Decision-Making by Machines: Is the ‘Law of Everything’ Enough?

Machines have moved from supporting decision-making processes of humans to making decisions for humans. This shift has been accompanied by concerns regarding the impact of decisions made by algorithms on individuals and society. Unsurprisingly, the delegation of important decisions to machines has therefore triggered a debate on how to regulate the automated decision-making practices. In Europe, policymakers have attempted to address these concerns through a combination of individual rights and due processes established in data protection law, which relies on other statutes, e.g., anti-discrimination law and restricting trade secret laws, to achieve certain goals. This article adds to the literature by disentangling the challenges arising from automated decision-making systems and focusing on ones arising without malevolence but merely as unwanted side-effects of increased automation. Such side-effects include ones arising from the internal processes leading to a decision (e.g., opacity), the impacts of decisions (e.g., discrimination), as well as the responsibility for decisions and have consequences on an individual and societal level. Upon this basis the article discusses the redress mechanisms provided in data protection law. It shows that the approaches within data protection law complement one another, but do not fully remedy the identified side-effects. This is particularly true for side-effects that lead to systemic societal shifts. To that end, new paradigms to guide future policymaking discourse are being explored.

Keywords: automated decision-making, artificial intelligence, data protection, transparency, fairness, due process

[1] Introduction

Machines have moved from supporting decision-making processes of humans to making decisions for humans. Such automated decision-making occurs not only in the private sector but also the public one. An often discussed example of decision-making by machines used by private actors is automated content removal software used to detect hate speech.¹ With the Covid-19 pandemic, we have also witnessed an increased reliance on automation in general and automated decision-making in particular by both private (e.g., employers) and public entities (e.g., public schools).² For instance, school officials have relied on software to determine final grades of

¹ Jennifer Cobbe and Jatinder Singh, ‘Regulating Recommending: Motivations, Considerations, and Principles’ (2019) *European Journal of Law and Technology* 10(3); Facebook uses the RoBERTa method (self-supervised natural language processing systems) cf. <<https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems/>> (last accessed 11 December 2020).

² Gemma Newlands, Christoph Lutz, Aurelia Tamò-Larrieux, Eduard Fosch Villaronga, Rehana Harasgame, and Gil Scheitlin, ‘Innovation under pressure: Implications for data privacy during the Covid-19 pandemic’ (2020) *Big Data & Society* 7(2).

students, which has led to massive public outcry.³ Of course, the increased use of automated decision-making has been under way before the pandemic. A recent report of Algorithm Watch on “Automating Society” maps the automating decision-making systems currently employed or in development within the European Union.⁴ From scraping social networking platforms to detect tax fraud, to the development of robot-judges in Estonia, the report ultimately warns from a techno-solutionist trap as the side-effects of such systems have yet to be addressed.⁵

Indeed, addressing the challenges of automated decision-making systems has become a key priority of policymakers. Yet, to unveil new regulatory mechanisms requires first the identification of the problems at hand and determining whether (and if not why) such issues are currently not sufficiently well addressed in existing regulation. In other words, when determining new regulatory strategies, the primary question should always be why is there a need for establishing new regulatory mechanisms. This question requires an analysis of the issues raised by automated decision-making with respect to their unwanted impact on the subject that policymakers intend to protect. Within the scope of this article, I focus on unintended side-effects of automated decision-making systems, meaning issues that arise even if the system does not malfunction. Side-effects are thus unintended harms on an individual or societal level distinct from malevolent actions by third parties. Side-effects occur partly because of the internal processing mechanisms of the technology itself, the social environment in which the technology is embedded, as well as the economic and political forces at play. Side-effects can be divided into ones that relate to the ways automated decision-making are reached, the impact of the decision-making process itself, and the question of responsibility for an automated decision. The effects can arise on an individual level (e.g., unfair treatment of a natural person) or on a societal level (e.g., echo-chamber effects on democracy). In other words, the unwanted impact of such side-effects can include the well-being of individuals and the society at large, the well-functioning of democracy and the economy, and the sustainability for the environment, among others.

In the case of automated decision-making within the private sector one core goal of policymakers is addressing the side-effects of automated decision-making that impact the well-being of individuals and society. This is not to say that the other subjects are less important; in fact, the balancing of conflicting interests plays a key role in addressing the side-effects of automated decision-making. One major regulation that provides legal tools to address said side-effects is the General Data Protection Regulation (GDPR). Through a combination of due processes, including its guiding principles, and individual rights, the GDPR regulates the

³ Cf. e.g., in the UK the Ofqual fiasco. Richard Adams, ‘Ofqual board minutes reveal tensions with DfE ahead of exam results fiasco’ *The Guardian* (London, 22 October 2020) <<https://www.theguardian.com/education/2020/oct/22/ofqual-board-minutes-reveal-tensions-with-dfe-ahead-of-exam-results-fiasco>> (last accessed 11 December 2020).

⁴ AlgorithmWatch, ‘Automating Society Report 2020’ (2020) <<https://automatingsociety.algorithmwatch.org/>> (last accessed 11 December 2020).

⁵ AlgorithmWatch, (above note 4).

development and deployment of automated decision-making processes and motivate the development of processes that take into account their impact on the right to data protection of the affected individuals. While time will tell how effective the means of the GDPR are as a remedy against the side-effects of automated decision-making, new paradigms have been proposed in the literature that policymakers should take into consideration.

Such newer proposals call for the establishment of *ex ante* legitimacy criteria for automated decision-making systems and a so-called “right to reasonable inferences.”⁶ It requires determining before a decision is reached what a reasonable decision-making by machines entails. Overall, such calls urge policymakers and ethicists to establish guidelines to determine the appropriateness of automated decision-making in various contexts. This first set of proposals move away from current approaches seen in the GDPR which focuses on the input data rather than the classification parameters and resulting decision-making processes. Similarly, other proposals move towards more collective protection mechanisms.⁷ Here too, arguments against a strict individual focus, dominant in data protection law, are brought forward and the social perspective and challenges raised by automated decision-making highlighted. Latter proposals also call for audits and investigation by appropriate authorities (e.g., data protection supervisory authorities) to support the individual rights-based approaches found in European law.

This article provides first a systemic analysis and literature review of the side-effects arising from automated decision-making, second, it elaborates upon how adequately those side-effects are being addressed by current data protection law, and third, it discusses future approaches to mitigate the still unaddressed side-effects of automated decision-making. To this end, this article is divided into six sections, including an introduction and conclusion: To set the stage, I start by describing the complexity of human and automated decision-making (Section 2). I then analyze the side-effects that automated decision-making systems pose to individuals, society, and regulators. These include issues relating to the quality of how automated decision-making reach conclusions, the concerns with respect to the actions that automated decision-making triggers, and issues relating to the lack of traceability of automated decisions (Section 3). Upon this basis, I discuss the tools the GDPR has to address these side-effects in practice and how different approaches within the GDPR complement one another (Section 4). Finally, I discuss new paradigms to address side-effects currently inadequately remedied within the current data protection framework (Section 5).

⁶ Sandra Wachter and Bernt Daniel Mittelstadt, ‘A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI’ (2019) *Columbia Business Law Review* 494, 515 et seq.; Giovanni Sartor and Francesca Lagioia, ‘The impact of the General Data Protection Regulation (GDPR) on artificial intelligence’ (2020) Study for the European Parliament, Brussels, <[https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf)> (last accessed 11 December 2020), 41.

⁷ Cf. e.g., Viktor Mayer-Schönberger, ‘Beyond Privacy, Beyond Rights—Toward a “Systems” Theory of Information Governance’ (2010) *California Law Review* 98, 1883; Dennis D. Hirsch, ‘From Individual Control to Social Protection: New Paradigms for Privacy Law in the Age of Predictive Analytics’ (2020) *Maryland Law Review* 79(2) 439; see for further references Section 5.

[2] The Complexity of (Automated) Decision-Making

This article focuses on the delegation of decision-making processes to algorithms, a core function of artificially intelligent systems.⁸ The term automated decision-making combines two distinct features: First, automated or algorithmic, which implies a scalable action, and second, decision-making, which implies that at least two options are available and that one is preferred over the other. The preference of one option over the other seems like a straightforward and trivial definition of a decision-making process. But the complexity of the decision-making process is already apparent when looking at human decision-making.⁹ Elements such as random occurrences (i.e., any situation that cannot be influenced by the decision-maker), and the goals, values, judgments, and biases of an individual affect his or her decision-making process.¹⁰ Likewise, human decision-making processes are influenced by emotions. Emotions occur in reaction to a specific interpretation of a situation, i.e., a specific appraisal of a situation which in turn triggers an emotion (frustration, anger, happiness). Emotions influence the decision-making process by providing an additional sort of information, and thereby enable quick responses; and moral emotions, which fulfil a social dimension (e.g., guilt, empathy), play an important factor in enabling altruistic (or egoistic) behavior.¹¹ The rich literature from various disciplines, from psychology, to philosophy, economics, to mathematics, unveiling the complex layers behind human decision-making processes, illustrate the not only the complexity of human decision-making but also the inherent biases and situational factors that create injustices.¹² In light of these side-effects of human decision-making, automated decision-making sometimes is seen as a potential remedy to the malleability (proneness to external factors) of a human mind. Yet these discussions, even if sometimes considering the side-effects of automated decision-making, assume the same scalability between human decisions and automated decisions or argue that on average, the side-effects of automated decision-making outweigh the ones of human

⁸ AlgorithmWatch, 'Automating Society: Taking Stock of Automated Decision-Making in the EU' (2019) <https://algorithmwatch.org/wp-content/uploads/2019/01/Automating_Society_Report_2019.pdf> (last accessed 11 December 2020), 9; Claude Castelluccia and Daniel Le Métayer, 'Understanding algorithmic decision-making: Opportunities and challenges' (2019) Study for the European Parliament, Brussel <[https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU\(2019\)624261_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf)> (last accessed 11 December 2020), 3-4.

