinarily considers, to think about fairness in a particular set of circumstances. Similarly, the subject of an automated decision should be able to explain why an algorithm's framing is not the full picture and to introduce individualizing, sometimes mitigating, factors an algorithm has not considered.

Take, for example, the United States Department of Agriculture ("USDA") algorithm that decided that Somali-owned grocery stores in Seattle would no longer be permitted to accept food stamps because

39. Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1119 ("[T]he personhood rationale can be converted to a more actionable legal issue . . . .").

40. Meg Leta Jones, *The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood*, 47 SOC. STUD. SCI. 216, 231 (2017); Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995, 1016–17 (2017).

41. *See* Zarsky, *supra* note 14, at 1545–50.

42. Thanks to Alon Harel for helping to clarify this point at the Tel Aviv University Faculty seminar.

43. *See* Citron, *supra* note 14, at 1304.

44. *See, e.g.*, Golden Press, Inc. v. Rylands, 235 P.2d 592, 596 (Colo. 1951) (en banc).

customer behavior purportedly indicated cash-for-stamp fraud.[45] The algorithm responded to "suspicious transactions," such as even dollar amounts and large purchases made in short time spans. But the grocers had credible contextual explanations for these purportedly "unusual" practices: the shopping patterns of East African immigrants, which include shopping in groups and shopping for meat by whole-dollar amounts at Halal butchers.[46] In addition to having instrumental implications—the algorithm incorrectly identified this behavior as fraud—the system's failure to consider contextualizing factors to which the algorithm was blind led to unfair outcomes that objectified decisional subjects in too-broad strokes ("I may look like a cheat but I am not").

This notion that a person is more than the sum of her abstracted traits resonates with a concern in the privacy literature over the notion of the "data double": a shadow self consisting of data points gathered about an individual, often without permission.[47] People constantly engage in the process of self-construction, determining both how they appear to others and who they are to themselves.[48] A data double objectifies an individual by taking this dynamic, participatory process and placing it in the hands of other entities and out of the hands of the individual.

Secret profiles and decisions made based on secret profiling can threaten personhood and thus dignity by proscribing active individual involvement in the construction of this objectified version of the self.[49] This again resonates with aspects of U.S. law, even given the absence of federal data privacy law: the right of publicity and appropriation torts, which permit individuals to sue to protect use of their likenesses without permission; protection against public disclosure of private fact; and defamation, which

45. *Somali Grocers: Feds Urged To Requalify Grocers for Food Stamps*, KITSAP SUN (Apr. 14, 2002), https://products.kitsapsun.com/archive/2002/04-14/0038_somali_grocers__feds_urged_to_req.html. Thanks to Deirdre Mulligan for pointing to this example. For an overview of the USDA Supplemental Nutrition Assistance Program algorithm, see H. Claire Brown, *How an Algorithm Kicks Small Businesses Out of the Food Stamps Program on Dubious Fraud Charges*, NEW FOOD ECON. (Oct. 8, 2018), http://newfoodeconomy.org/usda-algorithm-food-stamp-snap-fraud-small-businesses.

46. Chris McGann, *Somali Grocers Lose Right To Use Food Stamps*, SEATTLE POST-INTELLIGENCER (Apr. 8, 2002, 10:00 PM), https://www.seattlepi.com/news/article/Somali-grocers-lose-right-to-use-food-stamps-1084746.php.

47. *See, e.g.*, David Lyon, *Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique*, BIG DATA & SOC'Y, July–Dec. 2014, at 1, 6; *see also* Evelyn Ruppert, *The Government Topologies of Database Devices*, THEORY, CULT. & SOC'Y, Oct. 4, 2012, at 116, 123–26.

48. JULIE E. COHEN, CONFIGURING THE NETWORKED SELF 49–50 (2012); Hildebrandt, *supra* note 14, at 47–48.

49. Bygrave, *supra* note 14, at 18 (explaining that the " 'data shadows' . . . threaten to usurp the constitutive authority of the physical self despite their relatively attenuated and often misleading nature").

allows individuals to legally contest certain lies made about them.[50] The rights of correction or amendment in select sectoral U.S. privacy laws—for example, the Privacy Act[51] and the Health Insurance Portability and Accountability Act's ("HIPAA's") Privacy Rule[52]—also reflect this dignitary concern with individual participation in the creation of an object-self.

Finally, dignitary concerns include concerns (more familiar to Americans) about individual autonomy. Algorithmic decision-making founded on individual profiling limits the choices and, thus, the freedom a person will have.[53] A teacher's autonomy is circumscribed when the teacher is fired based on information and reasoning she cannot contest. There are fewer choices and, thus, less freedom in what employment opportunities the teacher can go on to pursue. Algorithms, too, determine the online ads we see, often on the basis of individual profiling.[54] Ads for higher-paid, executive level, and science, technology, engineering, and math jobs have been shown to target men over women.[55] Companies have been shown to target African Americans with ads for payday loans and credit cards with disadvantageous terms.[56] Limiting the choices we see—whether by failing to show opportunities or by offering only bad options—limits our freedom to make choices.

Failing to be transparent about the fact that individuals are being targeted or the reasons why they are targeted itself may threaten autonomy. Secret profiling and decision-making can lead to manipulation.[57] Without

50. RESTATEMENT (SECOND) OF TORTS §§ 558, 652A, 652C, 652E (AM. LAW. INST. 1977); William L. Prosser, *Privacy*, 48 CALIF. L. REV. 383, 389 (1960).

51. Privacy Act of 1974, Pub. L. 93-579, 88 Stat. 1896 (codified at 5 U.S.C. § 552a(d)(2) (2018)).

52. 45 C.F.R. § 164.526 (2019) (establishing the right to request amendment of records and right to submit a statement of disagreement if amendment is denied).

53. Tal Zarsky, *The Trouble with Algorithmic Decisions*: *An Analytic Road Map To Examine Efficiency and Fairness in Automated and Opaque Decision Making*, SCI., TECH., & HUM. VALUES 118, 129–30 (2016) [hereinafter Zarsky, *Trouble*]; Zarsky, *supra* note 14, at 1541–50.

54. Patrick Kulp, *Facebook Has Discriminated Against You, and It's Not Going To Stop*, MASHABLE (Nov. 12, 2016), https://mashable.com/2016/11/12/facebook-google-ad-discrimination.

55. Tom Simonite, *Probing the Dark Side of Google's Ad-Targeting System*, MIT TECH. REV. (July 6, 2015), https://www.technologyreview.com/s/539021/probing-the-dark-side-of-googles-ad-targeting-system.

56. Alvaro Bedoya & Clare Garvie, Center on Privacy & Technology at Georgetown Law, Comment Letter on Follow the Lead: An FTC Workshop on Lead Generation (Dec. 18, 2015), https://www.ftc.gov/system/files/documents/public_comments/2015/12/00017-99877.pdf; O'NEIL, *supra* note 1, at 157–58.

57. Zarsky, *supra* note 14, at 1541–53; *see also* FREDERIK J ZUIDERVEEN BORGESIUS, IMPROVING PRIVACY PROTECTION IN THE AREA OF BEHAVIORAL TARGETING 62 (2015) ("Privacy isn't merely about control. Privacy is about not *being* controlled. . . . Privacy as identity construction concerns protection against unreasonable steering or manipulation—by humans or by technology." (footnote omitted)).

knowing how we are being targeted or why, we can be manipulated into making choices that are not autonomous at all.[58] Concerns about autonomy and the potential for manipulation, to a great degree, motivated the indignation around Cambridge Analytica's targeted manipulation of U.S. voters prior to the 2016 election (and motivated the California legislature to enact the California Consumer Privacy Act in 2018).[59]

The third category of concerns about algorithmic decision-making, justificatory concerns, aims to ensure the legitimacy of a decisional system.[60] Justificatory concerns resonate strongly with calls for rule of law. Justificatory concerns about state action might be addressed through imposing individual due process or through creating a broad system of accountability like the Administrative Procedure Act.[61] Justificatory concerns are not solely about fixing errors or bias; fixing errors or bias is a byproduct of ensuring that a decisional system is fair, valid, and legitimate.

Rule-of-law values require not just explanations of decisions but justifications that are legitimate within a particular mode of reasoning.[62] For

58. *See, e.g.*, Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1031–34 (2014); Rory Van Loo, *Helping Buyers Beware: The Need for Supervision of Big Retail*, 163 U. PA. L. REV. 1311, 1326–31 (2015). Dignitary and autonomy concerns, in fact, share values with concerns about information asymmetries, or market failure. For a classic explanation of calls for regulation in the face of market failures, see Lobel, *The Renew Deal*, *supra* note 7, at 445 (describing market failures as occurring when there are "distributional inequities, unincorporated externalities, collective action failures and free rider problems, information asymmetries, cognitive biases, . . . scale inefficiencies[,] . . . national monopolies, . . . commons[,] . . . and 'anticommons' " (footnote omitted)).

59. Kevin Granville, *Facebook and Cambridge Analytica: What You Need To Know as Fallout Widens*, N.Y. TIMES (Mar. 19, 2018), https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html (explaining that in March 2018, it came to light that tens of millions of people had their personal data misused by a data mining firm called Cambridge Analytica; a series of congressional hearings highlighted that our personal information may be vulnerable to misuse when shared on the Internet; as a result, our desire for privacy controls and transparency in data practices is heightened).

60. Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1119 ("To the extent that the personhood rationale can be converted to a more actionable legal issue, it is reflected in the concept of 'procedural justice' . . . .").

61. Citron, *supra* note 14, at 1278–79 (noting the "separate, yet parallel, procedural regimes that govern individual adjudications and rulemaking" and that "computers both render decisions about important individual rights *and* engage in rulemaking").

62. Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1288 (2017) ("A key tenant of legality, separating lawful authority from ultra vires conduct, is the idea that not all explanations qualify as justifications."). The justificatory rationale importantly goes beyond enabling people to challenge a particular decision or enabling people to change their behavior so as to obtain a different decision the next time. Sandra Wachter et al., *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J.L. & TECH. 841, 880 (2018) ("[C]ounterfactual explanations do not attempt to clarify how decisions are made internally."). It requires that enough be visible of decision-making substance and process so that its systemic legitimacy can be assessed. The justificatory rationale also differs from the

example, a police officer may explain that he has pulled you over because you drive the same car as his ex-girlfriend. That is an explanation, but not a legally or socially acceptable justification. Whether a decision is legally justified is largely given content by what rights the legal system gives to an individual to challenge a decision. For example, if legislators decide to give an individual the right to sue prospective employers for using genetic information in hiring determinations, they have effectively decided that hiring decisions cannot be justified by reference to an individual's genetic background.[63] Transparency is often necessary for justification; nobody can challenge the validity or legitimacy of a decision if they cannot see the reasoning it was based on.[64]

Algorithmic decision-making triggers a particular set of justificatory concerns. When we replace human decision makers with nonhuman decision makers, we potentially eliminate important work that a human decision maker does to both fill in and circumscribe decisional context in a particular case.[65] Human decision makers fill in context by carrying with them cultural knowledge about what is or is not an appropriate decisional heuristic in a particular case ("You are driving too fast" versus "You remind me of my ex"). Human decision makers have the capacity to expand decisional context when it seems unfair to ignore information a machine might not know is relevant ("You are speeding on the way to the hospital"). Human decision makers might also circumscribe context by knowing when it seems unfair to rely on information that strikes them as too far afield ("Because you use an iOS operating system, I am going to give you a ticket based on the likelihood

computer science concept of accountability, which is concerned only with ensuring that a system applies the same rules to all individuals within it—that is, guaranteeing that there is no individual discrimination or arbitrariness. Desai & Kroll, *supra* note 14, at 9–12; Kroll et al., *supra* note 14, at 702–05. That version of accountability neither presents justifications for analysis nor creates a procedural hook or other kinds of oversight to ensure the legitimacy of the decisional system.

63. *See generally* Genetic Information Nondiscrimination Act of 2008, Pub. L. No. 110-233, 122 Stat. 881 (prohibiting discrimination on the basis of someone's genetic information). Similarly, as we see in Section III.B.1 below, the GDPR gives individuals transparency rights that are closely tied to their ability to contest a decision or correct information.

64. Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1120–22.

65. Thanks to Michael Birnhack for this point. Kiel Brennan-Marquez and Andrew Selbst, it turns out, have debated exactly this issue of algorithmic decision-making and context. Brennan-Marquez, *supra* note 62, at 27 ("[T]he whole point of a[n] [algorithmic] tool . . . is to make predictions from correlative variables *out* of context—a process that, by its nature, frustrates inquiry into the tool's case-by-case performance . . . ."); Andrew D. Selbst, *A Mild Defense of Our New Machine Overlords*, 70 VAND. L. REV. EN BANC 87, 92 (2017) (explaining that an algorithmic decision-making "system would not be able to give a useful answer . . . unless the system has some way to connect that information to the context it needs").

that you'll speed in the future").[66] When we use algorithmic decision-making, we take these human capacities and temporally remove them from the point at which an individual decision is made. Human decisions about how to treat context may be incorporated, to some extent, into the design of an algorithm, but they are absent at the end point when an algorithm is applied to a particular individual. Algorithms, as programmed entities fed both goals and datasets by humans who are more remote from a particular decision, are often or even inherently culturally or contextually incomplete. Algorithmic decision-making thus lacks certain capacities when compared to even a lousy human decision maker.

Human contextual knowledge can clearly have much-discussed negative effects—for example, by bringing human biases and discrimination into decision-making.[67] But it also serves an important role. It means that human decision makers will, for the most part, not produce particularly random explanations—that is, acontextual explanations that we are particularly likely to view as unjustified, such as connecting loan decisions to your smartphone choice or your choice about the color of your socks.[68] It also means that human decision makers will easily weed out certain kinds of extreme error, like the decision by an algorithm to kill the pilot in the flight simulator because it realized it could obtain a perfect landing score by crashing the plane.[69] Because algorithms both fail to import context and fail to circumscribe context as human decision makers do, we should ask for at least as much, if not more, justificatory transparency and process from algorithms as from human decision makers making the same decision.

The justificatory rationale leads to more than calls for transparency. Justification requires process. Justification is not just about showing one's

66. Louise Matsakis, *Your Smartphone Choice Could Determine if You'll Get a Loan*, WIRED (May 8, 2018, 11:00 AM), https://www.wired.com/story/your-smartphone-could-decide-whether-youll-get-a-loan.

67. *See* Jane Bambauer, *Hassle*, 113 MICH. L. REV. 461, 500 (2015). For a longer discussion of cognitive biases beyond discrimination, see generally DANIEL KAHNEMAN, *supra* note 16.

68. Ed Felten, *What Does It Mean To Ask for an "Explainable" Algorithm?*, FREEDOM TO TINKER (May 31, 2017), https://freedom-to-tinker.com/2017/05/31/what-does-it-mean-to-ask-for-an-explainable-algorithm.

69. Janelle Shane, *When Algorithms Surprise Us*, AI WEIRDNESS (Apr. 13, 2018, 10:59 AM), http://aiweirdness.com/post/172894792687/when-algorithms-surprise-us. Thanks to Christina Mulligan for this pointer. Another great example—"[A]n algorithm that was supposed to sort a list of numbers. Instead, it learned to delete the list, so that it was no longer technically unsorted." *Id.*; *see also* Daniela Hernandez & Ted Greenwald, *IBM Has a Watson Dilemma*, WALL ST. J. (Aug. 11, 2018, 12:19 AM), https://www.wsj.com/articles/ibm-bet-billions-that-watson-could-improve-cancer-treatment-it-hasnt-worked-1533961147 (quoting Dr. Lukas Wartman of Washington University School of Medicine on IBM's Watson cancer system: "The discomfort that I have—and that others have had with using it—has been the sense that you never know what you're really going to get").

reasoning, but about ensuring the legitimacy of a decisional system. Affected individuals are more likely to perceive a decisional system as legitimate "when they play a meaningful role in the process."[70] Different forms of process may apply to different kinds of algorithmic decision-making, as in the law. For example, procedural due process requires individualized notice and an opportunity to be heard, but only in some contexts.[71] The U.S. legal system also contains systemic justificatory systems beyond individual due process, such as the Administrative Procedure Act, which is less discussed in the algorithmic governance literature.[72]

In fact U.S. law is full of different kinds of justificatory obligations. A warrant requirement can be understood as a justificatory requirement, requiring the police to justify why they believe they have probable cause for search or seizure.[73] The legal standard for admitting expert evidence, the *Daubert* standard, can be understood as a justificatory requirement, requiring parties to justify bringing expert information that has authority outside of the law into the legal system.[74] Even open government law, such as the Freedom of Information Act and Federal Advisory Committee Act, can be characterized as justificatory in nature, demonstrating the legitimacy of government decision-making through both transparency and process requirements. In many areas of the law, a combination of substantive and procedural requirements serves to demonstrate that a decision-making process is both based on normatively and legally acceptable justifications and is procedurally fair.

Thus subscribing to the justificatory rationale for regulating algorithmic decision-making can lead to both calls for individual due process and calls for systemic oversight and accountability.[75] When an individual is subjected

70. Freeman, *supra* note 6, at 656 ("[P]arties are more likely to view outcomes as legitimate when they play a meaningful role in the process . . . [and are] included in the enterprise, taken seriously, and offered explanations for decisions."). Procedural justice is not an empty concept; individuals are more likely to accept a decision if they believe the process was fair. Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 30 CRIME & JUST. 283, 317–18 (2003).

71. Crawford & Schultz, *supra* note 5, at 113 ("[T]he level of due process required differs according to the gravity of the deprivation and the magnitude of the countervailing state interest . . . .").

72. The exception to this is Citron, *supra* note 14, at 1279, 1288–91 (emphasizing both individual due process and systemic accountability under the Administrative Procedure Act and noting that the use of automated systems "creates confusion about the procedures owed individuals, interfering with both due process guarantees and rulemaking procedures").

73. *See generally* Brennan-Marquez, *supra* note 62 (characterizing the Fourth Amendment's probable cause requirement as a "plausible cause" or justifiability requirement).

74. Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 598–99 (1993); *see also* Citron, *supra* note 14, at 1307 (discussing *Daubert*).

75. Zarsky, *supra* note 14, at 1519–21 (discussing the need for interpretability to go beyond correlations to understandings of causality).

to a decision that has significant effects on that individual, we traditionally require individualized justifications.[76] But when we make broad systemic decisions or write rules that will govern everyone, we often look to collective or proxy systems for justification. Because algorithmic decision-making is really both kinds of decisional systems (broad policies for many people coupled with individual decisions with deep effects on individuals), both approaches to justification may apply.[77] This results in calls for individual due process and explanations ("Let me show you how and why this decision was made about *you* and let you contest it") coupled with calls for third-party accountability ("Let me assure you that neutral experts are providing oversight to make sure this decision was made fairly and for fair reasons"). As discussed at greater length below,[78] in a binary system, systemic accountability measures also serve more than one purpose: they may bolster individual rights by providing oversight in the name of protecting individuals, and provide the accountability necessary to oversee collaborative governance, as companies create and implement rules.

The dignitary and justificatory rationales often significantly overlap. A decisional system flouts rule-of-law values if individuals do not have meaningful dignity or autonomy within it.[79] It lacks legitimacy if individuals cannot meaningfully invoke procedural safeguards or assess its logical underpinnings. Similarly, tools used to provide justification and procedural legitimacy can contribute substantially to improving a decision's impact on human dignity and autonomy by enabling individual participation and allowing challenges to decisions when a person has been characterized by that person's traits instead of treated as an individual.

Of the three concerns, however, the instrumental concern about algorithmic decision-making is by far the most prevalent in recent scholarship.[80] Perhaps this is because, as with privacy harms in general, dignity can be hard to pin down.[81] Or it may be that some scholars do not

76. *See infra* Section II.A.

77. Citron, *supra* note 14, at 1253 (noting that algorithmic decision-making combines individual adjudication with rulemaking).

78. *See infra* Section II.C.

79. Hildebrandt, *supra* note 14, at 48 ("[P]rivacy is also a public good that concerns a citizen's 'freedom from unreasonable constraints on the construction of her identity.' This freedom is a precondition for democracy and rule of law . . . ." (footnote omitted)).

80. *See, e.g.*, Kroll et. al, *supra* note 14, at 636. (listing three slightly different harms: "incorrect, unjustified, or unfair" (discriminatory) outcomes from computers); Zarsky, *supra* note 14, at 1569 (describing the use of machine learning decision-making as potentially "ineffective, error-ridden, generat[ing] chilling effects, lead[ing] to unfair discrimination, and . . . prone to enable or even encourage function creep").

81. *See, e.g.*, Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1119 ("Ultimately, that there is

subscribe to either the dignitary or justificatory arguments. Those scholars who reject individualized algorithmic due process and individualized transparency largely implicitly reject both the dignitary and justificatory rationales for them.[82] They understand regulating private-sector algorithms as being largely about correcting error or discrimination and bias. They view the problems of algorithmic decision-making as consumer protection problems, best addressed not through individual rights but through a broader, more systemic regulatory approach.

The next Part addresses the "what" of governing algorithmic decision-making:[83] What kinds of regulation do we put in place, given the above three rationales for regulating? But one final preliminary discussion is necessary before moving on to the regulatory framework. Just as not every government action triggers the same level of due process,[84] not every use of algorithms in decision-making might trigger a full regulatory regime. The threshold question is what uses of algorithms should be covered.

The EU's GDPR illustrates how this threshold question of coverage might be answered. For one, it might be a matter of how much human

inherent value in explanation is clear. But as a practical matter, those concerns are difficult to administer . . . .").

82. *See, e.g.*, Zarsky, *supra* note 14, at 1552 (describing the EU idea that dignitary harms justify a freedom from machine decision-making as "an anachronistic notion" stemming from "neo-Luddite" tendencies). Elsewhere Zarsky does recognize other dignitary interests. *See, e.g.*, Zarsky, *Trouble*, *supra* note 53, at 129. Dignity also sometimes gets outright ignored. *See generally* Kroll et al., *supra* note 14 (discussing this topic without mentioning dignity at all); Wachter et al., *supra* note 62 (same).

83. *See infra* Part II.

