RETHINKING EXPLAINABLE MACHINES: THE GDPR'S "RIGHT TO EXPLANATION" DEBATE AND THE RISE OF ALGORITHMIC AUDITS IN ENTERPRISE

Bryan Casey,[†] *Ashkon Farhangi*[††] *& Roland Vogl*[†††]

ABSTRACT

The public debate surrounding the General Data Protection Regulation's (GDPR) "right to explanation" has sparked a global conversation of profound social and economic significance. But from a practical perspective, the debate's participants have gotten ahead of themselves. In their search for a revolutionary new data protection within the provisions of a single chapter of the GDPR, many prominent contributors to the debate have lost sight of the most revolutionary change ushered in by the Regulation: the sweeping new enforcement powers given to European data protection authorities (DPAs) by Chapters 6 and 8 of the Regulation. Unlike the 1995 Data Protection Directive that it replaced, the GDPR's potent new investigatory, advisory, corrective, and punitive powers granted by Chapters 6 and 8 render DPAs de facto interpretive authorities of the Regulation's controversial "right to explanation." Now that the DPAs responsible for enforcing the right have officially weighed in, this Article argues that at least one matter of fierce public debate can be laid to rest. The GDPR provides a muscular "right to explanation" with sweeping legal implications for the design, prototyping, field testing, and deployment of automated data processing systems. The protections enshrined within the right may not mandate transparency in the form of a complete individualized explanation. But a holistic understanding of the interpretation by DPAs reveals that the right's true power derives from its synergistic effects when combined with the algorithmic auditing and "data protection by design" methodologies codified by the Regulation's subsequent chapters. Accordingly, this Article predicts that algorithmic auditing and "data protection by design" practices will likely become the new gold standard for enterprises deploying machine learning systems both inside and outside of the European Union.

DOI: https://doi.org/10.15779/Z38M32N986

[†] Bryan Casey is a Lecturer in Law at Stanford Law School, Legal Fellow at the Center for Automotive Research at Stanford, and an affiliate scholar at the Stanford Machine Learning Group and CodeX: The Stanford Center for Legal Informatics.

[††] Ashkon Farhangi is a Product Manager at Google and a fellow at CodeX: The Stanford Center for Legal Informatics.

[†††] Roland Vogl is the Executive Director of CodeX: The Stanford Center for Legal Informatics and the Stanford Program in Law, Science and Technology and a Lecturer in Law at Stanford Law School.

TABLE OF CONTENTS

I. INTRODUCTION

The year is 1995 and a spate of pioneering companies, including the upstarts Amazon.com and eBay, are staking their financial futures on an emerging technology that appears poised to forever transform the computing and communications worlds.[1] The technology, known among its proselytizers as the "Net," represents a new form of digital infrastructure that facilitates the worldwide sharing of data and communications without regard for geographic location.[2] Though adoption rates of this mysterious new technology remain relatively low, European anxieties surrounding its increasingly widespread use are already in full swing—precipitating the passage of legislation known as the Data Protection Directive (DPD) designed to grapple with the societal and technical complexities of a world on the cusp of a new digital era.[3]

Fast forward twenty years to the present and the Internet is, decidedly, old hat. But a technology equally alluring to the "Net" circa 1995 is enjoying a period of similarly rapid ascendance. The technology is known as "machine learning"[4]—or, for those of a more poetic bent, "artificial intelligence."[5] The level of optimism surrounding its potential to transform the world by turning

[1] *See* Harry McCracken, *1995: The Year Everything Changed*, FAST COMPANY (Dec. 30, 2015), https://www.fastcompany.com/3053055/1995-the-year-everything-changed [https://perma.cc/976P-XBMP]. eBay launched under the name of AuctionWeb at the time. *See id.*

[2] *See id.*

[3] *See* Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281) 31 [hereinafter DPD]; Press Release, European Comm'n, IP/14/650, Commission Proposes a Comprehensive Reform of Data Protection Rules to Increase Users' Control of Their Data and to Cut Costs for Businesses (Jan. 25, 2012), http://europa.eu/rapid/press-release_IP-12-46_en.htm [https://perma.cc/T2X5-2526] [hereinafter GDPR Proposal].

[4] Machine learning can be described as a field of computer science that gives computers the ability to solve problems without being explicitly programmed to do so (i.e., the ability to "learn" by progressively improving performance on specific tasks). For references to definitions proffered by EU data authorities, see, e.g., DATATILSYNET, ARTIFICIAL INTELLIGENCE AND PRIVACY (Jan. 2018) [hereinafter ARTIFICIAL INTELLIGENCE AND PRIVACY]; NAT'L RESEARCH COUNCIL ET AL., FRONTIERS IN MASSIVE DATA ANALYSIS 101 (2013).

[5] While it is not wholly accurate to define "machine learning" and "artificial intelligence" as coextensive, for practical purposes this Article adopts to the convention of treating the two terms as synonymous. *See* ARTIFICIAL INTELLIGENCE AND PRIVACY, *supra* note 4, at 5 (defining artificial intelligence as "the concept used to describe computer systems that are able to learn from their own experiences and solve complex problems in different situations – abilities we previously thought were unique to mankind"). "Artificial intelligence is an umbrella term that embraces many different types of machine learning." *Id.* at 6.

machines into "intelligent"[6] decision-makers is matched only by the level of anxiety felt by those who fear the potential for bias to infiltrate machine decision-making systems once humans are removed from the equation.[7]

As recently as a decade ago, concerns surrounding bias within these types of complex automated systems would likely have struck many observers as far-fetched. Ever since the birth of computation with Alan Turing, humans have ascribed a kind of perfect "objectivity" to the mechanistic processes underlying algorithmic decision-making—a propensity now known as "automation bias."[8] Indeed, study after study has documented an innate human tendency to assume the validity of decisions made by algorithms,[9] even when presented with information that directly contradicts the decision's apparent validity.[10] The drafters of Europe's DPD explicitly acknowledged this phenomenon in 1992. They were so worried that "machine[s] using more and more sophisticated software" might be perceived as having "an apparently objective and incontrovertible character" that they felt it necessary to legislate specific

[6] The word intelligent, here, is used in quotes because of the fraught definitional issues associated with the term. As the scholar, Ryan Calo, notes, "Few complex technologies have a single, stable, uncontested definition [and] [r]obots are no exception." Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 529 (2015). For stylistic purposes, this Article uses "machine learning" and "artificial intelligence" interchangeably. Both terms lack a universally accepted definition, but this Article uses them to refers broadly to any "computerized system that exhibits behavior that is commonly thought of as requiring intelligence." EXEC. OFFICE OF THE PRESIDENT NAT'L SCI. & TECH. COUNCIL COMM. ON TECH., PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE 6 (2016).

[7] *See infra* Section IV.C and accompanying text.

[8] *See, e.g.*, A HISTORY OF ALGORITHMS: FROM THE PEBBLE TO THE MICROCHIP (Evelyn Barbin & Jean-Luc Chabert eds., 1999) [hereinafter A HISTORY OF ALGORITHMS]; Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1271–72 (2008); Kate Goddard et al., *Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators*, 19 J. AM. MED. INFORMATICS ASS'N 121 (2012); Christian Sandvig, *Seeing the Sort: The Aesthetic and Industrial Defence of "the Algorithm"*, 10 J. NEW MEDIA CAUCUS 1 (2014); Linda J. Skitka et al., *Accountability and Automation Bias*, 52 INT'L J. HUMAN COMPUTER STUD. 701, 704 (2000); Mary Cummings, *Automation Bias in Intelligent Time Critical Decision Support Systems*, AIAA 1ST INT. SYS. TECHNICAL CONF. (2004).

[9] An "algorithm" can be defined as "a formally specified sequence of logical operations that provides step-by-step instructions for computers to act on data and thus automate decisions." Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674 n.10 (2016) (quoting SOLON BAROCAS ET AL., DATA & CIVIL RIGHTS: TECHNOLOGY PRIMER (2014)); *see* A HISTORY OF ALGORITHMS, *supra* note 8, at 2 (defining "algorithm" even more broadly as "any process of systematic calculation, that is a process that could be carried out automatically").

[10] *See* Cummings, *supra* note 8; Kathleen Mosier et al., *Automation Bias: Decision Making and Performance in High-Tech Cockpits*, 8 INT'L J. AVIATION PSYCHOL. 47, 47 (1997); Goddard et al., *supra* note 8, at 121.

measures guarding against it.[11]

In recent years, however, society's deferential attitude toward algorithmic objectivity has begun to wane—thanks, in no small part, to a flurry of influential publications examining bias within complex computational systems.[12] Particularly in the last five years, numerous studies across multiple industry sectors and social domains have revealed the potential for algorithmic systems to produce disparate real world impacts on vulnerable groups.[13] These

[11] *See Amended Proposal for a Council Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data*, at 26, COM (1992) 422 final—SYN 297 (Oct. 15, 1992).

[12] *See, e.g.*, Bart Custers, *Data Dilemmas in the Information Society: Introduction and Overview*, *in* DISCRIMINATION AND PRIVACY IN THE INFORMATION SOCIETY 3, 20 (Bart Custers et al. eds., 2013); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 101 (2014) (noting "housing providers could design an algorithm to predict the [race, gender, or religion] of potential buyers or renters and advertise the properties only to those who [meet certain] profiles"); Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55, 57 (2013); Brent Mittelstadt et al., *The Ethics of Algorithms: Mapping the Debate*, 3 BIG DATA & SOC'Y 1, at 7–9 (2016); Latanya Sweeney, *Discrimination in Online Ad Delivery*, 11 ACM Queue 10, 12–13 (2013); Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75 (2015); Solon Barocas, *Data Mining and the Discourse on Discrimination*, (2014) (unpublished manuscript), https://dataethics.github.io/proceedings/DataMiningandtheDiscourseOnDiscrimination.pdf [https://perma.cc/LQ6R-FJZQ]; *see also, e.g.*, Citron, *supra* note 8, at 1254 ("Although programmers building automated systems may not intend to engage in rulemaking, they in fact do so . . . . The resulting distorted rules effectively constitute new policy that can affect large numbers of people."); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 4, 13–16 (2014) ("Because human beings program predictive algorithms, their biases and values are embedded into the software's instructions . . . ."); Devah Pager & Hana Shepherd, *The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets*, 34 ANN. REV. SOC. 181, 184 (2008); Michal S. Gal, *Algorithms as Illegal Agreements*, 34 BERKELEY TECH. L.J. 67 (2019); Julia Angwin et al., *Facebook (Still) Letting Housing Advertisers Exclude Users by Race*, PROPUBLICA (Nov. 21, 2017), https://www.propublica.org/article/facebook-advertising-discrimination-housing-racesex-national-origin [https://perma.cc/5B5W-WYEH]; Julia Angwin & Terry Parris Jr., *Facebook Lets Advertisers Exclude Users by Race*, PROPUBLICA (Oct. 28, 2016), https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race [https://perma.cc/4QDV-HC92].

[13] *See, e.g.*, Bryan Casey, *Title 2.0: Discrimination in a Data Driven Society*, 2019 J.L. & MOBILITY 36 (2019); Christine L. Borgman, *Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier*, 33 BERKELEY TECH. L.J. 365 (2018); Kevin Werbach, *Trust, but Verify: Why the Blockchain Needs the Law*, 33 BERKELEY TECH. L.J. 487 (2018); Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV. (Apr. 1, 2013), https://hbr.org/2013/04/the-hidden-biases-in-big-data [https://perma.cc/E95C-TUQU]; Alistair Croll, *Big Data Is Our Generation's Civil Rights Issue, and We Don't Know It*, SOLVE FOR INTERESTING (July 31, 2012), http://solveforinteresting.com/big-data-is-our-generations-civil-rights-issue-and-we-dont-know-it [https://perma.cc/K77Z-PK3L]; Moritz Hardt, *How Big Data Is Unfair*, MEDIUM (Sept. 26, 2014), https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de

revelations, in turn, have had a pronounced effect on scholars, policymakers, industry leaders, and society *writ large*—often serving as a rallying cry for greater efforts to promote fairness, accountability, and transparency in the design and deployment of highly automated systems.[14]

Yet, despite society's recent shift in attitude toward these types of algorithmic systems, the inexorable march of machine learning "eating the world" is only accelerating.[15] Across a diverse array of industries—from private social networks to public sector courtrooms[16]—organizations are adopting

[https://perma.cc/ZTZ4-8EG5]; Nadya Labi, *Misfortune Teller*, ATLANTIC (Jan./Feb. 2012), http://www.theatlantic.com/magazine/archive/2012/01/misfortune-teller/308846 [https://perma.cc/V3VV-84YU]; Anders Sandberg, *Asking the Right Questions: Big Data and Civil Rights*, PRAC. ETHICS (Aug. 16, 2012), http://blog.practicalethics.ox.ac.uk/2012/08/asking-the-right-questions-big-data-and-civil-rights [https://perma.cc/V86T-9S2P]; Tanzina Vega, *New Ways Marketers Are Manipulating Data to Influence You*, N.Y. TIMES: BITS (June 19, 2013), https://bits.blogs.nytimes.com/2013/06/19/new-ways-marketers-are-manipulating-data-to-influence-you/ [https://perma.cc/P89Y-2967].