⁹ Interesting research in this field include studies that introduce uncertainty and analyze how decision-making processes evolve. Famous research in this field stems from John von Neumann and Oskar Morgenstern that established game theory, as well as research focusing on the limitations of rationality (bounded rationality) and limitations of human decision-making processes. Cf. John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (60th-anniversary edition, Princeton University Press, 2004); for an overview on the limitations of rationality cf. Daniel Kahnemann, *Thinking, Fast and Slow* (Penguin, 2011).

¹⁰ Hans-Rüdiger Pfister, Helmut Jungermann and Katrin Fischer, *Die Psychologie der Entscheidung: eine Einführung* (4th edition, Springer, 2017), 2 et seq.; John Zerilli, Alistair Knott, James Maclaurin, and Colin Cavaghan, 'Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?' (2019) 32 *Philosophy & Technology* 661, 674 et seq.

¹¹ Pfister, Jungermann and Fischer (above note 10), 301 et seq.; Jonathan Baron, *Thinking and Deciding* (4th edition, Cambridge University Press, 2008), 67 et seq., 250 et seq. and 451 et seq.

¹² An often-cited example here is the study of Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso, 'Extraneous factors in judicial decisions' (2011) *Proceedings of the National Academy of Sciences of the United States of America*, illustrating how judge rulings also depend on the time of the day in which the proceedings were heard and decided upon.

decision-making. Before dwelling upon the various side-effects arising from automated decision-making, let me first describe the term automated decision-making and the perception of individuals over of such processes.

As the word suggests, automated decision-making entails that the decision-making process is ‘taken over’ by an algorithm.¹³ An algorithm describes a specific computational procedure that achieve a desired input/output relationship.¹⁴ We typically understand an algorithm to be a set of mathematical rules, implemented into a technology in order to conduct a particular task.¹⁵ In that sense, they fulfill “certain human purposes in a specifiable and reproducible way.”¹⁶ While traditionally algorithms relied on individually defined rules and weights, today they rely more and more on machine learning capabilities. Learning algorithms are best described as general templates with modifiable parameters; the algorithm can adjust the parameters of the model “by optimizing a performance criterion defined on the data.”¹⁷ By doing so, such an algorithm can incrementally improve its performance over time, thus enabling it to come up with novel patterns and predictions.¹⁸ Such automated decision-making processes become especially interesting from an ethical and legal point of view when we consider algorithms that “make generally reliable (but subjective and not necessarily correct) decisions based upon complex rules that challenge or confound human capacities for action and comprehension.”¹⁹ Examples of such decision-making algorithms include filtering and recommendation systems, and online dispute resolution systems.²⁰

While the attractiveness of learning algorithms is their adaptability, the overall attractiveness of automated decision-making lies in their scalability and ability to replace mundane (currently human-performed) tasks. Even if automated decision-making can be seen as supplementing human decision-making, a key goal of automation has always been to delegate how to arrive to a decision to machines. Thus, it will be key to understand the perceptions of individuals being affected by a (automated) decision-making. Indeed, social scientists are currently conducting research on how automated decision-making is perceived in contrast to

¹³ Note that the discourse on automated decision-making is not new. Fundamental research has already been proposed a.o. by Helen Nissenbaum, ‘Accountability in a Computerized Society’ (1996) *Science and Engineering Ethics* 2(1), 25; Batya Friedman and Helen Nissenbaum, ‘Bias in Computer Systems’ (1996) *ACM Transaction on Information Systems* 14(3), 330.

¹⁴ Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest and Clifford Stein, *Introduction to algorithms* (3rd edition, MIT Press, 2009), 5; cf. also Harvey Brooks, ‘Technology, evolution, and purpose’ (1980) *Deedalus* 109(1) 65; Micheal Veale, ‘Governing Machine Learning that Matters’, (Doctoral thesis, University College London, 2019), 27 et seq.

¹⁵ Robin K. Hill, ‘What an Algorithm Is’ (2016) *Philosophy & Technology* 29, 35, 47.

¹⁶ Brooks (above note 14), 66; Veale (above note 14), 29.

¹⁷ Ethem Alpaydin, *Machine Learning* (MIT Press, 2016), 24-25.

¹⁸ Martijn van Otterlo, ‘A machine learning view on profiling’, in Mireille Hildebrandt and Katja de Vries, *Privacy, Due Process and the Computational Turn: The Philosophy of law meets the philosophy of technology* (Routledge, Taylor & Francis Group, 2013), 41-64.

¹⁹ Bernt Daniel Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi, ‘The Ethics of Algorithms: Mapping the Debate’ (2016) *Big Data & Society* 1, 3.

²⁰ Cobbe and Singh (above note 1); Mittelstadt et al. (above note 19).

human decision-making. With respect to the decision-maker, studies seem to indicate that “people will perceive algorithmic decision-makers as more rational, and less intentional and emotional than people.”²¹ The perception of how accurate an automated decision-making system is depends significantly on the characteristics of the task assigned to the machine.²² However, additional research on the perceptions of automated decision-making is required to gain a clearer and more robust picture of this phenomenon.

Like much of the literature on artificial intelligence, automated decision-making as a core functionality of intelligent systems, typically refers and makes parallels to human intelligence and human decision-making. This is not surprising, since often mimicking and augmenting human intelligence is seen as the overall goal for general artificial intelligence.²³ This is also why arguments for more automated decision-making highlight its benefits with respect to human decision-making. However, the projection of humanoid functions (such as thinking and judging when it comes to decision-making) loads automated decision-making with values. Instead of perceiving automated decision-making as results of complicated mathematical functions, we attach (autonomous) judgment to their decisions. This complicates the relationship between humans confronted with automated decision-making systems and alter the perceptions towards such systems. Importantly, attaching autonomous judgment, and thereby agency, to automated decision-making system alters our perception towards accountability. We might wrongly “blame” automated decision-making systems, instead of holding the institution in which they are developed responsible for their programming and data selection choices.²⁴ This is why it is key to understand the side-effects automated decision-making systems raise and how they can be addressed, in order to ensure that individual and societal needs are being taken care of.

[3] Side-Effects of Automated Decision-Making

Policy analysis requires the identification of the problem at hand that would justify a regulatory reaction. Such a justification exists if the issue is currently not addressed, or not sufficiently well addressed, in current regulation.²⁵ Within the scope of this article I am interested in the unintended side-effects of automated decision-making systems,²⁶ i.e., issues that arise even when the systems work as they are supposed to work, and impact individual and societal well-being. This restriction

²¹ Min Kyung Lee, ‘Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management’ (2018) *Big Data & Society* 1, with references to various studies in the field of perception of decisions by robots.

²² Lee (above note 21), 8.

²³ Cassio Pennachin and Ben Goertzel, ‘Contemporary Approaches to Artificial General Intelligence’ (2007) in B. Goertzel and C. Pennachin. *Artificial General Intelligence*, pp. 1-28, Springer.

²⁴ Alicia Solow-Niedermann, ‘Administering Artificial Intelligence’ (2019) 93 *South California Law Review* 633, 635.

²⁵ Cf. Eugene Bardach, *A Practical Guide for Policy Analysis: The Eightfold Path to More Effective Problem Solving* (3rd edition, CQ Press, 2009).

²⁶ In the literature side-effects are also being discussed as challenges, concerns, issues, harms, and risks; cf. on unintended consequences Anja Lambrecht and Catherine Tucker, ‘Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads’ (2019) *Management Science* 65(7), 2947.

of scope does not imply that intended harms, i.e., issues that arise because of malevolent actions by third parties, are not inherently problematic. And I acknowledge that clear-cut distinction between intended and unintended²⁷ issues and correct and incorrect working of a system might not always be possible. Nonetheless, I believe that the restriction of scope and framing of issues is necessary, as the regulatory means to address different problems vary. Within this article I thus only zoom into the problems that do not require malevolence of developers or third parties but occur irrespectively of the intention, partly because of the technology itself, the (social) environment the technology is embedded in, and economic forces at play.

The below discussed side-effects can be divided into ones that relate to the ways automated decision-making are reached (i.e., the internal processes), the question of explainability of and accountability for an automated decision, and the impact of the decision-making process itself. The effects vary depending on what level one focuses on: While on an individual level, the natural person affected by a decision, unfair treatment, and an affected person's feeling of being dehumanized are at the forefront of the discussion, on a societal level the effects of an ever increasing granularized world on the social fabric and democracy are central tenants of the debate. Latter effects lead to systemic shifts that, as will be shown below, are inadequately addressed by current regulations (Section 4).