84. Ambrosino v. Metro. Life Ins. Co., 899 F. Supp. 438, 446 (N.D. Cal. 1995) (holding that termination of membership based on a physician's previous drug addiction violated the common law right to fair procedures); St. Agnes Hosp., Inc. v. Riddick, 748 F. Supp. 319, 336–42 (D. Md. 1990); Cotran v. Rollings Hudig Hall Int'l, Inc., 948 P.2d 412, 422 (Cal. 1998) (stating that "good cause" in context of implied employment contract requires an "appropriate" investigation, which is "not arbitrary or pretextual," and includes "notice of the claimed misconduct and a chance for the employee to respond"); Pinsker v. Pac. Coast Soc'y of Orthodontists, 526 P.2d 253, 267–68 (Cal. 1974) (holding that orthodontist society must use fair process in rejecting application for membership); Silva v. Lucky Stores, Inc., 76 Cal. Rptr. 2d 382, 387 (Ct. App. 1998) ("[I]nvestigative fairness contemplates listening to both sides and providing employees a fair opportunity to present their position and to correct or contradict relevant statements prejudicial to their case, without the procedural formalities of a trial." (citing *Cotran*, 948 P.2d at 422)); Curl v. Pac. Home, 239 P.2d 481, 483–84 (Cal. Ct. App. 1952) (involving nursing home termination of residence and analogizing residence to membership in private organization); Falcone v. Middlesex Cty. Med. Soc'y, 170 A.2d 791, 799–800 (N.J. 1961) (holding that membership decisions by county medical society are subject to judicial review and may not be arbitrary, unreasonable, or contrary to public policy); Freeman, *supra* note 6, at 588–91 (describing examples of due-process-like protections applied to the private sector); *see also* Zechariah Chafee, Jr., *The Internal Affairs of Associations Not for Profit*, 43 HARV. L. REV. 993, 999–1010 (1930); F. Eric Fryar, Note, *Common-Law Due Process Rights in the Law of Contracts*, 66 TEX. L. REV. 1021, 1041–49 (1988).

involvement or oversight there is over the decision-making.[85] Only "solely" automated decisions are subject to the GDPR's requirements of explanation and contestation.[86] If a decision by an algorithm is subject to meaningful human review, meaning a human decision maker has the capacity and ability to contest that decision, then a number of the GDPR's individual rights do not apply. This suggests that in the EU, at least, human involvement in decision-making is seen as somehow mitigating the dignitary, justificatory, and even instrumental concerns about algorithmic decision-making.

Second, whether an algorithmic decision should be targeted for regulation might turn on the extent and nature of the impact the decision has, as with procedural due process. The GDPR regulates algorithmic decision-making differently from general data processing only if it produces "legal effects" or "similarly significant[]" effects.[87] The denial of a loan is a significant effect; whether being subjected to targeted advertising is a significant effect has been subject to some debate.[88]

Third, algorithmic decision-making might be regulated differently or trigger regulation at different thresholds in different policy contexts and against the backdrop of different areas of the law. For example, concerns about algorithmic decision-making are heightened in the criminal context, where state action is clear and the consequence of imprisonment is harsh and concrete.[89] By contrast, in the medical context, while the effects of a decision may be equally serious, we may be less concerned with a patient's ability to challenge or understand the rationale behind a decision (especially as patients often delegate decision-making to a trusted expert) and more concerned instrumentally with making sure the decision is unbiased, correct, and overseen and contextualized by a trusted medical professional. Existing laws, too, such as evidence and criminal procedure in the criminal context and fiduciary responsibilities and Food and Drug Administration ("FDA") regulations in the medical context, may interact with, affect, or even be

85. *See* Citron, *supra* note 14, at 1307; Meg Leta Jones, *The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles*, 18 VAND. J. ENT. & TECH. 77, 118 (2015).

86. Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 201 (2019).

87. Council Regulation 2016/679, art. 22, 2016 O.J. (L 119) 1, 46 (EU).

88. *See* Edwards & Veale, *supra* note 15, at 48; Kaminski, *supra* note 86, at 202.

89. There, too, the identity of the regulated actor may complicate the concept of collaborative governance. When the state acquires technology from private companies, as is often the case, collaborative governance might be applied to those companies, and procurement methods may be a vehicle for regulation. When the state builds its own technology, while many of the tools of collaborative governance may still work (impact assessments, expert boards, audits, etc.), the governance of a state actor is not usually considered within the purview of collaborative governance.

vehicles for introducing the type of regulatory regime discussed in Part II.[90]

It is beyond the scope of this Article to fully address threshold coverage questions or to contextualize the regulatory approach proposed below to specific legal contexts or applications. It may be that this binary approach is not the right fit everywhere. Or it may be the case that the binary approach is necessary but not sufficient, and that it will need to be calibrated along with substantive rights—for example, in antidiscrimination law. But in identifying the reasons behind calls for regulating algorithmic decision-making and tracing those reasons to regulatory design, this Article identifies common threads in the literature and aims to form the basis for future conversations about algorithmic decision-making in specific policy contexts.

II. THE BINARY APPROACH

Understanding that there are three different categories of concerns behind calls for regulating algorithmic decision-making explains why the proposed solutions thus far have involved such an array of regulatory tactics—from explanations of individual decisions, to expert oversight within companies, to specific design requirements for algorithms. Those who are concerned about dignitary harm or individualized justification of decisions call for individualized due process-like protections. Those who focus, instead, on instrumental goals or systemic justification see less of a value in individual rights because, as discussed below, individual rights are not the best way to achieve instrumental goals and broad systemic oversight might actually do more to legitimate a system than individual challenges.[91] Identifying these three categories of concerns lets us explore how to reconcile this conflict. It is possible to design a system of regulation to address all three.

To address all three categories of concerns about algorithmic decision-making (dignitary, justificatory, and instrumental), we need a regulatory system that employs a two-part, or binary, approach. An individual rights regime can address dignitary and individualized justificatory concerns but is less well suited to fix systemic problems such as bias or malfunction or provide systemic legitimacy. Instrumental goals are better addressed through

90. Similarly, discussions of the use of algorithmic decision-making for content moderation online trigger yet another set of policy concerns and a distinct legal landscape of First Amendment doctrine and related intermediary liability law. *See, e.g.*, Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1609–13, 1636–37, 1636 n.263 (2018) (especially note 263); Matthew Sag, *Internet Safe Harbors and the Transformation of Copyright Law*, 93 NOTRE DAME L. REV. 499, 511–13 (2017).

91. *See infra* Section II.A.

a systemic approach, likely using collaborative governance, to affect both the design of algorithms and the organizational design of the human systems around them. A systemic approach may also be necessary to produce a justified or legitimate decision-making system, given individual resource constraints and limited expertise. But a systemic approach alone will not be adequate, since it fails to capture significant dignitary and justificatory concerns—and in some cases, may miss out on the capacity of individual challenges to correct bias and error or provide accountability for the system as a whole. To effectively govern algorithmic decision-making, we need both.

Several existing proposals—most prominently from Danielle Citron, Frank Pasquale, Kate Crawford, and Jason Schultz—call for both approaches: individual rights combined with some form of systemic governance.[92] These authors, however, do not identify that systemic approaches to algorithmic accountability are functionally attempts at collaborative governance. This leads to some missed insights from the literature on private-public partnerships, discussed below.[93] Nor do these authors explore the interactions between individual rights and systemic governance. The binary approach explored here goes beyond porting individual due process and systemic oversight mechanisms from other areas of law, and from existing data privacy principles, to discuss overall regulatory design. This Part asks how these different regulatory features serve the goals of regulation, how they interact, where they might conflict, and how they work as a system.

A. AN INDIVIDUAL RIGHTS REGIME

An individual rights regime can address both dignitary and justificatory concerns about algorithmic decision-making. By giving individuals transparency and input into profiling or decision-making, we respect their dignity and prevent objectification. By providing individual transparency, explanation, and participation rights, we address justificatory concerns about the legitimacy of an algorithmic decisional system. Such a system would let someone like Wysocki, the fired teacher, receive notice that she was subject to algorithmic decision-making; receive an explanation of some kind of why the decision was made; and have some kind of right to challenge the decision.

92. Citron, *supra* note 14, at 1305, 1308–13 (addressing public sector use of algorithmic decision-making and calling for both individual rights and "[r]eplacing [r]ulemaking [p]rocedures" with testing, stakeholder involvement, and more); Citron & Pasquale, *supra* note 4, 1301–13; Crawford & Schultz, *supra* note 5, 125–28.

93. *See infra* Section II.B.

This idea of algorithmic due process is not new. Several scholars have argued at length for individual due process-like protections.[94] In the context of government actions, procedural due process requires that individuals deprived of a significant interest be provided both notice and an opportunity to be heard before a neutral arbiter before the deprivation occurs.[95] The algorithmic due process literature ports these requirements and others into the context of algorithmic decision-making.

The literature argues that individual due process protections should be put in place even when the algorithmic decision is made by a private, nonstate actor.[96] It largely analogizes from the regime of due-process-like rights established under the Fair Credit Reporting Act ("FCRA"), which applies to private sector "consumer reporting agencies." The FCRA requires disclosure of your credit score and the information a consumer reporting agency holds on you ("file disclosure"); creates a right to dispute incomplete or inaccurate information; and creates a right to be told if information in the file has been used against you, among other things.[97]

There are a number of examples available in U.S. law beyond the FCRA of due-process-like protections applied to decisions made by private parties with particularly significant consequences. For example, courts have put in place due process-like requirements for employment decisions, for membership decisions by professional organizations, and for decisions to terminate residence in a nursing home.[98] Congress, too, has enacted privacy laws that give individuals rights of access and correction for information held by private companies.[99] These laws and decisions suggest an intuition by U.S. courts and legislators that significant decisions made by private parties can be made subject to individual process "rights."

The algorithmic due process literature calls for various types of individualized "meaningful notice." These notice proposals include

94. Citron, *supra* note 14, at 1313; Citron & Pasquale, *supra* note 4, at 19–20; Crawford & Schultz, *supra* note 5, at 120–21.

95. Goldberg v. Kelly, 397 U.S. 254, 267–68 (1970); *see also* Houston Federation of Teachers, Local 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1172–73 (S.D. Tex. 2017); Crawford & Schultz, *supra* note 5, at 111.

96. Citron & Pasquale, *supra* note 4, at 19; Crawford & Schultz, *supra* note 5, at 121–28.

97. Citron & Pasquale, *supra* note 4, at 16–17; Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1101.

98. *See supra* note 84.

99. *See, e.g.*, 15 U.S.C. § 6502(b)(1)(B) (2018) (tasking the Federal Trade Commission with creating means for parents to access children's information gathered by websites); 47 U.S.C. § 551(d) (giving cable subscribers access to personally identifiable information collected and maintained by a cable operator and a "reasonable opportunity to correct any error in such information"); 16 C.F.R. § 312.6 (2019) (stating the Children's Online Privacy Protection Act rule).

notifying individuals that a decision is going to be made,[100] notifying individuals of the general sources of data inputs,[101] notifying individuals of privacy risks,[102] giving individuals the right to inspect data about them,[103] and establishing and disclosing audit trails that record both facts and rules supporting algorithmic decisions.[104] Many of these notice requirements are, in fact, aspects of the Fair Information Practices ("FIPs") on which many data privacy laws are based. To make notice effective and avoid overwhelming nonexperts, several scholars have suggested using technology that allows individuals to tinker with algorithmic decision-making to get a sense of how it works.[105] As discussed in Part III, there is a hearty debate ongoing in the EU over whether individuals should be afforded an explanation of an algorithmic decision and, if so, what kind of explanation they should be afforded.[106]

Algorithmic due process proposals also call for multiple kinds of opportunities to be heard. Proposals have called for individuals to have the right to correct inaccurate data.[107] This can involve allowing individuals to challenge not just erroneous personal data but erroneous inferences and scores.[108] Proposals have called for individuals to be given an opportunity to challenge a decision in front of a neutral third party, such as the Federal Trade Commission ("FTC").[109] In that process, individuals could examine evidence used against them, including the data, the algorithmic logic applied and the audit trails.[110] Or an opportunity to be heard could be less robust and less neutral, internalized within a company as the opportunity to obtain human involvement in an algorithmic decision, either during the process or ex post.[111]

100. Citron & Pasquale, *supra* note 4, at 28.

101. Citron & Pasquale, *supra* note 4, at 20; Crawford & Schultz, *supra* note 5, at 125.

102. Crawford & Schultz, *supra* note 5, at 126 (stating that notification happens in a way "reasonably calculated" to inform people of privacy risks).

103. Citron & Pasquale, *supra* note 4, at 20.

104. *See* Citron, *supra* note 14, at 1305.

105. Citron & Pasquale, *supra* note 4, at 28–30; *see also* Hildebrandt, *supra* note 14, at 53–54. For a similar suggestion in the copyright law context, see Maayan Perel & Niva Elkin-Koren, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, 69 FLA. L. REV. 181, 190–200 (2017).

106. Kaminski, *supra* note 86, at 198.

107. Crawford & Schultz, *supra* note 5, at 126–27; Kaminski, *supra* note 86, at 213 (discussing the right to correction in the GDPR).

108. Citron & Pasquale, *supra* note 4, at 28.

109. Crawford & Schultz, *supra* note 5, at 127.

110. *Id*. at 127–28.

111. Putting a human in the loop or on the loop raises the specter of automation bias—the human tendency to take as authoritative decisions made by machines. *See* Citron & Pasquale, *supra* note 4, at 6–8 (describing the levels of automation spoken about in autonomous weapons literature, including humans

Algorithmic due process would address both dignitary and justificatory concerns about algorithmic decision-making. Disclosing secret systems and allowing individuals to participate in and correct both profiling and decision-making respects human dignity.[112] Individualized procedural protections would also address justificatory concerns to the extent that they make visible the bases of decisions and create accountability through process.[113]

As has been argued in the literature, these individual rights would also have at least some instrumental value. Individual due process has been described as a form of system management, uncovering and correcting systemic errors, bias, and discrimination.[114] By allowing individuals the opportunity to correct both factual and legal errors in their own proceedings, individualized due process can lead to larger corrections in both the application and creation of rules.[115] It can decrease bias and outright discrimination both by permitting challenges to individual decisions and by exposing decisional systems to external evaluation for compliance with legal principles and societal norms.[116] In the context of expert decision-making,

in the loop, humans on the loop, and humans out of the loop). For alternative ways to classify automation levels, see NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP'T OF TRANSP., PRELIMINARY STATEMENT OF POLICY CONCERNING AUTOMATED VEHICLES 4-5 (2019), https://www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf (defining vehicle automation as having five levels: "No-Automation" (Level 0), "Function-specific Automation" (Level 1), "Combined Function Automation" (Level 2), "Limited Self-Driving Automation" (Level 3), and "Full Self-Driving Automation" (Level 4)); NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP'T OF TRANSP., AUTOMATED DRIVING SYSTEMS 4 (2017), https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf (relying on levels designated by the Society of Automotive Engineers to define automation); NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP'T OF TRANSP., FEDERAL AUTOMATED VEHICLES POLICY 9–10 (2016), https://www.transportation.gov/sites/dot.gov/files/docs/AV%20policy%20guidance%20PDF.pdf (same); see also Paul Godsmark, *The Definitive Guide to the Levels of Automation for Driverless Cars*, WONDER HOW TO, https://driverless.wonderhowto.com/news/definitive-guide-levels-automation-for-driverless-cars-0176009 (last updated Apr. 9, 2017, 9:51 PM). Humans could be trained on "automation bias" (the likelihood that a human will unquestioningly accept an automated decision) and be required to "explain, in detail, their reliance on the automated system's decision." Citron, *supra* note 14, at 1306–07.

112. Citron & Pasquale, *supra* note 4, at 27; Hildebrandt, *supra* note 14, at 47–48; Zarsky, *supra* note 14, at 1548–49.

113. *See* Mireille Hildebrandt, *Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning*, 20 THEORETICAL INQUIRIES L. 83, 119–20 (2019).

114. Crawford & Schultz, *supra* note 5, at 121.

115. Citron, *supra* note 14, at 1280, 1284–86; *see also* Crawford & Schultz, *supra* note 5, at 119 (listing the following due process values: "(1) accuracy; (2) the appearance of fairness; (3) equality of inputs into the process; (4) predictability, transparency, and rationality; (5) participation; (6) revelation; and (7) privacy-dignity" (footnotes omitted)).

116. Crawford & Schultz, *supra* note 5, at 120 (observing that due process is also a check and balance—"a core function of due process is to separate those who write the legal code from adjudicators who use it. . . . [In big data,] there is no system of checks and balances to ensure that biases are not present in the system, which is especially crucial to a system of enforcement" (footnote omitted)).

putting a human expert in the loop as a proxy or fiduciary for the affected individual could prevent error by ensuring that individualized context is included in the decision.

These are all good reasons for giving individuals transparency and process rights over machine decisions. The problem, however, is that individual rights are not a particularly good way to correct a complex, opaque, and evolving system.[117]

B. COLLABORATIVE GOVERNANCE

While individual rights can address important dignitary and justificatory concerns, there are better ways to identify and fix systemic problems in algorithmic decision-making. A growing body of literature calls for moving away from ex post, individualized transparency and due process or, at least, supplementing them with regulations that target algorithmic design at earlier stages and target the human systems around algorithms.[118] This second camp of algorithmic accountability proposals ranges from

117. This is why the algorithmic due process literature also calls for systemic accountability on top of individual due process. In addition to individual notice and an opportunity to be heard, the due process literature contains proposals for publicly releasing the source code, releasing information about the system's logics, releasing information about the datasets (but not the datasets themselves), oversight by a board of experts, third-party and government audits, and government agency oversight. Citron, *supra* note 14, at 1308–10, 12; Citron & Pasquale, *supra* note 4, at 20–28, 31.

118. One set of scholars calls for moving away from individual transparency and process rights. *See* Desai & Kroll, *supra* note 14, at 10; Edwards & Veale, *Slave to the Algorithm*, *supra* note 15, at 23; Kroll et al., *supra* note 14, at 636–42. Tal Zarsky weighs the pros and cons of individual transparency against a host of other interests. Zarsky, *supra* note 14, at 657–60. Zarsky and Jane Bambauer more recently discuss the problem of gaming raised by individual rights. Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1, 11–12 (2018). By contrast both Mike Ananny and Kate Crawford and Andrew Selbst and Solon Barocas can be understood as saying that individual transparency is sometimes necessary but not sufficient. Ananny & Crawford, *supra* note 10, at 982; Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1088. And as discussed in *supra* note 117, the algorithmic due process literature also emphasizes systemic accountability.

If there is a consensus between the divergent branches of the literature, it is that most call for something like systemic "accountability by design." Kroll and others call for implementing this earlier in the design process, with specific technical recommendations. Kroll et al., supra note 14. David Lehr and Paul Ohm similarly call for more of a focus on the training phase of machine learning. Lehr & Ohm, *supra* note 3, at 716–17. Ananny and Crawford suggest a greater focus on the human-machine assemblage rather than the machine itself. Ananny & Crawford, *supra* note 10, at 983–84. And Selbst and Barocas similarly emphasize recording requirements, which will capture more of what happens at earlier phases. Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1129–38. It is in fact a bit disingenuous of later literature to characterize algorithmic due process only as a set of individual transparency and process rights. Even the earlier-stage algorithmic due process literature called for instituting systemic accountability measures beyond individual due process. Citron, *supra* note 14, at 1308–13; Citron & Pasquale, *supra* note 4, at 20–27; Crawford & Schultz, *supra* note 5; *see also* FRANK PASQUALE, THE BLACK BOX SOCIETY 140–88 (2015) (calling for a system of layered "qualified transparency").

suggesting specific technical requirements for building algorithms[119] to calling for a patchwork of systemic accountability mechanisms, such as whistle-blower protections[120] or impact assessments and other forms of transparency.[121]

To some extent, this second camp is right about individual due process rights. Individual rights are a limited tool for finding and fixing system-wide problems such as error, bias, and discrimination in algorithms. In part this is because of the limited capacity of rights-bearing individuals, whether technical, behavioral, legal, or economic.[122] In part this is because uncovering systemic problems such as discrimination will be easier to do with an eye to the system's overall behavior, rather than by evaluating one-off decisions.[123] And in part this is because of timing. The ability to challenge individual decisions or determinations after-the-fact or even at regular intervals will not be as effective at fixing a machine learning algorithm as designing a good system from the onset[124] or monitoring for compliance on an ongoing basis.[125] This is because of the nature of machine learning algorithms: much of the work happens at the design and training stages, and it can be very hard or even impossible to assess and correct a running model.[126] Some algorithms, too, evolve over time. Trying to regulate the quality of these systems through individual challenges will constitute an elaborate game of whack-a-mole.[127] Thus, more recent literature on algorithmic accountability has either decried individual process and transparency as inadequate by itself[128] or rejected individual rights and called

119. Kroll et al., *supra* note 14, at 662–72.

120. Desai & Kroll, *supra* note 14, at 42–64; Katyal, *supra* note 8, at 99–141.

121. A. Michael Froomkin, *Regulating Mass Surveillance as Privacy Pollution: Learning from Environmental Impact Statements*, 2015 U. ILL. L. REV. 1713, 1757–58; Price, *supra* note 8, at 466 (describing a mandatory disclosure regime); Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 169–72 (2017); David Wright & Charles D. Raab, *Constructing a Surveillance Impact Assessment*, 28 COMPUTER L. & SECURITY REV. 613, 616–22 (2012).

122. Edwards & Veale, *supra* note 15, at 74–75.

123. Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1111–13.

124. Desai & Kroll, *supra* note 14, at 41–42 (discussing the limits of ex post testing); Kroll et al, *supra* note 14, at 656–78.

125. Ananny & Crawford, *supra* note 10, at 982 (discussing ongoing disclosure and noting that "[a]ny notion of transparency or auditing without temporal dimensions misses seeing previous iterations, understanding how they worked, why they changed, and how their interacting components actually constituted different systems").

126. Kroll et al., *supra* note 14, at 640; Lehr & Ohm, *supra* note 3, at 657.

127. Desai & Kroll, *supra* note 14, at 41 ("[D]ynamic systems . . . , especially highly dynamic systems, that are not designed for evaluation pose perhaps a larger problem . . . .").

128. Ananny & Crawford, *supra* note 10, at 984 ("[T]ransparency alone cannot create accountable systems.").

for different kinds of accountability mechanisms altogether.[129] More recent literature, too, has pushed back against public transparency, arguing instead for court and expert oversight. [130]

The literature has arrived at an important inflection point: it shows that individual due process is not the best way to address instrumental concerns about algorithmic decision-making. But the discussion is also incomplete. It fails to capture the bigger picture of what exactly the proposed array of accountability tools is meant to accomplish. I argue that what these proposals are driving at is collaborative governance of algorithmic decision-making.[131]

1. Collaborative Governance

Collaborative governance, or "new governance,"[132] deploys private-public partnerships towards public governance goals. Collaborative governance should not be confused with self-regulation, though it may include or even rely in substantial part on private governance. In its ideal form, collaborative governance is not hands-off or deregulatory.[133] It exists on a spectrum between traditional command-and-control regulation and private ordering, and may employ significant aspects of each.[134]

On one end of this spectrum are top-down, government-dominated

129. Desai & Kroll, *supra* note 14, at 56–64 (calling for whistleblower protection and a public interest cause of action); Edwards & Veale, *Slave to the Algorithm*, *supra* note 15, at 74–80 (describing the aspects of the GDPR that go beyond individual rights); Kroll et al, *supra* note 14, at 639-640.

130. Kroll et al, *supra* note 14, at 657–60.

131. Only a handful of scholars, as mentioned, have explicitly called for the collaborative governance of algorithms. *See* Guihot et al., *supra* note 8, at 456; Perel & Elkin-Koren, *supra* note 8, at 530; Price, *supra* note 8, at 421.