[14] *See, e.g.*, DEP'T FOR DIGITAL, CULTURE, MEDIA & SPORT, DATA ETHICS FRAMEWORK (Aug. 30, 2018), https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework [https://perma.cc/FS48-CPA3]; HOUSE OF COMMONS, SCIENCE AND TECHNOLOGY COMMITTEE , ALGORITHMS IN DECISION-MAKING INQUIRY LAUNCHED , 2018, HC 351 (UK); SPECIAL EUROBAROMETER 431, DATA PROTECTION (June 2015); EUROPEAN DATA PROTECTION SUPERVISOR (EDPS), MEETING THE CHALLENGES OF BIG DATA: A CALL FOR TRANSPARENCY, USER CONTROL, DATA PROTECTION BY DESIGN AND ACCOUNTABILITY (2015); *Report with Recommendations to the Commission on Civil Law Rules on Robotics* 2015/2103(INL) (Jan. 27, 2017), http://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html [https://perma.cc/UE64-BJA5]; HOUSE OF COMMONS, SCIENCE AND TECHNOLOGY COMMITTEE, ROBOTICS AND ARTIFICIAL INTELLIGENCE, 2016, HC 145 (UK); INFORMATION COMMISSIONER'S OFFICE, BIG DATA, ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DATA PROTECTION (2017) (UK); *see also* INFORMATION COMM'R'S OFFICE, OVERVIEW OF THE GENERAL DATA PROTECTION REGULATION (GDPR) (2017) (UK) [hereinafter ICO'S OVERVIEW OF GDPR]; NAT'L SCI. & TECH. COUNCIL, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (2016); THE ROYAL SOCIETY, MACHINE LEARNING: THE POWER AND PROMISE OF COMPUTERS THAT LEARN BY EXAMPLE (2017); WETENSCHAPPELIJKE RAAD VOOR HET REGERINGSBELEID [DUTCH SCIENTIFIC COUNCIL FOR GOVERNMENT POLICY (WRR)], BIG DATA IN EEN VRIJE EN VEILIGE SAMENLEVING [BIG DATA IN A FREE AND SAFE SOCIETY], WRR-Rapport 95 (2016); Sophie Curtis, *Google Photos Labels Black People as 'Gorillas'*, TELEGRAPH (May 4, 2017), http://www.telegraph.co.uk/technology/google/11710136/Google-Photos-assigns-gorilla-tag-to-photos-of-black-people.html [https://perma.cc/25QY-TR9L].

[15] *See* Tom Simonite, *Nvidia CEO: Software Is Eating the World, but AI Is Going to Eat Software*, MIT TECH. REV. (May 12, 2017), https://www.technologyreview.com/s/607831/nvidia-ceo-software-is-eating-the-world-but-ai-is-going-to-eat-software/ [https://perma.cc/VT63-YSTL].

[16] *See, e.g.*, Corbett-Davies et al., Algorithmic Decision Making and the Cost of Fairness (June 10, 2017) (unpublished manuscript), https://arxiv.org/pdf/1701.08230.pdf [https://perma.cc/329E-WYRD]; Nikolaj Tollenaar et al., *StatRec —Performance, Validation and Preservability of a Static Risk Prediction Instrument*, 129 BULL. SOC. METHODOLOGY 25 (2016)

machine learning systems at unprecedented rates due to the technology's ability to radically improve data-driven decision-making at a cost and scale incomparable to that of humans.[17] Today, many agree that machine learning algorithms processing vast troves of data will only continue to play an increasingly large role in regulating our lives.[18] The question, thus, becomes: how are we to regulate these algorithms?

In 2016, the European Union sought to become a global pioneer in answering this question by replacing its 1990s-era DPD with comprehensive reform legislation known as the General Data Protection Regulation (GDPR).[19] The numerous protections introduced by the GDPR included an update to the DPD's rights surrounding automated decision-making.[20] The update formally enshrined what has since come to be referred to as the "right to explanation."[21] The right mandates that entities handling the personal data of EU citizens "ensure fair and transparent processing."[22] This requires providing citizens with access to "meaningful information about the logic involved" in certain automated decision-making systems.[23]

Many view the GDPR's "right to explanation" as a promising new mechanism for promoting fairness, accountability, and transparency in a world pervaded by complex algorithmic systems that can be difficult for observers to understand.[24] But as is true of numerous other rights enshrined within the GDPR, the precise contours of the "right to explanation" protections are less than clear—leading some commenters to wonder exactly how it will impact

(detailing published UK and Dutch predictive models involving recidivism).

[17] *See* Corbett-Davies et al., *supra* note 16.

[18] *See, e.g.*, Gideon Lewis-Kraus, *The Great A.I. Awakening*, N.Y. TIMES MAG. (Dec. 14, 2016), https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html [https://perma.cc/KG5C-NAD4].

[19] *See* Regulation (EU) 2016/679, of the European Parliament and the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter GDPR].

[20] *See id.*

[21] *See infra* Part II and accompanying notes.

[22] GDPR, *supra* note 19, at Recital 71.

[23] *Id.* at art. 15.

[24] *See infra* Section IV.C and accompanying notes; *see also, e.g.*, EXEC. OFF. OF THE PRESIDENT NAT'L SCI. & TECH. COUNCIL, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE (Oct. 2016); Catherine Stupp, *Commission to Open Probe into Tech Companies' Algorithms Next Year*, EURACTIV (Nov. 7, 2016), https://www.euractiv.com/section/digital/news/commission-to-open-probe-into-tech-companies-algorithms-next-year/ [https://perma.cc/B4TE-EHNQ]; GOV'T OFF. FOR SCI., ARTIFICIAL INTELLIGENCE: OPPORTUNITIES AND IMPLICATIONS FOR THE FUTURE OF DECISION MAKING (2016).

the use of machine learning in enterprise.[25]

In the two years since the GDPR's official publication, this uncertainty has ignited a heated global debate surrounding the Regulation's actual substantive protections.[26] The debate has centered on a cluster of four provisions found in Chapter 3 of the Regulation that circumscribe the specific text giving rise to the right. Scholars, industry leaders, and media sources across the globe have scoured the language of these provisions, proffering various competing interpretations of what the GDPR's new, and potentially revolutionary, "right to explanation" entails.[27] But lost in the debate's focus on the text of the provision has been a recognition of the more revolutionary change ushered in by the GDPR: the sweeping new enforcement powers given to Europe's data protection authorities.[28]

Unlike the DPD that it replaced, the GDPR grants EU data authorities vastly enhanced investigatory powers, a broad corrective "tool kit," and the capacity to levy fines several thousand times larger than the previous maximum limit.[29] Thanks to the GDPR's introduction of these truly threatening administrative powers, EU data authorities will no longer be rendered the toothless watchdogs many companies have long viewed them to be.[30] Rather, these newly empowered authorities will play a weighty role in enforcing and,

[25] *See infra* Part II.B and accompanying notes.

[26] *See infra* Part II and accompanying notes.

[27] *See infra* Part III and accompanying notes; *see also, e.g.*, FRANCESCA ROSSI, ARTIFICIAL INTELLIGENCE: POTENTIAL BENEFITS AND ETHICAL CONSIDERATIONS (2016). For media perspectives, see Cade Metz, *Artificial Intelligence Is Setting Up the Internet for a Huge Clash With Europe*, WIRED (July 11, 2016), https://www.wired.com/2016/07/artificial-intelligence-setting-internet-huge-clash-europe/ [https://perma.cc/4JSZ-THTR]; Bernard Marr, *New Report: Revealing The Secrets of AI or Killing Machine Learning?*, FORBES (Jan. 12, 2017), https://www.forbes.com/sites/bernardmarr/2017/01/12/new-report-revealing-the-secrets-of-ai-or-killing-machine-learning/#35a503e543ef [https://perma.cc/K8UQ-Q3GA]; Liisa Jaakonsaari, *Who Sets the Agenda on Algorithmic Accountability?*, EURACTIV (Oct. 29, 2016), https://www.euractiv.com/section/digital/opinion/who-sets-the-agenda-on-algorithmic-accountability/ [https://perma.cc/938H-4TPR]; Nick Wallace, *EU's Right to Explanation: A Harmful Restriction on Artificial Intelligence*, TECHZONE360 (Jan. 25, 2017), http://www.techzone360.com/topics/techzone/articles/2017/01/25/429101-eus-right-explanation-harmful-restriction-artificial-intelligence.htm [https://perma.cc/7XEA-B834].

[28] *See* GDPR, *supra* note 19, at chs. 6, 8.

[29] *See id.* The exact multiple can vary depending on the company's annual turnover. *See infra* Part III.

[30] *See* Natasha Lomas, *WTF Is GDPR*, TECHCRUNCH (Jan. 2018), https://techcrunch.com/2018/01/20/wtf-is-gdpr/ [https://perma.cc/G9FD-LRQV] (noting that the "beefing up of enforcement that's baked into the new regime means there's a better opportunity for DPAs to start to bark and bite like proper watchdogs"); *infra* Part III and accompanying notes.

therefore, *interpreting* the GDPR's numerous protective mandates.[31]

Viewed through this lens, it becomes apparent that many disagreements surrounding the "right to explanation" may have clearer answers than the current state of debate suggests. While vocal observers on both sides have dominated the headlines, those tasked with actually enforcing the "right to explanation" have quietly gone to work.[32] In the last six months, these authorities have produced a richly detailed framework for companies seeking to promote compliance with the GDPR's "right to explanation."[33] Given that these are the very same authorities on the front lines of enforcing compliance, their interpretations merit careful consideration.

Now that the dust from this recent burst of activity by data authorities has begun to settle, this Article attempts to take stock of the new developments—just in time for the Regulation's recent effectuation. In doing so, this Article seeks to turn the page within the GDPR's fraught "right to explanation" debate by answering a question that has, thus far, gone almost entirely overlooked: What do those actually tasked with enforcing the right think it entails?

Stepping outside of the debate's focus on the text of the GDPR, this Article adopts a holistic approach to understanding the Regulation's somewhat loosely-worded mandate. This Article contextualizes the "right to explanation" provisions by setting them against the backdrop of the potent range of new administrative capabilities prescribed by subsequent provisions. These new provisions effectively render Europe's data protection agencies de facto interpretive authorities.[34] In adopting this approach, this Article takes particular pains to let the words of the Regulation and its downstream interpreters speak for themselves—making use of direct quotes or passages whenever possible.[35]

Through the words of the authorities in charge of enforcing the GDPR, this Article finds a muscular "right to explanation" enshrined within the Regulation—albeit one that is subtly different from the competing visions contemplated by some scholars and industry experts. Europe's data protection authorities consistently reveal that they envisage the "right to explanation" not only as an individual remedial mechanism but also as part and parcel of a broader form of oversight with broad implications for the design and

[31] *See id.*

[32] *See infra* Part IV.

[33] *See infra* Part IV and accompanying notes.

[34] *See infra* Part III and accompanying notes.

[35] The hope, here, is to minimize editorializing—not to bore the reader with block quotes.

deployment of automated systems that process personal data.[36]

This Article seeks to better understand this newly articulated "right to explanation" and, in doing so, hopes to shed light on how enterprises can prepare for, react to, and promote compliance with what will doubtless be one of the most influential data protection frameworks of the coming decades. The Article proceeds in five parts. Part II traces the history of the public debate surrounding the "right to explanation." It begins with the right's origins in the specific text of Chapter 3 and proceeds to overview several of the most prominent contributions to the public debate thus far. In highlighting the debate's merits and demerits, it argues that the participants' general failure to countenance the substantive changes to enforcement introduced by Chapters 6 and 8 of the Regulation represents a fundamental oversight—one that has hindered a genuine understanding of the right's substantive protections.

Part III turns the page in the debate by broadening its focus to include Chapters 6 and 8 of the GDPR. It contextualizes the newfound role that enforcement agencies will play by detailing their limitations under the DPD and outlining their vastly enhanced administrative powers granted by Chapters 6 and 8. It argues that these newly empowered data watchdogs will serve as *functional* interpretive authorities of the GDPR's "right to explanation," even if other legislative or judicial authorities may, theoretically, have the final say. Because these agencies will be on the front lines of enforcement, their interpretations will, of necessity, be the most relevant for enterprises seeking to comply with the GDPR. Fortunately, these very agencies have recently produced extensive guidance describing their interpretations of the "right to explanation" that offers powerful insights into the substantive protections afforded by the GDPR's vaguely-worded mandate.

Part IV details this newly issued guidance and summarizes its implications for companies seeking to better understand what compliance with the GDPR's "right to explanation" actually entails. It reveals that Europe's data authorities have repeatedly envisioned the "right to explanation" as a robust data protection whose true power lies in its synergistic combination with the "data protection by design" principles codified in the Regulation's subsequent chapters. As a result, this Article argues that data auditing methodologies designed to safeguard against algorithmic bias throughout the entire product life cycle will likely become the new norm for promoting compliance in automated systems. It further argues that this more general version of a "right to explanation" offers greater hope of promoting genuine "algorithmic accountability" than the individualized remedial mechanism many

[36] *See infra* Part III and accompanying notes.

commentators have presumed it to be.

Part V examines the GDPR's global implications for companies and countries grappling with compliance, both inside and outside of Europe. It argues that the Regulation will likely have an outsized extraterritorial impact due to the well-documented "Brussels Effect" and the introduction of several legal mechanisms that implicate entities operating outside of the EU. Thanks to the far-flung legal reach of the Regulation, it argues that the "right to explanation"—as envisioned by the GDPR's enforcement authorities—appears destined to become part of a new global data protection standard for companies handling personal information. The new standard will certainly pose its share of challenges for enterprises seeking to deploy sophisticated algorithms. But it also offers those who hope for a more fair, accountable, and transparent automated decision-making systems genuine reason for optimism.

II. DOES THE GDPR ENVISAGE A RIGHT TO EXPLANATION?

In January 2012, the European Commission made global headlines by submitting a proposal to "update and modernise the principles enshrined in the 1995 Data Protection Directive."[37] For seventeen years, the DPD had reigned as Europe's preeminent legislation governing the processing of digital data. But after nearly two decades, the longstanding Directive was beginning to show signs of age. The DPD was originally passed when "less than 1% of Europeans used the internet."[38] Since then, the Commission noted, "[t]echnological progress . . . [had] profoundly changed the way [] data is collected, accessed and used."[39]

The press release accompanying the Commission's announcement set the stage for "a comprehensive reform of [the DPD's] data protection rules."[40] The Commission called for rules to be designed "to increase users' control of their data," to "provide[] for increased responsibility and accountability for those processing personal data," and to create a "single set of rules" that would be "valid across the EU."[41] More than three years of negotiations followed the preliminary proposal, eventually culminating in the formal adoption of the General Data Protections Regulation (GDPR) in April of 2016.[42] The finalized

[37] GDPR Proposal, *supra* note 3.