On the Quality of Decisions Made by Algorithms

Today algorithms can draw conclusions from data using machine-learning or inferential statistics processes. These processes lead to “probable yet inevitably uncertain” conclusions,²⁸ which “can only be as reliable (but also as neutral) as the data they are based on,” meaning that the quality and quantity²⁹ of the input data is key for ‘reliable’ and ‘neutral’ decisions (informally known as the “garbage in, garbage out” principle).³⁰ But even if the data itself is ‘reliable,’ instances arise where correlating certain data points against others is just not reasonable to do.³¹ Of course, while the validity of a correlation based on two data points seems straightforward, automated decision-making systems complicate the issue by blurring the mechanisms of how multiple data points contributed to the conclusion.

²⁷ For instance, a developer might not have wanted to develop a system that discriminates based on sex but did not take steps to ensure that the dataset used to train a system was representative. In this scenario it is thus highly questionable whether the action of the developer should be considered unintended or, put differently, how to demark the line between intended and unintended.

²⁸ Mittelstadt et al. (above note 19), 4.

²⁹ E.g., decision-making systems that abstract the decision-making process to weigh only certain data points against each other are likely to neglect the broader aspects that should be accounted within their decision-making process.

³⁰ Mittelstadt et al. (above note 19), 5.

³¹ Alpaydin (above note 17), 25 who argues that developers must first reflect if it makes sense to assume underlying rules within a data set, stating that employing machine learning to “learn” about the relationship of phone numbers and names would make no sense.

A decision made by an algorithm can lead to unjustified and biased actions, as such actions are triggered without relying on a causal relationship.³² Algorithms do not have to rely on causality to trigger an action, but merely “establish a sufficiently reliable evidence base to drive action.”³³ In particular with learning-algorithms the evidence base can be updated with new data which makes it even harder to debate whether the categories relied upon for a decision are sensible. In other words, the adaptability of a system challenges the very notion of how we currently justify individual actions or determine the quality of a decision.

Moreover, the development of automated decision-making system is not neutral (as choices with respect of the inner workings have to be made by developers) and the algorithms have to be trained with datasets of past or present data at best.³⁴ While social biases can be embedded in system design either consciously or unconsciously,³⁵ other biases can arise from technological constraints or actual flaws in the design of the system. Unjustified and biased actions can be hard to detect, as categories determined by algorithms typically are accepted by individuals as true and discourage the exploration of alternative models.³⁶ As we rely on algorithms and defer decisions to machines, we start to more and more accept said decisions as correct, and do not feel responsible for the outcome.³⁷ The lack of challenging automated decision-making and sometimes even the lack of feeling responsible on an individual level ties in with the discussion on the perceptions of automated decision-making described above (see Section 2) and impacts society overall. Interestingly, we see that on the one hand, automated decision-making is seen as superior to human decision-making and thus more strongly relied upon; and on the other hand, the lack of ‘humanness’ challenges how individuals feel about an automated decision-making process.

Ultimately, these effects are linked to the value-laden descriptions of automated decision-making systems and the parallels made to human decision-making (see Section 2). In fact, on an individual level, mimicking human decision-making would require algorithms to ensure that individuals are given the feeling of being heard and listened to. These feelings are inherently related to a relational dimension in human-to-human interaction. Yet, automated decision-making processes are decoupled from human discretion. While this may reduce arbitrariness and

³² Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, (John Murray, 2013), 50 et seq.; cf. also Emre Bayamlioglu, ‘Contesting Automated Decisions: A View of Transparency Implications’ EDPL 2018(4), 433.

³³ Mittelstadt et al. (above note 19), 5.

³⁴ Cf. Castelluccia and Métayer (above note 8), 7 with further references; cf. also Friedman and Nissenbaum (above note 13), 330 et seq.; Zerilli et al. (above note 10), 661 et seq.

³⁵ Friedman and Nissenbaum (above note 13), 330 et seq.

³⁶ Mike Ananny, ‘Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness’ (2015) *Science, Technology, & Human Values* 41(1), 93, 100 et seq.

³⁷ Tal Zarsky, ‘The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making’ (2016) *Science, Technology, & Human Values* 41(1), 118, 121 with reference to Danielle Keats Citron, ‘Technological Due Process’ (2007) *Washington University Law Review* 85(6), 1249; Isak Mendoza and Lee A. Bygrave, ‘The Right Not to Be Subject to Automated Decisions Based on Profiling’ in Tatiana-Eleni Synodinou, Philippe Jougoux, Christiana Markou, and Thalia Prastitou (eds.), *EU Internet Law: Regulation and Enforcement* (Springer, 2017), 77-98.

(conscious/sub-conscious) prejudices, the virtues of human judgment in a decision-making are also eliminated. These virtues include values we attach to human-to-human relationships, such as the ability to listen to each other and empathize with others. Through automated decision-making the relational dimension of a decision-making process, between the decision-maker and affected individual or group, and the experience of being listened to as well as the societal value that arises from hearing, listening, and empathizing with an individual circumstance are removed.³⁸ On a societal level, automated decision-making major side-effect is its potential to amplify unjust and biased decisions. While an individual can only make a finite number of decisions in his lifetime, algorithms make decisions within milliseconds and impact the life of millions of individuals. Moreover, because of the architectural vulnerabilities of technical systems, they are being prone to attacks.³⁹

Traceability and Responsibility of Automated Decision-Making

Automated decision-making systems are often opaque and unpredictable, which make them difficult to oversee. As different underlying learning models are used to determine the relationship between a certain input and output (or the observation of interests), various factors can contribute to the opaqueness and unpredictability of such decision-making systems, which are often portrayed as black boxes.⁴⁰ One such factor is often the sheer amount of parameters within the model (e.g., a linear model with multiple parameters).⁴¹ Another factor is the ability of machine learning algorithms to evolve and modify rules on how to cluster the data inputs received, meaning that the decision-making logic can change with every new input.⁴² This makes machine learning models become inscrutable and nonintuitive.⁴³ Aside from these technical underpinnings, legal frameworks (in particular trade secret laws) enable companies employing machine learning algorithms to maintain their competitive advantage and keep the complex logic of the decision-making systems secret.⁴⁴ But even if the logic of the decision-making systems was accessible, it would most likely not lead to any advantage for non-tech-savvy users.

Scholars have argued that transparency is not the savior against the black box phenomenon and that accessibility to the functioning of automated decision-making processes

³⁸ Karen Yeung, 'Why Worry about Decision-Making by Machine?' in Yeung K. & Lodge M. (eds.), *Algorithmic Regulation* (Oxford: University Press, 2019), 21-48, 29 et seq.

³⁹ Castelluccia and Métayer (above note 8), 7 et seq.

⁴⁰ Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2015).

⁴¹ Zachary C. Lipton, 'The Mythos of Model Interpretability' (2018) *ACM Queue*, 16(3), 1, 14.

⁴² Jena Burrell, 'How the machine 'thinks': Understanding opacity in machine learning algorithms' (2016) *Big Data & Society*, 1, 5.

⁴³ Andrew D. Selbst and Solon Barocas, 'The Intuitive Appeal of Explainable Machines' (2018) *Fordham Law Review* 87(3), 1085, 1090; Bayamlioglu (above note 32), 433 et seq.

⁴⁴ Burrell (above note 42), 3 et seq.; Rob Kitchin, 'Thinking critically about researching algorithms' (2017) *Information, Communication and Society* 20(1), 14, 20; Selbst and Barocas (above note 43), 1091.

might lead tech-savvy users to employ their knowhow to game the system in their favor.⁴⁵ In fact, relying merely on the informational dimension of transparency and the ideal of individual control (i.e., decide if one wants to be subject to automated decision-making such as recommender systems) has shown that it has its limits. Not only because putting the onus on the individuals in such respect seem negligent in light of the difficulties to weigh the individual as well as societal risks against more short-term benefits, but also due to the side-effects arising on a societal level resulting from the lack of accountability. While some argue that transparency disclosures could be useful if they were targeted to an oversight board trained in understanding automated decision-making processes,⁴⁶ others are more critical, stating that even experts will not be able to oversee complex machine-learning systems, as not even the developers sometimes can trace back how decisions were reached. Moreover, the question of who would examine automated decision-making systems and how they would be evaluated has still to be debated. What has become clear though, is that the mere focus on one-size fits-all transparency as information or verifiability is too short-sided. A more nuanced approach which considers the target groups of information and their respective needs is necessary. More focus should also be put on the performative element of transparency (i.e., practices of how transparency is being performed or not performed by means of dark patterns or the conscious provisions of an overwhelming amount to information to discourage from engaging with the information), as we need to ensure that there is no meaningful confrontation and engagement with the provided information.

The question of transparency is closely tied to the inhibition of oversight, which in turn raises concerns of accountability.⁴⁷ The inhibition of oversight fundamentally implies that informational and procedural justice cannot be guaranteed by algorithmic decision-making processes.⁴⁸ We typically assume that a developer understands and can reflect upon the impact of a system and make design choices. In a perfectly accountable world, such reflections and design choices are traceable, in order to be able to hold developers responsible for their actions. While such an allocation of responsibility might be conceivable when decision-making rules are ‘hand-coded’ for non-learning algorithms, traceability and responsibility are harder to come by for algorithms with learning capacities.⁴⁹ The lack of traceability might lead to distrust in the process and a decrease in confidence that a system is legitimate and reasonable.⁵⁰ The concerns of an

⁴⁵ Zarsky (above note 37), 124; Mike Ananny and Kate Crawford, ‘Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability’ (2018) *New Media & Society* 20(3), 973.