132. *See* Lobel, *The Renew Deal*, *supra* note 7, at 346–47. As Orly Lobel has explained,

> [T]he governance model emerges from a myriad of recent scholarly theories including . . . reflexive law, soft law, collaborative governance, democratic experimentalism, responsive regulation, outsourcing regulation, reconstitutive law, post-regulatory law, revitalizing regulation, regulatory pluralism, decentering regulation, meta-regulation, contractarian law, communicative governance, negotiated governance, destabilization rights, cooperative implementation, interactive compliance, public laboratories, deepened democracy and empowered participatory governance, pragmatic lawyering, nonrival partnership, and a daring legal system.

Id. (footnotes omitted) (internal quotation marks omitted).

133. Lobel, *New Governance*, *supra* note 7, at 69 ("These approaches . . . do not, however, entail a complete shift from command-and-control regulation to self-regulation. . . . [T]he central challenge . . . is to maintain an effective role for law and regulation amidst the shifts to more private efforts of governance.").

134. Bamberger, *supra* note 9, at 396 ("[T]he delegation of decisionmaking discretion to private firms does not change their nature as parties who must follow the law. . . . [I]mposing civil or criminal penalties provides an important means of ex post punishment . . . ."); *see also* Emily S. Bremer, *Private Complements to Public Governance*, 81 MO. L. REV. 1115, 1116–22 (2016) (illustrating a spectrum of public-private approaches).

approaches, in which government institutions dictate and enforce the law's substance. On the other end is private ordering, in which private sector actors voluntarily undertake self-regulation in the absence of government action.[135] Collaborative governance represents a hybrid or, in the case of "responsive regulation," an escalating approach.[136] It may include but is not limited to formal coregulation through the adoption of codes of conduct or certification mechanisms.[137] Collaborative governance is, at best, a highly tailored, site-calibrated regulatory system that aims to pull inputs from, obtain buy-in from, and affect the internal institutional structures and decision-making heuristics of the private sector, while maintaining the legitimacy, efficacy, and public-interest orientation of public sector governance.

There are a number of much-touted benefits to collaborative governance compared to command-and-control regulation on the one hand and self-regulation on the other.[138] Proponents of purely private governance (self-regulation) point out that some regulatory problems are ill-suited to a command-and-control approach.[139] This may be because the private sector has technical expertise the government cannot obtain[140] or because the technology at issue is evolving at a rate top-down regulation cannot keep up

135. Bremer, *supra* note 134, at 1116.

136. AYRES & BRAITHWAITE, *supra* note 12, at 4–7; William McGeveran, *Friending the Privacy Regulators*, 58 ARIZ. L. REV. 959, 983–85 (2016).

137. Formal coregulation—that is, the use of industry codes of conduct to set substantive standards in the law and the sharing of government responsibility with industry for enforcement—can be understood as a version of collaborative governance or a set of tools in the collaborative toolkit, although some authors use the terms "coregulation" and "collaborative governance" interchangeably. CHRISTOPHER T. MARSDEN, INTERNET CO-REGULATION 59–63 (2011) (contrasting a wide range of forms of collaborative governance, which Christopher Marsden calls coregulation, from self-regulation, to "Potemkin" regulators that have little to no power, to collaborative governance with significant teeth); Dennis D. Hirsch, *The Law and Policy of Online Privacy: Regulation Self-Regulation, or Co-Regulation?*, 34 SEATTLE U. L. REV. 439, 465–66 (2011). Europe and countries like Australia have been experimenting with versions of both formal coregulation and a spectrum of collaborative governance techniques for years, especially in the Internet sector. Hirsch, *supra* note 137, at 442–43; *see also* IAN BROWN & CHRISTOPHER T. MARSDEN, REGULATING CODE 3–4 (2013). There has been comparatively little discussion of collaborative governance as applied to technology law in the U.S. landscape. Ian Brown & Chris Marsden, Regulating Code: Towards Prosumer Law? 4 (Feb. 25, 2013) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2224263 ("Adoption of [coregulation] in the United States has been slow . . . ."). *But see* Philip J. Weiser, *The Future of Internet Regulation*, 43 U.C. DAVIS L. REV. 529, 552–84 (2009).

138. Bremer, *supra* note 134, at 1122–24 (suggesting looking to "the concept of comparative institutional advantage" to decide when to turn to private governance); Hirsch, *supra* note 137, at 439–44 (discussing the arguments for and against coregulation).

139. Bremer, *supra* note 134, at 1123–24; Guihot et al., *supra* note 8, at 435–36; Hirsch, *supra* note 137, at 455, 458–59.

140. Bremer, *supra* note 134, at 1123; Guihot et al., *supra* note 8, at 465.

with (the so-called "pacing problem").[141] Even short of the pacing problem, private institutions may be more nimble and efficient,[142] self-regulation may better approximate a market-driven optimum,[143] and involving the private sector in its own regulation may lead to greater buy-in and adherence to voluntary rules over time.[144]

The central problem with self-regulation is, of course, capture. Inherent in the notion of self-regulation is that the regulated set the substance of their own regulation and participate in enforcement and compliance, often without input by representatives of the public interest or competing voices from the private sector.[145] Even where the substance of self-regulation is publicly oriented, self-regulating companies can lack mechanisms to enforce compliance, and the risk of free riding means every actor has a strong incentive not to comply.[146] Private governance, too, is ill-suited for areas with a high degree of divergence between interests, which may prevent substantive agreement, thus thwarting the purported benefits of voluntary buy-in and nimbleness or efficiency.[147]

Collaborative governance is a middle ground, a third way, that aims to harness the benefits of self-regulation without its pitfalls.[148] The government stays significantly involved as a backdrop threat to nudge private sector involvement, as a forum for convening and empowering conflicting voices, as an arbiter or certifier in the name of the public interest, and as a hammer that can come down to enforce compliance. Often in collaborative governance, the government does some version of all of the above.[149] The toolkit of collaborative governance ranges from formal delegation to the

141. Guihot et al., *supra* note 8, at 421; Meg Leta Jones, *Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw*, 2018 J.L., TECH. & POL'Y. 249, 277–81; Scherer, *supra* note 14, at 367–73; *see also* Kenneth A. Bamberger & Deirdre K. Mulligan, *New Governance, Chief Privacy Officers, and the Corporate Management of Information Privacy in the United States: An Initial Inquiry*, 33 L. & POL'Y 477, 483 (2011) ("As the pace of technological and market change accelerated, both rule-based and purely self-regulatory approaches have become increasingly less relevant to the protection of privacy."); Michael Birnhack, *Reverse Engineering Informational Privacy Law*, 15 YALE J.L. & TECH. 24, 38–39 (2012); Freeman, *supra* note 7, at 28 (describing collaborative governance as "a flexible, adaptive system capable of responding to advances in science, technology, knowledge . . . [when] rules are shaped by, and responsive to, the particular contexts in which they are deployed").

142. Bremer, *supra* note 134, at 1123.

143. Hirsch, *supra* note 137, at 455–57.

144. Bremer, *supra* note 134, at 1123.

145. Hirsch, *supra* note 137, at 458–59.

146. *Id.*

147. Bremer, *supra* note 134, at 1123–24.

148. Hirsch, *supra* note 137, at 465.

149. *See* Freeman, *supra* note 7, at 559–60, 671–73; Lobel, *The Renew Deal*, *supra* note 7, at 371–404.

private sector on one end—for example, in the form of government-certified codes of conduct—to informal convening, nudging, and "responsive regulation" on the other. To some degree, many systems that might at first be characterized as hard-law regulation are, in fact, systems of collaborative governance.

By involving the private sector and other third parties, collaborative governance can purportedly (1) increase the amount of private sector expertise in governance, (2) contribute to the perceived legitimacy of governance, thus contributing to increased compliance, (3) harness nongovernmental mechanisms towards compliance and enforcement, thus increasing the state's enforcement capacity, and (4) solve pacing problems by shifting from onetime, specific rules to an ongoing, iterative system of monitoring and compliance.[150] Collaborative governance can thus, in theory at least, address many of the regulatory challenges posed by algorithmic decision-making.[151] It has been touted as well suited to governing complex, changing, and risky systems, including those in the field of privacy law.[152] Rather than attempting to dictate in specific, detailed legal rules precisely how to prevent the problems of algorithmic error, bias, and discrimination, the state could structure, against a backdrop of significant state enforcement resources, a system of delegation to and oversight over the private sector in coming up with contextually appropriate solutions and maintaining ongoing compliance with the regulatory regime.[153] These solutions could address problems that arise at an early design stage and could create ongoing oversight and refinement.

150. *See, e.g.*, Hirsch, *supra* note 137, at 467.

151. Guihot and others make a similar argument with respect to responsive risk regulation, and W. Nicholson Price II has argued this in the context of black box medicine. *See supra* note 8.

152. *See* KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, PRIVACY ON THE GROUND 12–13 (2015); COLIN J. BENNETT & CHARLES D. RAAB, THE GOVERNANCE OF PRIVACY 121–22 (Ashgate Publ'g Ltd. 2003); Bamberger & Mulligan, *supra* note 141, at 480; Dennis D. Hirsch, *Going Dutch? Collaborative Dutch Privacy Regulation and the Lessons It Holds for U.S. Privacy Law*, 2013 MICH. ST. L. REV. 83, 98–99, 157; Dennis D. Hirsch, *supra* note 137, at 464–67; Margot E. Kaminski, *When the Default Is No Penalty: Negotiating Privacy at the NTIA*, 94 DENVER L. REV. 925, 927 n.11, 947–48 (2016); Ira S. Rubinstein, *Privacy and Regulatory Innovation: Moving Beyond Voluntary Codes*, 6 I/S: J.L. POL'Y INFO. SOC'Y 355, 413–14 (2011); David Thaw, *Enlightened Regulatory Capture*, 89 WASH. L. REV. 329, 343 n.69, 367–70 (2014). *See generally* COLIN J. BENNETT, REGULATING PRIVACY: DATA PROTECTION AND PUBLIC POLICY IN EUROPE AND THE UNITED STATES (1992) (comparing computer privacy laws in the United States, Britain, West Germany, and Sweden); Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247 (2011) (discussing the convergence of corporate privacy practices); Lauren E. Willis, *Performance-Based Consumer Law*, 82 U. CHI. L. REV. 1309 (2015) (suggesting a performance-based consumer law approach).

153. Michael P. Vandenbergh, *The Private Life of Public Law*, 105 COLUM. L. REV. 2029, 2038 (2005) (describing three stages: "the setting, implementation, and enforcement (including monitoring) of standards").

However, collaborative governance is subject to many of the same capture concerns as self-regulation. Accountability thus is the key to a successful collaborative governance regime. I argue that characterizing the systemic governance prong of algorithmic accountability as collaborative governance more fully captures the kinds of accountability necessary for addressing instrumental concerns about algorithmic decision-making. It also highlights that while most scholars have been focusing on a first-order accountability problem (ensuring that an algorithm is being overseen by impartial but expert outsiders), they have failed to identify the second-order accountability problem (ensuring that these delegated processes of governance remain accountable to society at large).[154]

Accountability in a collaborative governance regime looks different and serves different goals from the individualized transparency of due process. Rather than protecting individual dignity or providing individualized justification, accountability in a collaborative governance regime attempts to ensure that private sector involvement in public governance both substantively serves public goals and is procedurally legitimate.[155]

In other words, the problem of algorithmic accountability is not just about letting a team of external computer scientists see the source code and data set and play with the algorithm. It is also about, for example, putting in place independence requirements that ensure those engineers do not get later hired by that company, putting in place processes for those engineers to report problems to a regulator or to the public, and putting in place opportunities for stakeholders to have oversight over, or at least input into, decisions made by these engineers. The first kind of accountability (letting engineers see the source code) is concerned about oversight over how the algorithm functions. The second (requiring independence or putting in place ways for those engineers to trigger public reactions or regulatory enforcement) is concerned with the legitimacy of the collaborative governance system—of delegating substantive decisions about policy to private actors—as a whole.

154. Kaminski, *supra* note 12, (manuscript at 21–23) (describing "first order" and "second order" accountability issues).

155. This resonates with Danielle Citron's important point that algorithmic governance is not just about individual due process but also about creating overall accountability towards setting the rules of an algorithmic system. *See* Citron, *supra* note 14, at 1302–03.

2. The Collaborative Governance Toolkit

As discussed, collaborative governance is not the same thing as self-regulation or deregulation.[156] Collaborative governance shifts from commanding private actors to structuring both collaboration with and delegation to them.[157] I discuss a number of the available tools here.

First there is, crucially, still room for command-and-control government in collaborative governance.[158] State-set prohibitions and parameters can ensure that both lawmaking and compliance do not move too far from the public good.[159] Additionally, the state can and should use traditional enforcement when necessary to incentivize private participation in both rule setting and compliance against the background threat of a regulatory hammer (what I and others have referred to as a "penalty default," drawing on the literature on private ordering).[160] Thus, for example, the state might set a broad rule that algorithmic decision-making should not discriminate and issue significant fines when a particular algorithmic decision-making system produces discriminatory results.

Coupled with a traditional command-and-control approach, a collaborative governance regime mixes soft and hard law along multiple axes.[161] The goal is to retain both enough give in the legal system that private actors will be incentivized to participate and enough strength in the system that they are motivated to do so towards the public good. Collaborative governance tools range from the more to the less formal and deploy more or less government involvement. More formal "coregulatory" methods include negotiated rulemaking,[162] legal safe harbors to encourage the adoption of

156. *See supra* Introduction.

157. Bamberger, *supra* note 9, at 386 ("Regulators no longer command, they delegate."); Lobel, *The Renew Deal*, *supra* note 7, at 377 ("[T]he role of government changes from regulator and controller to facilitator, and law becomes a shared problem-solving process rather than an ordering activity.").

158. *See* Lobel, *The Renew Deal*, *supra* note 7, at 452.

159. *See id.* at 371.

160. Kaminski, *supra* note 152, at 943–45; Price, *supra* note 8, at 466. This idea appears throughout the literature but is not explicitly identified as a penalty default. David Dana, *The New "Contractarian" Paradigm in Environmental Regulation*, 2000 U. ILL. L. REV. 35, 47; Freeman, *supra* note 6, at 666 ("The background threat of regulation by an agency can provide the necessary motivation for effective and credible self-regulation."); *see also* Lobel, *The Renew Deal*, *supra* note 7, at 452 (quoting Dana, *supra*) ("[C]ommand-and-control regulation is a precondition for contractarian regulation" as "actors that recognize the possibility of regulation . . . have an incentive to voluntarily reach a cooperative agreement.") (referring to David Dana's "contractarian regulation").

161. *See* Lobel, *The Renew Deal*, *supra* note 7, at 388–95.

162. *Cf.* Philip J. Harter, *Negotiating Regulations: A Cure for Malaise*, 71 GEO. L.J. 1, 34–35, 42–51 (1982) (stating that coregulatory methods are less formal than rulemaking).

industry codes of conduct,[163] and the incorporation by reference of technical standards.[164] In these formal methods of collaborative governance, it is clear that the government's goal is structured delegation to or collaboration with the private sector.

Less formal methods of collaborative governance include audited self-regulation,[165] informally "delegating" the interpretation of broader standards to private actors,[166] applying performance standards instead of specific rules,[167] and encouraging private ordering in response to unappealing defaults crafted in the law.[168] These less formal methods may escape identification as collaborative governance tactics or, indeed, may not be deliberately collaborative on the part of the government at all. But they perform collaborative functions by structuring the involvement of the private sector in determining the substance or implementation of regulation and in overseeing compliance with it.

There is no one-size-fits-all version of collaborative governance. A particular regime will need to be tailored to a particular sector.[169] The state ideally calibrates a particular collaborative governance system to the features of a particular industry, evaluating whether that industry contains repeat players, whether its actors are motivated by professional reputation, how firms are internally organized, including whether there is an established compliance culture, and what kinds of network links exist between actors. Extrinsic forms of motivation and accountability can lead to the creation of very different kinds of collaborative legal regimes. So can other features of the regulatory environment, including whether civil society players are well established and well resourced and how great the technical barriers are to

163. *See* Rubinstein, *supra* note 152, at 356–60.

164. Emily S. Bremer, *Incorporation by Reference in an Open-Government Age*, 36 HARV. J.L. & PUB. POL'Y 131, 133–39 (2013).

165. Bremer, *supra* note 134, at 1118–19.

166. Bamberger, *supra* note 9, at 386 ("Regulators no longer command, they delegate."); Lobel, *New Governance*, *supra* note 7, at 71 ("[R]egulations are often deliberately ambiguous. Instead of regulating the details of behavior, agencies increasingly use broad policy goals such as 'risk management' and allow the regulated industries to implement and interpret these mandates.").

167. *See, e.g.*, Willis, *supra* note 152, at 1330. *But see* Bamberger, *supra* note 9, at 389 ("Certain public problems . . . lend themselves to neither specific behavioral commands nor measurable outcomes.").

168. *See* Bremer, *supra* note 134, at 1119 (calling these "second-order regulatory agreements . . . 'agreements entered into between regulated firms and other private actors in the shadow of public regulations'" (quoting Vandenbergh, *supra* note 153, at 2030)); *see also* Kristelia A. García, *Penalty Default Licenses: A Case for Uncertainty*, 89 N.Y.U. L. REV. 1117, 1122 (2014) (identifying "penalty default licenses" and penalty defaults in general as mechanisms for inducing private ordering).

169. Jody Freeman refers to this approach as "microinstitutional[ism]." Freeman, *supra* note 6, at 674.

participation by third parties.

i. The Problem of Accountability

Relying on the private sector and private third parties for rule-setting, implementation, and enforcement, however, creates a significant problem: a lack of accountability.[170] Critics observe that such a regime can easily become subject to collusion or capture.[171] Collaborative governance in its worst form becomes a way for powerful companies to cement their power and evade regulation and oversight.[172] Governance mechanisms can become an "enablement paradigm" instead of an "empowerment paradigm," with purported accountability tactics providing a fig leaf that legitimizes bad behavior and does not mitigate it.[173]

Accountability is thus the central problem of collaborative governance.[174] Accountability can mean many things.[175] Accountability is not synonymous with transparency. Transparency involves structuring information flows—to the public, to experts, and to affected third parties. But an accountable collaborative governance regime is not necessarily a publicly transparent regime; nor is public transparency alone sufficient for establishing accountability. The collaborative governance literature is filled with calls for public transparency.[176] But it also contains skepticism,

170. Bamberger, *supra* note 9, at 384 ("[D]ecisions assigned to regulated firms should also be viewed through an accountability lens.").

171. *See, e.g.*, AYRES & BRAITHWAITE, *supra* note 12, at 55–56.

172. Lobel, *The Renew Deal*, *supra* note 7, at 458–59 ("A central challenge for the governance model is therefore to understand how collaborative environments can be nurtured to produce equitable results, especially in settings where vast power imbalances exist.").

173. Bamberger, *supra* note 9, at 429 ("[T]he establishment of auditable controls often provides firms with ways to signal legitimacy without addressing deeper problems inherent in existing routines and structures."); Lobel, *The Renew Deal*, *supra* note 7, at 385 (fearing that collaborative techniques will be "used by management merely as mechanisms for monitoring, controlling, and exerting additional pressures on workers").

174. Martha Minow, *Public and Private Partnerships: Accounting for the New Religion*, 116 HARV. L. REV. 1229, 1236–37 (2003) (identifying "accountability as the central issue requiring inventive work" in public-private roles).

175. Bamberger, *supra* note 9, at 404; Desai & Kroll, *supra* note 14, at 6–23 (providing an overview of accountability problems in both the public and private sectors); Minow, *supra* note 174, at 1260 ("Accountability . . . means being answerable to authority that can mandate desirable conduct and sanction conduct that breaches identified obligations.").

176. Bamberger, *supra* note 9, at 407 ("[A] certain level of decisional transparency [is] essential to permitting meaningful review."); Bremer, *supra* note 134, at 1124 ("[A] private governance system should be transparent, so that both its participants and products are knowable to a public that may be affected by that system."); Freeman, *supra* note 6, at 635 ("As in other contexts, mandatory disclosure could serve as one among other accountability mechanisms."); Lobel, *The Renew Deal*, *supra* note 7, at 454 ("[G]overnance embraces the essential significance of transparency and information disclosure."); Minow, *supra* note 174, at 1263 ("Democratic governments promise accountability through transparency,

including acknowledgments that accountability necessarily involves both enforcement mechanisms and expert analysis, in addition to public knowledge and engagement.[177] This observation resonates with the literature on algorithmic accountability, which more recently downplays public transparency in favor of expert oversight.[178] But it also pushes back against more recent attempts to establish regulatory or expert oversight without any public or third-party transparency.

Accountability in the context of collaborative governance has been characterized as "checks on decision making" targeted at producing legitimacy—that is, the public acceptability of a system.[179] Accountability involves both substantive and procedural goals. An accountable regime produces substantively good regulation, in the name of the public good rather than private interests.[180] At the same time, an accountable collaborative governance regime maintains procedural norms about fair decision-making to produce public acceptance of private involvement in governance.[181] This can be understood as a concern about justification, in that it addresses the perceived and real legitimacy of a decision-making system—not just the algorithm, but the public-private partnerships around it.

In practice, accountability can take a variety of forms. Accountability can take the form of participation and process requirements at the rulemaking stage—for example, by requiring that an industry consult with civil society

a trendy term for public disclosure of key decisions and the information necessary to assess those decisions.").

177. Lobel, *The Renew Deal*, *supra* note 7, at 455–57 ("The Renew Deal vision must resist the illusion of information and transparency—that the information age, through its own mechanisms, can solve all problems. . . . At its best, the governance model should aim to combine expertise and experience—involving representatives in many avenues while recognizing the importance of direct engagement.").

178. PASQUALE, *supra* note 118, at 153–56; Ananny & Crawford, *supra* note 10, at 984 ("If transparency requires professional expertise to be understood and acted upon, then models of accountability might examine how the system develops and deploys different types of authority and specialized ways of seeing.").

179. Freeman, *supra* note 6, at 664–66.

180. Bamberger, *supra* note 9, at 403 (defining accountability in terms of regulatory goals as promoting the following: "*rationality* in choosing between solutions; *responsiveness* to public interests; and *reviewability* by others").