[38] *Id.*

[39] *Id.*

[40] *Id.*

[41] *Id.*; CONSOLIDATED VERSION OF THE TREATY ON THE FUNCTIONING OF THE EUROPEAN UNION art. 288, 2008 O.J. C 115/47.

[42] *See* GDPR, *supra* note 19.

Regulation constituted a major overhaul of European data processing standards. By enumerating a litany of powerful protections, the new Regulation intended to make the EU bloc "fit for the digital age."[43]

One such protection—located within Chapter 3 of the GDPR—sets forth what the Regulation describes as the "right not to be subject to a decision based solely on automated processing."[44] The protection establishes a number of safeguards designed to ensure the "fair and transparent processing" of personal data, including an obligation that entities provide "meaningful information about the logic involved" in certain types of highly automated decision-making systems.[45] The protection's requirement that "meaningful information" be made available to data subjects has led it to be variously characterized as enshrining a "right to information," a "right to be informed," or, most commonly, a "right to explanation."[46]

As the first piece of European Union Regulation to explicitly gesture toward such a right,[47] the substantive protections that eventually flow from it will set a precedent with ramifications extending far beyond the technology sector. While the usual suspects, such as Facebook, may have grabbed global headlines by announcing millions of dollars spent toward promoting GDPR compliance, the rapid proliferation of machine learning technology across diverse industries indicates that vast swaths of the private sector will soon be forced to take action. Depending on how the protection is eventually applied

[43] EUROPEAN COMM'N, REFORM OF EU DATA PROTECTION RULES (2018), http://ec.europa.eu/justice/data-protection/reform/index_en.htm [https://perma.cc/JH2B-YRMU].

[44] GDPR, *supra* note 19, at art. 22; *see id.* at arts. 13(2)(f), 14(2)(g), 15(1)(h). As is likely obvious, this phrasing leaves open considerable room for ambiguity.

[45] *See id.* at arts. 14(2), 14(2)(g).

[46] *See, e.g.*, Bryce Goodman & Seth Flaxman, EU Regulations on Algorithmic Decision Making and "a Right to an Explanation" (June 28, 2016) (unpublished manuscript), https://ora.ox.ac.uk/objects/uuid:593169ee-0457-4051-9337-e007064cf67c/download_file?safe_filename=euregs.pdf&file_format=application%2Fpdf&type_of_work=Journal+article [https://perma.cc/C6UP-DZQE]; Sandra Wachter et al., *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INT'L DATA PRIVACY L. 76 (2017); Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, 7 INT'L DATA PRIVACY L. 233 (2017); *Data Subjects' Rights*, RADBOUD U., https://www.ru.nl/privacy/english/protection-personal-data/data-subjects-rights/#hf4dfc431-41bd-452c-8cac-3f98083db3b1 [https://perma.cc/8KQ2-WUFM]; ARTICLE 29 WORKING PARTY, GUIDELINES ON AUTOMATED INDIVIDUAL DECISION-MAKING AND PROFILING FOR THE PURPOSES OF REGULATION 2016/679, 9 (2017) [hereinafter A29WP Automated Decision-Making Guidelines].

[47] This could be more precisely phrased as the European Union Regulation to mandate this right in the context of *automated systems with a meaningful threat of enforcement*—a nuance that is covered in greater detail in Part III *infra*.

in practice, it could have profound implications for the use of some of the most powerful computational techniques available to modern enterprises. But, as is true of many protections enshrined within the legislative text of the GDPR, the precise reach of the right is far from certain. A careful examination of the language provides a useful starting point for understanding and contextualizing it.

A. Specific Text Giving Rise to the "Right to Explanation"

Article 22 of the GDPR grants all data subjects[48] a rebuttable[49] "right not to be subject to a decision based solely[50] on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."[51] The GDPR defines "processing" as follows:

> [A]ny operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction[.][52]

The GDPR's use of the term "profiling" introduces a relatively novel concept under EU data protection law.[53] The regulation defines "profiling" as

> any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements[.][54]

[48] *See* GDPR, *supra* note 19, at art. 4. The GDPR defines a "data subject" as "an identified or identifiable natural person" and "an identifiable natural person" as "one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." *Id.*

[49] *See id.* at art. 22(2)–(4) (specifying limited circumstances where automated decision-making is permitted, and providing for different data safeguards).

[50] This term has recently been subject to clarification. *See infra* Part IV.

[51] GDPR, *supra* note 19, at art. 22(1).

[52] *Id.* at art. 4(2).

[53] *See* Frederike Kaltheuner & Elettra Bietti, *Data is Power: Towards Additional Guidance on Profiling and Automated Decision-Making in the GDPR*, 2 J. INFO. RIGHTS, POL'Y & PRACTICE (2018).

[54] GDPR, *supra* note 19, at art. 4(4). Recital 71 of the GDPR adds:

> Such processing includes 'profiling' that consists of any form of automated

Article 22(2) enumerates a limited number of circumstances in which companies[55] processing personal data are exempt from its prohibitions—including when automated decision-making is done consensually or is necessary for contracting.[56] But even in such instances, Article 22 requires that companies nevertheless "implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests."[57] This requirement, at a minimum, includes the subject's "right to obtain human intervention on the part of the [company], to express his or her point of view and to contest the decision."[58]

Article 22's protections are buttressed by those located within Articles 13–15, pertaining to the rights of data subjects whose personal information is directly or indirectly implicated by automated processing techniques. These Articles are intended to "provide the data subject with the . . . information necessary to ensure fair and transparent processing."[59] In fulfilling this goal, Articles 13(2)(f), 14(2)(g), and 15(1)(h) mandate that companies provide subjects with information regarding "the existence of automated decision-making, including profiling, referred to in Article 22 . . . 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."[60]

In addition to the text of the GDPR, the accompanying nonbinding Recital

> processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her.

See GDPR, *supra* note 19, at Recital 71; *see also* Mireille Hildebrandt, *Defining Profiling: A New Type of Knowledge?*, *in* PROFILING THE EUROPEAN CITIZEN 17 (Mireille Hildebrandt & Serge Gutwirth eds., Springer 2008) (exploring the difference between organic and machine profiling).

[55] *See* GDPR, *supra* note 19, at art. 4(7). The GDPR does not single out companies, but instead uses the term "controller" which "means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law." *Id.*

[56] *See id.*, at art. 22(2)–(4).

[57] *Id.* at art. 22.

[58] *Id.*

[59] *Id.* at arts. 13, 14.

[60] *Id.* at arts. 13(2)(f), 14(2)(g), 15(1)(h) (emphasis added). This disclosure requirement extends even to data subjects whose personal information has not been directly obtained by a company.

71 offers further clarification regarding the Regulation's protections pertaining to automated decision-making.[61] The Recital states that the data processing techniques implicating personal data "should be subject to suitable safeguards, which should include [the provision of] specific information to the data subject[,]" as well as the rights "to obtain human intervention," "to express his or her point of view," "to *obtain an explanation of the decision reached* after such assessment," and "to challenge the decision."[62] The Recital further stipulates:

> In order to ensure fair and transparent processing . . . [companies] should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject, and prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or processing that results in measures having such an effect.[63]

While the authority of the Recital is nonbinding under EU law, it nonetheless provides a critical reference point for future interpretations by data protection agencies as well as for co-determinations of positive law that

[61] *See* Tadas Klimas & Jurate Vaiciukaite, *The Law of Recitals in European Community Legislation*, 15 ILSA J. INT'L & COMP. L. 61, 62, 92 (2008). Recitals in EU law lack "independent legal value, but they can expand an ambiguous provision's scope. They cannot, however, restrict an unambiguous provision's scope, but they can be used to determine the nature of a provision, and this can have a restrictive effect." *Id.* at 63. "Recitals explain the background to the legislation and the aims and objectives of the legislation. They are, therefore, important to an understanding of the legislation which follows." COMMISSION OF THE EUROPEAN COMMUNITIES, GUIDE TO THE APPROXIMATION OF EUROPEAN UNION ENVIRONMENTAL LEGISLATION 115 (2017); *see* Case C-355/95 P, Textilwerke Deggendorf GmbH v. Comm'n, 1997 E.C.R. I-02549 ("In that regard, it should be stated that the operative part of an act is indissociably linked to the statement of reasons for it, so that, when it has to be interpreted, account must be taken of the reasons which led to its adoption."). European Court of Justice (ECJ) jurisprudence reveals that the role of Recitals is "to dissolve ambiguity in the operative text of a framework." Wachter et al., *supra* note 46, at 80. According to the ECJ: "Whilst a recital in the preamble to a regulation may cast light on the interpretation to be given to a legal rule, it cannot in itself constitute such a rule." Case 215/88, Casa Fleischhandels-GmbH v. Bundesanstalt fur Landwirtschaftliche Marktordnung, 1989 E.C.R 02789; *see* Roberto Baratta, *Complexity of EU Law in the Domestic Implementing Process*, *in* 2 THE THEORY AND PRACTICE OF LEGISLATION 293 (2014) (highlighting how the complexity of EU law can cause difficulties at the national level); Klimas & Vaiciukaite, at 62.

[62] GDPR, *supra* note 19, at Recital 71 (emphasis added).

[63] *Id.*

may be made by legislators, courts, or other authorities.[64]

B. THE "RIGHT TO EXPLANATION" DEBATE

Despite the GDPR's concerted efforts to detail the protections enshrined under Articles 13, 14, 15, and 22, much uncertainty continues to shroud the Regulation's so-called "right to explanation." This phenomenon owes, in large part, to the GDPR's somewhat fuzzy mandate that entities "ensure fair and transparent processing" by providing "meaningful information about the logic involved" in automated decision-making systems. At a minimum, the protection appears to envisage a limited right for data subjects to understand and verify the basic functionality of certain automated decision-making systems. But beyond that minimum threshold, the precise contours of the "right to explanation" have been the subject of much speculation—giving rise to an "explosive" public debate.[65]

Among the most prominent contributions to the debate, thus far, have been three distinct perspectives originating from scholars within the U.K. and the U.S.[66] Their claims and critiques are set forth below.

1. *The Original Claim*

Goodman's and Flaxman's conference paper—*European Union Regulations on Algorithmic Decision-making and a "Right to Explanation"*—first popularized the knotty, sometimes vexing, issues at the heart of the GDPR's "right to explanation."[67] Published just two months after the Regulation's official release, the piece drew widespread attention to the technical and societal challenges inherent in "explain[ing] an algorithm's decision" made by machine learning algorithms.[68] Goodman and Flaxman observed that, unlike algorithms

[64] These authorities, amongst others, include the GDPR's designated "Supervisory Authorities," the Article 29 Working Party, the European Data Protection Board, the European Data Protection Supervisor, and the European Data Protection Supervisor's Ethics Advisory Group.

[65] *See infra* Section II.B.

[66] Many other contributors beyond these three have also thrown their hats in the ring.

[67] *See* Goodman & Flaxman, *supra* note 46. It should be noted that this paper was subsequently revised.

[68] *See* FRANK PASQUALE, THE BLACK BOX SOCIETY 3–4 (2015); Brenda Reddix-Smalls, *Credit Scoring and Trade Secrecy: An Algorithmic Quagmire or How the Lack of Transparency in Complex Financial Models Scuttled the Finance Market*, 12 U.C. DAVIS BUS. L. J. 87 (2011); Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. ON TELECOMM. & HIGH TECH. L. 235, 237 (2011); Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473, 482 (2016); *see generally* NICHOLAS DIAKOPOULOS, ALGORITHMIC ACCOUNTABILITY REPORTING: ON THE INVESTIGATION OF BLACK BOXES (Tow Centre for Digital Journalism, 2013); Goodman & Flaxman, *supra* note 46.

of past decades,[69] machine learning systems in increasingly widespread usage were "alone on the spectrum in their lack of interpretability."[70] The scholars noted an inherent "tradeoff between the representational capacity of a model and its interpretability"—one that sometimes rendered the underlying decision-making process of the most powerful systems an uninterpretable "black box."[71]

While these types of "black box" algorithms had existed in research labs since the 1980s, Goodman and Flaxman made the prescient observations that their recent proliferation throughout industry presented many challenges for companies and governments seeking to comply with the GDPR.[72] The scholars discussed how numerous factors—including potentially biased training sets, uneven "data quality," the complexity of the most powerful predictive models, and the steep barriers to technical fluency—could pose significant challenges for modern enterprises seeking to comply with the GDPR's mandate of algorithmic explicability.[73]

Although the scholars' work was widely credited with sparking the "right to explanation" debate,[74] their piece was less a legal treatise than a technical primer. Their analysis offered relatively little commentary regarding the right's substantive protections and made only a passing reference to the GDPR's

[69] I.e., those which relied on explicit, rules-based logic for processing information.

[70] *See* Goodman & Flaxman, *supra* note 46, at 6 (quoting PAULO J. G. LISBOA, INTERPRETABILITY IN MACHINE LEARNING PRINCIPLES AND PRACTICE 1521 (2013)).

[71] *See id.* Machine learning techniques that explicitly encode logic do exist—particularly in the natural language processing and bioinformatics realms—but are not focused on for purposes of concision.

[72] *See* Robert D. Hof, *Deep Learning*, MIT TECH. REV. (2013), https://www.technologyreview.com/s/513696/deep-learning [https://perma.cc/Y822-QJC9] (noting that in the mid-80s, "[scientists] spark[ed] a revival of interest in neural networks with so-called "deep" models that made better use of many layers of software neurons"); Goodman & Flaxman, *supra* note 46.