⁴⁶ Tal Zarsky, ‘Transparent Predictions’ (2013) *University of Illinois Law Review*, 1504, 1533 et seq.

⁴⁷ Burrell (above note 42), 1 et seq.; Selbst and Barocas (above note 43), 1093 et seq.

⁴⁸ Cf. Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao and Nigel Shadbolt, ‘It’s Reducing a Human Being to a Percentage’; Perceptions of Justice in Algorithmic Decisions’ (2018) *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, USA, Paper 377, 1 with reference to Jason A. Colquitt, Donald E. Conlon, Michael J. Wesson, Christopher O. L. H. Porter and K. Yee Ng, ‘Justice at the Millennium: A Meta-Analytic Review of 25 Years of Organizational Justice Research’ (2001) *Journal of Applied Psychology* 86(3), 425.

⁴⁹ Mittelstadt et al, (above note 19), 6; cf. also Andreas Matthias, ‘The responsibility gap: Ascribing responsibility for the actions of learning automata’ (2004) *Ethics and Information Technology*, 6, 175, 181.

⁵⁰ Zerilli et al. (above note 10), 663.

accountability gap are further fueled by the rise of information asymmetries between a few data-rich tech companies and the users of their services.⁵¹ It becomes clear from the discussion that allocating accountability does not only impact an individual person subject of an automated decision-making process but has much broader societal consequences. These societal consequences arise when confidence in the overall automated decision-making system's legitimacy and reasonableness dwindles. The lack of confidence can result in social unrest (see Section 1 and 5) and potentially lead to an increased distrust towards automated decision-making. It will become a key challenge to ensure trust in automation. While policy initiatives promoting trust in automated decision-making systems⁵² have gained momentum, we are still at the forefront of understanding how regulation (soft and hard) can impact the development of automated decision-making systems that are reasonable and take the societal impacts into account (see Section 5).

On the Impact of Decisions Made by Algorithms

Prominent concerns raised within the literature are the discriminatory effects of decisions (i.e., decisions that disproportionately, negatively impact a specific group of individuals). Even if strategies to mitigate such side-effects exist –e.g. having an algorithm exclude specific types of data (such as data related to gender or ethnicity) from consideration⁵³—it remains difficult to predict the proxies of such protected attributes.⁵⁴ In fact, multiple researchers have pointed out that even supposedly ‘neutral’ data types can overlap with protected data categories.⁵⁵ From a data protection perspective ‘neutral’ data types might still be considered personal data, providing thus the data subject with some rights (see Section 4). Yet, ‘neutral’ data types could also include transiently processed data, anonymously processed data via cryptographic tools (e.g., homomorphic encryption), and non-personal data altogether. While the processing of such data can still overlap with protected data categories and thus lead to biased automated decision-making, individuals are left with no ability to voice opposition to said processing under data protection law, which in turn impacts groups of individuals in potentially negative ways. These

⁵¹ Bruno Lepri, Nuria Oliver, Emmanuel Letouzé, Alex Pentland and Patrick Vinck, ‘Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges’ (2018) *Philosophy and Technology* 31(4), 611, 619 et seq.

⁵² High-Level Expert Group on Artificial Intelligence, ‘Ethics Guidelines for Trustworthy AI’ (2019) available <<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>> (last accessed December 11, 2020)

⁵³ Cf. Solon Barocas and Andrew D. Selbst, ‘Big Data’s Disparate Impact’ (2016) *California Law Review* 104(3), 671, 721 et seqq.; Toon Calders, Faisal Kamiran and Mykola Pechenizkiy, ‘Building classifiers with independency constraints’ (2009) *IEEE International Conference on Data Mining Workshops* 13; Faisal Kamiran and Toon Calders, ‘Data preprocessing techniques for classification without discrimination’ (2012) *Knowledge and Information Systems* 33, 1; Bart W. Schermer, ‘The limits of privacy in automated profiling and data mining’ (2011) *Computer Law & Security Review* 27(1), 45, 49.

⁵⁴ Monique Mann and Tobias Matzner, ‘Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination’ (2019) *Big Data & Society* 1; Zarsky (above note 37), 123.

⁵⁵ Mann and Matzner (above note 54), 1 et seq.; Mittelstadt et al. (above note 19), 8; Wachter and Mittelstadt (above note 6) 515 et seq.; Zarsky (above note 37), 118 et seq.

processes show the shift from individual to societal side-effects: Even if automated decision-making is not ‘interested’ in the individual recipient of a decision, larger societal challenges arise that have yet to be addressed by policymakers (see Section 5).

In addition, inferences raise the fear of objectification, i.e., the reduction of individual traits into digital profiles and statistical equations. While the reduction of individuals into profiles has the ‘advantage’ of enabling systems to decide what personalized information to provide individuals instead of overloading them with ‘irrelevant’ content, but simultaneously reduces information diversity and serendipity.⁵⁶ Information diversity prevents the creation of ‘echo-chambers’ and reduces the likelihood of influencing (or radicalizing) individuals towards one cause. Determining what information is relevant to a specific group of people is “inherently subjective;” if the information provided to an individual reflects more the interests of third parties (e.g., advertisers or trolls), then the autonomy of the individual in making his or her own decision is at the very least devalued.⁵⁷ Disrespecting an individual’s autonomy becomes particularly problematic when algorithmic decision-making is not merely used to nudge individuals (e.g., towards a healthier lifestyle), but becomes a hidden influence that subverts an individual’s decision-making power by exploiting said individual’s (cognitive or affective) weaknesses.⁵⁸

Moreover, being constantly surrounded by data processing devices—from wearables to smartphones—and the resulting steering of decision-making through algorithms, could lead to the chilling of certain behaviors both in online and offline contexts; this may occur in the form of self-censorship, self-restraint, or as silencing effects.⁵⁹ It has not been established by empirical research if profiling activities of corporations by means of algorithmic systems create chilling effects. Nonetheless, evidence from state surveillance studies suggests that individuals indeed alter their behaviors online in light of perceived fear of tracking.⁶⁰ We seem to be only at the beginning of understanding such larger societal challenges arising from the impacts of automated decision-making. While a lot of academic focus has rested for instance on echo-chambers and (to a lesser extent) the formation of chilling effects, the policy discourse is still ongoing. This is not surprising, as moving beyond an individual control approach will require a substantial reframing of the current policy strategy. Yet, the described characteristics of automated decision-making

⁵⁶ Christoph Lutz, Christian Pieter Hoffmann and Miriam Meckel, ‘Online serendipity: A contextual differentiation of antecedents and outcomes’ (2017) *Journal of the Association for Information Science and Technology* 68(7), 1698; Sue Newell and Marco Marabelli, ‘Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of ‘datafication’ (2015) *Journal of Strategic Information Systems* 24, 3; Eli Pariser, *The Filter Bubble: What The Internet Is Hiding From You*, (Penguin, 2011).

⁵⁷ Mittelstadt et al. (above note 19), 9.

⁵⁸ Daniel Susser, Beate Roessler and Helen Niessenbaum, ‘Online Manipulation: Hidden Influences in a Digital World’ (2019) *Georgetown Law Technology Review* 4, 1, 12 et seqq.

⁵⁹ Moritz Büchi, Eduard Fosch-Villaronga, Christoph Lutz, Aurelia Tamò-Larrieux, Shruthi Velidi and Salome Viljoen, ‘The chilling effects of algorithmic profiling: Mapping the issues’ (2020) *Computer Law & Security Review* 36.

⁶⁰ Büchi et al. (above note 59) with a literature review on the chilling effects caused by state and corporate surveillance.

(see Section 2) will necessitate such a larger discussion that takes societal impacts such as deterrence effects and behavior manipulation into account.

[4] “The Law of Everything”⁶¹ to the Rescue?

With the replacement of the Data Protection Directive⁶² of 1995 (Directive 95/46/EC) by the General Data Protection Regulation⁶³ (GDPR), one notices an increase in awareness of policymakers regarding automated decision-making. This focus is not surprising when one considers the overall evolution of data protection law in Europe.⁶⁴ In fact, despite policymakers’ intent for data protection law to be technologically neutral, it is not possible to decouple the establishment of new data protection norms from the technical and societal environment present at the time they were drafted. The rise of automated decision-making systems therefore led regulators to include norms within the revised data protection act that specifically target the unwanted side-effects of automated decision-making. The following sub-sections take a closer look at how the GDPR, in conjunction with other legislation, addresses these side-effects described throughout Section 3 in practice. To do so, I will elaborate on different provisions within the GDPR that contain redress mechanisms to the side-effects mentioned above. As will be seen, data protection law, which relies on other statutes, e.g., anti-discrimination law and restricting trade secret laws, cannot address all side-effects mentioned. As it is outside the scope of this article to explore further redress mechanisms (such as educational and technical tools), side-effects, for instance, arising from a wrongful perception on automated decision-making (e.g., expecting superior decision-making by machines), which would necessitate digital literacy measures will not be addressed.