181. *Id*. at 404 ("[U]understand[ing] accountability . . . as 'checks on decision making' intended to channel discretion so as to promote both effective and legitimate regulatory decisions." (quoting Freeman, *supra* note 6, at 664)). Bremer's calls for accountability echo this attention to both substance and procedural legitimacy. Emily Bremer, *supra* note 134, at 1124 (stating that transparency is necessary so that "both [a private governance system's] participants and products are knowable to a public that may be affected by that system" and that "[r]ules to ensure openness and participation by a balanced range of affected interests may similarly be necessary to preserve legitimacy and protect the substantive validity of a private governance regime").

or external experts in coming up with policies or codes of conduct.[182] For example, the company Axon (formerly TASER), the world's largest producer of body-worn police cameras, established an AI Ethics Board featuring academics and private sector actors, in the shadow of efforts to regulate algorithmic decision-making in New York City.[183] Absent input by affected stakeholders and public transparency, the efficacy and legitimacy of this board has been much questioned, as have other attempts at AI ethics boards at other companies.[184]

A central aim of collaborative governance is to produce a more effective compliance culture.[185] Thus accountability can also take the form of structuring industry self-assessment,[186] by requiring the appointment of independent compliance officers[187] and requiring ongoing internal reports.[188] Accountability can also involve independent oversight over a company's behavior, such as third-party auditors or civil society assessments of whether a company is actually behaving as it said it would.[189]

Collaborative governance emphasizes systemic accountability, or aggregate accountability, which looks at the interplay between different accountability mechanisms, over time.[190] It does not rely, as the algorithmic

182. For the process used by the NTIA in coming up with codes of conduct, see generally Harter, *supra* note 162; Kaminski, *supra* note 152.

183. *Axon Ethics Board*, POLICING PROJECT, https://www.policingproject.org/axon-ethics-board (last visited Aug. 4, 2019); Julia Powles, *New York City's Bold, Flawed Attempt To Make Algorithms Accountable*, NEW YORKER: ANNALS OF TECHNOLOGY (Dec. 20, 2017), https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable.

184. Tom Simonite, *Tech Firms Move To Put Ethical Guard Rails Around AI*, WIRED (May 16, 2018, 4:32 PM), https://www.wired.com/story/tech-firms-move-to-put-ethical-guard-rails-around-ai (noting an open letter by more than forty academic, civil rights, and community groups criticizing Axon for "omit[ing] representatives from the heavily policed communities most likely to suffer the downsides of new police technology") .

185. Bamberger, *supra* note 9, at 446 ("[I]nternal firm decisionmaking is rendered more accountable by outside influence on decision processes and structural design."); Freeman, *supra* note 6, at 644–45 (stating that self-monitoring is used to "inculcate management reforms").

186. Freeman, *supra* note 6, at 663 (observing that management reform can occur through (1) producing commitments and impact assessments before a program is implemented and (2) periodically self-assessing compliance with those commitments and assessments over time); *see also* Bamberger, *supra* note 9, at 448 (describing section 404 of the Sarbanes-Oxley Act as producing an annual set of internal commitments with legal force).

187. AYRES & BRAITHWAITE, *supra* note 12, at 106; Lobel, *The Renew Deal*, *supra* note 7, at 463–64 (providing the example of professional safety engineers in a firm); *see* McGeveran, *supra* note 136, at 988.

188. Bamberger, *supra* note 9, at 451–52 (requiring yearly benchmarks, an explanation of alternative approaches that were rejected, and specific factors to be considered in risk assessment).

189. AYRES & BRAITHWAITE, *supra* note 12, at 102; Freeman, *supra* note 6, at 551–52; Lobel, *New Governance*, *supra* note 7, at 16.

190. Freeman, *supra* note 6.

accountability literature largely does, on just one or two individual accountability mechanisms. Enacting whistleblower protections by itself is not enough. Requiring an algorithmic impact assessment by itself is not enough. To achieve aggregate accountability in a collaborative governance regime, regulators must put mechanisms in place to ensure balanced and expert input and oversight over all three stages of governance: rule-setting, implementation, and enforcement. And the designer of the regulatory system must be particularly aware of how these mechanisms interact as a whole. This requires designing multiple accountability mechanisms so that they feed back into each other at different stages of regulation.[191]

For example, self-assessment or third-party auditing to ensure that an algorithmic decision-making system is not biased might be released to a regulator that in turn summarizes such results and publishes them to the public.[192] The public could then respond by avoiding a particular industry or company (by deploying both shaming and a market mechanism as soft enforcement) or by putting pressure on the regulator to enforce the law. Or the public could respond to composite reports of particularly egregious behavior by a sector of industry by voting to create new laws that will then be newly implemented by a company, which in turn would be assessed by third-party auditors. Thus the governance system iterates. Merely putting in place a stand-alone auditing requirement would not achieve this regulatory iteration, nor achieve accountability over the use of private-public partnerships to govern.

Recharacterizing the conversation about algorithmic accountability as a conversation about collaborative governance thus lets us draw on work that the literature has already done. Rather than debating the value of individualized transparency or of public disclosure, we can turn efforts towards structuring an effective and accountable governance regime qua regime. And rather than deploying accountability mechanisms one at a time, we can more seriously begin to address how they interact in a system of governance as a whole.

191. Lobel, *The Renew Deal*, *supra* note 7, at 425 n.351.

192. Freeman, *supra* note 6, at 644 (explaining that "[a] typical self-regulatory initiative" includes the publication of "reports (sometimes to regulators only, sometimes both to regulators and to the general public)."); Lobel, *The Renew Deal*, *supra* note 7, at 426 (describing a disclosure mechanism whereby firms report to an agency, "which then releases the data in a yearly report for industries, consumers, and nongovernmental stakeholders"). A similar concept of "stepped disclosure" can be invoked in an enforcement context, requiring a report or disclosure to the government when something bad occurs, followed by disclosure to the public if a firm fails to mitigate harm. *See* Bamberger, *supra* note 9, at 466 (describing stepped disclosure, although not using the term, as a system with in which disclosure is made to regulator first and then to public).

3. The Collaborative Governance of Algorithms

There are aspects of algorithmic decision-making that make it well suited to collaborative governance.[193] Algorithmic decision-making involves complex systems that are better suited to regulation at earlier stages of design,[194] are constantly changing over time, require a high level of technical expertise, have goals that are often articulated in terms of risk management, and have regulatory goals that are not easily measured or precisely defined. Moreover, algorithmic decision-making challenges at least some aspects of our legal system such that, absent substantive legal changes, individual challenges under existing laws may fail.[195]

Much of the literature on algorithmic accountability ports in various techniques from collaborative governance, but without recognizing that it is deploying collaborative governance and without an eye to building systemic accountability for the governance regime itself. This means that it too lightly dismisses public transparency, stakeholder input, and individual process as aspects of producing a legitimate systemic regime.

This Article produces two important insights for the existing literature on algorithmic governance. First many of the policy recommendations in that literature are, in fact, techniques from the collaborative governance toolkit. Auditing,[196] expert input both individually and through expert boards,[197] impact assessments,[198] documentation,[199] creating public interest causes of action,[200] and whistle-blower protections[201] are, in fact, techniques deployed in collaborative governance. Situating this conversation in the collaborative governance literature both broadens the possible toolkit and makes us take a

193. Only a few scholars have called for the use of collaborative governance in regulating algorithmic decision-making. *See supra* note 7. Most have done so in the particular contexts of copyright and public health law, rather than contemplating what such a regime might look like for algorithmic decision-making writ large. And none have coupled their suggestions with a system of individual rights.

194. Desai & Kroll, *supra* note 14, at 41–42; *see also* Kroll et al., *supra* note 14, at 695; Lehr & Ohm, *supra* note 3, at 658.

195. Barocas & Selbst, *Disparate Impact*, *supra* note 4, at 701–12. For an alternative viewpoint, see generally Grimmelmann & Westreich, *supra* note 4 (proposing how antidiscimination law might handle machine learning algorithms).

196. Kim, *supra* note 14, at 190.

197. Kroll et al., *supra* note 14, at 703 (calling for appointing expert special masters); *see also* Citron, *supra* note 14, at 1312.

198. *See supra* note 121; *see also* AI Now Inst., *Algorithmic Impact Assessments: Towards Accountable Algorithms in Public Agencies*, MEDIUM (Feb. 21, 2018), https://medium.com/@AINow Institute/algorithmic-impact-assessments-toward-accountable-automation-in-public-agencies-bd9856e6 fdde.

199. Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1129–38.

200. Desai & Kroll, *supra* note 14, at 56.

201. *Id.*; *see also* Katyal, s*upra* note 8, at 126–29.

systemic and governance-based point of view instead of deploying these tools one or several at a time.

Second, understanding that the conversation about accountability in algorithmic decision-making is, in fact, largely (though, not entirely) a conversation about collaborative governance better frames how we talk about transparency. Rather than arguing over the instrumental value of individual notice or of publicly releasing source code, we should be discussing how to obtain aggregate accountability across a firm's decision-making, over time.

That is, most proposals to date largely delegate the governance of algorithmic error, bias, and discrimination to private firms. We need to establish accountability over how these firms set rules (determine what constitutes "fairness" or "bias" and how to address it) and how they implement those rules (ensuring firms deploy the tools they say are necessary to solve the problems and that the deployed tools in fact address the problems).

To some extent, analogies to individual procedural due process have obscured this point. We impose accountability frameworks on companies that deploy decision-making algorithms not just because those technologies make decisions that look like the kind of serious decisions that traditionally invoked procedural due process. We impose accountability because we effectively rely on those firms to come up with and comply with substantive rules preventing error, bias, and discrimination.[202] Transparency, including individual transparency and public transparency of some kind, is a necessary component of accountability. But by itself it is not enough.[203] Stakeholder input, expert oversight, and a real threat of enforcement are also crucial aspects of an accountable regime.

i. Designing Collaborative Algorithmic Governance

Collaborative governance is highly context dependent.[204] We will need to learn far more about the players in algorithmic decision-making, in particular contexts, to structure an effective regulatory system. This Section represents an initial attempt to outline both a portrait of the sector and what

202. Citron, *supra* note 14, at 1288 ("Computer programmers inevitably engage in rulemaking when they construct an automated system's code.").

203. Ananny & Crawford, *supra* note 10, at 984 ("If we recognize that transparency alone cannot create accountable systems and engaging with the reasons behind this limitation, we may be able to *use* the limits of transparency as conceptual tools for understanding how algorithmic assemblages might be held accountable.").

204. *See* Freeman, *supra* note 7, at 32–33, 98.

an effective collaborative governance regime might look like, with the caveat that the development and use of algorithmic decision-making in may look significantly different in different sectors.

First there are indications both of self-organizing and professionalism within the industry and related academic circles.[205] For example, the Institute of Electrical and Electronics Engineers ("IEEE"), an international standard-setting organization, established the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.[206] This initiative recently launched the second edition of a treatise on ethics in autonomous systems.[207] Based on this treatise, the IEEE Standards Association is actively working on new standards related to AI (including a project to help users certify how to eliminate bias in algorithms) and also now hosts a number of courses on ethics and AI.

Outside of the IEEE, computer scientists are having robust conversations about algorithmic accountability and explainability and how to make such systems intelligible to individuals. One recent paper identified nearly 300 core papers on explainable systems, with over 12,000 citing papers in the literature.[208] A growing annual, multidisciplinary conference on fairness, accountability, and transparency in machine learning[209] has, as of 2018, joined the Association for Computing Machinery, an international computing society and resource for computing professionals, embedding it in the computing profession. One key question, however, is how much of algorithmic decision-making will be built by established players and professionals versus be deployed by smaller startups or by users potentially less situated in the developing professional community.

205. Guihot and others note that the Partnership on AI led by Google, the Future of Life Institute's twenty-three Asilomar principles, and the IEEE's self-regulatory efforts in the form of discussion papers and projects, are related to AI and autonomous systems. Guihot et al., *supra* note 8, at 432–34. Katyal, too, notes private sector standards from the Association for the Advancement of AI, the Association of Computing Machinery ("ACM"), IEEE, and the British Computer Society. Katyal, *supra* note 8, at 109–10.

206. *See Ethics in Action*, IEEE, https://ethicsinaction.ieee.org (last visited Aug. 4, 2019).

207. IEEE GLOB. INITIATIVE ON ETHICS OF AUTONOMOUS & INTELLIGENT SYS., IEEE, ETHICALLY ALIGNED DESIGN (2017), http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf.

208. Ashraf Abdul et al., *Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda*, CHI 2018, Apr. 21–26, at 1, 1 ("We investigate how HCI researchers can help to develop accountable systems by performing a literature analysis of 289 core papers on explanations and explainable systems, as well as 12,412 citing papers.").

209. *ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*)*, ACM FAT* CONF., https://fatconference.org (last visited Sept. 3, 2019). This author has, since writing this Article, been named a cochair of the law track of the 2020 ACM Conference on Fairness, Accountability, and Transparency.

There is a growing array of civil society actors concerned with algorithmic decision-making, at least in the United States. Well-established civil liberties organizations, such as the Electronic Frontier Foundation, Center for Democracy and Technology, or the Electronic Privacy Information Center, could participate in rule-setting or monitoring around algorithmic decision-making. So could existing nonprofits, such as the Southern Poverty Law Center or the American Civil Liberties Union, that address related substantive concerns, including discrimination and civil rights. Newer AI-focused organizations also have the potential for involvement in a collaborative governance regime.[210] These third-party actors could be used in oversight mechanisms like expert boards or to double-check auditing or in more formal negotiations of codes of conduct. They will, however, likely need additional support and resources for developing technological expertise and capacity.[211]

Given both the increasing professionalization of the field and the array of possible third-party actors, a collaborative approach to algorithmic accountability might work in practice. How, then, might the regime be structured?

First we would need to establish clear liability for the kinds of failures or systemic problems that we want to avoid—establishing it in broader standards rather than specific rules, but bounding those standards to remain tethered to the public interest, so companies cannot exploit limitless flexibility.[212] We would need to create a forceful mechanism for enforcing these standards, whether through an existing agency, such as the FTC (which currently addresses privacy harms through a consumer protection approach), or through private rights of action or actions by federal prosecutors or state attorneys general.[213] There would need to be a strong "penalty default" in place: a real threat of significant fines, like the GDPR's 4 percent of worldwide revenue. And the government agency responsible for enforcement would have to issue fines often enough to scare private industry

210. *See, e.g.*, *About*, AI NOW INST., https://ainowinstitute.org/about.html (last visited Sept. 3, 2019); *About Us*, PARTNERSHIP ON AI, https://www.partnershiponai.org/about (last visited Sept. 3, 2019).

211. *See* Freeman, *supra* note 7, at 32 & n.81 (noting that even the formal involvement of third parties alone might not be enough, as third parties such as nongovernmental organizations often have a lower bandwidth and fewer resources than companies whose interests are in tension with them).

212. The literature on rules versus standards has an eye to limiting what can be deregulated. Lobel, *The Renew Deal*, *supra* note 7, at 468 (quoting Roderick A. Macdonald, *Metaphors of Multiplicity: Civil Society, Regimes and Legal Pluralism*, 15 ARIZ. J. INT'L & COMP. L. 69, 77 (1998)) ("The idea that core substantive arrangements are left open becomes, under certain conditions, insufficiently value-oriented. We do not want a paradigm in which 'conceptions of justice are . . . infinitely plural' . . . .").

213. *See* Citron & Pasquale, *supra* note 4, at 23; Crawford & Schultz, *supra* note 5, at 126–27.

to the negotiating table or into establishing a real compliance regime. The worldwide rush of companies to create internal compliance regimes to comply with the GDPR suggests that its level of fines might be working at doing just that. In the U.S. context, something would need to change. As I have noted elsewhere, "[W]hat penalties there are in U.S. privacy law are not high enough, or likely enough to be enforced against a particular industry actor, to drive participation by most of the industry actors with whom the government wants to co-regulate."[214]

Against the backdrop of potential enforcement, we could either formally create the opportunity to negotiate safe harbors in the form of codes of conduct, with transparency to and input from at least third-party stakeholders if not the public. Or, more informally, we could encourage private actors to fill in the details of compliance in their particular context, subject to independent monitoring (what Kenneth Bamberger refers to as "regulation as delegation").[215] We could encourage the creation of certification mechanisms to create private standards and systems of compliance[216] and have agencies issue guidance and convene workshops to establish best practices in the field.[217]

We could require the establishment of professional and independent "algorithmic decision-making officers"—that is, compliance officers—in companies that deploy algorithmic decision-making. Or we could craft a safe harbor from liability for companies that install such officers.[218] We could require companies to both make substantive commitments about their systems and create impact assessments before their deployment.[219] We could require those impact assessments to clearly describe the system, detail decisions around its design, address risk-mitigation measures, establish potential benchmarks, and detail considered but rejected alternatives.[220] We could require periodic reporting or performance assessments, either to a

214. Kaminski, *supra* note 152, at 946.

215. *See generally* Bamberger, *supra* note 9 (exploring the delegation of decision-making in various contexts and offering suggestions to improve accountability).

216. *But see* Edwards & Veale, *supra* note 15, at 80 (discussing the risks of self-regulation).

217. Guihot et al., *supra* note 8, at 397 n.43 (discussing the Obama-era White House's early forms of convening and guidance on AI).

218. *See* Lobel, *The Renew Deal*, *supra* note 7, at 421 (citing Burlington Indus., Inc. v. Ellerth, 524 U.S. 742, 764 (1998)) (discussing how the Supreme Court has "recognized" that "the adoption of internal antiharassment policies by firms" may be "a defense to sexual harassment suits").

219. *See supra* note 121.

220. AI Now Inst., *The 10 Top Recommendations for the AI Field in 2017*, MEDIUM (Oct. 18, 2017), https://medium.com/@AINowInstitute/the-10-top-recommendations-for-the-ai-field-in-2017-b3253624a7.

government agency, to an independent expert board, or to the public.[221] We could instruct a government agency, such as the FTC, to release regular industry-wide assessments to the public.

We could establish mandatory third-party auditing both during training and as a company runs the model.[222] We could require companies to include nongovernmental organizations ("NGOs") or other experts on internal oversight panels, revealing deeper information to these internal boards than to the public, potentially including both source code and data sets.[223] This is particularly important because the instrumental goals of algorithmic governance are not just technical, but have a significant normative-legal component; engineers should not be defining "discrimination" or "fairness" without extensive conversation with lawyers and impacted community members.[224] We could establish whistleblower protections for employees who reveal bad actions in case corporate compliance culture misaligns with public goals or outright fails.[225]

To encourage and enable private actor enforcement, we could explicitly award standing to third parties for a cause of action against the behavior we want to prevent.[226] We could award attorneys' fees or create other kinds of monetary incentives.[227] Federal agencies, again such as the FTC, could serve in a capacity-enhancing role, hiring technologists and conducting relevant research to aid NGOs and other private actors.

If these proposals sound familiar, it is because they have been peppered across more recent contributions to the algorithmic governance literature. They have not, however, been brought together as a whole. Nor has there been much conversation about how different levers of accountability at different stages—rule-setting versus implementation versus compliance or enforcement—might feed back into each other.[228] For example, a company with an internal compliance officer might produce reports to an agency that could then publish summary reports to the public, which might respond by avoiding the company in the market, pressuring the agency to issue new best

221. Zarsky, *supra* note 14, at 1529.

222. Citron & Pasquale, *supra* note 4, at 21, 24–25, 28; Kim, *supra* note 14, at 190–91.

223. PASQUALE, *supra* note 118, at 160–65.

224. *See* Kim, *supra* note 14, at 193 ("[O]ne challenge is the ongoing disagreement about the meaning of discrimination."); Lehr & Ohm, *supra* note 3, at 705 n.187 (explaining that there are different possible mathematical definitions of fairness and "optimizing for one precludes achieving another").

225. Desai & Kroll, *supra* note 14, at 56–60.

226. *Id*. at 60–64.

227. *Id*. at 63–64.

228. For one exception, see generally Price, *supra* note 8 (arguing that the FDA should take a collaborative approach to regulating medical algorithms).

practices, pressuring legislators to produce new laws, or putting resources behind civil society groups that could end up participating on oversight boards of companies and contributing to the impact assessment process. If there is no disclosure at any point to persons outside the company or regulatory body, then the aspect of collaborative governance that harnesses or even relies on the power of third parties in both setting publicly oriented rules and producing compliance with them will fail.

In designing a system of collaborative governance for algorithmic decision-making, then, we should be considering the system as a whole, not introducing standalone mechanisms. Tools such as audits without a threat of enforcement, impact assessments with no public transparency, even oversight boards without the ability to enforce or, at least, report out (whether to a regulator or the public) that are deployed individually do not create an effective or legitimate governance regime.

ii. Potential Pitfalls

There are significant hurdles and costs to such a system, beyond the more general problem of passing legislation through Congress, and beyond the problem of capture discussed at length above.[229] Some of these apply to collaborative governance systems in general, but others may be more specific to algorithmic governance.

There are problems to creating and enforcing broad, behavioral standards rather than specific rules.[230] Companies may mishandle the delegation implied in standards, either intentionally or unintentionally. Standards can be in tension with rule-of-law values about fair notice of prohibited conduct by the law.[231] We would have to be wary of producing unfettered agency discretion.[232] And there are costs to delaying the creation of specific rules proscribing bad conduct, as companies escape liability by arguing that a standard is too vague.[233]

Collaborative governance can favor incumbents who can afford to

229. *See supra* Section II.B.1.

230. *See* Birnhack, *supra* note 141, at 38–45; Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557, 561–62 (1992) ("One can think of the choice between rules and standards as involving the extent to which a given aspect of a legal command should be resolved in advance or left to an enforcement authority to consider."); David A. Super, *Against Flexibility*, 96 CORNELL L. REV. 1375, 1411–17 (2011).

231. Bamberger, *supra* note 9 at 434 (stating that standards can be in tension with "rule-of-law values [that] demand that law—particularly law enforceable by criminal sanction—include sufficient detail to provide fair notice of prohibited conduct").

232. Freeman, *Collaborative Governance*, *supra* note 7, at 85–87.

233. *See* Super, *supra* note 230, at 1417–18.

participate in rule-setting, potentially producing rules that favor them and disfavor newer industry actors. It may also disfavor smaller enterprises that can ill afford to take on compliance costs and culture. And if the field indeed proves to be highly heterogeneous, then creating a collaborative governance regime tailored to different subsectors may end up proving extraordinarily costly to administer.[234]

One of the paradoxes of collaborative governance is that precisely those subject matter areas that the approach would be best suited to—newer technologies, complex and evolving technologies—are those areas that for other reasons may be least suited to it.[235] If there is a lack of a culture of professionalism, big gaps between established players and new entrants, and a lack of consensus within and between industry and outside players, collaborative governance can be extremely challenging to implement.

A final and important criticism of collaborative governance regimes is that they are ill-suited to determining the content of human rights. Unless we are very, very careful to delineate the outer limits of what may be negotiated, we may find private companies reasoning away rights protections in their rule-setting or implementation of standards. This brings us back to why binary governance is binary—why we need not just collaborative governance but a system of individual rights as well.

C. COMBINING THE TWO APPROACHES

A systemic collaborative governance regime has important benefits over other regulatory approaches in addressing instrumental concerns about algorithmic decision-making. It addresses some but not all classes of justificatory concerns about the legitimacy of the system. It does not, however, adequately address dignitary concerns, nor address justificatory concerns about particular individual decisions. For that reason, we need to additionally establish a system of individual rights.