[73] "Data quality" is a broadly construed term whose components include "accuracy, precision, completeness, consistency, validity, and timeliness, though this catalog of features is far from settled." *See* Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 684 n.47 (2016); 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, 21 (2017); *see also, e.g.*, Luciano Floridi, *Information Quality*, 26 PHIL. & TECH. 1 (2013); Richard Y. Wang & Diane M. Strong, *Beyond Accuracy: What Data Quality Means to Data Consumers*, 12 J. MGMT. INFO. SYS. 5 (1996); LARRY P. ENGLISH, INFORMATION QUALITY APPLIED (2009).

[74] *See* Michelle Menting, *EU GDPR: The Impact on the Use of Machine Learning*, ABI RES. (Sept. 17, 2018), https://www.abiresearch.com/blogs/eu-gdpr-impact-use-machine-learning/ [https://perma.cc/8SXQ-V9V8] (crediting Goodman and Flaxman with initiating the debate); Selbst & Powles, *supra* note 46, at 234 (noting that the most "most prominent contributions" to the debate are Goodman and Flaxman's piece and Wachter et al.'s response).

newly introduced enforcement provisions. When the piece did discuss the "right to explanation" directly, Goodman and Flaxman construed the protection as relatively narrow. Aside from a single loosely-worded sentence in the paper's abstract that received outsized attention, the scholars suggested that the "right to explanation" could be satisfied relatively easily. They indicated that simply answering questions such as: "*Is the model more or less likely to recommend a loan if the applicant is a minority?*" or "*Which features play the largest role in prediction?*" could suffice.[75]

2. *The Response*

In response to the widespread attention garnered by Goodman's and Flaxman's conference paper, Wachter et al. entered into the public arena with the provocatively titled piece, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation.*[76] The scholars wasted no time going on the offensive, immediately calling into doubt both the legal existence and the technical feasibility of what Goodman and Flaxman referred to as the GDPR's "right to explanation." Wachter et al.'s contribution offered a richly detailed tour of the Regulation's relevant text and associated Recital—one that reached greater analytic depths than the technically-oriented conference paper it criticized. The scholars articulated a powerful framework for distinguishing questions of algorithmic explicability along chronological and functional dimensions—an important contribution that has since been replicated by numerous researchers.[77]

But as thorough as Wachter et al.'s analysis may have been, their focus was also highly selective. Several of their arguments all but ignored key terms within Articles 13, 14, 15, and 22. In particular, Wachter et al. disregarded the word "meaningful" as applied to a substantive analysis of the phrase "meaningful information about the logic involved" in automated decision-making.[78] Just as importantly, their piece paid short shrift to the Regulation's powerful new administrative capabilities. Instead, their discussion of the GDPR's new

[75] Goodman & Flaxman, *supra* note 46 (emphasis added). The scholars offered virtually no substantive support for their argument that the right could be satisfied with these types of explanations.

[76] Wachter et al., *supra* note 46.

[77] *See, e.g.*, Edwards & Veale, *supra* note 73. Wachter et al.'s framework distinguishes between explanations describing "system functionality" and "specific decisions," and also distinguishes between explanations that occur *before* a data-subject's information has been processed and those that occur *after*. *See* Wachter et al., *supra* note 46, at 78–79.

[78] *See* Wachter et al., *supra* note 46, at 84. The scholars also made a few claims of astonishing scope, including one assertion that, "There are no ambiguities in the language [of the GDPR] that would require further interpretation with regard to the minimum requirements that must be met by data controllers." *Id.* at 80.

enforcement capabilities was limited to a single footnote.[79] Most strikingly of all, the central thesis they advanced was outright contradicted by their own subsequent analysis. After electing to title their work *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*,[80] the scholars went on to repeatedly acknowledge that just such a right existed—noting, for example, that the Regulation could mandate "an explanation when automated decisions have (i) legal or similarly significant effects, and (ii) are based solely on automated processes."[81]

Rather than calling it a "right to explanation," however, the scholars instead sought to replace it with a phrase of narrower implications. They insisted that "the GDPR does not . . . implement a right to explanation, but rather [a] 'right to be informed.' "[82] The scholars, however, went on to note that this mandate provided data-subjects, at minimum, "a right to explanation of system functionality . . . [subject to] restrict[ions] by the interests of data controllers and future interpretations."[83] As such, their insistence on calling it a "right to be informed" appeared to be a distinction of little more than semantic significance.[84]

3. *The Rebuttal*

In November 2017—with the GDPR just six months away and the "right to explanation" debate rapidly rising to a fevered pitch—Selbst and Powles entered into the fray with a point-by-point takedown of Wachter et al. in *Meaningful Information and the Right to Explanation*. Their contribution sought to address what they described as the numerous "unfounded assumptions and unsettling implications of [Wachter et al.'s] analytical frame."[85] In doing so, Selbst and Powles "offer[ed] a positive conception of the right [to explanation] located in the text and purpose of the GDPR."[86] They convincingly argued that it "should be interpreted functionally, flexibly, and should, at a minimum, enable a data subject to exercise his or her rights under the GDPR and human

[79] *See id.* at 99 n.130.

[80] *See id.* The scholars Selbst and Powles correctly noted that this tactic was "not only disingenuous but dangerous, as it invites less scrupulous or more time-pressed advocates to cite the paper for the proposition that there is no right to explanation, which is not even what the paper argues in substance." Selbst & Powles, *supra* note 46, at 238.

[81] Wachter et al., *supra* note 46, at 95.

[82] *Id.* at 77.

[83] *Id.* at 96.

[84] *See* Selbst & Powles, *supra* note 46, at 239.

[85] *See id.*; *infra* Section II.B.2. Many of these criticisms are outlined in the section above.

[86] Selbst & Powles, *supra* note 46, at 234.

rights law."[87]

Selbst's and Powles's piece represented a vital course correction in a public debate that had begun to more closely resemble a rebranding effort than an actual refutation of the substantive right itself.[88] But their contribution occurred in advance of Europe's most influential data protection authorities releasing extensive guidance which provided much needed clarity on the hotly contested topic.[89] Accordingly, the actual language of EU data protection authorities that emerged immediately after its publication did not ground Selbst and Powles's piece. Further, the piece did little to underscore the GDPR's newly-invigorated enforcement measures, as well as the practical implications that flow from them, which are discussed below.

C. LOST IN THE FOG OF BATTLE

Since its origins with Goodman and Flaxman, the GDPR's "right to explanation" debate has fostered a conversation of profound global significance—exploring the economic benefits, technical feasibility, and social tradeoffs of applying "algorithmic accountability" practices in enterprise and government.[90] The contributions of Goodman, Flaxman, Selbst, Powles, and Wachter et al. constitute just a tiny sample of the vast and impressively diverse array of perspectives on this issue.[91] Over a period of just eighteen months, countless industry leaders, media sources, and researchers of various backgrounds have also contributed their unique perspectives.[92] But

[87] *Id.* at 242.

[88] Watcher et al.'s piece continues to enjoy widespread popularity among more casual observers—with many remaining unaware of the important counterweight provided by Selbst & Powles.

[89] *See* Selbst & Powles, *supra* note 46.

[90] *See infra* Section IV.C for a more detailed description of the literature on "algorithmic accountability."

[91] Mendoza and Bygrave, who argue that the "right to explanation" arises as a necessary precondition to Article 22(3)'s "right to contest" could also be added to this list, but are not discussed in detail for purposes of concision. Izak Mendoza & Lee A. Bygrave, *The Right Not to Be Subject to Automated Decisions Based on Profiling*, *in* EU INTERNET LAW: REGULATION AND ENFORCEMENT 77 (T.-E. Synodinou et al. eds., Springer 2017).

[92] *See, e.g.*, Rich Caruana et al., *Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-Day Readmission*, *in* KDD '15PROCEEDINGS OF THE 21ST ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 1721 (2015); David Bamman, Interpretability in Human-Centered Data Science (2016) (unpublished manuscript), https://cscw2016hcds.files.wordpress.com/2015/10/bamman_hcds.pdf [https://perma.cc/3KLR-8MDY]; Michael Gleicher, *A Framework for Considering Comprehensibility in Modeling*, 4 BIG DATA 75 (2016); Finale Doshi-Valez & Been Kim, A Roadmap for a Rigorous Science of Interpretability (2017) (unpublished manuscript), https://arxiv.org/abs/1702.08608 [https://perma.cc/Q9K4-PZYM]; Eric Horvitz, Presentation at the Berkeley Center for Law & Technology: On the Meaningful Understanding

mystifyingly, many of the most distinguished contributions to this multifaceted debate have largely overlooked what is potentially the most profound change of all heralded by the GDPR: the sweeping new enforcement powers granted to EU data protection authorities by the new Regulation.

Beyond the "right to explanation" debate's narrow focus on Articles 13, 14, 15, and 22, there lies a series of provisions that appear destined to forever change the practical reality of enforcement by data protection authorities. These Articles—contained in Chapters 6 and 8 of the Regulation—grant vast new administrative powers to EU watchdog agencies that have long been viewed as toothless under the DPD.[93] Failing to elucidate the profound new role that these freshly empowered agencies will play in enforcing and, therefore, *interpreting* the GDPR's "right to explanation" currently represents a major blind-spot within the public debate. If left unaddressed, this blind spot risks allowing the public debate to move in an unproductive and unnecessarily adversarial direction.

III. TURNING THE PAGE IN THE "RIGHT TO EXPLANATION" DEBATE

Although the introduction of the GDPR will represent the largest overhaul of EU data protection laws in twenty years, the Regulation's most revolutionary change actually involves the addition of a host of new legal mechanisms for promoting enforcement.[94] After all, the EU has long boasted an extensive list of rules[95] that set a high bar for data protection, including

of the Logic of Automated Decision Making (Mar. 24 2017); Ethan Chiel, *EU Citizens Might Get a 'Right to Explanation' About the Decisions Algorithms Make*, SPLINTER (July 5, 2016), http://fusion.kinja.com/eu-citizens-might-get-aright-to-explanation-about-the-1793859992 [https://perma.cc/23TL-TUXP]; Cade Metz, *Artificial Intelligence Is Setting Up the Internet for a Huge Clash With Europe*, WIRED (July 11, 2016), https://www.wired.com/2016/07/artificial-intelligence-setting-internet-huge-clash-europe/ [https://perma.cc/GFY4-D4SR]; Ian Sample, *AI Watchdog Needed to Regulate Automated Decision-making, Say Experts*, GUARDIAN (Jan. 27, 2017), https://www.theguardian.com/technology/2017/jan/27/ai-artificial-intelligence-watchdog-needed-to-prevent-discriminatory-automated-decisions [https://perma.cc/J4KB-WVEL]; Matt Burgess, *Watching Them, Watching Us: Can We Trust Big Tech to Regulate Itself?*, CREATIVE REV., (Apr. 2017), https://www.creativereview.co.uk/watching-watching-us/ [https://perma.cc/85ZF-KWLA]; ACM U.S. Pub. Policy Council, Statement on Algorithmic Transparency and Accountability (May 25, 2017).

[93] *See* GDPR, *supra* note 19, at chs. 6, 8; Lomas, *supra* note 30 (noting that the "beefing up of enforcement that's baked into the new regime means there's a better opportunity for DPAs to start to bark and bite like proper watchdogs").

[94] *See* Lomas, *supra* note 30.

[95] In addition to the DPD, there are numerous other regulations that allude to rights involving automated decision-making explicability. "For example, the public sector is subject

rights that specifically address automated decision-making.[96] What these rules have lacked, however, is a meaningful threat of enforcement.[97]

Under the DPD, EU agencies tasked with carrying out its mandate were highly limited in their capacity to levy financial penalties against entities breaching the DPD.[98] Before the GDPR, the UK's Information Commissioner's Office (ICO), for example, was capped at a maximum fine of just £500,000 for violations.[99] Facebook's annual revenue for the 2017 fiscal year, by comparison, topped $40B.[100] Therefore, at most, the ICO could only hope to impose a fine representing a paltry percentage of the company's annual revenue.

Moreover, replacing the DPD with the GDPR represents an instance of an EU Regulation replacing a Directive. While directives "set out general rules to be transferred into national law by each country as they deem appropriate," regulations constitute a single, uniform law that is "directly applicable" to all

to the Public Administration Act that requires, *inter alia*, individual decisions to be substantiated. The person concerned has the right to be informed of the regulations and the actual circumstances underpinning a decision, as well as the main considerations that have been decisive." ARTIFICIAL INTELLIGENCE AND PRIVACY, *supra* note 4, at 22 (quoting Public Administration Act Sections 24 and 25). The EU also explicitly treats privacy protection as a fundamental right.

[96] *See* DPD, *supra* note 3; *see also, e.g.*, Mendoza & Bygrave, *supra* note 91; Lee A. Bygrave, *Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, 17 COMPUTER L. & SECURITY REV. 17 (2001); Alfred Kobsa, *Tailoring Privacy to Users' Needs*, *in* PROCEEDINGS OF THE 8TH INTERNATIONAL CONFERENCE ON USER MODELING 303 (M. Bauer et al. eds., 2001); Mireille Hildebrandt, *Profiling and the Rule of Law*, 1 IDENTITY IN INFO. SOC'Y 55, 55 (2008). Wachter, et al. actually discuss this phenomenon, noting: "Interestingly, despite years of negotiations, the final wording of the GDPR concerning protections against profiling and automated decision-making hardly changed from the relevant Articles and Recitals of the Data Protection Directive [of] 1995." Wachter et al., *supra* note 46, at 81. But their failure to address the enhanced enforcement powers introduced by the GDPR renders moot their underlying argument that the new provisions will do little to change the current regulatory landscape.

[97] *See* Mendoza & Bygrave, *supra* note 91, at 78 (describing art. 15 as "a second-class data protection right: it is rarely enforced, poorly understood and easily circumvented").

[98] *See id.*; DPD, *supra* note 3.