The following sub-sections mirror the side-effects described above. With respect to the quality of decisions made by algorithms, the fairness and accuracy principle will be discussed as means to address unjust and biased decision-making by machines. Linked to it is the transparency principle and information duties provided within the law that can address opacity issues surrounding automated decision-making and the question of traceability and responsibility. Lastly, individual rights and the mix between the control paradigm and oversight mechanisms might help address some of the impacts of decisions made by algorithms. The discussion will

⁶¹ Nadezhda Purtova, ‘The law of everything. Broad concept of personal data and future of EU data protection law’ (2018) *Law, Innovation and Technology* 10, 40.

⁶² EU Directive 95/46: 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, OJ 1995 L 281/31.

⁶³ EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of 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), OJ 2016 L 119/1.

⁶⁴ Aurelia Tamò-Larrieux, *Designing for Privacy and its Legal Framework: Data Protection by Design and Default for the Internet of Things* (Law, Governance and Technology Series, Springer, 2018), 78 et seqq. relying on an analysis by Viktor Mayer-Schönberger, ‘Generational Development of Data Protection in Europe’ in Philip E. Agre & Mark Rotenberg (eds.), *Technology and privacy: the new landscape* (Cambridge: MIT Press, 1997), 219-236.

show that while the GDPR has a palette of tools to address the side-effects, some of the fundamental principles would need to be interpreted in broader terms to provide adequate measures.

Fair Automated Decision-Making

The principle of fairness is inherently connected to the principle of lawfulness and transparency.⁶⁵ Unlike lawfulness, which is specified in Article 6 of the GDPR, and transparency, which is also established through numerous articles (see below Section on Transparent Automated Decision-Making), fairness is less clearly defined and has thus been subject of numerous research.⁶⁶

Fairness is intuitively understood as the lack of biased differential treatment.⁶⁷ The link between fairness and non-discrimination is now established within Recital 71 of the GDPR and through opinions of data protection authorities.⁶⁸ Anti-discrimination law in the EU is heterogenous. Next to Art. 21 of the European Charter for Fundamental Rights (EUCFR), norms in multiple directives that set a minimum standard for national legislation in the different EU member states.⁶⁹ These legislations protect from both direct discrimination (i.e., treating an individual differently based on at least one protected class information), and indirect discrimination (i.e., rules that apply based on seemingly neutral criteria, yet affect certain groups with a protected characteristic adversely). Indirect discrimination is a newer concept, and less case law exists regarding when indirect discrimination has been the ground to abolish a certain policy practice.⁷⁰ In general though, a claimant must be able to demonstrate the harm that occurred because of the (direct or indirect) discrimination, how the harm manifests itself within the protected group, and in the case of indirect discrimination, show the disproportionality of the harm. Furthermore, an indirect discrimination claim requires a group level comparison, rather

⁶⁵ In fact, Art. 5(1)(a) GDPR reads “processed lawfully, fairly and in a transparent manner (...)”; Gianclaudio Malgieri, ‘The concept of fairness in the GDPR: A linguistic and contextual interpretation’ (2020) *Proceeding of the Conference on Fairness, Accountability, and Transparency (FAT*)*, 154, 155; cf. also Article 29 Working Party, ‘Guidelines on consent under Regulation 2016/679’ (WP 259, rev.01, 10 April 2018), 12 et seq.

⁶⁶ Cf. a.o. Lepri et al. (above note 51), 611 et seq.; Malgieri, (above note 65), 154 et seq.; Sandra Wachter, Berndt Daniel Mittelstadt and Chris Russell, ‘Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI’ (2019) unpublished version available on SSRN <<https://ssrn.com/abstract=3547922>>, 1 et seq.

⁶⁷ Lepri et al. (above note 51), 615.

⁶⁸ Article 29 Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of the Regulation 2016/679’ (WP 251, rev.01, 6 February 2018), 10; Commission Nationale de l’Informatique et des Libertés (CNIL), ‘Algorithms and artificial intelligence: CNIL’s report on the ethical issues’ 25 May 2018 <<https://www.cnil.fr/en/algorithms-and-artificial-intelligence-cnils-report-ethical-issues>> (last accessed 11 December 2020).

⁶⁹ E.g., Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation, OJ L 303, 2.12.2000, p. 16–22; Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, OJ L 180, 19.7.2000, p. 22–26; Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services, OJ L 373, 21.12.2004, p. 37–43; Directive 2006/54/EC of the European Parliament and of the Council of 5 July 2006 on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (recast), OJ L 204, 26.7.2006, p. 23–36; Proposal for a Council Directive on implementing the principle of equal treatment between persons irrespective of religion or belief, disability, age or sexual orientation {SEC(2008) 2180} {SEC(2008) 2181} /* COM/2008/0426 final - CNS 2008/0140 */.

⁷⁰ Wachter, Mittelstadt and Russell (above note 66), 15.

than an individual level one.⁷¹ The understanding of fairness as non-discrimination has also motivated technical researchers to design systems that can test for discrimination and mitigate discriminatory impact.⁷² Yet, at least from a European law perspective, these tools seem to be unable to take into account the highly contextual and flexible interpretation of anti-discrimination law provided by courts.⁷³

As mentioned above, the historical analysis of the fairness principle also shows its inherent link with the lawfulness and transparency and principles stated in Article 5(1)(a) of the GDPR.⁷⁴ An analysis of different language versions of EUCFR even unveils how depending on the language of the member state the term fair in Art. 8(2) EUCFR overlaps with the term lawful.⁷⁵ When combined with the principle of lawfulness, fairness can be interpreted as procedural fairness (i.e. procedures ascribing a level of burden of care to the processing).⁷⁶ Procedural fairness requires the set in place of abstract yet clear enough procedures upon which one relies and regulate instances when the procedures are disregarded. Such an interpretation provides leeway to call for the establishment of more procedural guidelines. Such guidelines could be targeted at developers and companies employing automated decision-making systems and move beyond the mere establishment of a legal ground. Such guidelines could include more high-level steps (e.g., general steps to be followed irrespective of the sector and context of employment of automated decision-making systems) combined with more concrete, potentially sector-specific guidance. Best practices of tests to be conducted to detect common biases and unfair decision-making could also be part of the guidelines. Overall, this would lay the foundation to determine whether developers of or companies employing and implementing automated decision-making systems adhere to a certain duty of care (see Section 5).

Moreover, fairness can be interpreted as a transparent and proportional balancing, that takes into account the effects of certain automated decision-making processes on individual or

⁷¹ Wachter, Mittelstadt and Russell (above note 66), 16 et seq. with reference to Lilla Farkas and Orlagh O'Farrell, 'Reversing the burden of proof: Practical dilemmas at the European and national level, European Commission, 2015, 44 et seq.; Elisa Holmes, 'Anti-Discrimination Rights without Equality' (2005) *The Modern Law Review* 68(2), 175, 184; Micheal P. Foran, 'Discrimination as an Individual Wrong' (2019) *Oxford Journal of Legal Studies* 39(4), 901, 908 et seq.

⁷² Cf. a.o. Zachary Litpon, Julian McAuley, and Alexandra Chouldechova, 'Does mitigating ML's impact disparity require treatment disparity?' (2018) *Advances in Neural Information Processing Systems* 31, 8125 testing for disparate impacts; Aniya Aggarwal, Pranay Lohia, Seema Nagar, Kuntal Dey and Diptikalyan Saha, 'Black Box Fairness Testing of Machine Learning Models' (2019) *Proceedings of the ACM Joint Meeting on European Software Engineering Conference and Symposium on the foundations of Software Engineering*, 625; Christian Sandvig, Kevin Hamilton, Karrie Karahalios and Cedric Langbort, 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms' (2014) a preconference at the 64th Annual Meeting of the International Communication Association, Seattle, WA, USA; Rachel K. E. Bellamy et al., 'AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias' (2019) *IBM Journal of Research and Development*, 63(4/5), 1.

⁷³ Wachter, Mittelstadt and Russell (above note 66), 1 et seq.

⁷⁴ Malgieri (above note 65), 155 et seq.; Article 29 Working Party, 'Guidelines on Transparency under Regulation 2016/679' (WP 260, rev.01, 11 April 2018). This link is also indicated in multiple recitals and articles of the GDPR that mention both fairness and transparency together: Recs. 39, 60, 71, Arts. 13(2), 14(2), 40(2) GDPR; cf. below section on Transparent Automated Decision-Making.

⁷⁵ Malgieri (above note 65), 155 et seq. pointing to the Greek and Swedish version of Art. 8 EUCFR.

⁷⁶ Damian Clifford and Jef Ausloos, 'Data Protection and the Role of Fairness' (2018) *Yearbook of European Law* 37, 130, 140 et seqq.

group of individuals and the expectations of said individuals or group.⁷⁷ Achieving a just balance requires taking context into account with the goal to prevent unjust “outcomes” or “impacts.”⁷⁸ Such an interpretation further calls for impact assessments that move beyond assessing privacy risks (see below Combining two Approaches: Individual Rights and Due Processes) taking into account the impact of automated decision-making on an individual and societal level. To do so requires establishing the major side-effects that occur through the employment of automated decision-making systems and the reasons why they occur. It also requires understanding which side-effect is most adequately addressed by what means, taking into account educational and technical tools as well.