While collaborative governance is centrally concerned with accountability, the kinds of accountability that it produces are not always coextensive with the kinds of accountability we require in the context of due process owed to individuals. Collaborative governance may produce a regime that is adequately accountable and legitimate when it comes to producing rules that govern algorithmic systems, but not adequately accountable and legitimate when it comes to justifying individual decisions.

234. Bamberger, *supra* note 9, at 398 ("[T]he more a regulation prescribes broad policy goals, rather than specific behavior as a measurable outcome, the more difficult it is to monitor compliance.").

235. Kaminski, *supra* note 152, at 948–49.

A system of governance through third-party audits, expert boards, government inspection and enforcement, and performance reports might produce better and more legitimate algorithms, but it might still not produce a justificatory system that would be acceptable from the perspective of an individual affected by a particular decision.

Yet as noted above,[236] individual rights, especially individual rights invoked ex post and triggered by a particular decision, will not be the best way to create more accountable companies, improve management culture, and correct algorithmic design. Thus to effectively govern algorithmic decision-making, we need both. The two parts of this system, however, will interact in complex ways.

1. The Two Systems as Complementary

In some aspects, the two parts of a binary system of governance will be complementary. Individual rights can produce instrumental contributions, be an important component of systemic accountability, give substance to the rules in a collaborative governance regime, and address one of the problems of using collaborative governance in this particular area—that oversight and transparency can produce additional individual privacy harms. Collaborative governance can, in turn, contribute deeper layers of accountability towards justificatory goals than a system reliant just on individual rights. Collaborative governance could help us define how rights will be implemented in a particular technological setting.

First, individual rights can complement collaborative governance by addressing individualized error, bias, and discrimination. As discussed above,[237] individual rights can help uncover and fix instrumental problems with algorithmic decision-making. Individuals are best situated to know when profiles contain factual error or erroneous inferences. Additionally, individual narratives about discrimination or bias may be more palatable to the public than agency-produced reports or statistics, and could feed back into collaborative governance by contributing to ongoing policy conversations about the broader governance regime.

Second, individual rights can be an important aspect of systemic accountability. Proponents of collaborative governance might, in fact, be surprised to see individual participation and process characterized as distinct from aggregate accountability.[238] Individual enforcement is quintessentially

236. *See supra* Section II.B.
237. *See id.*
238. *See, e.g.*, Freeman, *supra* note 6, at 587, 622 (referring to individual hearings).

collaborative—it spreads the cost of obtaining compliance from the government to private actors. And putting in place an individual right to contest an algorithmic decision, or to obtain an explanation, or to obtain a human in the loop could contribute to management reform within a company, encouraging the establishment of compliance personnel or complimentary review processes.

Third, individual rights can provide a check on substantive rule-setting by private actors. The law might require, for example, that individuals be given an explanation of an algorithmic decision, but leave it to private actors to implement precisely what that explanation constitutes in practice. Making this an individual right, rather than just a company duty, could ensure that there is individual or even judicial review of any implementation. Individual rights, in other words, can provide the substantive backstop for company rule-setting, by subjecting rules or implementations to challenges, potentially subject to judicial review. This potentially brings courts into collaborative governance as important players that are more insulated from politics and take into account both human rights regimes and fairness concerns.

Fourth, an individualized disclosure regime could address one of the central challenges to deploying collaborative governance in this space: that releasing large amounts of information to the public threatens individual privacy, among other informational interests.[239] Collaborative governance requires extensive monitoring and disclosure, but broader public disclosure of profiles and personal inferences raises serious privacy concerns. Allowing individuals to obtain this information produces oversight without necessarily passing personal information on to the public. Relatedly, using collaborative governance instead of command-and-control governance means that information may be passed on to private third parties for oversight purposes rather than to the government itself—which matters to the extent that we want to protect individual privacy from government surveillance.

An accountable collaborative governance regime can also complement individual procedural rights. Establishing systemic accountability in a collaborative governance regime can bolster individual rights by providing oversight in the name of affected individuals. Making a collaborative governance regime systemically accountable also does work towards making it individually justifiable.

This interaction of types of accountability envisions several types of

239. *See* Kroll et al., *supra* note 14, at 658.

differently purposed but complementary information flows.[240] Individuals may be provided with a simplified and understandable explanation of algorithmic decision-making, giving them insight into whether a decision is justified in their respective cases, at the same time that third-party auditors assure them that the system, as a whole, is not biased or unjustifiable. Moreover, a system of collaborative governance could bolster the ability of individuals to enforce rights by producing an aggregate picture that can root out systemic bias, when one-off, individual narratives might not.[241]

Finally, we could use collaborative governance to implement individual rights. Several scholars have, for example, suggested that algorithmic due process should require notice to individuals that allows them to tinker with an algorithmic decision-making system through an easily comprehensible interface.[242] This is quintessentially a collaborative governance problem: we would not want lawmakers to write specific and soon-outdated rules dictating what such a system must look like. Instead, the law might put in place a broad notice requirement, coupled with agency guidance, or a safe harbor for those coming up and complying with a government-approved industry code of conduct. As discussed in Part III, there is some evidence that just such a process is beginning to take place in the EU.[243]

2. The Two Systems in Tension

The two systems will also, however, be in tension. An individual rights regime can conflict with system-wide accuracy and bias. Second, a strong individual rights regime may restrict regulators' abilities to deploy "responsive regulation"—that is, to calibrate enforcement measures up or down in order to properly incentivize private collaboration.[244] Third, establishing accountability in a collaborative governance regime can raise individual privacy concerns. Fourth, where systemic and individual accountability can be complementary, there is also a danger of confusing one kind of accountability for another and crafting a system that is accountable along only one axis. And finally, while collaborative governance can do

240. *See, e.g.*, Ananny & Crawford, *supra* note 10, at 983 ("Holding an assemblage accountable requires not just seeing inside any one component . . . but understanding how it works as a system."); Citron & Pasquale, *supra* note 4, at 18–29; Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 711.

241. Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1130–33.

242. *See supra* note 105.

243. Jef Ausloos et al., Position Paper Presented at CHI 2018: Algorithmic Transparency and Accountability in Practice (Jan. 2018), https://uploads-ssl.webflow.com/5a2007a24a11ce000164d272/5ac883392c10d1baaa4358f2_Algorithmic_Transparency_and_Accountability_in_Practice_CameraReady.pdf.

244. *See* AYRES & BRAITHWAITE, *supra* note 12, 4–7.

important work in the implementation of fundamental rights or human rights, there is also danger in letting private parties do interpretative work instead of courts or lawmakers.

A system of individual rights can conflict with system-wide accuracy and system-wide concerns about bias. Individuals could use correction and erasure rights to intentionally game a decision-making algorithm, by changing or eliminating negative information about themselves.[245] Even just opting out of a system without actively introducing inaccuracies can affect system bias. For example, if only well-educated, socioeconomically elite individuals opt out, then the contours of a machine learning system would come to reflect those who are less empowered and remain within it. Allowing individuals to alter or erase their information changes the data set, which changes the algorithm going forward. There are thus real tensions between attending to individual dignitary or autonomy concerns on the one hand, and addressing systemic bias and discrimination on the other.[246]

Second, a strong individual rights regime may prove to be too much of a hammer. If collaborative governance relies on the ability of regulators to soften regulation as an incentive for coming to the table, then giving enforceable rights to individuals may remove that incentive. Companies might decide not to take on the cost of participating in collaborative governance, since they may end up facing costly individual legal challenges after all.

A third point of tension, mentioned above,[247] is that collaborative governance centrally requires extensive monitoring and information disclosure. Individual privacy concerns push back against this. So do other informational interests, ranging from business interests in trade secrets to security concerns.[248] But there is a central problem with hinging a collaborative governance regime not on disclosure to the public but on disclosure to individuals. Not all individuals will invoke their individual rights; individually targeted disclosure will function selectively based on access to justice, cost, and both legal and technical expertise. A collaborative governance regime with minimal public disclosure would, as we see below in the example of the GDPR,[249] need to carefully calibrate all of its other accountability mechanisms in order to function.

245. *See* Bambauer & Zarsky, *supra* note 118, at 34–43.
246. *Id*. at 23–32; Hildebrandt, *supra* note 14, at 48–49; *see* Zarsky, *supra* note 14, at 1505–10.
247. *See supra* Section II.C.1.
248. PASQUALE, *supra* note 118, at 12, 193; Kroll et al., *supra* note 14, at 639.
249. *See infra* Part III.

Fourth, there is a danger of confusing the two kinds of justificatory requirements entwined in a dual system. Individual justification requires information that is legible to an ordinary individual.[250] But systemic accountability requires other kinds of checks, including deep disclosure to experts and monitoring by third parties.[251] Just because a binary system is "transparent" along one axis does not mean it is transparent along others. A system might adequately justify itself for purposes of producing individual decisions but not for purposes of producing legitimate rules about preventing system-wide discrimination, or vice versa. A binary system should, thus, be assessed towards both individual and systemic accountability goals.

Finally, collaborative governance can be dangerous when applied to the substance of fundamental rights.[252] Companies are not courts; nor are they agents of the federal government. They do not have individuals' best interests in mind, nor are they bound by constitutional norms. When addressing serious human rights concerns, we must be careful not to use collaborative governance tactics towards producing accountability measures that function as enablement regimes.

III. THE TWO FACES OF THE GDPR

In the United States, despite repeated calls for transparency and accountability,[253] the policy landscape around algorithmic decision-making remains largely a blank slate.[254] Some requirements apply to some government use of algorithms,[255] but the United States lacks general law to govern private sector and many government uses of algorithmic decision-

250. *See, e.g.*, Gianclaudio Malgieri & Giovanni Comandé, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 243, 248–50 (2017); *see also* Kaminski, *supra* note 86, at 213.

251. PASQUALE, *supra* note 118, at 56–58.

252. *See, e.g.*, CHRISTINA ANGELOPOULOS ET AL., INST. FOR INFO. LAW, STUDY OF FUNDAMENTAL RIGHTS LIMITATIONS FOR ONLINE ENFORCEMENT THROUGH SELF-REGULATION 52 (2015), https://pure.uva.nl/ws/files/8763808/IVIR_Study_Online_enforcement_through_self_regulation.pdf.

253. *See, e.g.*, *Publications*, AI NOW INST., https://ainowinstitute.org/reports.html (last visited Sept. 3, 2019).

254. Margot E. Kaminski & Andrew D. Selbst, Opinion, *The Legislation That Targets the Racist Impacts of Tech*, N.Y. TIMES: THE PRIVACY PROJECT (May 7, 2019), https://www.nytimes.com/2019/05/07/opinion/tech-racism-algorithms.html (discussing Senator Wyden, Senator Booker, and Representative Clarke's proposal); DJ Pangburn, *Washington Could Be the First State To Rein In Automated Decision-Making*, FAST CO. (Feb. 8 2019), https://www.fastcompany.com/90302465/washington-introduces-landmark-algorithmic-accountability-laws (discussing proposed Washington state law, now off the table); Powles, *supra* note 183 (discussing a New York attempt at legislation, later watered down); *see also supra* note 4.

255. Zarsky, *supra* note 14, at 1507–09.

making.[256]

In Europe, however, this is not the case. The GDPR went into direct effect on member states in May 2018. The GDPR governs processing of personal data and applies to both the government and the private sector. The GDPR contains an elaborate algorithmic accountability regime.

For the most part, the focus of the conversation about algorithmic accountability in the GDPR has been on the individual rights regime and, even more narrowly, on the so-called "right to explanation" of individual algorithmically made decisions.[257] This Article calls attention instead to the binary nature of the GDPR. I argue that the GDPR is both a system of individual rights and a complex compliance regime that, when applied to the private sector, is constituted through collaborative governance.[258] The GDPR relies on both formal and informal tactics to create public-private partnerships in governing algorithmic decision-making.[259]

256. Crawford & Schultz, *supra* note 5.

257. Edwards & Veale, *supra* note 15, at 44 ("In 2016, to the surprise of some EU data protection lawyers, and to considerable global attention, Goodman and Flaxman asserted . . . that the GDPR contained a 'right to an explanation' of algorithmic decision making. As Wachter et al. have comprehensively pointed out, the truth is not quite that simple." (footnote omitted)); *see also* Emre Bayamlioglu, *Contesting Automated Decisions: A View of Transparency Implications*, 4 EUR. DATA PROTECTION L. REV. 433, 444 (2018); Maja Brkan, *Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond*, 27 INT'L J.L. & INFO. TECH. 91, 110–19 (2019); Bryan Casey et al., *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L.J. 145, 159–64 (2019); Malgieri & Comandé, *supra* note 250, 246–50; Mendoza & Bygrave, *supra* note 14, at 92–94; Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, 7 INT'L DATA PRIVACY L. 233, 237–42 (2017); Wachter et. al, *supra* note 62, at 860–61; Sandra Wachter et al., *Why a Right to Explanation of Automated Decisionmaking Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76, 78–70 (2017); Bryce Goodman & Seth Flaxman, *European Union Regulations on Algorithmic Decision Making and a "Right to Explanation,"* AI MAG., Fall 2017, at 50, 55–56.

258. I am not the first to observe that the GDPR has a dual nature. *See, e.g.*, Roger Taylor, *No Privacy Without Transparency*, *in* DATA PROTECTION AND PRIVACY 63, 74–76 (Ronald Leenes et al. eds., 2017); Claudia Quelle, The Data Subject as a "Small-Scale Sovereign" – The Duties of Controllers and the Rights and Powers of Data Subjects under the GDPR 1 (Dec. 2017) (unpublished manuscript) (on file with author); *see also* Frederik Zuiderveen Borgesius, *Informed Consent: We Can Do Better To Defend Privacy*, 13 IEEE COMPUTER & RELIABILITY SOC'YS, Mar.–Apr. 2015, at 103, 103–07 (distinguishing between (i) protecting and (ii) empowering people). Even as early as 2003, the European Commission indicated its interest in private-public partnerships. *See First Report on the Implementation of the Data Protection Directive (95/46/EC)*, at 26, COM (2003) 265 final (May 15, 2003) ("[The Commission] believes that self-regulation, and in particular codes of conduct should play an important role in the future development of data protection in the EU and outside, not least in order to avoid excessively detailed legislation.").

259. Irene Kamara discusses formal coregulatory mechanisms, focusing on the incorporation of technical standards in privacy by design. She does not discuss the overall collaborative system nor the role of transparency in it. Irene Kamara, *Co-regulation in EU Personal Data Protection: The Case of*

This observation matters for three reasons. First, it matters because the GDPR contains not just a system of individual rights that can be invoked after an algorithm is deployed, but an approach to governing algorithmic system design from the onset, including the ability to affect the management structure in companies.[260] Second, this characterization shifts the conversation about transparency in the GDPR from being about whether individual transparency is necessary or useful to how to structure systemic accountability. The question is not just how to make algorithms accountable to affected individuals; it is how to hold private companies accountable in developing the norms, policies, and technical tools involved in and providing oversight over algorithmic decision-making.

Third, the GDPR's binary governance approach provides an example of how the two prongs of a binary governance system might interact. The two systems in the GDPR—individual rights and collaborative governance—are, as anticipated, often complementary. But they also may prove to be in tension and will require careful calibration, in conversation with each other, to be effective at governing algorithmic decision-making. Intriguingly the GDPR's absence of public-facing and stakeholder-facing accountability suggests that individual transparency rights may have to serve a crucial accountability role in its system of collaborative governance. Thus, even for those focused on instrumental rather than dignitary or justificatory goals, individual rights in the GDPR may be necessary for producing effective *systemic* regulation, too.

The GDPR thus provides an illustration of the binary approach to algorithmic accountability, in action. As the GDPR is enforced and policies develop further over time, the regime will be a useful source of evidence about when and whether this approach works in practice. Even now, at early stages of implementation, the GDPR's approach to governing algorithmic decision-making provides lessons for the rest of the world.

Technical Standards and the Privacy by Design Standardisation 'Mandate,' 8 EUR. J.L. & TECH., no. 1, 2017, at 1, 15. Ed Lee has framed the "right to be forgotten" under the GDPR's predecessor, the Data Protection Directive, as coregulatory in nature, but has not extended this description to other rights in the GDPR. *See* Edward Lee, *Recognizing Rights in Real Time: The Role of Google in the EU Right To Be Forgotten*, 49 U.C. DAVIS L. REV. 1017, 1066 (2016) (calling Google a "private administrative agency" of the right to be forgotten). Others have also referred to Google's role after the Google Spain decision as a form of privatization of rights. ANGELOPOULOS ET AL., *supra* note 252, at 23–24.

260. *See, e.g.*, Casey et al., *supra* note 257, at 39; Edwards & Veale, *supra* note 15, at 82.

A. A PRIMER ON THE GDPR

First, a note about the GDPR for U.S. readers unfamiliar with the regime: The GDPR, like its predecessor the Data Protection Directive, creates an extensive data protection regime built primarily around the Fair Information Practices ("FIPs").[261] The FIPs originated in the United States but have become the international standard for data protection, under the Organisation for Economic Co-operation and Development, and now form the basis for many countries' data protection laws.[262] Traditionally the FIPs establish a number of individual rights, including access, disclosure, and correction rights, along with general obligations respecting data gathering, storage, and use. The GDPR thus contains a broad assortment of individual rights and company obligations that apply beyond algorithmic decision-making, to personal data processing in general.[263]

This Article covers only those aspects of the GDPR that might apply to or affect algorithmic decision-making. For a U.S. audience, however, it is crucial to understand that the GDPR's individual rights respecting algorithmic decision-making do not stand in isolation. They exist against the backdrop of more generally applicable law that applies to data processing.[264] Thus, even if the specific rights that apply to algorithmic decision-making prove in practice to be toothless (which does not appear to be the case given regulators' recent interpretations[265]), other aspects of the GDPR give individuals substantial abilities to influence both profiling and consequent decision-making, algorithmic or not.[266]

261. Article 5 of the GDPR lays out its version of the FIPs principles, which are further elaborated in Articles 12 to 23 and elsewhere in the GDPR. These include the following: "lawfulness, fairness and transparency"; "purpose limitation"; "data minimisation"; "accuracy"; "storage limitation"; "integrity and confidentiality"; and "accountability." Council Regulation 2016/679, *supra* note 13, art. 5, at 35–36; *see also id.* at 7; Paul M. Schwartz & Karl-Nikolaus Peifer, *Transatlantic Data Privacy Law*, 106 GEO. L.J. 115, 128 (2017).

262. SEC'YS ADVISORY COMM. ON AUTOMATED PERS. DATA SYS., U.S. DEP'T OF HEALTH, EDUC. & WELFARE, DHEW PUB. NO. (OS)73-94, RECORDS COMPUTERS AND THE RIGHTS OF CITIZENS xxiii (1973); *Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, ORG. ECON. CO-OPERATION & DEV. (Sept. 23, 1980), http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm. *But see* BORGESIUS, *supra* note 57, at 58–59 (suggesting FIPs-like principles arose around the world in different places around the same time).

263. *See generally* Schwartz & Peifer, *supra* note 261 (describing European data privacy law).

264. *See, e.g.*, Edwards & Veale, *supra* note 15, at 23, 38, 74, 82.

265. Kaminski, *supra* note 86, at 193, 209–17. For the recent guidelines, see generally Article 29 Data Prot. Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679*, WP251 rev.01 (Feb. 6, 2018) [hereinafter Article 29 Data Prot. Working Party, *Decision-Making*].

266. *See generally* Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265

Additionally, U.S. readers may be unfamiliar with the various interpretative documents that surround the GDPR and their relative strength. This, too, is worth clarifying before delving into substance. The GDPR, as regulatory regime, consists of text (the Articles), a preamble (the Recitals), and Guidelines from the European Data Protection Board (formerly the Article 29 Working Party). Technically the GDPR's text is the actual law.[267] The Recitals may not create new law; they may, however, be used by courts and regulators to interpret the text.[268] The European Data Protection Board, which consists of data protection authorities (regulators and enforcers) from around the EU, issues Guidelines.[269] Arguably these Guidelines serve both an interpretative and a harmonizing role: they interpret and clarify GDPR requirements for affected entities while also indicating how regulators around the EU will act.[270] Various individual Member State regulators, too, issue guidance on how to interpret the GDPR.[271]

Throughout this Part, I indicate which of these sources I am citing in support of arguments about what the GDPR does or does not require. A good amount of the GDPR's requirements for algorithmic accountability come from the Recitals and Guidelines and not directly from the text. This does not, however, mean that these requirements are legally toothless, as they are used by courts, regulators, and companies to interpret what the text means. Additionally, understanding the nature of these sources is necessary for understanding the GDPR's approach of combining layers of harder and softer law—that is, the GDPR's approach to collaborative governance.

B. INDIVIDUAL RIGHTS

The GDPR contains a number of generally applicable individual rights that affect algorithmic decision-making and related profiling, in addition to several individual rights specific to algorithmic decision-making. This Section provides an overview of the GDPR's individual rights, dividing them into three categories: notification and access rights, checks on data use and retention, and Article 22's version of algorithmic "due process" for "solely

(clarifying the GDPR provisions on profiling and automated decision-making).

267. *See* Kaminski, *supra* note 86, at 193–95.

268. *See id.* at 193–94.

269. *Members*, EUR. DATA PROTECTION BOARD, https://edpb.europa.eu/about-edpb/board/members_en (last visited Sept. 3, 2019).

270. *Role of the EDPB*, EUR. DATA PROTECTION BOARD, https://edpb.europa.eu/role-edpb_en (last visited Sept. 3, 2019).

271. *See, e.g.*, *Guide to Data Protection*, INFO. COMMISIONER'S OFF., https://ico.org.uk/for-organisations/guide-to-data-protection (last visited Sept. 3, 2019).

automated" decisions.[272]

1. Notification and Access Rights

The GDPR contains a number of individual notification and access rights. These are both established in the text of the GDPR[273] and interpreted into the GDPR, in connection to its consent requirements.[274] I will explain how these rights and requirements apply in the context of algorithmic decision-making and related profiling.

The GDPR establishes a system of generally applicable notification and access rights.[275] Upon collecting personal information from an individual, a company must provide the purpose for which data is gathered, the recipients of the data, and the retention period of the data, among other things.[276] Nearly identical information must be disclosed if a company obtains personal data not directly from an individual but from another party.[277]

Additionally, the GDPR contains affirmative access rights for individuals, including access to the source of the data and a copy of the data itself.[278] Access rights may be invoked at intervals and give individuals the ability to regularly check in about what information a company has about them, beyond the moment at which data has originally been obtained. The GDPR also addresses how information must be communicated.[279] Information must be communicated "in a concise, transparent, intelligible and easily accessible form, using clear and plain language . . . ."[280]

These general notification and access rights have ramifications for algorithmic accountability. Calls for algorithmic due process, discussed in Section II.A above, include calls for access to one's personal data[281] and for

272. Council Regulation 2016/679, *supra* note 13, art. 22, at 46.

273. The notification and access rights are under the GDPR's requirements of transparent communication, notification, and access in Articles 12, 13, and 14. Council Regulation 2016/679, *supra* note 13, art. 22, at 8, 39–42; Article 29 Data Protection Working Party, *Guidelines on Transparency Under Regulation 2016/679*, WP260, at 6, 13–23 (April 11, 2018).