[99] *Facebook Faces £500,000 Fine from UK Data Watchdog*, BBC NEWS (July 11, 2018), https://www.bbc.com/news/technology-44785151 [https://perma.cc/P6E7-QNNB]. The GDPR specifies the monetary sanctions available to DPAs, unlike the DPD which left it to countries to set their own sanctions. *See* DPD, *supra* note 3, at art. 24 (leaving it to "Member States [to] adopt suitable measures to ensure the full implementation of the provisions of this Directive and shall in particular lay down the sanctions to be imposed in case of infringement of the provisions adopted pursuant to this Directive").

[100] *See* Press Release, Facebook Investor Relations, Facebook Reports Fourth Quarter and Full Year 2017 Results (Jan. 31, 2018) [hereinafter Facebook Press Release].

EU Member States.[101] The differences between these two paths to legislative implementation may seem trivial to outsiders looking in, but their practical effects are not. Unlike the GDPR, the DPD is subject to twenty-eight different interpretations and enforcement regimes—leading to differences that can foment confusion and inconsistency among industry leaders and data protection authorities alike. Coupled with the limited fines available under the DPD, these inconsistencies exacerbated enforcement problems for data protection authorities.

The combined effect of these DPD enforcement limitations produced a pack of EU data watchdogs tethered to a markedly short regulatory leash. For over two decades, the Directive set a high standard for data protection for companies handling the personal information of EU citizens. But those responsible of upholding these protections have long been perceived as lacking a genuine threat of enforcement.

Viewed through this lens, it is easy to understand why the debate surrounding the "right to explanation" has seen comparatively little attention paid to the authorities that will actually be tasked with enforcing it. For if the past were prologue, they could be expected to play a peripheral role in carrying out the right's protective mandate. However, with the passage of the GDPR, all of that is set to change. Chapters 6 and 8 of the Regulation grant data authorities vastly increased investigatory powers, an enhanced "enforcement tool kit," and the capacity to levy far greater financial penalties against entities in breach.[102]

EU data authorities will no longer be constrained by the limited range of enforcement options available under the DPD. Instead, these authorities will have far-reaching investigatory and corrective powers that allow them to issue sanctions against data protection violations that are "effective, proportionate," and, most importantly, "dissuasive."[103] Whereas data authorities under the DPD were limited to six-figure fines or sternly-worded letters, companies now will live under the threat of corrective measures that may be on orders of magnitude more potent.[104] Under this new reality, some commentators have

[101] KAREN DAVIES, UNDERSTANDING EUROPEAN UNION LAW (6th ed., 2016). Art. 288 of the Treaty on the Functioning of the European Union provides that: "A directive shall be binding, as to the result to be achieved, upon each Member State to which it is addressed, but shall leave to the national authorities the choice of form and methods." Consolidated Version of the Treaty on European Union art. 288, 2006 O.J. C 321 E/5. at 126. Article 288 states that a regulation, on the other hand, "shall be binding in its entirety and directly applicable in all Member States." *Id.* at 125.

[102] *See infra* Section III.A and accompanying notes.

[103] *See* GDPR, *supra* note 19, at art. 83.

[104] *See supra* note 99 and accompanying text.

asserted that the transition from the DPD to the GDPR should be understood as less about "individual EU Member States . . . getting stronger privacy laws" and more about EU data authorities finally starting "to bark and bite like proper watchdogs."[105]

The following subparts describe the specific enforcement powers that the GDPR provides European data authorities, as well as some of the practical implications of this power shift for downstream enterprises.

A. THE ASCENT OF ENFORCEMENT

Chapter 6 of the GDPR provides for the appointment, by each Member State, of "one or more independent public authorities to be responsible for monitoring [its] application . . . ."[106] The legislation endows these agencies—which it terms "supervisory authorities" (SAs)—with broad "investigatory," "advisory," and "corrective" powers of far greater scope than those currently available under the DPD.[107] According to Chapter 6, these powers ensure the "consistent application" of the GDPR throughout the EU and include, among many other provisions, the ability: (1) "to obtain . . . access to all personal data [belonging to a company] and to all information necessary for the performance of [investigatory] tasks," (2) "to carry out investigations in the form of data protection audits,"[108] (3) "to issue warnings [or] reprimands to a [company]," (4) "to impose a temporary or definitive limitation [against companies] including a ban on processing," and (5) "to order the suspension of data flows to a recipient in a third country[109] or to an international organisation."[110]

Chapter 6's expansive set of investigatory and corrective powers are buttressed by an equally expansive set of remedial powers laid out in Chapter 8. These powers provide supervisory agencies with the authority to impose administrative fines that are "effective, proportionate, and dissuasive."[111] Under Chapter 8, SAs can fine companies that violate the GDPR's basic administrative or technical requirements up to €10 million or up to 2% of the companies' total annual revenue for the preceding financial year, "whichever

[105] Lomas, *supra* note 30.

[106] *See* GDPR, *supra* note 19, at art. 51.

[107] *See id.* at art. 58.

[108] Data protection audits are discussed in greater detail in Section IV.C *infra*.

[109] This term is discussed in detail in *infra* Part V.

[110] GDPR, *supra* note 19, at art. 58.

[111] *See* GDPR, *supra* note 19, at art. 83. The DPD, by contrast, places authority for adopting "suitable measures to ensure the full implementation of the provisions" with individual Member States. This has led to highly limited enforcement capabilities. DPD, *supra* note 3, at art. 24; *see supra* notes 99–100 and accompanying text.

is higher."[112] For violations of provisions more fundamental to the GDPR's data protection mandate[113]—including Articles 13, 14, 15, and 22—the maximum allowable fine increases precipitously. SAs can punish infringers of these provisions with fines of up to €20 million, or up to 4% of the companies' total annual revenue for the preceding financial year—again, "which[ever] is higher."[114]

The operative adjective, in both such instances, is the word "higher." To return to the example of the tech giant Facebook, whose annual revenues approximate €40 billion, a fine of 4% of annual turnover could total €1.6 billion, more than 3,200 times larger than the maximum fine available in the UK under the DPD.[115] This switch from proportional, as opposed to fixed, financial penalties ensures that even the titans of industry will not be immune from enforcement.

But for any in-house practitioners whose pulse doubled at the sight of such a multiple, the Regulation also provides cause for relief. First, the GDPR makes clear that punishment for breaches should be individualized and proportionate. The GDPR does not mandate the use of fines for all enforcement actions.[116] Article 83 outlines an extensive list of considerations for SAs seeking to ensure that their punishments are commensurate with the alleged violation.[117] These factors shift the administrative focus to the actual impacts of the violation, including the number of individuals affected, the actual damages suffered, and the sensitivity of the personal data at root.[118] Also, the GDPR stipulates that good faith efforts to proactively implement protective policies, ensure transparency, notify enforcement agencies, and cooperate with SA oversight will further reduce the likelihood of companies facing serious sanctions.[119]

[112] *See* GDPR, *supra* note 19, at art. 83.

[113] "Examples that fall under this category are non-adherence to the core principles of processing personal data, infringement of the rights of data subjects and the transfer of personal data to third countries or international organizations that do not ensure an adequate level of data protection." *GDPR: Guidelines and Consequences for Non-Compliance*, GDPR:REPORT (June 16, 2017), https://gdpr.report/news/2017/06/16/gdpr-guidelines-consequences-non-compliance/ [https://perma.cc/J756-M5XD]. *See* GDPR, *supra* note 19, at art. 84.

[114] *See* GDPR, *supra* note 19, at art. 83.

[115] *See supra* notes 99–100 and accompanying text.

[116] *See* GDPR, *supra* note 19, at art. 83.

[117] *See id.*

[118] *See id.*

[119] *See id.*

B. THE IMPORTANCE OF UNDERSTANDING WHEN THE WATCHDOGS MIGHT BITE

With great power, of course, comes great interpretive responsibility. After all, what better source of guidance could there be for companies seeking to ensure compliance with the GDPR's "right to explanation" than the data authorities likeliest to bring enforcement action against them? Any agency action will, of course, be subject to the slower-burning process of judicial clarification through national and international litigation. But while any such activity percolates through the EU's multi-layered legal system, the de facto interpretive authorities of the "right to explanation" will be those whose primary responsibility it is to investigate and punish companies that breach the GDPR.

Data protection authorities have already begun to signal their anticipated ascendance by flexing additional regulatory muscle in the lead up to the GDPR's effectuation.[120] According to a recent report, the total monetary value of fines the UK's ICO levied doubled in 2016—coinciding with a steep uptick in the number of enforcement notices issued by the agency and a nearly 100% increase in the size of its fines.[121] This increased enforcement activity also came amid calls by the agency to increase its staff size in advance of the GDPR's May 2018 effectuation.[122]

[120] *See* Max Metzger, *Sharp Rise in ICO Fines and Enforcement Notices as GDPR Races Closer*, SC MEDIA (June 1, 2017), https://www.scmagazineuk.com/sharp-rise-in-ico-fines-and-enforcement-notices-as-gdpr-races-closer/article/665466/ [https://perma.cc/PT6D-EXMU]; Elizabeth Denham, the residing commissioner, remarked:

> In this world of big data, AI and machine learning, my office is more relevant than ever. I oversee legislation that demands fair, accurate and non-discriminatory use of personal data; legislation that also gives me the power to conduct audits, order corrective action and issue monetary penalties. Furthermore, under the GDPR my office will be working hard to improve standards in the use of personal data through the implementation of privacy seals and certification schemes. We're uniquely placed to provide the right framework for the regulation of big data, AI and machine learning, and I strongly believe that our efficient, joined-up and co-regulatory approach is exactly what is needed to pull back the curtain in this space.

Elizabeth Denham, *Information Commissioner's Foreword*, *in* BIG DATA, ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DATA PROTECTION 3 (2017); *see also* Jamie Doward et al., *Watchdog to Launch Inquiry into Misuse of Data in Politics*, GUARDIAN (Mar. 4, 2017), https://www.theguardian.com/technology/2017/mar/04/cambridge-analytics-data-brexit-trump [https://perma.cc/B3ST-5H5H].

[121] *See* Metzger, *supra* note 120.

[122] *See id.*

IV. THE NEXT CHAPTER IN THE DEBATE: SA ENFORCEMENT AND THE RISE OF DATA AUDITS

Viewed against the backdrop of Chapter 6's and 8's vastly enhanced enforcement powers, it becomes immediately apparent that the public debate over the "right to explanation" can no longer be confined exclusively to the text of the GDPR. Instead, the right articulated by the Regulation must be understood holistically with a newfound deference owed to the downstream interpretations by the EU data watchdogs whose regulatory bark and bite will soon become far costlier for companies to ignore. Fortunately, a recent burst of activity by these very data authorities has provided extensive guidance for enterprises seeking to better understand what meaningful compliance with the GDPR's controversial "right to explanation" entails in practice.

The following subparts detail these new activities, relying on the words of the data authorities themselves whenever possible in order to minimize the likelihood of editorializing. Subpart A details the recent activity by the Article 29 Data Protection Working Party, a European body charged with a senior advisory role in the GDPR's implementation. Subpart B then takes the interpretation of a single data protection authority, the UK's Information Commissioner's Office (ICO), as a case study for understanding the scope of the "right to explanation" in practice.

A. THE INTERPRETATION OF THE ARTICLE 29 DATA PROTECTION WORKING PARTY

In October 2017, the Article 29 Data Protection Working Party (A29WP) published its official "Guidelines on Automated Individual Decision-Making and Profiling" for the GDPR.[123] The A29WP "is the European Commission's most senior advisory body on data protection and information security matters" and serves as a central authority for all EU data protection agencies.[124] Although its guidelines are nonbinding, they constitute a vital reference point for the individual SAs appointed by EU Member States and are, therefore,

[123] *See* A29WP Automated Decision-Making Guidelines, *supra* note 46.

[124] ARTIFICIAL INTELLIGENCE AND PRIVACY, *supra* note 4, at 4. The A29WP, which launched in 1996, derives its name from Article 29 of the DPD setting out its composition and purpose. *See Glossary A*, EUROPEAN DATA PROTECTION SUPERVISOR, https://edps.europa.eu/data-protection/data-protection/glossary/a_en [https://perma.cc/3CTD-8T8E] (noting the " 'Article 29 Working Party' is the short name of the Data Protection Working Party established by Article 29 of Directive 95/46/EC"). It is a representative body composed of data protection authorities from each EU Member State, and it also includes the European Data Protection Supervisor and the European Commission. Since the GDPR took effect, it has been replaced by the "European Data Protection Board." *See* GDPR *supra* note 19, at art. 68.

critical to understanding how those authorities should interpret the GDPR.

The A29WP's guidance on automated decision-making included numerous provisions intended to clarify the "right to explanation"—stemming from a collection of rights that the A29WP referred to as the rights "to be informed," "to obtain human intervention," and "to challenge [a] decision" made by certain automated systems.[125] According to the A29WP, the "complexity of machine-learning" algorithms used in such systems "can make it challenging to understand how an automated decision-making process or profiling works."[126] But such complexity, it insisted, "is no excuse for failing to provide information" to data subjects.[127] The A29WP instructed that companies making automated decisions that fall under Article 22(1) "should find simple ways to tell the data subject about the rationale behind, or the criteria relied on in reaching the decision"—albeit "without necessarily always attempting a complex explanation of the algorithms used or [a] disclosure of the full algorithm."[128] In doing so, the A29WP stipulated that companies must:

- "[T]ell the data subject that they are engaging in this type of activity;
- [P]rovide meaningful information about the logic involved; and
- [E]xplain the significance and envisaged consequences of the processing."[129]

The A29WP further clarified that the phrase "[m]eaningful information about the logic involved will in most cases require controllers to provide details such as":

- "[T]he information used in the automated decision-making process, including the categories of data used in a profile;
- [T]he source of that information;
- [H]ow any profile used in the automated decision-making process is built, including any statistics used in the analysis;
- [W]hy this profile is relevant to the automated decision-making process; and

[125] *See* A29WP Automated Decision-Making Guidelines, *supra* note 46, at 9.
[126] *Id.* at 14.
[127] *Id.* at 14 n.12.
[128] *Id.* at 14.
[129] *Id.* at 13–14.