Understanding the fairness principle as fulfilling multiple ideals opens the door to including further due processes obligations occurring before an automated decision is made. However, as with a broader understanding of the accuracy principle (see below From Accurate Data to Competent Automated Decision-Making Systems) expansions of current terms to address the side-effects of automated decision-making require a broader public as well as policy discourse.

From Accurate Data to Competent Automated Decision-Making Systems

The principle of accuracy is a fundamental component of informational privacy. This principle originally stems from the USA and seems to align more so than other principles—say the principle of data minimization—with the interests of data controllers.⁷⁹ It states that personal data must be “accurate and, where necessary, kept up to date” (Article 5(1)(d)), otherwise the data should be erased or rectified. Data must be factually correct in order to enable the faithful representation of reality. Based upon said data, one must be able to reconstruct a particular situation or set of facts about a person. In this sense, the accuracy of data is linked to its integrity (i.e., the assurance that the data is maintained as it was originally collected), as well as its trueness or veracity, and thus has a technical component.⁸⁰ This definition of scope is quite narrow, focusing on the correctness of personal data used as an input. Yet, the term accuracy was always understood to be dependent on a given context and is therefore a relative term.⁸¹ The relativity is also highlighted in the ECJ’s *Google v. Spain* decision in which the court states that “even initially lawful processing of

⁷⁷ Information Commissioner’s Office, ‘Big data, artificial intelligence, machine learning, and data protection’ (2017) available at <<https://ico.org.uk/media/fororganisations/documents/2013559/big-data-ai-ml-and-dataprotection.pdf>> (last accessed 11 December 2020), 19 et seq.; Article 29 Working Party (above note 68), 16.

⁷⁸ Malgieri (above note 65), 159 with reference to CNIL (above note 65).

⁷⁹ Thomas Hoeren, ‘Big Data and Datenqualität – ein Blick auf die DSGVO’ (2016) ZD, 459; Gloria Gonzáles Fuster, ‘Inaccuracy as a privacy-enhancing tool’ (2010) *Ethics of Information Technologies*, 12, 87; the principle of accuracy is also a guiding principle in the OECD 1980 and now 2013 Guidelines.

⁸⁰ Hoeren, (above note 79), 459 with reference to the ISO Standard 5725-1: 1994; cf. also OECD Guidelines and Roger Clarke’s commentary on the OECD guidelines, <<http://www.rogerclarke.com/DV/PaperOECD.html>> (last accessed 11 December 2020).

⁸¹ Gonzáles Fuster (above note 79), 88.

accurate data may, in the course of time, become incompatible with the directive where those data are no longer necessary in the light of the purposes for which they were collected or processed.”⁸² The court’s reasoning seems to align with the understanding of the OECD’s data quality principle, which states that data used outside of its context can cause misleading conclusions, and thus, in order to avoid misunderstandings, such data should not be processed.⁸³ Implicitly, a link between the accuracy and fairness principles is acknowledged in both the ECJ decision and the data quality principle, as decisions that are based on out of context information can be considered unfair.⁸⁴

That being said, the principle of accuracy could also (teleologically) be interpreted more broadly as a principle of competence, i.e., moving away from the focus on personal data as an “accurate” input towards a “competent” processing system itself. Such an interpretation would take into account the fact that automated decision-making systems are goal-oriented and that society expects a certain “‘normativity’ in the system’s output,” focusing less so on the data input, but rather on the norms surrounding the decision-making processes (such as the underlying motives of the system, decisional criteria, assumptions) and its resulting decisions.⁸⁵ Accuracy understood as competence would hence require a paradigm shift, focusing on the output of a system and the overall proper and adequate functioning of a system.⁸⁶ Such an understanding would align with current calls to grant data protection rights to data subjects, not based merely upon the processing of their personal data, but on the impact such processing has on them.⁸⁷ Such an understanding would however require a policy debate on the characteristics of a “competent” processing system, how alleged “errors” in the systems are challenged, and how rights and duties are distributed. Moreover, such an understanding might also require weakening the current individual data protection rights, as individual objection and erasure can impact the overall accuracy (understood as competence) of a system.⁸⁸

Transparent Automated Decision-Making

The Madrid Resolution of 2009, which resolved that data processing has to occur “on the basis of transparent policies, informing adequately the data subjects and without any arbitrary

⁸² ECLI:EU:C:2014:317, Case C-131/12., Para. 93.

⁸³ Cf. The OECD Privacy Framework, available at <https://www.oecd.org/internet/ieconomy/oecd_privacy_framework.pdf> (last accessed 11 December 2020), 56; cf. also Clarke (above note 80).

⁸⁴ Rolf H. Weber and Simon Henseler, ‘Regulierung von Algorithmen in der EU und in der Schweiz: Überlegungen zu ausgewählten Regulierungsthemen’ (2020) EUZ, 28, 31.

⁸⁵ Bayamlıoglu (above note 32), 434 et seq. with further references below on the aspect of normativity.

⁸⁶ Cf. Wachter and Mittelstadt (above note 6), 615 et seq.

⁸⁷ Wachter and Mittelstadt (above note 6), 616 who align with the Article 29 Working Party, Opinion 4/2007 on the concept of personal data, WP136, 20 June 2007, 8 et seq. broad understanding of the term personal data; note that this author has criticized this broad notion of personal data as challenges legal certainty and legitimate basis, cf. Damian George, Kento Reutimann and Aurelia Tamò-Larrieux, ‘GDPR bypass by design? Transient processing of data under the GDPR’ (2019) International Data Privacy Law 9(4), 285.

⁸⁸ Cf. also Margot E. Kaminski, ‘Binary Governance: Lessons From the GDPR’s Approach to Algorithmic Accountability’ (2019) Southern California Law Review 92, 1529, 1580 et seq.

discrimination against them,”⁸⁹ was the basis for introducing a new concept into the GDPR, namely the principle of transparency.⁹⁰ This principle includes both a prospective and retrospective element. The prospective element includes information about the systems functionality that data controllers must provide before automated decisions are reached; the retrospective element refers “to the possibility to trace back how and why a particular decision was reached.”⁹¹

The prospective element of transparency is specified within Articles 13 and 14 of the GDPR, regarding the information duties of data controllers, and requires them to provide “easily accessible and easy to understand” information to the data subjects on how their personal data is being processed.⁹² Information on the identity of the controller and the explicit purposes of processing must be provided. Furthermore, data subjects “should be made aware of risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights in relation to such processing.”⁹³ The approach we see within the GDPR aligns with the premise that it is only possible to control a system if we (first) observe it. In this context, control means that individuals have a certain autonomy over a process, an assumption which is often referred to as the “autonomy-through-openness” assumption.⁹⁴ Yet, transparency itself is only a passive element, providing information on the inner-workings of a system.⁹⁵ This element is often seen as a cornerstone for accountability, which requires an active explanation of how and why a system works.⁹⁶ The credo that transparency is a means to achieve accountable (automated decision-making) systems, is not only found in data protection law, but also in consumer protection domains (e.g., through the inclusion of nutrition labels) and environmental law (e.g., through environmental impact records). As openness enables us to identify errors and impacts of a system, it enables us to shame a party for unjustified or biased actions within their decision-making systems. In theory, such shaming can affect the market by punishing companies that

⁸⁹ Madrid Resolution: International Standard on the Protection of Personal Data and Privacy (adopted at the International Conference on Data Protection and Privacy Commissioners), 2009, available at https://www.dataprotectionauthority.be/sites/privacycommission/files/documents/international_standards_madrid_2009.pdf (last accessed 11 December 2020).

⁹⁰ The term transparency was already included in the 2012 GDPR Draft, cf. draft available at [http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2012/0011/COM_COM\(2012\)0011_EN.pdf](http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2012/0011/COM_COM(2012)0011_EN.pdf) (last accessed 11 December 2020) and remained unmodified throughout the revision process. Cf. for an overview of the establishment of the GDPR: https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en (last accessed 11 December 2020).

⁹¹ Heike Felzmann, Eduard Fosch Villaronga, Christoph Lutz and Aurelia Tamò-Larrieux, ‘Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns’ (2019) *Big Data & Society*, 1, 3.

⁹² Rec. 39 GDPR.

⁹³ Rec. 39 GDPR.

⁹⁴ Ananny and Crawford, (above note 45), 975; Note that this assumption has been criticized in particular as too much openness or information overflow tend to make decisions more difficult rather than easier, cf. Mario Martini, *Regulating Algorithms – How to De-Mystify the Alchemy of Code?*, in Martin Ebers and Susana Navas (eds.), *Algorithms and Law* (Cambridge University Press, 2020), forthcoming in July 2020, unpublished version available on SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3458261 (last accessed 11 December 2020).

⁹⁵ Maranke Wieringa, ‘What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability’ (2020) *Proceeding of the Conference on Fairness, Accountability, and Transparency (FAT*)*, 1, 4.