274. Council Regulation 2016/679, *supra* note 13, art. 7, at 37; *see also* Article 29 Data Protection Working Party, *Guidelines on Consent Under Regulation 2016/679*, WP 259 rev.01 (Apr. 10, 2018) [hereinafter Article 29 Data Protection Working Party, *Consent*]; Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 12–13, 24 (discussing consent and explicit consent).

275. Council Regulation 2016/679, *supra* note 13, arts. 13–15, at 40–43.

276. *Id.* art. 13(1)(c), at 40, arts. 13(1)(e), 13(2)(a)–(b), at 41.

277. *Id.* art. 14, at 42 (providing that the controller will provide certain specified information "within a reasonable period after obtaining the personal data, but at the latest within one month . . . .").

278. *Id.* arts. 15(1)(g), 15(3), at 43.

279. *Id.* art. 12, at 39.

280. *Id.* art. 12(1).

281. Citron & Pasquale, *supra* note 4, at 20.

identifying the sources of personal information.[282] Whether obtained through notification or through an access request, these kinds of disclosures can help address all three categories of concerns about algorithmic decision-making: dignitary (they enable an individual to begin the process of protesting objectification); justificatory (they can reveal if not the details of a reasoning process the factual basis for it); and instrumental (they can help uncover error, bias, discrimination, or other kinds of unfairness).

Profiling as a subcategory of data processing triggers additional rights under the GDPR.[283] Individuals are entitled to be informed of the existence of profiling.[284] They are also entitled to be given some information about how profiling works (a right to an explanation of profiling, if you will).[285] They are entitled to request the data used as an input into their profiles, to request information in the profiles, and to request information on how they have been categorized—that is, inferences made about them.[286] According to the GDPR Guidelines, an individual must be informed of the existence of decision-making based on profiling, regardless of whether or not that decision-making is solely automated.[287]

282. *Id.*; Crawford & Schultz, *supra* note 5, at 125.

283. Profiling is defined under the GDPR as an automated form of processing, carried out on personal data, to evaluate personal aspects about a natural person. Council Regulation 2016/679, *supra* note 13, art. 4(4), at 33; Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 6–7. It does not need to be s*olely* automated, unlike the decisions referred to in Article 22. Council Regulation 2016/679, *supra* note 13, art. 22, at 46. Nor do decisions based on profiling need to have significant effects to fall under this provision. *See id.*; Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 6–8 (distinguishing automated decision-making from profiling).

284. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 17 ("The data subject has a right *to be informed* by the controller about and, in certain circumstances, a right *to object to* 'profiling', *regardless* of whether solely automated individual decision based on profiling takes place."); *see also* Council Regulation 2016/679, *supra* note 13, at 12.

285. Transparency of processing of personal data is a requirement of the GDPR for all kinds of processing. *See* Council Regulation 2016/679, *supra* note 13, art. 5(1)(a), at 35, art. 12(1), at 39; Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 9–11.

286. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 17 ("In addition to general information about the processing, pursuant to Article 15(3), the controller has a duty to make available the data used as input to create the profile as well as access to information on the profile and details of which segments the data subject has been placed into.").

287. Council Regulation 2016/679, *supra* note 13, art. 13(1)(c), at 40, art. 14(1)(c), at 41 (requiring disclosure of the purposes of processing); *id.* at 12 ("[T]he data subject should be informed of the existence of profiling and the consequences of such profiling."); *id.* ("Every data subject should therefore have the right to know and obtain communication in particular with regard to the purposes for which the personal data are processed . . . and, at least when based on profiling, the consequences of such processing."); *see also* Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 16 ("[W]here the processing involves profiling-based decision making (irrespective of whether it is caught by Article 22 provisions), then the fact that the processing is for the purposes of both (a) profiling and (b) making a decision based on the profile generated, must be made clear to the data subject.").

Thus even if no automated decision-making takes place, individuals have more substantial notification and access rights with respect to profiling than they do for data processing writ large. Because personal profiles are often the basis of algorithmic decision-making, this has the effect of governing algorithmic accountability whether or not a decision has been reached. It also has the effect of governing algorithmic accountability regardless of how automated the decision-making is.

Algorithmic decision-making itself, then, triggers yet more notification and access rights. A company must proactively notify individuals of the existence of solely automated decision-making.[288] This is precisely what the algorithmic due process literature has called for.[289] A company must, additionally, with respect to solely automated decision-making provide an explanation of the algorithm: "[m]eaningful information about the logic involved, as well as the significance and envisaged consequences of such processing for the data subject."[290] An individual also has a right to affirmatively request access to "meaningful information about the logic involved" and information about its significance and consequences.[291]

What this information will constitute in practice has been subject to an already extensive scholarly debate.[292] I note here only that these particular disclosure rights appear to be individual-centric, rather than system-oriented.[293] They aim at giving individuals meaningful transparency in order to enable other individual rights under the GDPR, such as the right of correction or right of contestation.[294] They are not expert-centric or aimed at providing oversight over the construction, administration, or development of an algorithm. But neither are they free of substance, and they appear, according to related Guidelines, to require significantly more explanatory depth than some have suggested.[295]

There is another source of individualized transparency in the GDPR: its

288. Council Regulation 2016/679, *supra* note 13, arts. 13(2)(f), 14(2)(g), at 41–42 ("[T]he existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.").

289. Citron & Pasquale, *supra* note 4, at 28.

290. Council Regulation 2016/679, *supra* note 13, art. 14(2)(g), at 42.

291. *Id.* art. 15(1)(h), at 43.

292. *See supra* note 257.

293. For more extensive discussion, see generally Kaminski, *supra* note 86.

294. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 27.

295. *Compare* Wachter et al., *supra* note 62, at 843, *with* Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 25–26.

consent requirements.[296] Explicit consent is one of the three exceptions to the GDPR's prohibition of automated decision-making.[297] For consent to be valid, individuals must be making an informed choice; they must "understand exactly what they are consenting to . . . ."[298] If a company does not adequately communicate to an individual both the purpose of data processing and information about the use of data for automated decisions, then consent may be deemed invalid.[299]

This again incentivizes disclosure of a particular kind: the kind individuals can meaningfully understand that contributes to individuals' ability to give or withdraw consent under the GDPR. This type of transparency is not necessarily conducive to expert oversight over algorithmic decision-making systems, but that is not its purpose. While the GDPR's individually oriented transparency provisions may have instrumental consequences, their primary objective is to address dignitary and justificatory concerns.

2. Other Checks on Data Processing

The GDPR does not just afford individuals transparency rights. It also includes substantive prohibitions on particular uses of data, heightened protections for certain kinds of data, and a number of FIPs-related individual rights beyond transparency. I again discuss these measures here with respect to automated decision-making and related profiling.

First, the GDPR contains substantive prohibitions on certain behavior—prohibitions that can be characterized as individual rights to *not* be subjected to something. One criticism of the FIPs is that they can be without substance, providing individuals the illusion of control, while in practice allowing companies to do nearly anything as long as they have gotten individuals to click through an agreement.[300] The GDPR attempts to provide backstops beyond individual control.

296. *See, e.g.*, Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 12–13.

297. Council Regulation 2016/679, *supra* note 13, art. 22(2)(c), at 46. Regular consent already requires a "clear affirmative act[]"; explicit consent is even more stringent, requiring "an express statement of consent," often though not always in writing. Article 29 Data Protection Working Party, *Consent*, *supra* note 274, at 18.

298. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 13.

299. Article 29 Data Protection Working Party, *Consent*, *supra* note 274, at 13 ("If the controller does not provide accessible information, user control becomes illusory and consent will be an invalid basis for processing."); *see also* Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 12–13.

300. Woodrow Hartzog, *The Inadequate, Invaluable Fair Information Practices*, 76 MD. L. REV. 952, 964–77 (2017).

The GDPR, in fact, prohibits solely algorithmic decision-making.[301] The three exceptions are when such decision-making is *necessary* to a contract (not merely subject to a contract), when member state law specifically addresses automated decision-making, and when an individual has explicitly consented to it.[302] Outside of these three contexts, an individual has a right not to be subject to solely algorithmic decision-making.

There is some discussion of whether the GDPR wholesale prohibits automated decision-making about children.[303] The applicable Guidelines toe an interpretative line, by stating that data protection authorities will protect children's data more strongly.[304] The Guidelines suggest that children's data will be subject to fewer than the above three exceptions to the general ban on automated decision-making.

The GDPR throughout creates heightened requirements for certain kinds of information, including sensitive information, such as race or biometric data,[305] referred to as "special categories of personal data . . . ."[306] Special categories of data are, like children's data, subject to fewer exceptions to the ban on automated decision-making. They can be processed only subject to explicit consent to processing "for one or more specified purposes"[307] or when "necessary for reasons of substantial public interest" and, even then, only subject to proportionality assessment.[308]

Finally, individuals have numerous FIPs-based rights in the GDPR that apply to data processing in general and, thus, to algorithmic decision-making and profiling.[309] For example, an individual has a right to rectification, whereby he or she can request the correction of inaccurate personal data,

301. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 23; *see also* Kaminski, *supra* note 86, at 196–98 (discussing algorithmic accountability and the GDPR).

302. Council Regulation 2016/679, *supra* note 13, art. 22(2), at 46; Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 23, (interpreting Article 22 as a prohibition).

303. The GDPR's text does not prohibit automated decision-making about children, but Recital 71 suggests that it nonetheless does ("Such measure should not concern a child"). Council Regulation 2016/679, *supra* note 13, at 14.

304. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 28–29.

305. Council Regulation 2016/679, *supra* note 13, art. 9(2), at 38–39.

306. *Id.* art. 6(4)(c), at 37.

307. *Id*. art. 9(2)(a), at 38, art 22(3), at 46.

308. *Id*. art. 9(2)(g), at 38.

309. *See, e.g*., Lilian Edwards, *Data Protection: Enter the General Data Protection Regulation*, *in* LAW, POLICY AND THE INTERNET 77, 77–119 (Lilian Edwards ed., 2018); Bart van der Sloot & Frederik Zuiderveen Borgesius, The EU General Data Protection Regulation: A New Global Standard for Information Privacy 15–17 (Apr. 15, 2018) (unpublished manuscript) (available at https://bartvander sloot.com/onewebmedia/SSRN-id3162987.pdf).

including inaccurate inferences.[310] Individual rights also include, under some circumstances, the infamous right to erasure (or, the "right to be forgotten"),[311] the right to restriction of processing,[312] the right to object,[313] and in the case of information processed subject to consent, the right to under certain circumstances withdraw consent to processing.

The algorithmic due process literature, in fact, refers to a number of these rights as necessary aspects of an individual's opportunity to be heard.[314] These may not be rights to a hearing in a traditional sense, but they give individuals the ability to intervene in data processing—and not just solely automated processing—in ways familiar to those steeped in the algorithmic due process literature.

Once again these individual rights serve both dignitary and justificatory ends. They allow individuals to participate in the formation of their "data double," including through correction and sometimes deletion of information. In allowing this participation, these rights potentially create something like a dialogue between the individual and a company about the rationale behind algorithmic decisions and other forms of personal profiling and data processing. They serve instrumental purposes, particularly towards correcting factual and inferential errors. But they do not allow for the intervention of third-party experts or create a systemic governance regime.

3. Individual Algorithmic "Due Process" Under Article 22

The GDPR explicitly contains a version of algorithmic due process. This significant development has been thus far overshadowed by the debate about the so-called "right to explanation" of individual algorithmic decisions. The GDPR's Article 22 establishes not just the much-discussed "right to explanation" but less discussed due process-like rights to human intervention, to express an opinion, and to contest an algorithmic decision.[315] Elsewhere, I have referred to Article 22 as stronger, broader, and deeper than its predecessor in the Data Protection Directive, Article 15.[316]

310. Council Regulation 2016/679, *supra* note 13, art. 16, at 43 (establishing the rectification right); Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 17–18 (discussing profiling on output data versus input data).

311. Council Regulation 2016/679, *supra* note 13, art. 17, at 43–44.

312. *Id*. art. 18, at 44–45.

313. *Id*. art. 21, at 45–46.

314. *See, e.g*., Citron & Pasquale, *supra* note 4, at 28; Crawford & Schultz, *supra* note 5, at 126–27.

315. Council Regulation 2016/679, *supra* note 13, art. 22, at 46.

316. Kaminski, *supra* note 86, at 201.

First, in short, Article 22 does establish an individual "right to an explanation" of an algorithmic decision.[317] The text requires companies that deploy "solely" automated decision-making to adopt "suitable measures" to safeguard the rights of individuals.[318] Recital 71 explains that these measures include a right "to obtain an explanation of the decision reached after such assessment."[319] The apparent discrepancy between text and Recital spurred heated debate.[320] The Guidelines on algorithmic decision-making appear, however, to resolve this.[321] The Guidelines reason that the Recital's right to explanation of an individual decision stems from an individual's right to contest a decision or express that individual's view, both of which are established in the GDPR's text.[322]

Article 22 establishes not just a right to an explanation but an opportunity to be heard. The GDPR requires that companies that use "solely" automated decision-making institute (1) a right to obtain human intervention, (2) a way to express one's point of view, and (3) a way to contest a decision.[323] This is a version of algorithmic due process, combining notice with several forms of an opportunity to be heard.

How this will work in practice is an open question. The Article 22 rights around contestation appear, thus far, to be fairly weak. The GDPR does not provide for a neutral arbiter.[324] Applicable Guidelines suggest the right to contest may be an internal company process.[325] There are no guidelines as to what this process must entail. This raises the question of whether a company

317. *Id*. at 204.

318. Council Regulation 2016/679, *supra* note 13, at art. 22, at 46.

319. *Id*. at 14.

320. *See generally supra* note 257.

321. Kaminski, *supra* note 86, at 204 (citing Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 27)

322. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 27 ("The data subject will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis.").

323. Council Regulation 2016/679, *supra* note 13, art. 22(3), at 46. Note that these requirements may or may not apply to member state law exception. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 23–24.

324. *Compare* Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 32 ("[A suggested human intervention mechanism is] for example providing a link to an appeals process at the point the automated decision is delivered to the data subject, with agreed timescales for the review and a named contact point for any queries."), *with* Crawford & Schultz, *supra* note 5, at 127 (calling for a neutral arbiter).

325. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 32. This seems to resemble some features of copyright's notice-and-takedown process in the United States. *See* Margot E. Kaminski & Jennifer Urban, The Right to Contestation 24–25 (Aug. 14, 2019) (unpublished manuscript) (on file with author). *See generally* 17 U.S.C. § 512 (2018) (establishing "limitations on liability related to material online").

whose interests do not always align with its users' will be capable of providing adequate process and fair results.[326] There is room for substantially more policy development in fleshing out this contestation right.[327]

The GDPR also requires the right to "obtain human intervention" in an algorithmic decision.[328] There is no explanation of what this means nor how it is related to the rights to contest or to express an opinion. While the applicable Guidelines discuss the extent of human involvement necessary for decision-making to fall *outside* of Article 22,[329] they do not discuss what level of human involvement constitutes "human intervention" to satisfy Article 22. As Danielle Citron has pointed out, due to "automation bias," human intervention by itself may be inadequate for addressing concerns about algorithmic decision-making.[330] Humans are inclined to accept what algorithms tell them as true unless they are trained otherwise. And it is unclear whether human intervention will, in fact, serve the goals of algorithmic due process. There may be dignitary benefits to putting a human on the loop, but there are also potential costs and dangers, including a possible increase in systemic bias. Additionally, with a human in the loop, blame can be misdirected at the intervening human rather than properly directed at overall system design.[331]

Article 22, like its predecessor in the Data Protection Directive, contains significant potential loopholes.[332] Article 22 applies only to *solely*

326. This is similar to the conversation in the United States about collateral censorship. Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 298–304 (2011). And the conversation in the EU about the right to be forgotten. *See* ANGELOPOULOS ET AL., *supra* note 252.

327. *See* Kaminski & Urban, *supra* note 326.

328. Council Regulation 2016/679, *supra* note 13, art. 22(3), at 46. For discussions of the human in the loop, see *infra* note 331.

329. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 21 (discussing meaningful human involvement).

330. Citron, *supra* note 14, at 1271–72 (describing automation bias).

331. There are potential dangers to trying to solve the problems of automation by reinserting a human in or on the loop. Requiring a system to be built for human intervention or even for human oversight can affect its design, in ways that negatively impact other values, such as accuracy or even bias correction. *See* Zarsky, *supra* note 40. Putting a human in or on the loop can result in moral—or legal—blame being directed at that human, rather than focusing enforcement efforts more effectively on overall system design. Madeleine Clare Elish, *Moral Crumple Zones: Cautionary Tales in Human-Robot*, 5 ENGAGING SCI., TECH., & SOC'Y 40, 42 (2019). Meg Jones has argued that moving a human in or out of the loop can be "both ineffective . . . and dangerous." Jones, *supra* note 85, at 134. She calls instead for general automation design principles. *Id.* at 82–83. The GDPR's insertion of a human into the loop of automated decision-making is a significant policy move that should prompt far more discussion than it has.

332. Bygrave, *supra* note 14, at 21–22.

automated decision-making.[333] It requires the decision to have legal effects or similarly significant effects.[334] Trade secret exceptions apply both to the right to explanation and to the related notification and access rights (disclosure of "meaningful information about the logic involved"), discussed above.[335] However, the Guidelines have weighed in on the scope and strength of Article 22 and largely interpreted it to close or limit many of these loopholes.[336] Thus the GDPR establishes a version of algorithmic due process oriented towards individuals and creates both notice and an opportunity to be heard.

C. COLLABORATIVE GOVERNANCE

The GDPR, however, does not limit its governance of algorithmic decision-making to individual rights. The GDPR also uses collaborative governance.

The collaborative governance side of the GDPR has been overlooked, particularly in the context of algorithmic decision-making. To the extent that the literature on algorithmic accountability and the GDPR has looked past individual rights, its focus has been on particular requirements companies must meet, such as third-party audits and data protection impact assessments,[337] rather than on the fact that the GDPR as a whole uses collaborative governance to address algorithmic decision-making.[338]

333. Council Regulation 2016/679, *supra* note 13, art. 22(1), at 46; Edwards & Veale, *supra* note 15, at 45; *see also* Mendoza & Bygrave, *supra* note 14, at 83; Selbst & Powles, *supra* note 257, at 234–35; Wachter et al., *supra* note 257, at 88.

334. Council Regulation 2016/679, *supra* note 13, art. 22(1), at 46; *see also* Mendoza & Bygrave, *supra* note 14, at 83 (quoting Council Regulation 2016/679, *supra* note 13, art. 22, at 46); Edwards & Veale, *supra* note 15, at 46; Selbst & Powles, *supra* note 257, at 234–35; Wachter et al., *supra* note 257, at 88 n.66.

335. Council Regulation 2016/679, *supra* note 13, art. 13–15, at 40–43; *see also id.* at 12 (providing in Recital 63 that the right of access "should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property"); Wachter et al., *supra* note 257, at 79 n.13, 84, 89; *supra* Section III.B.1. *But see* Brkan, *supra* note 257, (manuscript at 21–23) (citing Council Directive 2016/943, art. 5(d), 2016 O.J. (L 157) 1, 11 (EU)) (calling trade secrets and other IP concerns a "paper tiger", and noting that EU law "provides for an exception that allows for suspension of a trade secret 'for the purpose of protecting a legitimate interest recognized by Union or national law'" (citation omitted)).

336. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 19–28.

337. Casey et al., *supra* note 257, at 170–84; Edwards & Veale, *supra* note 15, at 78 ("[T]he new Article 35 is compulsory . . . and its definitions of 'high risk' technologies are almost certain to capture many if not most ML systems.").

338. One source evaluates the GDPR's impact assessment requirement through the lens of regulatory theory, identifying it as "meta-regulation," a subcategory of collaborative governance. Reuben Binns, *Data Protection Impact Assessments: A Meta-regulatory Approach*, 7 INT'L DATA PRIVACY L. 22, 29–30 (2017).

Focusing on these individual requirements misses the forest for the trees. Characterizing the GDPR as a system of collaborative governance changes the conversation about accountability and transparency from a discussion of individual tools to a system-wide evaluation that reveals both strengths and weaknesses in the GDPR's design, including in how its system of collaborative governance interacts with its individual rights.

The GDPR contains not just individual rights but duties imposed on companies. These duties are often voiced in broad terms that will be given meaning and effect over time through a variety of collaborative governance mechanisms—both formal, such as codes of conduct, and informal, such as standards coupled with guidance coupled with interpretation by internal company data protection officers ("DPOs") in conversation with regulators.[339]

I break down the toolkit of the GDPR into such formal and informal collaborative modes below.[340] The formally coregulatory aspects of the GDPR might never be realized; codes of conduct have, thus far, rarely been used in the EU.[341] I argue that the informally collaborative nature of the GDPR is both more overlooked and more practically relevant. First, however, I begin with an overall assessment of the GDPR as a collaborative governance regime.

1. The GDPR as Collaborative Governance

Rather than asking whether individual or institutional accountability is more effective, we should be asking whether the GDPR works as collaborative governance. Does it contain the right balance between hard and soft law—between command-and-control mechanisms and responsive regulation?[342] Does it effectively delegate to the private sector while bounding the level of deregulation a private company can achieve? Does it create a deep enough system of structured accountability to make its governance legitimate, both in the sense of eventually producing good

339. This feature of regulatory design is likely deliberate. *See First Report on the Implementation of the Data Protection Directive (95/46/EC)*, *supra* note 258, at 26.

340. *See infra* Sections III.C.2–.3.

341. *See, e.g.*, Dennis D. Hirsch, *In Search of the Holy Grail: Achieving Global Privacy Rules Through Sector-Based Codes of Conduct*, 74 OHIO ST. L.J. 1029, 1057 (2013); McGeveran, *supra* note 136, at 961 n.2 (noting codes of conduct are rarely used in Ireland). The European Commission noted as early as 2003 its frustration with the failure of codes of conduct. *First Report on the Implementation of the Data Protection Directive (95/46/EC)*, *supra* note 258, at 26 ("The Commission is disappointed that so few organisations have come forward with sectoral Codes of Conduct . . . .").

342. *See* AYRES & BRAITHWAITE, *supra* note 12, at 101–33; *see also* McGeveran, *supra* note 136, at 979–82.

substantive law and in the sense of appearing legitimate as a system?