- [H]ow it is used for a decision concerning the data subject."[130]

The A29WP added that it was "good practice [for companies] to provide the above information *whether or not* the processing falls within the narrow Article 22(1) definition."[131] The agency also insisted that companies could not avoid Article 22 by simply "fabricating" *de minimus* human involvement in decision-making.[132] According to the A29WP, companies must ensure that any human "oversight of [a] decision is meaningful, rather than just a token gesture" if they intend for their systems to fall outside the scope of Article 22's provisions pertaining to decisions "based *solely* on automated processing."[133]

In addition to the specific explanatory measures outlined above, the A29WP also recommended that companies introduce more general "procedures and measures to prevent errors, inaccuracies or discrimination" in data processing.[134] The guidelines suggested that companies "carry out frequent assessments on the data sets they process to check for any bias, and develop ways to address any prejudicial elements, including any over-reliance on correlations."[135] According to the A29WP, these assessments should be conducted "on a cyclical basis; not only at the design stage, but also continuously, as the profiling is applied to individuals," so that the "outcome of such testing [can] feed back into the system design."[136]

One such safeguard the A29WP repeatedly invoked involves the use of the "Data Protection Impact Assessment" (DPIA), originating under Article 35 of

[130] *Id.* at 28.

[131] *Id.* at 13 (emphasis added). This justification stemmed, in part, from GDPR Recital 60 stating:

> The controller should provide the data subject with any further information necessary to ensure fair and transparent processing taking into account the specific circumstances and context in which the personal data are processed. Furthermore, the data subject should be informed of the existence of profiling and the consequences of such profiling.

GDPR *supra* note 19, at Recital 60.

[132] A29WP Automated Decision-Making Guidelines, *supra* note 46, at 10.

[133] *Id.* at 13 (emphasis added). This question, too, has been the subject of heated debate due to Article 22's use of the phrase "solely" in its provisions related to automated decision-making. *See, e.g.*, Wachter et al., *supra* note 46, at 88; Selbst & Powles, *supra* note 46, at 5–6. The A29WP further clarified that: "[i]t should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the available input and output data." A29WP Automated Decision-Making Guidelines, *supra* note 46, at 10.

[134] A29WP Automated Decision-Making Guidelines, *supra* note 46, at 17.

[135] *Id.*

[136] *Id.*

the GDPR.[137] Although the GDPR does not formally define the concept of the DPIA, the A29WP described it as "a process for building and demonstrating" compliance by systematically examining automated processing techniques to determine the measures necessary to "manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data."[138]

While noting that the GDPR provides companies with considerable "flexibility to determine the precise structure and form of the DPIA," the A29WP stipulated that the DPIA represented a fundamentally "iterative process" with "common criteria" for carrying it out.[139] According to the A29WP, these criteria were best understood as falling within the GDPR's broader "data protection by design" principles, which apply at all stages of a system's life cycle.[140]

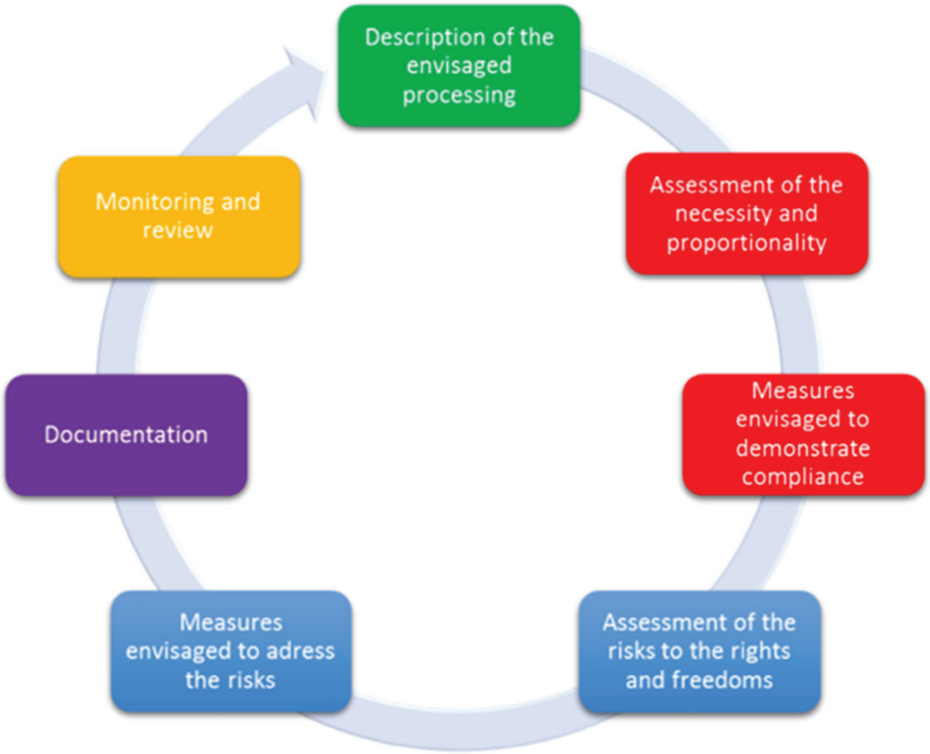
[137] *Id.* at 27.

[138] *See* ARTICLE 29 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 4 (2017) [hereinafter A29WP DPIA Guidelines].

[139] The A29WP DPIA Guidelines Annexes 1 and 2 provide additional details regarding these requirements. *See id.* at 21–22.

[140] *See id.* at 14.

Figure I: The Iterative DPIA Process[141]


Description of the envisaged processing
Assessment of the necessity and proportionality
Measures envisaged to demonstrate compliance
Assessment of the risks to the rights and freedoms
Measures envisaged to adress the risks
Documentation
Monitoring and review

Under the GDPR's "data protection by design" mandate, companies must "[t]ak[e] into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by [] processing."[142] The GDPR recommends DPIAs as a means of proactively identifying and addressing these considerations so that companies can effectively "implement appropriate technical and organisational . . . safeguards into the[ir] processing [operations]."[143]

1. *When Are DPIAs More Than Mere Recommendations?*

The A29WP's guidance stresses that, in many circumstances, DPIAs are not merely recommended as a matter of best practices but are compulsory. In

[141] *Id.* at 16.

[142] GDPR, *supra* note 19, at art. 25.

[143] *Id.* The GDPR explicitly recommends "measures, such as pseudonymisation, which are designed to implement data-protection principles, [and] data minimisation." *Id.*

determining whether a DPIA is or is not compulsory, Article 35(1) of the GDPR relies, primarily, on the heuristic of so-called "high risk" data processing operations.[144] According to the Regulation, DPIAs are mandatory "[w]here a type of processing . . . taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons . . . ."[145] Article 35 establishes a non-exhaustive list of scenarios likely to be deemed high risk, including when operations involve:

> a) [A] systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;
>
> b) [P]rocessing on a large scale of special categories of data referred to in Article 9(1),[146] or of personal data relating to criminal convictions and offences referred to in Article 10;[147] or
>
> c) [A] systematic monitoring of a publicly accessible area on a large scale.[148]

The A29WP's guidance elaborates on this list by enumerating ten specific scenarios that "provide a more concrete" set of criteria for determining

[144] *See id.* at art. 35.

[145] *Id.*

[146] *See id.* Article 9(1) states:

> Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

Id. at art. 9.

[147] *See id.* at art. 35. Article 10 states:

> Processing of personal data relating to criminal convictions and offences or related security measures based on Article 6(1) shall be carried out only under the control of official authority or when the processing is authorised by Union or Member State law providing for appropriate safeguards for the rights and freedoms of data subjects. Any comprehensive register of criminal convictions shall be kept only under the control of official authority.

Id. at art. 10. Article 6(1) includes a list of criteria for establishing the lawfulness of processing. *See id.* at art. 6(1).

[148] *Id.* at art. 35. The GDPR notes that the use of "new technologies" is "particularly" likely to produce high risks. *See id.*

whether operations are "high risk." These include instances where processing involves: (1) evaluating or scoring, (2) automated decision-making with legal or similarly significant effects, (3) systematic monitoring, (4) sensitive data, (5) data processed on a large scale, (6) datasets that have been matched or combined, (7) data concerning vulnerable data subjects, (8) innovative use or applying technological or organizational solutions, (9) data transfer across borders outside the European Union, and (10) processing that inherently "prevents data subjects from exercising a right or using a service or a contract."[149]

Although the A29WP emphasized that DPIAs are not obligatory "for every processing operation which may result in risks," the GDPR's requirement that an *ex ante* assessment be conducted for all processing operations produces a distinctly circular effect.[150] In cases where it is unclear whether a given operation requires a DPIA, carrying out a preliminary DPIA to assess the risks may be the best means of ensuring compliance. In other words, demonstrating that a DPIA is not necessary will, in many instances, itself require a DPIA.[151] This somewhat circular effect will likely incentivize companies to err on the side of caution with DPIAs. Companies may implement them even if the intent in doing so is to simply document or investigate whether more robust explanatory measures are required.

Crucially, these *ex ante* assessments are required even when the GDPR's provisions pertaining to decision-making "based *solely* on automated processing" are not directly implicated.[152] The A29WP repeatedly highlighted that Article 35(3)(a)'s deliberate exclusion of the word "solely" meant that the Article "appl[ied] in the case of decision-making including profiling with legal or similarly significant effects that is *not wholly automated*, as well as solely automated decision-making defined in Article 22(1)."[153]

[149] *See* A29WP DPIA Guidelines, *supra* note 138, at 9–11.

[150] *See id.* at 8.

[151] The A29WP DPIA Guidelines stressed that:

> In order to enhance compliance with this Regulation where processing operations are likely to result in a high risk to the rights and freedoms of natural persons, the controller should be responsible for the carrying-out of a data protection impact assessment to evaluate, in particular, the origin, nature, particularity and severity of that risk.

Id. at 4.

[152] *See* A29WP Automated Decision-Making Guidelines, *supra* note 46, at 10 (emphasis added).

[153] *See id.* at 29 (emphasis added).

2. *What Kinds of Documented Explanations Do DPIAs Require?*

As a means of promoting additional transparency through DPIAs, the A29WP instructed that when data "processing is wholly or partly performed by a [company]," the company should assist SAs "in carrying out [a] DPIA and provide any necessary information" to them.[154] Moreover, the A29WP emphasized that, under Article 35(9), companies are required, "where appropriate," to actively "seek the views of data subjects or their representatives" during the DPIA process.[155] In fulfilling this obligation, the A29WP stated that the views of data subjects could be solicited by a variety of means "depending on the context," including "an internal or external study related to the purpose and means of the processing operation," "a formal question" directed to the relevant stakeholders, or "a survey sent to the data controller's future customers."[156] The A29WP also noted that when a company's "final decision" to proceed with a particular process operation "differ[ed] from the views of the data subjects, its reasons for going ahead or not should be [also] documented."[157] Even in instances where a company has decided that soliciting the views of data subjects is not appropriate, the A29WP insisted that the company should nonetheless document "its justification for not seeking the views of data subjects."[158]

Article 35(7) of the GDPR specifically enumerates four basic features that all DPIAs must, at a minimum, contain:

> 1. [A] systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;
> 2. [A]n assessment of the necessity and proportionality of the processing operations in relation to the purposes;
> 3. [A]n assessment of the risks to the rights and freedoms of data subjects[; and]
> 4. [T]he measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the

[154] A29WP DPIA Guidelines, *supra* note 138, at 15.

[155] *Id.*

[156] Damiana Lesce, Paola Lonigro & Valeria de Lucia, *Privacy. Data Protection Impact Assessment (DPIA). The Art. 29 Data Protection Working Party Guidelines*, LEXOLOGY, https://www.lexology.com/library/detail.aspx?g=b7e8d97f-dd45-48de-8796-c68c2e5bf0a9 [https://perma.cc/WC7L-HWFB].

[157] A29WP DPIA Guidelines, *supra* note 138, at 15.

[158] *Id.*

> protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.[159]

Finally, the A29WP added that while publicly releasing "a DPIA is not a legal requirement of the GDPR," companies "should consider publishing . . . their DPIA[s]" either in full or in part.[160] The A29WP stated that the "purpose of such a process would be to help foster trust in the controller's processing operations, and demonstrate accountability and transparency"—particularly "where members of the public are affected by the processing operation."[161] According to the institution, 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" and "could even consist of just a summary of the DPIA's main findings."[162]

B. From the A29WP to Supervisory Authorities

From the central guidance provided by the A29WP come the specific downstream interpretations of EU data authorities. Although the individual interpretations of these SAs are, by design, the furthest from the textual wellspring of the GDPR, they are by far the most relevant for companies seeking to promote compliance. As the agencies on the front lines of overseeing investigations and issuing sanctions, the interpretations they provide will constitute the clearest signals for companies attempting to understand the substantive protections afforded by the GDPR's "right to explanation."

1. *Why the ICO?*

The analysis that follows focuses on one such authority—the UK's Information Commissioner's Office (ICO). The reasons for this focus on the ICO are twofold. First, surveying all twenty-eight agencies would be needlessly exhaustive, as each agency's interpretation draws directly from the GDPR as opposed to drawing indirectly from twenty-eight individual legislative enactments, as was the case under the DPD. Second, and most importantly, the UK's imminent exit from the EU makes the ICO a particularly informative example. Despite the imminent separation from the European bloc, the country seeks to continue the free flow of data with Continental Europe by

[159] *Id* at 4.

[160] *Id* at 18.