⁹⁶ Wieringa (above note 95), 4.

enable unjustified or biased automated decision-making to take place. The effect of both, the shaming of and impact on the market depend on how strongly the public cares about the identified practices and potential errors of the system.⁹⁷ In other words, shaming only works if there are social norms in place regarding what is acceptable and what is not.⁹⁸

The transparency requirements under the GDPR are limited by scope of the law (i.e., the processing of personal data) and thus only serve data subjects. Further criticisms with respect to the practical impact of the GDPR's transparency norms on regulating the side-effects of automated decision-making have been raised: First, the GDPR does not require data controllers to differentiate between information they provide to the general public, an affected individual, or a data protection authority. While the GDPR focuses on the individual who is identifiable and thereby assumed to be affected by the data processing, it neglects to account for the fact that, with automated decision-making systems, the affected individuals could be different from the data subjects themselves.⁹⁹ In addition, the information duties could be directed to data protection authorities or other expert institutions that have the expertise to understand more complex technical information.¹⁰⁰ Another point of criticism surrounds the lack of differentiation between the various stages of processing, i.e., the life cycle of data,¹⁰¹ and calls for a layered approach of information provisions. During the collection and aggregation stage, transparency should require that the general public and individuals affected by a decision are provided with information "regarding the kinds and forms of data and databases used in the analysis."¹⁰² The prospective transparency element within the GDPR, which focus only on the collection phase, neither mandates the provision of information on how a database is formulated, nor the methods used to aggregate data into one dataset. Likewise, the GDPR does not require, not even to fulfill the retrospective transparency element,¹⁰³ a disclosure of the algorithms source code.¹⁰⁴ In fact, such disclosures could conflict with other rights, in particular trade secrets, as such source code is arguably valuable only as long it remains secret.¹⁰⁵ Nevertheless, the protection of trade secrets should not undermine fundamental rights, such as the right to privacy and data protection.¹⁰⁶ Even

⁹⁷ Zarsky (above note 46), 1534 et seq.; cf. also Wieringa (above note 95), 5.

⁹⁸ Zarsky (above note 46), 1536.

⁹⁹ An example thereof is the use of personal data for the training of an algorithm and the generation of a decision tree or decision pattern, yet, never actually affecting the individual whose data was used within the training set but affecting other individuals who provided only limited and potentially non-identifiable data.

¹⁰⁰ Zarsky (above note 46), 1532 et seq.

¹⁰¹ Zarsky (above note 46), 1532 et seq. differentiating the collection and aggregation of data, its analysis, and its usage; cf. also Tamò-Larrieux (above note 64), 6.

¹⁰² Zarsky (above note 46), 1523.

¹⁰³ Cf. below section on Retrospective Element.

¹⁰⁴ Martini (above note 94), 16 et seq. cf. for a differing opinion (not with respect to the scope of the retrospective element within the GDPR but on a more general note): Zarsky (above note 46), 1524.

¹⁰⁵ Art. 2(1)(b) Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, OJ L 157, 15.6.2016, 1–18 (Trade Secrets Directive).

¹⁰⁶ Rec. 21 and 34 Trade Secrets Directive.

if the recitals of the Trade Secret Directive indicate that in the case of a conflict between trade secrets rights and fundamental rights (such as the right to privacy and data protection), the latter should prevail,¹⁰⁷ the GDPR's Recital 63 itself, could be interpreted as tipping the balance in the other direction, as it states with respect to the right of access to data that such a right "should not adversely affect the rights or freedoms of others, including trade secrets (...)." Although both laws affirm a non-prevalence of their own rules,¹⁰⁸ the recital of the GDPR also states that even if trade secrets may impede a right of access, this impediment cannot take the form of a "refusal to provide all information to the data subject," meaning that parts of the request must be complied with. How granular such information can be to fulfil both the needs for access to the decision-making process, and the need for protecting information ensuring a competitive advantage, must be determined on a case-by-case basis. According to Diakopoulos there are several elements that could be revealed without breaching IP rights and trade secrets, such as information regarding the human involvement, the data, the model and its inferences, as well as the algorithmic presence.¹⁰⁹ Additionally, the basis for relying upon the findings of the systems or applying automated decisions could be disclosed.¹¹⁰ Finally, with respect to the implementation stage, transparency refers to a disclosure of the "actual strategies and practices for using the data,"¹¹¹ and reasonable assurances with respect to the precision and effectiveness of the automated decision-making process should be set in place (which ultimately relates to the retrospective element of transparency).

The retrospective element of transparency is linked to the inspectability and explainability of an automated decision-making process.¹¹² Inspectability refers to the decomposition of the algorithm performing the decision-making.¹¹³ In order to better understand the details of the algorithm, access to the model structure itself is necessary,¹¹⁴ as well as access to the parameters and learning dataset in the case of automated decision-making processes that rely on machine learning.¹¹⁵ Explainability refers to a "human-understandable approximation of

¹⁰⁷ Gianclaudio Malgieri, 'Trade Secrets v Personal Data' (2016) International Data Privacy Law, 6(2), 102, 103; differing opinion: Peter Hustinx, European Data Protection Supervisor, Opinion on the Trade Secrets Directive, available at <https://edps.europa.eu/sites/edp/files/publication/14-03-12_trade_secrets_en.pdf> (March 2014) (last accessed 11 December 2020). Note that the Trade Secrets Directive introduces a minimal harmonized standard in the EU and that the trade secrets regimes and corresponding case law vary within the individual member states. Cf. for an analysis Francesco Banterle, 'The Interface Between Data Protection and IP Law: The Case of Trade Secrets and the Database sui generis Right in Marketing Operations, and the Ownership of Raw Data in Big Data Analysis' in Mor Bakhoun, Beatriz Conde Gallego, Mark-Oliver Mackenrodt and Gintarė Surblytė-Namavičienė (eds.) *Personal Data in Competition, Consumer Protection and Intellectual Property Law: Towards a Holistic Approach?* (MPI Studies on Intellectual Property and Competition Law, 28, Springer, 2018) 441-443.

¹⁰⁸ Malgieri (above note 107), 105 stating that a case-by-case approach is necessary.

¹⁰⁹ Nicholas Diakopoulos, 'Accountability in Algorithmic Decision Making' (2016) Communication of the ACM 59(2), 56 et seq.

¹¹⁰ Zarsky (above note 46), 1525.

¹¹¹ Zarsky (above note 46), 1526; cf. Section New Paradigms to Address the Side-Effects of Automated Decision-Making.

¹¹² Cf. on the concept of explanation: Tim Miller, 'Explanation in artificial intelligence: Insights from the social sciences', (2019) Artificial Intelligence 267.

¹¹³ Lillian Edwards and Michael Veale, 'Slave to the algorithm? Why a right to an explanation is probably not the remedy you are looking for' (2017) Duke Law and Technology Review 16, 18, 64; cf. also Zerilli et al. (above note 10), 665.

¹¹⁴ Edwards and Veale (above note 113), 64 et seq.

¹¹⁵ Castelluccia and Le Métayer (above note 8), 26.

the algorithm,”¹¹⁶ or a model of a decision. The topic of explainability has gained more traction since the introduction of the GDPR. Legal scholars¹¹⁷ have debated whether a retrospective right to explainability can be inferred from the right to access established in Article 15(1)(h) of the GDPR (read in conjunction with Article 13(2)(f), Article 14(2)(g), Article 22(1) and (4), and Recital 71 of the GDPR). The transparency and right of access articles mentioned, enable data subjects to receive information about “the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing to the data subject.”¹¹⁸ The “significance and the envisaged consequences” relate only to an individual subject to an automated decision-making and do not take into account broader, societal impacts (see Section 5). Moreover, the automated decision-making that the GDPR refers to relates (at least¹¹⁹) to any processing or profiling that leads to decisions that produce “legal effects” (e.g., court sentence, contract cancellation) or “similarly significantly” (e.g., being denied a loan, being denied a specific means of payment) affect a data subject (Article 22(1) of the GDPR).¹²⁰ Bygrave notes that the difference in wording between the Directive 95/46/EC and the GDPR (i.e., the use of the word similarly before significantly affects) indicates that these “effects must have a degree of significance that is roughly equivalent to that of legal effects (although they do not need to be legal in nature).”¹²¹ An automated decision must thus fully or partly influence an individual’s legal rights and duties or have “consequences that have a seriously adverse impact on a person’s welfare.”¹²² A certain, rather high, threshold of the impact must thus be crossed, for the provision to apply. However, to pin down the precise threshold of when “similarly significantly” effects arise is difficult and remains to be determined on a case-by-case

¹¹⁶ Sandra Wachter, Bernt Daniel Mittelstadt and Chris Russel, ‘Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR’ (2018) *Harvard Journal of Law & Technology*, 31(2), 841, 850.

¹¹⁷ Cf. a.o., Edwards and Veale, (above note 113); Bryan Casey, Ashkon Farhangi and Roland Vogl, ‘Rethinking Explainable Machines: The GDPR’s ‘Right to Explanation’ Debate and the Rise of Algorithmic Audits in Enterprise’ (2019) *Berkeley Technology Law Journal*, 34, 143; Bryce Goodman and Seth Flaxman, ‘European Union Regulations on Algorithmic Decision-Making and a “Right to Explanation”’ (2017) *AI Magazine*, 38(3), 50; Sandra Wachter, Bernt Daniel Mittelstadt and Luciano Floridi, ‘Why a Right to Explanation of Automated Decision-making does not exist in the General Data Protection Regulation’ (2017) *International Data Privacy Law* 7(2), 76; Margot E. Kaminski, ‘The Right to Explanation, Explained’ (2019) *Berkeley Technology Law Journal* 34(1), 189; Andrew D. Selbst and Julia Powles, ‘Meaningful information and the right to explanation’ (2017) *International Data Privacy Law* 7(4), 233; Martini (above note 94).