The GDPR is, along certain dimensions, fundamentally a hard law regime. Law can be harder or softer along a spectrum that runs across multiple dimensions: how precise a rule is versus how vague; how obligatory a rule is versus how optional or advisory; and what enforcement mechanisms exist.[343] While the GDPR contains a number of what I would argue are deliberately vague rules, it is for the most part hard law: an obligatory regime, coupled with strong enforcement powers. A GDPR violation can famously trigger administrative fines of up to 4 percent of worldwide revenue.[344] The GDPR consolidates and further empowers an extensive national and transnational system of government regulators.[345] It creates broader enforcement powers for individuals—for example, allowing individuals to authorize nonprofits to lodge complaints on their behalf.[346] And the GDPR is backed by an increasingly involved court, the European Court of Justice ("ECJ"), which has in recent years repeatedly ruled in favor of data protection rights.[347] In other words, the GDPR has potentially very serious teeth. There are, thus, real incentives for companies to participate in collaborative efforts to flesh out broader rules and to voluntarily comply with the outcomes of those efforts to avoid government sanctions.

The GDPR is also hardish law when it comes to its formal requirements, even where its substance is vague on its face. The vagueness of the GDPR is still highly bounded.[348] For example, the text of Article 22 does not conclusively define what "suitable measures" are,[349] but between the text,

343. See, for example, the three dimensions of hard and soft law defined by Kenneth Abbott and others, Kenneth W. Abbott et al., *The Concept of Legalization*, 54 INT'L ORG. 401, 401 (2000); see also Gunther F. Handl et al., *A Hard Look at Soft Law*, 82 AM. SOC'Y INT'L L. PROC. 371, 374–75 (1988) (discussing the dimensions of content, signals of authority, and communications of intent to make the law effective); Kal Raustiala & Anne-Marie Slaughter, *International Law, International Relations and Compliance*, *in* HANDBOOK OF INTERNATIONAL RELATIONS 538, 552 (Walter Carlsnaes et al. eds., Sage Publ'ns 2002) (considering "the form of the agreement; the substance of the agreement . . . ; and the structure for review of performance");

344. Council Regulation 2016/679, *supra* note 13, art. 83(5), at 83.

345. *Id.* art. arts 51–59, at 65–70.

346. *Id.* art. 80, at 81.

347. Case C-131/12, Google Spain SL v. Agencia Española de Protección de Datos (May 13, 2014), http://curia.europa.eu/juris/document/document.jsf?text=&docid=152065&doclang=EN; Joined Cases C-293/12 & C-594/12, Digital Rights Ir. Ltd v. Minister for Commc'ns, Marine and Nat. Res. (Apr. 8, 2014), http://curia.europa.eu/juris/document/document.jsf?docid=150642&doclang=EN; *see also* European Court of Human Rights, Zakharov v. Russia, App. No. 47143/06 (Dec. 4, 2015), http://hudoc.echr.coe.int/eng?i=001-159324. While the GDPR will primarily be interpreted by the ECJ, European Court of Human Rights case law forms a backstop of human rights protection in the EU.

348. *See* Kaminski, s*upra* note 152, at 946 (discussing bounded versus unbounded uncertainty in collaborative governance).

349. *See* Council Regulation 2016/679, *supra* note 13, art. 22, at 46.

the Recitals, and the Guidelines, the GDPR qua regime provides a slew of specific requirements, examples, and outer limits.

This is true of nearly every GDPR requirement. On the face of the text, the GDPR is vague; coupled with its interpretative documents or earlier law and practices arising out of the predecessor Data Protection Directive, there is only so much room for companies to maneuver. On the one hand, this potentially constricts collaborative governance, limiting the types of solutions the private sector might offer or discouraging the private sector from collaborating when it cannot change the rules. On the other hand, it bounds private-sector lawmaking and creates a substantive backstop to private-sector negotiations, in the name of protecting fundamental rights.

How the softer and harder aspects of the GDPR play out and interact in practice will do a great deal to determine whether its attempts at collaborative governance are effective. Authorities will need to show enough strength to incentivize companies to meaningfully participate but enough gentleness to discourage adversarial posturing.[350] The GDPR risks being too hard in some places and too soft in others, effectively encouraging companies to route around harder law to seek out less regulated spaces.[351]

There is potential for similar arbitrage with respect to member state variations. While the GDPR purports to harmonize EU data protection law, it contains a number of opportunities for member states to vary their legal systems, including in governing algorithmic accountability.[352] There are both costs and benefits to state-by-state variation. The potential benefit is that variation could allow for policy experimentation, as federalism does in the United States. But state-by-state variation can also increase compliance costs and incentivize companies to target their resources at lobbying individual Member State legislatures, rather than collaborating with data

350. It is possible that regulators' opening gambits in negotiations with private companies are too weak/conciliatory already, in an effort not to be seen as business killing. *See, e.g.*, Lobel, *The Renew Deal*, *supra* note 7, at 451 (discussing "positive slippage," with standards as opening salvo).

351. For example, Facebook's counsel discussed avoiding using consent for compliance and instead taking the less onerous business interests route. Caroline Spiezio, *In-House Leaders from Facebook, Uber and Others Discuss the Complexity of Consent in GDPR*, LAW.COM: CORPORATE COUNSEL (Apr. 20, 2018, 3:16 PM), https://www.law.com/corpcounsel/2018/04/20/in-house-leaders-from-facebook-uber-and-others-discuss-the-complexity-of-consent-in-gdpr/?slreturn=20180326083826.

352. Council Regulation 2016/679, *supra* note 13, art. 22(2)(b), at 46 (discussing that a person's right not to be subject to solely automated decisions does not apply when the decision "is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests"); *see also* Gianclaudio Malgieri, *Automated Decisionmaking in the EU Member States: The Right to Explanation and Other "Suitable Measures" in the National Legislations*, COMPUTER L. & SECURITY REV., July 19, 2019, at 1, 5–18.

protection authorities. From a managerial perspective, too, a more harmonized regime can result in stronger centralized compliance culture in a company, while a less harmonized regime may lead to a more fractured, less effective internal compliance structure.

The GDPR heavily relies on government regulators—perhaps too heavily, for a purportedly collaborative regime. Data protection authorities and the European Data Protection Board are envisioned as active, independent regulators, responsible for a wide variety of tasks.[353] If the GDPR's regulators are too command-and-control minded, they may override the collaborative nature of the system and eliminate envisioned benefits from private sector involvement. If, on the other hand, they are not strong or involved enough, then the GDPR has numerous potential weaknesses that companies can exploit to effectively deregulate. In practice, in the recent past, data protection authorities have faced limited resources.[354]

While the GDPR heavily relies on traditional government regulators, it minimally invokes participation by third parties. This is perhaps the weakest point in the entire system, and one I discuss at greater length below.[355] The GDPR for the most part envisions collaboration as taking place between regulators and regulated private parties, to the exclusion of third parties, such as civil society or external experts. This threatens both the substance and legitimacy of the regime. To some extent, this design flaw may reflect the relative weakness of civil society in the EU.

Before I return to this central problem of structured accountability, however, I outline the details of the GDPR as a collaborative system. The GDPR consists of both formal and informal collaborative mechanisms that together create the outlines of an extensive collaborative governance regime.

2. Formal Coregulation

The GDPR contains a number of formal mechanisms for policy collaboration between companies and regulators, including codes of conduct

353. Council Regulation 2016/679, *supra* note 13, at 22, art. 52, at 66, art. 69, at 76.

354. *See, e.g.*, J. TREVOR HUGHES, INT'L ASS'N OF PRIVACY PROF'LS, DATA PROTECTION AUTHORITIES 10 (2011), https://iapp.org/media/pdf/knowledge_center/DPA11_Survey_final.pdf (stating the average European budget is under €5 million); *see also* DAVID BARNARD-WILLIS & DAVID WRIGHT, TRILATERAL RESEARCH & CONSULTING, U.K., CO-ORDINATION AND CO-OPERATION BETWEEN DATA PROTECTION AUTHORITIES 9, 143 (2014), http://www.phaedra-project.eu/wp-content/uploads/PHAEDRA-D1-30-Dec-2014.pdf ("The consortium recognises that many [data protection authorities] face constraints, by way of human and/or budgetary shortages, institutional and legislative rules and other factors.") (citing limited budget or human resources as constraining cooperation).

355. *See infra* Section III.C.4.

and certification.[356] Both of these formal coregulatory mechanisms are cited in the Guidelines on algorithmic decision-making.[357] And even though the incorporation of technical standards is not yet explicitly part of the GDPR's approach to algorithmic decision-making, I discuss it briefly here.[358]

First, the GDPR might formally coregulate algorithmic decision-making through codes of conduct.[359] Industry groups are encouraged to prepare codes of conduct to clarify the application of the GDPR in sector-specific or even technology-specific areas.[360] Codes of conduct act as safe harbors from the GDPR: once a code has been approved by the relevant government authority, a company that follows it can be assured it will not be held liable.[361] Some codes may even be eventually implemented as EU-wide law.[362] Thus companies could come together to create a code of conduct for auditing machine-learning algorithms[363] or a code of conduct outlining other suitable measures to be applied to prevent algorithmic bias or discrimination or privacy violations.

Certification is a softer coregulatory mechanism.[364] It does not create a safe harbor from GDPR enforcement[365] but instead seeks to use market measures to incentivize industry participation. It works as follows: groups of companies would create certification standards and consumers would seek out those companies that are certified, like purchasers who search for goods or companies certified by the Better Business Bureau or Certified Humane.[366] While it does not create a safe harbor, certification potentially reduces enforcement risks for companies. Supervisory authorities or the

356. Council Regulation 2016/679, *supra* note 13, arts. 40–42, at 56–59.

357. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 30.

358. *Id*. at 32 (referring to "agreed standards," but leaving unclear whether this refers formally to technical standards).

359. Council Regulation 2016/679, *supra* note 13, arts. 40–41, at 56–58.

360. *Id*. art. 40(2), at 56–57.

361. *Id*. art. 40(5), at 57 ("The supervisory authority shall provide an opinion on whether the draft code, amendment or extension complies with this Regulation and shall approve that draft code, amendment or extension if it finds that it provides sufficient appropriate safeguards.") (describing how the Board shall issue an opinion if the activity applies in more than one Member State).

362. *Id*. art. 40(9), at 57 ("The Commission may, by way of implementing acts, decide that the approved code of conduct, amendment or extension submitted to it pursuant to paragraph 8 of this Article have general validity within the Union.").

363. *See* Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 32.

364. *See* Council Regulation 2016/679, *supra* note 13, art. 42, at 58–59.

365. *Id*. art. 42(4), at 59 ("A certification pursuant to [Article 42] does not reduce the responsibility of the controller or the processor for compliance with this Regulation and is without prejudice to the tasks and powers of the supervisory authorities . . . .").

366. Lilian Edwards & Michael Veale, *Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?*, 16 IEEE SECURITY & PRIVACY, May-June 2018, at 46, 52 ("There is also little proof users regard seals and certificates as indicators of trust . . . .").

Board are involved in the creation of certification criteria.[367] Since those same authorities are responsible for GDPR enforcement, they are unlikely to pursue a company that is compliant with certification standards.

A third formal coregulatory mechanism involves the development of technical standards and the incorporation of them by reference into the regulation.[368] There is in fact a formal process in the EU whereby the European Commission can issue a request to the European Standardization Organizations to establish "an agreed way of meeting legal requirements," through technical standards.[369] There is evidence that this process will be used for a number of GDPR requirements.[370] Several articles of the GDPR explicitly reference technical standards,[371] and other provisions leave space for the process.[372]

With respect to algorithmic decision-making, there are several possible hooks for the incorporation of technical standards. The Guidance on algorithmic decision-making references the use of "agreed standards,"[373] and the Recital language on profiling suggests that companies use "technical and organisational measures" to prevent inaccuracies and error.[374]

3. Informal Collaborative Governance

In addition to these formal coregulatory mechanisms, the GDPR contains a number of informal tools that appear to be central to its system of governing algorithmic decision-making. These include broad legal standards

367. Council Regulation 2016/679, *supra* note 13, art. 42(5), at 59.

368. Kamara, *supra* note 259, at 15; *see also* Bremer, *supra* note 164, at 147–50.

369. Kamara, *supra* note 259, at 7 (citation omitted).

370. *Id.* at 14 (discussing standards on "how to address and manage privacy and personal data protection issues during the design and development and the production and service provision processes of security technologies" and standards that "specify[] the privacy and personal data protection management processes with an explanation how [sic] to realise them"). In 2015 the European Commission issued a request for the development of standards addressing privacy by design under the Data Protection Directive; this process is ongoing.

371. Council Regulation 2016/679, *supra* note 13, art. 21(5), at 46 ("[T]he data subject may exercise his or her right to object by automated means using technical specifications.") (discussing the right to object and certification); *id.* art. 43(9), at 60 ("The Commission may adopt implementing acts laying down technical standards for certification mechanisms . . . .").

372. Kamara, *supra* note 259, at 8 ("[S]everal provisions of the GDPR could be the basis for development of technical standards in the field. One prominent example is the provisions that establish *technology design obligations*, such data protection [sic] by design and by default . . . .").

373. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 32 (discussing "agreed standards").

374. Council Regulation 2016/679, *supra* note 13, at 14 ("[T]he controller should use appropriate mathematical or statistical procedures for the profiling [and] implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimized . . . .").

interpreted over time through softer regulatory guidance and private sector efforts; management reform, such as the requirement that companies appoint Data Protection Officers ("DPOs") and abide by reporting requirements; and requirements of third-party oversight, such as audits or expert boards.

In its governance of algorithmic decision-making, as elsewhere,[375] the GDPR uses broad legal standards to articulate company duties and fills them in through collaborative mechanisms. One such standard is Article 22's "suitable measures" requirement. Article 22 tasks companies that conduct algorithmic decision-making with implementing "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests . . . ."[376] This is a broad, technology-neutral standard.[377] The GDPR itself does not indicate how a company operating in a specific sector or using a specific technology might comply. The text of the GDPR provides some specific requirements, discussed above. But this list is not exhaustive, as the debate over the right to explanation makes clear.

The Guidelines, as part of the collaborative governance ecosystem, begin to fill out what this broad legal standard will look like in practice. Over time, enforcement by data protection authorities or even court decisions may create more specific standards or even specific rules. Until then, "suitable measures" will largely be constituted in part by hard law, in part by guidance, and in part by internal company practices.

The GDPR's principle of "fairness" is another example. The GDPR establishes the broad principle of "fairness" in its hard-law text. Accompanying softer law documents interpret the broad requirement of "fairness" to include at least accuracy and non-discrimination; this interpretation applies to algorithmic decision-making.[378] The Guidelines several times refer to fairness, non-discrimination, and accuracy in the same breath.[379]

375. *See e.g.*, Lee, *supra* note 259, 1027–29.

376. Council Regulation 2016/679, *supra* note 13, art. 22(3), at 46.

377. Kamara, *supra* note 259. at 10–11 (discussing "technology neutrality").

378. Council Regulation 2016/679, *supra* note 13, at 14, art. 5(1)(a), at 35 ("In order to ensure fair and transparent processing . . . , [companies must] ensure . . . that factors which result in inaccuracies in personal data are corrected and the risk of errors is [sic] minimized . . . .") (discussing the application of principles of lawfulness, fairness, and transparency to personal data processing). To meet the GDPR's broad requirement of fairness, companies must "prevent[] . . . 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 . . . ." *Id.* at 14. The GDPR Guidelines echo this interpretation. The Guidelines refer to inaccuracy, discrimination, and the perpetuation of existing stereotypes as harms caused by algorithmic decision-making. *See* Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 5–6.

379. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 14 (discussing

These softer law documents explicitly turn to the private sector to determine how these goals will be met in practice. Recital 71 calls for algorithmic error and discrimination to be addressed through "appropriate mathematical or statistical procedures" and the implementation of "technical and organisational measures," the precise nature of which will be determined by the private sector or perhaps through technical standards setting, discussed above.[380] The Guidelines call for companies to address these goals through a robust combination of management reform,[381] technical measures that include regular testing,[382] third-party auditing,[383] and expert review boards.[384]

Thus the broad standards established in the GDPR's text feed into an elaborate system of collaborate governance. As part of this system, the GDPR repeatedly attempts to institute internal management reform. It requires, for example, certain companies to hire an internal but independent DPO[385] tasked with monitoring compliance with the GDPR.[386] Most companies using algorithmic decision-making will be subject to this requirement.[387]

The GDPR additionally attempts to influence management reform through recording requirements, impact assessments for some kinds of processing (including most algorithmic decision-making), and the suggestion that companies perform regular quality checks and internal auditing. The GDPR generally requires companies (with some exceptions

"safeguards aimed at ensuring fairness, non-discrimination and accuracy in the profiling process"). *See also id.* at 16 ("Controllers should introduce appropriate procedures and measures to prevent errors, inaccuracies or discrimination on the basis of special category data." (footnote omitted)).

380. Council Regulation 2016/679, *supra* note 13, at 14.

381. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 29–31 (discussing data protection impact assessments ("DPIAs")).

382. *Id*.

383. *Id*. at 30.

384. *Id*.

385. Council Regulation 2016/679, *supra* note 13, art. 38, at 55–56.

386. *Id*. art. 39(1)(b), at 56.

387. Companies who use algorithmic decision-making include those companies whose core activities involve processing especially sensitive personal data (such as racial or ethnic origin, biometric data, and health data) and those companies whose business model involves systemic large-scale monitoring. *See id*. art. 9, at 38–39, art. 37(1)(b)–(c), at 55. The DPO is described as being "a person with expert knowledge of data protection law and practices . . . ." *Id.* at 18. And the DPO is responsible for training staff about their GDPR responsibilities. *Id.* art. 39(1)(b), at 56.

for small companies)[388] to keep records of data processing.[389] These reports are not necessarily geared towards creating detailed external oversight, though they are accessible to regulators upon inspection.[390] Recording requirements may by themselves instigate reform within a company, as technical experts are required to assess and describe their systems for these records.[391]

In addition to keeping records for outside inspection, most companies deploying algorithmic decision-making will have to create impact assessments.[392] Impact assessments are mandatory under the GDPR in some circumstances, including in algorithmic decision-making.[393] The Guidelines envision impact assessments as an iterative, ongoing process that includes documentation, monitoring, and review.[394] This is largely an internal

388. *Id*, art. 30(5), at 51. Article 30 explains that certain record keeping obligations

> shall not apply to an enterprise or an organisation employing fewer than 250 persons unless the processing it carries out is likely to result in a risk to the rights and freedoms of data subjects, the processing is not occasional, or the processing includes special categories of data as referred to in Article 9(1) or personal data relating to criminal convictions and offences referred to in Article 10.

Id.

389. *Id*. art. 30, at 50–51. These records must include the purposes of processing, a description of the categories of individuals and data, and a description of where the data goes, among other things. *Id.*

390. These records might not be detailed; for example, cybersecurity records need only give "a general description" of measures employed. *Id.* art. 30(1)(g), at 51.

391. *See also* Selbst & Barocas, *Intuitive Appeal*, *supra* note 4, at 1129 (suggesting recording requirements as reform).

392. Edwards & Veale, *supra* note 15, at 77–80.

393. The GDPR mandates internal impact assessments when a company engages in "high risk" processing, which includes personal evaluations "based on automated processing" and with significant effects. Council Regulation 2016/679, *supra* note 13, art. 35, at 53–54 (stating that a DPIA is required only if processing is "likely to result in a high risk to the rights and freedoms of natural persons"). The Working Party Guidelines read this as including the solely automated processing covered by Article 22, in addition to algorithms that involve more substantial human oversight. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 29–31. Impact assessments are also required when there is "systematic monitoring of a publicly accessible area on a large scale." Council Regulation 2016/679, *supra* note 13, art. 35(3)(c), at 53. This could reach a good amount of sensor processing (say processing information gathered by smart cars or drones) regardless of whether it involves a decision with significant effects on a person or is solely automated. *Id.* at 18–19, art. 35(3)(a), at 53. In fact, Recital 91 provides the following:

> A data protection impact assessment should also be made where personal data are processed for taking decisions regarding specific natural persons following any systematic and extensive evaluation of personal aspects relating to natural persons based on profiling those data or following the processing of special categories of personal data, biometric data, or data on criminal convictions and offences or related security measures.

Id. at 18–19.

394. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 29-30; Article 29 Data Prot. 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*, WP248

process. Only "high risk" assessments require consultation with the government.[395] For the rest, if a company has a DPO, the impact assessment must involve the DPO.[396] While this is not equivalent to direct government involvement, it may put in place both internal oversight and internal compliance efforts.

For high-risk activity, the GDPR's impact assessment process could be characterized as a soft version of premarket approval: requiring a company to be in conversation with the government and to adjust its risk-management process before releasing automated decision-making on the public.[397] Even for non-high-risk impact assessments, the government still plays a role in ensuring accountability because impact assessments are subject to retention and updating requirements and potentially to government disclosure.[398]

Impact assessments may play a potential role in rule setting, as well. They do not create an industry-wide standard. Over time, however, impact assessments may end up affecting general compliance standards, as the government repeatedly assesses individual use cases.[399]

---

rev.01, at 16 (Apr. 4, 2017) [hereinafter Article 29 Data Prot. Working Party, *DPIA*] (providing a chart of impact assessment processes).

395. Council Regulation 2016/679, *supra* note 13, art. 36, at 54–55 ("The controller shall consult the supervisory authority prior to processing where a data protection impact assessment under Article 35 indicates that the processing would result in a high risk in the absence of measures taken by the controller to mitigate the risk."). Unfortunately the guidelines give little explanation of what activity will require prior government consultation. Instead, the Guidelines provide the following:

> It is in cases where the identified risks cannot be sufficiently addressed by the data controller (i.e. the residual risks remains high) then the data controller must consult the supervisory authority. An example of an unacceptable high residual risk includes instances where the data subjects may encounter significant, or even irreversible, consequences, which they may not overcome . . . and/or when it seems obvious that the risk will occur . . . . Whenever the data controller cannot find sufficient measures . . . (i.e. the residual risks are still high), consultation with the supervisory authority is required.

Article 29 Data Prot. Working Party, *DPIA*, *supra* note 394, at 18–19 (footnote omitted); *see also* Council Regulation 2016/679, *supra* note 13, at 18.

396. Companies whose core activities involve processing especially sensitive personal data (such as racial or ethnic origin data, biometric data, and health data) or which "by virtue of their nature, their scope and/or their purposes, require regular and systematic monitoring of [individuals] on a large scale" must appoint a DPO. *See* Council Regulation 2016/679, *supra* note 13, art. 9, at 38–39, art. 37(1)(b)–(c), at 55. If a company has a DPO, then it must consult with that officer when it conducts impact assessments. *Id.* art. 35(2), at 53.

397. Price, *supra* note 8, at 43, at 449–51 (describing the FDA Class III premarket approval for medical devices); Tutt, *supra* note 14, at 111 (calling for premarket approval of algorithms).

398. Council Regulation 2016/679, *supra* note 13, art. 36(3)(e), at 54–55; Article 29 Data Prot. Working Party, *DPIA*, *supra* note 394, at 18 ("[R]egardless of whether or not consultation with the supervisory is required based on the level of residual risk then the obligations of retaining a record of the DPIA and updating the DPIA in due course remain.").