[161] *Id.*

[162] *Id.*

promoting domestic compliance with the GDPR. Thus, the fact that the ICO is, in one sense, a bad example makes it an especially good one. The agency, after all, will be particularly attuned to ensuring its framework is coextensive with the rest of the EU's.

2. *The ICO's Guidance*

Since the A29WP's release of its GDPR guidance in October 2017, the ICO, along with every other EU data authority, published extensive guidelines for organizations seeking to comply with the GDPR's requirements.[163] The agency describes these guidelines as a "living document" subject to elaboration or alteration on an ongoing basis.[164] Among the ICO's many provisions interpreting the GDPR are those pertaining to the data subjects' "rights related to automated decision making including profiling."[165] According to the ICO, companies processing data "must identify whether any of [their] processing falls under Article 22 and, if so, make sure that" they:

- "[G]ive individuals information about the processing;
- [I]ntroduce simple ways for them to request human intervention or challenge a decision;
- [C]arry out regular checks to make sure that your systems are working as intended."[166]

When processing operations fall under Article 22's specific purview,[167] the ICO also requires that companies carry out a DPIA "to identify the risks to individuals," to "show how [they] are going to deal with them," and to

[163] *See generally Guide to the General Data Protection Regulation (GDPR)*, INFO. COMMISSIONER'S OFF., https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/ [https://perma.cc/2GC6-4GEC] (last visited Apr. 2, 2019). The UK Government has also issued new data protection legislation that will implement the standards set forth by the GDPR. *See GDPR Fact Sheet*, BENEFACTO, https://benefacto.org/gdpr-fact-sheet/ [https://perma.cc/5KJD-T8C3] (last visited Apr. 2, 2019). These laws include a number of additional protections going above and beyond the baseline set by the GDPR which extend to "journalists, scientific and historical researchers, and anti-doping agencies who handle people's personal information." *Id.*

[164] *See* ICO'S OVERVIEW OF GDPR, *supra* note 14, at 3.

[165] *See Rights Related to Automated Decision Making Including Profiling*, INFO. COMMISSIONER'S OFF., https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling/ [https://perma.cc/6SEH-DNKD] [hereinafter ICO Automated Decision Making Guidelines].

[166] *Id.* Notably, this mandate is coextensive with the A29WP's own non-binding recommendation, which the ICO appears to be diligently replicating.

[167] *See id.* Some instances do not apply. *See supra* Part II.A.

demonstrate the "measures [they] have in place to meet GDPR requirements."[168]

Even when processing operations fall outside of Article 22, the ICO's guidelines explicitly endorse the use of a DPIA as part of a broader compliance tool kit based on the same principles of "data protection by design" (DPbD) identified by the A29WP.[169] In addition to the comprehensive set of recommendations involving DPbD detailed in its public discussion paper,[170] the ICO states that companies "have a general obligation to implement technical and organisational measures to show that [they] have considered and integrated data protection into [their] processing activities."[171]

C. THE RISE OF THE DPIA AND DATA PROTECTION BY DESIGN

From the guidance set forth by the A29WP and the ICO, one fact is overwhelmingly clear: the GDPR's "right to explanation" is no mere remedial mechanism to be invoked by data subjects on an individual basis, but it implies a more general form of oversight with broad implications for the design, prototyping, field testing, and deployment of data processing systems. The "right to explanation" may not require that companies pry open their "black boxes" per se, but it does require that they evaluate the interests of relevant stakeholders, understand how their systems process data, and establish policies for documenting and justifying key design features throughout a system's life cycle. Not only must companies convey many of these details directly to downstream data subjects,[172] but they must also document and explain the safeguards in place for managing data processing risks either through a DPIA as described in Article 35 or through a substantively similar mechanism. Indeed, it is perhaps no coincidence that the formulation of Article 35(1) bears such a striking similarity to that of Article 22(1). Taken together, these two mandates produce a powerful synergistic effect that promotes the kinds of prophylactic DPbD principles prevalent throughout the GDPR.[173] As a

[168] *Id.* Even in instances where Article 22's requirements do not apply, the ICO recommends that companies nonetheless "carry out a DPIA to consider and address the risks before [they] start any new automated decision-making or profiling" and "tell [] customers about the profiling and automated decision-making [they] carry out, what information [they] use to create the profiles and where [they] get this information from." *Id.*

[169] *See Data Protection by Design and Default*, INFO. COMMISSIONER'S OFF., https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/ [https://perma.cc/E9FS-J4NM] (last visited Apr. 2, 2019).

[170] *See* ICO'S OVERVIEW OF GDPR, *supra* note 14, at 32–37.

[171] *Id.* at 32.

[172] *See supra* Section IV.A.

[173] *See* GDPR, *supra* note 19, at art. 25, Recital 78.

consequence, it now appears that *ex ante* DPIAs—as opposed to *ex post* invocations of an individual "right to explanation"—are destined to "become the required norm for algorithmic systems, especially where sensitive personal data, such as race or political opinion, is processed on a large scale."[174]

The advantages of shifting the dialogue surrounding the GDPR's "right to explanation" from one involving individual remedies to one involving more general DPbD principles are manifold. First, mere algorithmic explicability is not the panacea it is often presumed to be.[175] As numerous experts of diverse backgrounds have noted, the reliance on transparency as an individualized mechanism often places excessive burdens on resource-constrained users to "seek out information about a system, interpret it, and determine its significance, only then to find out they have little power to change things anyway, being disconnected from power."[176] Though transparency may often feel like a robust solution intuitively, explainable artificial intelligence—or

[174] *See* Edwards & Veale, *supra* note 73, at 78 (quoting GDPR, art. 35(3)(b)) (internal quotations omitted) (arguing that DPIAs will soon become mainstream in enterprise); *see also, e.g.*, A29WP DPIA Guidelines, *supra* note 138. This prediction involving the rise of data auditing methodologies is also supported by additional legal mechanisms within the GDPR that, for purposes of concision, are not addressed by this Article. *See, e.g.*, GDPR, *supra* note 19, at art. 42 (requiring "the establishment of data protection certification mechanisms and of data protection seals and marks . . . available via a process that is transparent" and subject to regular review); *id.* at art. 40 (recommending that companies "prepare codes of conduct . . . such as with regard to . . . fair and transparent processing" and "to carry out the mandatory monitoring of compliance").

[175] *But see* Kasper Lippert-Rasmussen, *"We Are All Different": Statistical Discrimination and the Right to Be Treated as an Individual*, 15 J. ETHICS 47, 54 (2011)

> [O]btaining information is costly, so it is morally justified, all things considered, to treat people on the basis of statistical generalizations even though one knows that, in effect, this will mean that one will treat some people in ways, for better or worse, that they do not deserve to be treated.

See, e.g., Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085 (2018) (describing increasingly vocal pushes for transparency due to the intuitive, but not always correct notion, that explanations will resolve unfairness within algorithms).

[176] *See* Edwards & Veale, *supra* note 73, at 67 (quoting Mike Annany & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, NEW MEDIA & SOC'Y 1, 5 (2018)) (internal quotations omitted); *see also, e.g.*, FRANK PASQUALE, *supra* note 68 (arguing that transparency in and of itself does not translate to accountability in many contexts); Joshua Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 638 (2017) (rejecting transparency as a true remedy for promoting accountability); Brendan Van Alsenoy et al., *Privacy Notices Versus Informational Self-Determination: Minding The Gap*, 28 INT'L REV. L., COMPUTERS & TECH. 185, 185 (2014) (arguing that privacy notices don't necessarily achieve the accountability goals that many expect they will).

"XAI"[177] as it is increasingly called—is especially unlikely to provide significant remedial utility to individuals in instances where the discrimination involved is only observable at the statistical scale. Moreover, some commentators have convincingly argued that too great a focus on individualized explanations—as opposed to broader, multi-methodological design practices for mitigating unfairness—could "nurture a new kind of transparency fallacy . . . ."[178] Indeed, providing a basic explanation to individual users could provide false cover for companies whose processing operations may be biased for other reasons.

Second, providing enterprises a broader range of compliance options could allow them greater flexibility when deploying machine learning systems that may make more conventional forms of explicability impractical or impossible.[179] Under the current state of the art, many of the highest performing machine learning algorithms pose significant "tradeoff[s] between the representational capacity of a model and its interpretability."[180] Techniques capable of achieving the richest predictive results tend to do so through the use of aggregation, averaging, or multilayered techniques which, in turn, make it difficult to determine the exact features that play the largest predictive role.[181] Depending on the circumstances, performance losses associated with adopting a more explicable approach could prove far costlier than the social utility of providing individualized explanations.[182] Particularly in instances where the leading techniques far outpace the remedial options available to data subjects, a one-size-fits-all approach to oversight could lead to unnecessary bureaucratic

[177] *See* Tim Miller, Explanation in Artificial Intelligence: Insights from the Social Sciences (June 22, 2017) (unpublished manuscript).

[178] *See* Edwards & Veale, *supra* note 73, at 81 (internal quotations omitted); *see also, e.g.*, Toon Calders & Indrė Žliobaitė, *Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures*, *in* DISCRIMINATION AND PRIVACY IN THE INFORMATION SOCIETY 43, 46 (2013) ("[T]he selection of attributes by which people are described in [a] database may be incomplete.").

[179] *See supra* notes 68–71 and accompanying text.

[180] *See* Goodman & Flaxman, *supra* note 46, at 6. "Representational capacity" here refers, roughly, to the ability of an algorithm to make predictions that account for complex patterns, phenomenon, or inputs. Machine learning systems, especially those using deep neural networks, can give rise to models so complex that humans are unable to understanding precisely how the system arrives at a given decision or prediction.

[181] *See* Wojciech Samek et al., *Evaluating the Visualization of What a Deep Neural Network Has Learned*, 28 IEEE TRANSACTIONS ON NEURAL NETWORKS & LEARNING SYS. 2660, 2666–67 (2017); Marco Tulio Ribeiro et al., *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*, PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING 1135 (2016); Jon Kleinberg et al., *Human Decisions and Machine Predictions* (Nat'l Bureau of Econ. Research, Working Paper No. 23180, 2017).

[182] This, however, may eventually prove to be a moving target.

roadblocks for technologies with massively beneficial social impacts.[183]

Finally, and perhaps most importantly, system-wide audits of the type envisioned by DPIAs already have a well-documented track record of detecting and combating algorithmic discrimination in otherwise opaque systems. As Sandvig et al. note, audit studies are "the most prevalent social scientific methods for the detection of discrimination" in complex computational systems.[184] In recent years, these auditing techniques have been used by researchers and journalists to successfully detect and document algorithmic bias across diverse industry sectors and social domains.[185] Further, this approach includes the added benefit of allowing outside entities that may have more resources than individuals to scrutinize the integrity of complex computational systems. Regulators, NGOs, media outlets, and public interest organizations that specialize in this area will be able to invest in the expertise

[183] *See, e.g.*, Toon Calders & Sicco Verwer, *Three Naive Bayes Approaches for Discrimination-Free Classification*, 21 DATA MINING & KNOWLEDGE DISCOVERY 277 (2010) (describing trade-off between discrimination removal and classifier performance); Faisal Kamiran & Toon Calders, *Data Preprocessing Techniques for Classification Without Discrimination*, 33 KNOWLEDGE & INFO. SYS. 1 (2012) (describing trade-off between discrimination removal and classifier performance); Jagriti Singh & S. S. Sane, *Preprocessing Technique for Discrimination Prevention in Data Mining*, 4 INT'L J. ENGINEERING RES. & APPLICATIONS 54 (2014) (noting inherent trade-offs in the current state-of-the-art); Sam Corbett-Davies et al., Algorithmic Decision Making and the Cost of Fairness (June 2017) (unpublished manuscript). These tradeoffs will likely be a moving target. Indeed, Edwards & Veale note that the inevitability of these tradeoffs may only be "an interim conclusion" and are "convinced that recent research in ML explanations shows promise" for reducing or eliminating some of these tradeoffs. *See* Edwards & Veale, *supra* note 73, at 81.

[184] Christian Sandvig et al., Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms 5, 16 (May 22, 2014) (unpublished manuscript) (noting that the "audit study" is "the most prevalent social scientific method for the detection of discrimination" and that it is "considered to be the most rigorous way to test for discrimination in housing and employment"); Andrea Romei & Salvatore Ruggieri, *Discrimination Data Analysis: A Multi-Disciplinary Bibliography*, *in* DISCRIMINATION AND PRIVACY IN THE INFORMATION SOCIETY 109, 120 (2013); Faisal Kamiran, Toon Calders & Mykola Pechenizkiy, *Techniques for Discrimination Free Predictive Models*, *in* DISCRIMINATION AND PRIVACY IN THE INFORMATION SOCIETY 223, 223–24 (2013).

[185] *See generally, e.g.*, James Grimmelmann & Daniel Westreich, *Incomprehensible Discrimination*, 7 CALIF. L. REV. ONLINE 164, 173 (2017); FRANK PASQUALE, *supra* note 68; Mireille Hildebrandt, *The New Imbroglio - Living with Machine Algorithms*, *in* THE ART OF ETHICS IN THE INFORMATION SOCIETY 55 (Liisa Janssens ed., 2016); Kiel Brennan-Marquez, *"Plausible Cause": Explanatory Standards in the Age of Powerful Machines*, 70 VAND. L. REV. 1249, 1287 (2017); Andrew D. Selbst, *A Mild Defense of Our New Machine Overlords*, 70 VAND. L. REV. EN BANC 87 (2017); Reuben Binns, *Algorithmic Accountability and Public Reason*, 31 PHIL. & TECH. 543 (2018); Katherine Strandburg, N.Y. Univ. School of Law, Presentation at The Human Use of Machine Learning: An Interdisciplinary Workshop, Venice: Decision-Making, Machine Learning and the Value of Explanation (Dec. 16, 2016).

necessary not only to provide data subjects with the right answers but also to ensure that the right questions are asked.