399. Antoni Roig, *Safeguards for the Right Not To Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)*, 8 EUR. J.L. & TECH., no. 3, 2017, at 1, 2 (describing DPIAs

Beyond appointing a DPO, meeting reporting requirements, and conducting impact assessments, the Guidelines suggest a wealth of additional internal company practices geared at instrumental goals. The Guidelines caution that as part of "suitable measures" to protect individuals' rights, companies should perform "regular quality assurance checks of their systems to make sure that individuals are being treated fairly and not discriminated against . . . ."[400] Companies should also perform "algorithmic auditing," regularly testing algorithms to ensure they are "not producing discriminatory, erroneous or unjustified results . . . ."[401]

Finally, at least in the context of algorithmic decision-making, the GDPR's softer law guidance envisions substantial third-party oversight. It is possible to interpret Recital 71 to require auditing.[402] The Guidelines more explicitly suggest deploying third-party audits and establishing ethical review boards.[403] These suggested accountability requirements, although not in the GDPR text itself, are likely to have a significant practical impact on the industry as it searches for guidance on how to comply with the GDPR.[404]

The GDPR's collaborative governance regime is aimed not just at protecting privacy or ensuring accountability but at what should now be familiar instrumental goals of preventing error, bias, and discrimination in algorithmic decision-making.[405] The GDPR's required "safeguards" that must be applied to automated decision-making are not just individual due process rights but an iterative system of management reform and third-party oversight.[406] Thus the text of the GDPR—both its formal coregulatory

as "data generator[] for policy purposes").

400. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 32.

401. *Id*.

402. Malgieri & Comandé, *supra* note 250, at 258 ("[W]e argue that Articles 13(2)(f), 14(2)(g) and 15(1)(h) state a duty to perform an *auditing* of decision-making algorithms . . . .").

403. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 32 (suggesting audits "where decision-making based on profiling has a high impact on individuals"). For third-party auditing, the Working Group envisions a deeper form of transparency, explaining that it will be good practice to "provide the auditor with all necessary information about how the algorithm or machine learning system works . . . ." *Id.* The Working Party additionally envisions harnessing companies themselves as oversight, suggesting that companies contractually require third parties to conduct auditing and testing and ensure compliance with "agreed standards." *Id.*

404. Casey et al., *supra* note 257, at 171–74.

405. Article 29 Data Prot. Working Party, *Decision-Making*, *supra* note 265, at 6 ("The GDPR introduces new provisions to address the risks arising from profiling and automated decision-making, notably, but not limited to, privacy."); *id.* at 10 ("Profiling may be unfair and create discrimination . . . ."); *id.* at 14 (referring to "the safeguards aimed at ensuring fairness, non-discrimination and accuracy in the profiling process").

406. *Id*. at 28 ("Controllers should introduce appropriate . . . measures to prevent errors, inaccuracies or discrimination on the basis of special category data. These measures should be used on a cyclical basis; not only at the design stage, but also continuously . . . . The outcome of such testing should

provisions and its less formal collaborative mechanisms—combined with the context of interpretative guidance together establishes a binary approach to regulating algorithmic decision-making under the GDPR.

4. The GDPR's Accountability Problem

It is important to understand the GDPR as a system of collaborative governance, because this reveals its potentially significant weakness. The GDPR, for all its coregulatory and collaborative measures, does not establish adequate public-facing or even expert-facing accountability. The lack of public transparency coupled with a lack of mechanisms for third-party involvement, both expert and stakeholder, threatens both the substantive output and procedural legitimacy of the GDPR as a collaborative governance regime. This Section points out the gaps in the GDPR's system of structured accountability, starting with several of its formal coregulatory mechanisms and then turning to the example of impact assessments.

The GDPR's process for establishing codes of conduct does not require public transparency or the involvement of third parties. The GDPR primarily envisions a back-and-forth between companies and government authorities.[407] (Recital 99 suggests that companies should consult stakeholders when drawing up codes of conduct but does not require it.)[408] Codes of conduct are to be published after they are completed.[409] The GDPR then envisions using third parties to help monitor compliance,[410] imagining that they will lodge complaints with a compliance body.[411] But these third parties have no information-finding powers over companies. It is unclear how they will be meaningfully capable of identifying violations of codes of conduct if they have no way to see what companies are doing internally.

The GDPR's certification process similarly lacks public accountability.[412] Again the process of creating certification standards does

feed back into the system design." (footnotes omitted)) (discussing a special category data).

407. Council Regulation 2016/679, *supra* note 13, art. 40(5), (7)–(8), at 57 (describing the processes in individual Member States, involving the Board and Commission in the case of several Member States).

408. *Id*. at 19 ("[Drafters of a code of conduct should] have regard to submissions received and views expressed in response to such consultations.").

409. *Id*. art. 40(6), at 57.

410. *Id*. art. 41(2)(c), at 58 ("[A compliance body must set up] established procedures and structures to handle complaints about infringements of the code or the manner in which the code has been, or is being, implemented by a controller or processor, and to make those procedures and structures transparent to data subjects and the public . . . .").

411. *Id*. art. 41, at 58.

412. Recital 100 seems to characterize certification itself as a form of transparency, as shorthand for info data subjects can use "allowing data subjects to quickly assess the level of data protection of relevant products and services." *Id.* at 19.

not appear open to the public or to third-party participation. Again, as with codes of conduct, the GDPR envisions that third parties will be able to complain about a lack of compliance with certification standards.[413] But they are not involved in setting standards nor in ongoing monitoring or oversight.

Impact assessments, too, only minimally involve third parties at the creation phase. Unlike with codes of conduct or certification, companies are required to consult with third parties in forming impact assessments. But this consultation is required only "[w]here appropriate" and with an eye to guarding commercial secrets.[414] The Guidelines further explain that third-party views can be sought in a variety of ways, including through studies or questionnaires or surveys, rather than giving third parties a seat at the table.[415] These modes of consultation do not necessarily bring meaningful external oversight into the process of creating the substance of an impact assessment.

The most striking gap in public and third-party accountability in the GDPR is its approach to releasing—or not releasing—algorithmic impact assessments. While the GDPR's impact assessments have been heralded as a model for algorithmic accountability,[416] the process does not in fact involve releasing information to the public. A company is merely encouraged, not required, to publicly release its impact assessments.[417] Even where publication is encouraged, the Guidelines envision only partial release or release of a summary.[418] This differs crucially from the model of impact assessments usually employed in collaborative governance literature.[419] Publicly disclosed impact assessments are used as a soft form of regulation

413. *Id*. art. 43(2)(d), at 60–61 (providing that to be accredited a certification body must "establish[] procedures and structures to handle complaints about infringements of the certification or the manner in which the certification has been, or is being, implemented by the controller or processor, and to make those procedures and structures transparent to data subjects and the public").

414. *Id*. art. 35(9), at 54.

415. Article 29 Data Prot. Working Party, *DPIA*, *supra* note 394, at 15. A company should document its reasons for not seeking third-party views or for making a decision that diverges from the outcome of its surveys

416. AI Now Inst., *supra* note 198.

417. Article 29 Data Prot. Working Party, *DPIA*, *supra* note 394, at 18 ("Publishing a DPIA is not a legal requirement of the GDPR, it is the controller´s decision to do so. However, controllers should consider publishing at least parts, such as a summary or a conclusion of their DPIA.").

418. *Id*. ("The published DPIA does not need to contain the whole assessment, especially when the DPIA could present specific information concerning security risks for the data controller or give away trade secrets or commercially sensitive information . . . [T]he published version could consist of just a summary of the DPIA's main findings . . . .").

419. Freeman, *supra* note 6, at 663 ("When publicly disclosed, the commitments in the FPA might serve, moreover, as the benchmark against which wholly independent third-party monitors could hold both the agency and the firm to account.").

to trigger market mechanisms and other forms of third-party oversight and feedback. An impact assessment, in other words, is supposed to be a tripartite conversation between a regulated entity, the regulator, and third parties such as impacted persons or civil society organizations.[420] In the GDPR, it is largely used internally or, at most, in conversation with regulators.

It is possible that despite the GDPR's lack of public transparency and input by third parties, impact assessments and maybe draft codes of conduct will be made available to the public through other means. Freedom of information law might be used to obtain public disclosure. These efforts will encounter various exceptions—not to the GDPR but to local freedom of information law—including for trade secrets. But by making the "how" of transparency in this space a pull mechanism (freedom of information law) rather than a push mechanism (required public release), the GDPR increases transparency's costs and lowers the likelihood that it will be achieved.

This, then, is the central question about the GDPR as a system of collaborative governance: Are the various forms of both government and third-party oversight outlined in the GDPR's text, Recitals, and Guidelines adequate to ensure high-quality collaborative governance in the absence of true public transparency?[421]

The answer will depend on a number of factors. It will depend on the government's independence and resources—how effective, in practice, data protection authorities will be. It will depend on how government authorities enforce the GDPR, including how they interpret individual disclosure requirements and whether they receive adequate legal training to push back on overclaims of corporate secrecy. It will depend on the ability of NGOs and other policy advocates to harness the GDPR's system of individual transparency rights coupled possibly with freedom of information laws to obtain both enforcement and transparency.[422] To accomplish meaningful oversight, both NGOs and the press will need to link individual disclosures into politically effective group narratives, revealing what is going on over an algorithmic decision-making system as a whole. This will be costly and time-consuming and will involve much coordination. Still, it may be possible over time.

The GDPR's effectiveness at collaborative governance will also depend

420. AYRES & BRAITHWAITE, *supra* note 12.

421. Thanks to Sarah Eskens for helping to formulate the question.

422. There is no class action mechanism in Europe, generally speaking, but the GDPR envisions forms of third-party representation that could do some of this work. Council Regulation 2016/679, *supra* note 13, art. 80, at 81.

on companies themselves. As industries come together to determine codes of conduct and certification criteria and relatedly the content of appropriate technological design, will they (voluntarily) engage external stakeholders, including members of impacted communities? Will they use these systems to try to constrain rogue bad actors within industry?[423]

Perhaps the GDPR's extensive individual transparency rights could go some way to providing access to third parties. Although the GDPR contains significant subject access rights,[424] it is not a general-purpose freedom of information law. The GDPR's transparency measures are derived from the FIPs and oriented towards individuals, not the public. But because the GDPR relies on collaborative governance, its attempts at individual transparency must serve a dual role. Absent policy changes, the GDPR's individual transparency provisions will need to serve both individual dignitary and justificatory ends and as a crucial element of structured accountability in its collaborative governance regime. It is not clear that, as currently interpreted to provide information useful to individuals but not to experts, they will be capable of this dual function.

There thus remain significant gaps in the structured accountability of the GDPR regime. Individual disclosure of an individually explanatory nature will likely fail to trigger market mechanisms or drive new policy efforts. It will also likely fail to incorporate external expertise, because information of an explanatory nature cannot be effectively evaluated by outside experts with a view to auditing a system. The GDPR's third-party audit mechanisms or expert boards (which are Guideline suggestions, not GDPR text) conversely allow for expert oversight but envision no way for the public to be alerted by regulators to problems with a particular algorithmic system once experts identify it. These gaps in the GDPR's system of structured accountability mean that the rules governing companies, implementation of these rules, and compliance with them will lack public and third-party involvement and oversight. This will hamper both expert input into the system and the substantive and procedural legitimacy of the regime.

423. *See* Thaw, *supra* note 152, at 371–74 (noting it can be in the self-interest of an industry to monitor risk across an industry as a whole to prevent political backlash when bad things happen because of the behavior of a rogue company).

424. Jef Ausloos & Pierre Dewitte, *Shattering One-Way Mirrors—Data Subject Access Rights in Practice*, 8 INT'L DATA PRIVACY L. 4, 28 (2019) (reviewing subject access rights in practice and arguing for their importance in providing checks and balances both individually and collectively).

D. INTERACTION BETWEEN THE TWO PRONGS

This brings us to potential interactions between the two prongs of binary governance in the GDPR. The GDPR, as discussed, is primarily hard law and is strongly individual rights-oriented. At the same time, the GDPR's collaborative face will play a significant role in constituting individual rights and relatedly determining what duties companies owe. Depending on how the GDPR is implemented in practice, these two prongs can strengthen or weaken each other and may in places inevitably conflict.

1. Where the Two Prongs Are Complementary

The GDPR's nature as hard law—backed by serious enforcement penalties and a rights-protective ECJ—will in some ways create a complementary relationship between individual rights enforcement and collaborative governance. As companies fear both lawsuits[425] and investigations sparked by individual complaints,[426] they may be more likely to come together with regulators to help to define what is and is not feasible in their particular sector.

For example, a company that fears being fined for failing to put in place suitable safeguards in automated decision-making is more likely to negotiate a code of conduct or put in place compliance procedures and infrastructure (for example, audits, DPOs, impact assessments, and so forth) that match the Guideline suggestions, in order to credibly argue that it is in compliance. Individual litigants enforcing their individual rights can, thus, serve as a penalty default, a credible threat that will drive companies to collaborate and negotiate. By participating in collaborative governance, companies can effectively help to create their own safe harbors from the GDPR's extensive regime of individual rights.

The GDPR's system of individual rights can relatedly complement its collaborative regime by spreading the cost of compliance from the government to individual actors and their delegated NGOs. Cost is a huge problem for the GDPR. Audits will be expensive, and someone will have to pay for them or determine a system of reputational benefits to incentivize algorithmic auditing.[427] Bringing technical expertise into the government will be expensive. Monitoring companies will be expensive. If the backstop of the GDPR's system of collaborative governance of algorithms is the government, the government must be well resourced—and it is far from clear

425. *See* Council Regulation 2016/679, *supra* note 13, art. 79, at 80.
426. *See id*. art 77, at 80.
427. Thanks to Natali Helberger for this point.

that it will be.[428] This means that individual rights bearers and their proxies may have to do a good deal of monitoring and even enforcement work—including by invoking judicial remedies when regulators fail to act.[429]

The GDPR's system of individual rights will in some cases make instrumental contributions that will also serve the goals of its collaborative regime. The individual right of correction, for example, may help make algorithmic decision-making systems less erroneous, by correcting incorrect facts and inferences. Making the right of correction an individual right puts correction in the hands of the least cost avoider—the people who best know what information is correct about themselves. Other individual rights—of access and notification, for example—may help address other systemic problems, such as discrimination, by revealing individual instances of unjustifiable decision-making.

Individual transparency rights can also be an important component of structured accountability in collaborative governance. This is important especially if regulators do not otherwise address the lack of third-party and public accountability mechanisms in the GDPR. If aggregated by third parties—the media or civil society—individual stories can trigger soft accountability mechanisms, like market responses or naming-and-shaming. The question is whether the individual transparency produced in the GDPR's individual rights vindication provisions will be adequate to serve those functions in its collaborative governance regime.

Similarly, the GDPR's individual due process rights can function as a component of structured accountability for purposes of collaborative governance. By allowing individuals to challenge individual algorithmic decisions, the GDPR potentially makes companies accountable to an external force. How useful this is will depend a great deal on how substantive Article 22's right to contest an algorithmic decision ends up being in practice.

The GDPR's collaborative regime may, conversely, strengthen its individual rights regime. First, collaborative approaches may lead to more effective systemic accountability, by imposing audits and third-party oversight rather than relying on individual challenges alone. This may, if it

428. Matt Reynolds, *Lords AI Report Warns of 'Big Five' Data Grabs and Ethical Failures*, WIRED (Apr. 16, 2018), https://www.wired.co.uk/article/house-of-lords-artificial-intelligence-report-ethics-monopolies (quoting Michael Veale: "With those bodies, you wonder if they're spreading too thinly.").

429. Council Regulation 2016/679, *supra* note 13, arts. 78–79, at 80 (providing individuals a right to lodge a complaint against a supervisory authority and a right to judicial remedies generally).

works, create better systemic protection of individuals' rights, without the cost of constant individual challenges.

Second, collaborative governance will in fact give substance to individual rights—and this may in places be a good thing. By using collaborative governance, the GDPR may end up creating more workable solutions, rather than solutions that look good on paper but do not function in reality. For example, the requirement that companies build products with individual rights in mind from the onset—data protection by design and by default—is textually broad or even without substance. Over time and collaboration, however, companies and standards bodies may in conversation with regulators be able to come up with workable and concrete requirements that will apply to protect everyone and not rely on individuals' capacity to withhold consent or raise challenges.

2. Where the Two Prongs Are in Tension

The GDPR's dual systems of individual rights and collaborative governance will also, however, run into tension with each other. Some of the same features that could be complementary may also end up creating conflict between the two systems. How or whether these tensions will be resolved is a story that will play out over time and court decisions.

Structurally the hard-law nature of the GDPR could create problems for its envisioned collaborative relationships. One aspect of the collaborative governance toolkit—known as "responsive regulation"—is that regulators be able to graduate their responses to violations in order to incentivize and encourage good faith behavior by companies.[430] Under the GDPR's new system of administrative fines, there may still be room for this kind of calibrated reaction.[431] But the rights provided to individuals—including the right to a judicial remedy against government actors for nonaction and the right to a judicial remedy against a company directly[432]—will create penalties even when regulators choose not to act and may push regulators towards implementing harsher penalties or towards more investigations to begin with.[433] While from an individual rights perspective this is a good

430. McGeveran, *supra* note 136, at 979–88.

431. *Id*. at 1019 (citing Council Regulation 2016/679, *supra* note 13, art. 83, at 82–83) (discussing GDPR graduated administrative fine system).

432. Council Regulation 2016/679, *supra* note 13, arts. 78–79, at 80.

433. *Id*. art. 78(2), at 80 (providing a right to judicial remedy "where the supervisory authority which is competent pursuant to Articles 55 and 56 does not handle a complaint or does not inform the data subject within three months on the progress or outcome of the complaint lodged pursuant to Article 77"). Thanks to Comandé for pointing this out.

thing, it may from a collaborative governance perspective make it less possible for regulators to create safe-harbor-like rewards for collaborating companies and thus lead to lower degrees of voluntary collaboration.

The GDPR's system of strong individual rights may, too, favor individuals at the cost of correcting systemic problems, running counter to the instrumental goals of its collaborative governance system. Allowing individual correction and deletion can lead to gaming, which can distort the data set and the algorithm.[434] Even when done in good faith, individual correction and deletion can make an overall decision-making system less accurate (if individuals take themselves out of the data pool), less fair (as it skews the data set), and even discriminatory (reflecting access to justice rather than fact). The same is true of the right to human intervention in the case of algorithmic decision-making—it may be the case that adding a human in the loop will respect individual dignity but could make accuracy of the overall system worse, thus negatively impacting other individuals subject to the algorithm.

The debate over the right to explanation reveals another tension: that the kind of transparency that is necessary for one part of the system (individual rights) may not be adequate for the other (collaborative governance). Under the GDPR, individual transparency is intended to be understandable to and actionable by individuals.[435] That makes it less useful for other kinds of oversight, including expert oversight. The kind of information that individuals need to make choices and invoke their respective rights is not the same as the kind of information that experts need to assess whether an algorithmic decision-making system is functioning correctly. But making individualized transparency more useful to experts may make it less useful to individuals.

Individual rights also run in tension with efforts to solve the accountability gap in the GDPR's collaborative regime, by releasing more information to the general public. If we attempt to solve the GDPR's accountability gaps by increasing public transparency, we risk exposing more individuals' information and threatening individual privacy. Even if we increase structured accountability by adding in layers of third-party oversight, that comes at the cost of distributing personal information to more parties.

Perhaps the biggest source of tension between the two systems,

434. Bambauer & Zarsky, *supra* note 118, at 25–28.

435. Council Regulation 2016/679, *supra* note 13, art. 12, at 39–40.

however, is the tension that arises over trying to constitute individual rights through a collaborative regime. The GDPR is aimed at protecting fundamental rights. Yet, its requirements are largely given substance in collaboration with private companies. The question in the EU is particularly loaded because of the fundamental rights backdrop to the GDPR regime: Is it appropriate to use collaborative governance to constitute fundamental individual rights?[436]

Private companies are not reliable rights guardians, and their interests often misalign with the interests of individuals. The privatization of fundamental rights protection is prevalent throughout the GDPR. It is not just that companies are charged with protecting the rights of citizens.[437] It is that companies are charged with codetermining with the government the actual content of rights.[438] There is a substantial question of whether the ECJ, given its recent attention to data privacy, will find the GDPR's collaborative approach to be adequately protective of fundamental rights.[439] Alternatively, recent case law suggests the ECJ may end up functioning as an aspect of the collaborative regime, constraining companies' behavior but still buying in to company participation in determining how fundamental rights are implemented in specific contexts.[440]

CONCLUSION

Governing algorithmic decision-making is hard. The technology is complex and opaque and a fast-moving target. But in significant part, solving the governance problem is hard because we cannot agree on *why* to regulate. A growing literature now focuses on regulating algorithmic decision-making in order to solve problems such as error, bias, and discrimination, but ignores or brushes over legitimate dignitary and justificatory reasons for regulating.

436. ANGELOPOULOS ET AL., *supra* note 252, at 3–4. *But see* Claudia Quelle, Does the Risk-based Approach to Data Protection Conflict with the Protection of Fundamental Rights on a Conceptual Level? 3–4 (Feb. 3, 2013) (unpublished manuscript) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2726073) (arguing that there is no conflict).

437. Schwartz & Pfeifer, supra note 261, at 126 ("[T]hese rights have 'horizontal' effects; that is, these interests reach within 'private-on-private' relations as contrasted with merely 'vertical' applications that concern 'government-on-private' matters." (citing Case C-144/04, Mangold v. Helm, 2005 E.C.R. I-10013)).

438. *See* Lee, *supra* note 259, at 1066.

439. Thanks to Nico van Eijk for this point.

440. The ECJ itself appears to have not just endorsed but established collaborative governance over fundamental rights in its recent "right to be forgotten" decision in *Google Spain*, delegating implementation of the right to Google. Lee, *supra* note 259, at 1023–25. For the decision, see generally Case C-131/12, Google Spain SL v. Agencia Española de Protección de Datos (AEPD) (May 13, 2014), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62012CJ0131.

The law's role is not just instrumental; it legitimizes and delegitimizes, validates or invalidates other decisional systems, and protects individual rights even against private actors.

To serve all three goals, I propose a binary approach to algorithmic accountability that couples individual rights with systemic collaborative governance. The EU's GDPR, at least on paper, comes close to realizing such a regime. The GDPR reveals, however, that in building a dual system we must constantly evaluate the role of a particular tool in both systems.

The devil, as always, will be in the details: creating the right balance between hard law and soft, between flexibility and accountability, between bounded rights and room for private innovation. But binary governance is the scaffolding on which those details should be built. If we take only an individual rights approach, we risk failing to correct serious systemic problems with algorithmic decision-making. If we take only a systemic approach, we disregard real concerns about dignity and justification in such systems. The future of good algorithmic governance is a binary system of governance—one that may slide more towards one pole or the other, depending on the subsector-specific features or consequences of a particular type of decision—but one that addresses dignity and autonomy, systemic and individual legitimacy concerns, in addition to error and bias in algorithmic decision-making.