Although data audit and DPbD methodologies come with their own unique set of challenges,[186] the multifaceted advantages[187] offered by these approaches present exciting new possibilities for fostering genuine algorithmic accountability in enterprises without stifling technological and business advances.[188] In contrast to a remedial "right to explanation" invoked on an individual basis by downstream data subjects, properly implemented auditing and DPbD can provide the evidence necessary to inform and vet the design and deployment of more fair, accountable, and transparent algorithmic systems.[189]

V. EXPORTING THE "RIGHT TO EXPLANATION": THE BRUSSELS EFFECT AND THE GDPR'S LONG TENTACLES

Although the EU is sometimes maligned as a declining force on the world stage, numerous recent studies have demonstrated that it actually exercises "unprecedented global power . . . through its legal institutions and standards that it successfully exports to the rest of the world . . . ."[190] This "export" effect

[186] *See* Bryce Goodman, A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection 7 (2017) (unpublished manuscript)

> [A] process that passes a safety audit may fail for other reasons (e.g., inefficiency). Passing a safety audit does not mean that all risk is eliminated but, rather, that risk is reduced to an acceptable level. Choosing an acceptable level of risk depends in turn on the process evaluated and, in particular, both the likelihood and severity of a failure.

See also Lior Jacob Strahilevitz, *Privacy Versus Antidiscrimination*, 75 U. CHI. L. REV. 363, 364 (2008).

[187] The list enumerated above is, of necessity, far from exhaustive.

[188] *See* Goodman, *supra* note 186, at 7.

[189] *See id.*; *see also* Anupam Datta et al., *Algorithmic Transparency via Quantitative Input Influence*, *in* TRANSPARENT DATA MINING FOR BIG AND SMALL DATA 71, 87–89 (Tania Cerquitelli et al. eds., Springer 2017).

[190] Anu Bradford, *The Brussels Effect*, 107 NW. U. L. REV. 1, 64 (2012); *see* Case COMP/M.5984, Intel/McAfee, SG-Greffe (2011) D/1407, C(2011) 529, EUR-Lex 32011M5984 (Jan. 26, 2011); *see also, e.g.*, Christopher Kuner, *The Internet and the Global Reach of EU Law* (LSE Legal Studies, Working Papers No. 4/2017, 2017); David Scheer, *Europe's New High-Tech Role: Playing Privacy Cop to the World*, WALL ST. J. (Oct. 10, 2003), https://www.wsj.com/articles/SB106574949477122300 [https://perma.cc/9LZK-XCZB]; Brandon Mitchener, *Rules, Regulations of Global Economy Are Increasingly Being Set in Brussels*, WALL ST. J. (Apr. 23, 2002), https://www.wsj.com/articles/SB1019521240262845360 [https://perma.cc/J8MS-DREP]; *Regulatory Imperialism*, WALL ST. J. (Oct. 26, 2007),

occurs through the process of "unilateral regulatory globalization." This entails a process whereby "a single state is able to externalize its laws and regulations outside its borders through market mechanisms, resulting in the globalization of standards."[191] Particularly in the last decades, the EU has evinced "a strong and growing ability to promulgate regulations that become entrenched in the legal frameworks of developed and developing markets alike" without relying on international institutions or intergovernmental negotiations.[192] This phenomenon has since come to be described as the "Brussels Effect."[193]

The following subparts explore this effect on enterprises seeking to comply with the EU's data protection mandate. Section A describes the DPD's influence as a global "gold standard" since 1995 as well as the potential consequences of this phenomenon for the GDPR's own global legacy. Section B then details the implications of the GDPR's "Brussels Effect" for individual enterprises and concludes by documenting some of the real-world impacts technology companies have already experienced.

A. DATA PROTECTION AND THE "BRUSSELS EFFECT"

There is, perhaps, no better exemplar of the "Brussels Effect" in action than the DPD itself, which has become a de facto standard for data privacy protection across the globe.[194] Since its enactment in 1995, more than thirty

http://online.wsj.com/article/SB119334720539572002.html [https://perma.cc/KT8A-RNCZ].

[191] Bradford, *supra* note 190, at 3, 18; *see, e.g.*, Daniel W. Drezner, *Globalization, Harmonization, and Competition: The Different Pathways to Policy Convergence*, 12 J. EUROPEAN PUB. POL'Y 841, 841–59 (2005) ("[A] . . . reasonable conjecture would be to say that the public good benefits from regulatory coordination depend upon the size of the newly opened market."); Beth Simmons, *The International Politics of Harmonization: The Case of Capital Market Regulation*, *in* DYNAMICS OF REGULATORY CHANGE: HOW GLOBALIZATION AFFECTS NATIONAL REGULATORY POLICIES 42, 50–52 (2001); David A. Wirth, *The EU's New Impact on U.S. Environmental Regulation*, 31 FLETCHER F. WORLD AFF. 91, 96 (2007) ("If [a] jurisdiction's market share is sufficiently large, [its] regulatory requirements can affect an even larger area, including those under the control of other sovereign authorities.").

> This process can be distinguished from political globalization of regulatory standards where regulatory convergence results from negotiated standards, including international treaties or agreements among states or regulatory authorities. It is also different from unilateral coercion, where one jurisdiction imposes its rules on others through threats or sanctions. Unilateral regulatory globalization is a development where a law of one jurisdiction migrates into another in the absence of the former actively imposing it or the latter willingly adopting it."

Bradford *supra* note 190, at 4.

[192] *See* Bradford, *supra* note 190, at 1.

[193] *See id.* at 3.

[194] *See id.*

countries have heeded Brussels' call by "adopt[ing] EU-type privacy laws, including most countries participating in the Organization for Economic Cooperation and Development."[195]

According to those who have studied the "Brussels Effect" closely, its underlying mechanics are relatively intuitive. Countries confronted with the EU regulations' stringent standards face a stark choice. They can either revise their own domestic policies to reflect those within Europe or risk breaking economic ties with the world's largest trading bloc.[196] For most, the decision requires little more than a moment's contemplation. Aside from a few notable outliers—such as the United States,[197] Russia, and China—most countries simply make the rational calculation that the costs of exclusion from a market consisting of 500 million of the globe's most affluent inhabitants far outweigh the costs of complying with Europe's higher standards.[198]

And lest those powerful incentives prove to be insufficient, the GDPR also includes a number of notable changes intended to promote extraterritorial compliance that are likely to extend its regulatory reach above and beyond the baseline already established by the "Brussels Effect." The most significant changes, in this realm, are those involving the Regulation's "adequacy decision" used to determine whether "third countries" (i.e., countries outside of the EU) have sufficient protections in place to warrant the transfer of personal data between themselves and EU Member States.[199] Once a country is deemed "adequate" through an assessment by the European Commission, data can flow freely without the need for additional protective measures.[200] But unlike the DPD, adequacy decisions made under the GDPR will be subject to a periodic review at least once every four years and will also be subject to

[195] *See id.* at 23.

[196] *See* David Bach & Abraham L. Newman, *The European Regulatory State and Global Public Policy: Micro-Institutions, Macro-Influence*, 14 J. EUROPEAN PUB. POL'Y 827, 831 (2007); Bradford, *supra* note 190, at 11–28. There are, of course, other factors that contribute to this effect. *See id.* at 11–19.

[197] *See* Bradford, *supra* note 190, at 13, 15.

[198] The EU's population exceeds 500 million, and its GDP per capita exceeds $35,000. *See Living in the EU*, EUROPEAN UNION, https://europa.eu/european-union/about-eu/figures/living_en [https://perma.cc/YD47-J682] (last visited Apr. 3, 2019); *European Union GDP Per Capita Ppp*, TRADING ECON., https://tradingeconomics.com/european-union/gdp-per-capita-ppp [https://perma.cc/QU8X-K74N] (last visited Apr. 3, 2019).

[199] *See* GDPR, *supra* note 19, at art. 45.

[200] *See id*; *see also, e.g.*, Press Release, European Comm'n, Questions & Answers on the Japan Adequacy Decision (July 17, 2018) (describing an adequacy decision as "a decision taken by the European Commission establishing that a third country provides a comparable level of protection of personal data to that in the European Union, through its domestic law or its international commitments").

repeal, amendment, or suspension on an ongoing basis.[201]

Thanks to the introduction of these far-reaching forms of regulatory oversight, the GDPR is already showing signs of its global standard-setting authority. Countries such as Israel, New Zealand, Argentina, and Japan have all recently undergone efforts to receive EU "adequacy" certifications by ensuring that their domestic data protections rise to the level of Europe's.[202] "Other countries, from Colombia to South Korea to the tiny island nation of Bermuda, are similarly rebooting [their] domestic legislation . . . [which at times] involves adopting European rules almost word for word."[203]

B. THE GDPR'S EFFECTS ON GLOBAL ENTERPRISE

Though the "Europeanization" of global regulatory standards is often most pronounced at the national level, a phenomenon like the one occurring on the global scale due to the "Brussels Effect" is also taking place within individual enterprises. According to a recent headline-grabbing announcement by Facebook, "[d]ozens of people at [the company] are working full time on" GDPR compliance—requiring upwards of a 250% increase in staffing related to EU data protection.[204] A company spokesperson noted:

> It is hard for us to put an exact figure on it, but when you take into account the time spent by our existing teams, the research and legal assessments and the fact that we have had to pull in teams from product and engineering, it is likely to be millions of dollars.[205]

Recent reporting by *The Financial Times* provided even further confirmation of this phenomenon. The media outlet—which contacted twenty "of the largest social media, software, financial technology and internet companies with EU operations"—noted that its inquiries "revealed that the sector is scrambling to hire new staff and redesign products as it faces millions of dollars in higher costs and lost revenues."[206] And while not every company has quite the multinational reach of the average tech giant, this extraterritorial effect is

[201] *See* GDPR, *supra* note 19, at art. 45.

[202] *See* Mark Scott & Laurens Cerulus, *Europe's New Data Protection Rules Export Privacy Standards Worldwide*, POLITICO (Jan. 31, 2018), https://www.politico.eu/article/europe-data-protection-privacy-standards-gdpr-general-protection-data-regulation/ [https://perma.cc/DRX2-Y9BZ].

[203] *Id.*

[204] Aliya Ram, *Tech Sector Struggles to Prepare for New EU Data Protection Laws*, FIN. TIMES (Aug. 29, 2017), https://www.ft.com/content/5365c1fa-8369-11e7-94e2-c5b903247afd [https://perma.cc/S6GS-RXPW].

[205] *Id.*

[206] *Id.* This phenomenon has led some experts to speculate that the "GDPR could be one of the most expensive pieces of regulation in the [technology] sector's history." *Id.*

made all the more pronounced by the GDPR's applicability to *any* company processing the data of EU citizens, not just those companies actually located within the EU itself.[207]

For some companies operating outside of the GDPR's immediate purview, it may be feasible to fragment their internal processing pipelines by treating data originating in Europe differently from that of other geographies. But doing so could prove administratively onerous and require multiple, separate handling processes for data flowing through any given enterprise. Moreover, this type of maneuver may also be perceived as a public relations risk for companies concerned about being "outed as deliberately offering a lower privacy standard to [their] home users [versus] customers abroad."[208] Thus, just as is true at the national level, the path of least resistance for many companies will likely entail treating the GDPR as the new "gold standard." Ultimately, the Regulation enforcement agencies will effectively dictate the way companies handle all personal data, regardless of geography.[209] While the precise contours of this new gold standard may be continuously revised, it is now clear that it includes a muscular "right to explanation" with sweeping implications for companies and countries throughout the world. As one commentator working to promote GDPR compliance as far away as South Africa recently noted, any entity not currently addressing it will soon realize that the "GDPR has long tentacles."[210]

VI. CONCLUSION

Now that the data protection authorities responsible for enforcing the GDPR's "right to explanation" have weighed in, at least one matter of fierce public debate appears closer to resolution. The GPDR's enforcement

[207] *See* GDPR, *supra* note 19, at art. 3 (describing the territorial scope of the Regulation as applying to any entities "processing . . . personal data of data subjects who are in the Union"); *see also, e.g.*, Goodman & Flaxman, *supra* note 46, at 2 (commenting that the GDPR's "requirements do not just apply to companies that are headquartered in the EU but, rather, to any companies processing EU residents' personal data . . . [thus] [f]or the purposes of determining jurisdiction, it is irrelevant whether that data is processed within the EU territory, or abroad"); Lomas, *supra* note 30 (noting "that GDPR does not merely apply to EU businesses; any entities processing the personal data of EU citizens need to comply").

[208] *See* Lomas, *supra* note 30.

[209] *See* GDPR, *supra* note 19, at art. 3 (describing the territorial scope of the Regulation as applying to any entities "processing . . . personal data of data subjects who are in the Union"); *The History of the General Data Protection Regulation*, EUROPEAN DATA PROTECTION SUPERVISOR, https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en [https://perma.cc/2SZ9-Y4ZP].

[210] Scott & Cerulus, *supra* note 202.

authorities envision a muscular "right to explanation" with sweeping legal implications for the design, prototyping, field testing, and deployment of automated data processing systems. Failing to countenance this right could subject enterprises to economic sanctions of truly historic magnitudes—a threat that simply did not exist under the GDPR's predecessor.

Although the protections enshrined by the right may not mandate transparency in the form of a complete individualized explanation, a holistic examination of the Regulation reveals that the right's true power derives from its synergies with other DPbD practices codified by the Regulation's subsequent chapters. While these new design standards will undoubtedly pose significant challenges for the enterprises that fall within the GDPR's purview, the speed and scale of the global response thus far are cause for genuine optimism. Indeed, there is perhaps no more hopeful bookend to this profoundly important debate than the recent words of Bryce Goodman, one of the authors responsible for first sparking the controversy: "In the past, companies have devoted immense resources to improving algorithmic performance. Going forward, one hopes to see similar investments in promoting fair and accountable algorithms."[211]

[211] Bryce Goodman, *supra* note 186, at 7.