¹¹⁸ Art. 13(2)(f), Art. 14(2)(g), and Art. 15(1)(h) of the GDPR.

¹¹⁹ Cf. Art. 13(2)(f), Art. 14(2)(g), and Art. 15(1)(h) GDPR; Martini (above note 94) criticizing the possibility of an obligation to inform data subjects about the decision-making processes beyond the scope of Art. 22 GDPR, as unjustifiable under the rule of law. At the same time though, the author argues for establishing an extension of labelling obligations and information duties, 18. The term “at least” does indeed lead to less legal certainty. Already the term “similarly significant” requires a flexible threshold that is difficult to pin down in practice. It remains to be seen how courts will interpret these terminologies and how their scope will be able to provide redress mechanisms for the side-effects arising on an individual and societal level describe above (see Section 3).

¹²⁰ Mario Martini, ‘DS-GVO Art. 22 Automatisierte Entscheidungen im Einzelfall einschliesslich Profiling’ in Boris P Paal and Daniel Pauly (eds.), *Datenschutz-Grundverordnung* (3rd edn, beck-online, Munich, 2021) Rn 25-28 with further references for the examples provided; Article 29 Working Party (above note 68), 21 et seq.

¹²¹ Lee A. Bygrave, ‘Commentary to Article 22 of the GDPR’ in Christopher Kuner, Lee A. Bygrave, Christopher Docksey (eds.), *The EU General Data Protection Regulation* (Oxford University Press, 2020), 534.

¹²² Bygrave (above note 121), 534.

basis.¹²³ While focused on the data subject, i.e., the effect on an individual, the provision does enable to take into account group characteristics, thereby addressing some side-effects arising on a societal level. For instance, if an automated decision-making impact a data subject that is part of a vulnerable group (e.g., children), it will enable courts to lower the threshold of when a significant effect arises.¹²⁴

While the right of access grants data subjects affected by automated decision-making systems fulfilling the criteria of Article 22 of the GDPR the right to obtain meaningful information about the logic involved in the decision-making, the European policymakers did not specify when information duties beyond the outlined scope should reasonably apply (i.e., when the “at least” parameter is fulfilled), and how information about the logic involved must be structured like in practice. With respect to the scope of obtaining meaningful information, Martini argues that information duties should be extended to algorithms that support human decision-making (so-called “human-in-the-loop” and “human-on-the-loop” decision-making systems).¹²⁵ As opinions on the scope of the retrospective transparency element from “human-out-of-the-loop” systems to those with humans-in/-on-the-loop will differ, so do understandings of what information regarding the logic involved refers to. Does it reference an understanding of how an automated decision was reached, should it enable an affected individual to contest a certain decision, or should it enable him or her to change his or her behavior to obtain a different result in the future?

Technically speaking, explanations of automated decision-making systems can be operational, logical, or causal, as well as global (whole algorithm) or local (about a specific result).¹²⁶ Different approaches to explainability exist: Explanations can be constructed from observations (i.e., analysis of the relationship of output to input without ‘opening’ the black box);¹²⁷ the code of the automated decision-making system can be accessed (so-called white box approach);¹²⁸ and counterfactual explanations can “provide a minimal amount of information capable of altering a decision” without requiring “the data subject to understand any of the internal logic of a model in order to make use of it.”¹²⁹ The target audience of the explanation will determine the granularity and technicality that the explanation must fulfil. The inclusion of the wording “for the data subjects” in Article 15(1)(h) of the GDPR suggests that the meaningfulness

¹²³ Article 29 Working Party (above note 68), 21, stating that the threshold may be met if the decision has the potential to “significantly affect the circumstances, behavior or choices of the individuals concerned, have a prolonged or permanent impact on the data subject, or at its more extreme, led to the exclusion or discrimination of individuals.”

¹²⁴ Bygrave (above note 121) 535 with reference to Rec. 71 and 38 of the GDPR.

¹²⁵ Martini (above note 94), 18; cf. on typology of human-in/on/out of-the-loop, Wieringa (above note 95), 3 with further references. A human-in-the-loop-system is one that gives suggestions to the human on how to make a decision, a human-on-the-loop-system is one that is monitored by humans but whereby default the machines makes a decision, and a human-out-of-the-loop-system is a fully automated processes without any kind of human involvement.

¹²⁶ Castelluccia and Le Métayer (above note 8), 47 et seqq.; cf. also Zerilli et al. (above note 10), 663 et seq.

¹²⁷ Castelluccia and Le Métayer, (above note 8), 47 et seqq.

¹²⁸ Castelluccia and Le Métayer, (above note 8), 47 et seqq.

¹²⁹ Wachter, Mittelstadt and Russel, (above note 116), 851.

of the information must be assessed from the perspective of the individual, and that the information must be relevant to the decision that affects said individual.¹³⁰

As the discussions on the prospective and retrospective element of transparency show, the GDPR is too narrowly focused on informational transparency or transparency as verifiability. This approach enables controllers of automated decision-making systems to rely on an ‘inform and consent’-framework without differentiating the ways in which transparency is performed or the target groups of information and their respective needs. The lack of a more nuanced approach has not only consequences on an individual level (e.g., untailored information about employed automated decision-making systems), but also triggers societal side-effects (e.g., performing transparency in a way that will ensure that there is no meaningful confrontation with the provided information). While it is sensible to move beyond the informational dimension of transparency and better understand its performative element (i.e., practices of how transparency is being performed or not performed by means of dark patterns or the conscious provisions of an overwhelming amount of information to discourage from engaging with the information) as well as promote a relational approach (i.e., drawing on the concept of trustworthiness, and duty of care principles),¹³¹ those aspirations have yet to resonate in the policymaking discourse (see Section 5).

Forbidding Automated Decision-Making?

The data subject has “the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her”¹³² (see above Transparent Automated Decision-Making). Three exceptions to this rule exist, namely, when the automated processing is necessary to enter or perform a contract, when the data subject has given his or her explicit consent to the automated processing, and when legislation in the member state where the processing occurs allows for such processing and is subject to suitable measures safeguarding the data subject’s rights and freedoms.¹³³ If one of the first two exceptions is applicable, the data subject still has the “right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.”¹³⁴

While objecting to processing under Article 21 of the GDPR is undisputedly an individual right that a data subject has to exercise, the literature on forbidding automated decision-making

¹³⁰ Dimitra Kamarinou, Christopher Millard, and Jatinder Singh, ‘Machine Learning with Personal Data’, in Ronald Leenes, Rosamunde van Brakel, Serge Gutwirth and Paul de Hert (eds.), *Data Protection and Privacy: The Age of Intelligent Machines* (Oxford Hart Publishing 2017), 89-114, 107 et seq.

¹³¹ Felzmann, Fosch Villaronga, Lutz and Tamò-Larrieux (above note 91), 3.

¹³² Art. 22(1) GDPR.

¹³³ Art. 22(2) GDPR.

¹³⁴ Art. 22(3) GDPR.

under Article 22 of the GDPR is divided into two camps, understanding it as a prohibition¹³⁵ or a right.¹³⁶ The rationales for introducing this mix of right and prohibition within Article 22 of the GDPR are similar to those that gave rise to the introduction of Article 15 of the Directive 95/46/EC, which was interpreted differently (as a right or prohibition) in various member states.¹³⁷ Even though the goal of the GDPR is to harmonize data protection law, the cultural embeddedness and interpretation of data protection norms and the resulting variances in application of said norms is reflected in the debates on how to understand Article 22 of the GDPR. Interestingly, while proponents of understanding the article as an individual right acknowledge the side-effects of automated decision-making occurring on an individual and societal level, their remedy against those effects remain user focused. This aligns with aim to strengthen the position of individuals to “exercise influence over decision-making processes” that affect them and ensure that data subjects “maintain the primary role in ‘constituting’ themselves.”¹³⁸ While strengthening the user rights are welcomed developments, putting the onus on individuals to address the side-effects of automated decision-making seems only sufficient when targeting side-effect on an individual level (see below Combining two Approaches: Individual Rights and Due Process). Yet, proponents of a prohibition have a broader outlook on the societal side-effects caused by the digital market and the implementation of automated decision-making. This is also where the discussion between automated decision-making and profiling becomes relevant. In fact, with the introduction of the GDPR, the focus broadened beyond automated decision-making to include profiling.¹³⁹ Profiling can either be seen as an inherent element of automated decision-making

¹³⁵ Mendoza and Bygrave (above note 37), 86 et seq. also commenting on the Directive 95/46/EC; cf. Martini (above note 120), 29; Article 29 Working Party (above note 68), 21 et seq. and 39 et seq.; Frederike Kaltheuner and Elettra Bietti, ‘Data is power: Towards additional guidance on profiling and automated decision-making in the GDPR’ (2018) *Journal of Information Rights, Policy and Practice*, 2(2), 1, 10; Guido Noto la Diega, ‘Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property Data Protection, and Freedom of Information’ (2018) *Journal of Intellectual Property, Information Technology and E-Commerce Law* 9(1), 3, 17; Maja Brkan, ‘Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond’ (2019) *International Journal of Law and Information Technology*, 27, 91, 97.

¹³⁶ The wording of Art. 22 GDPR as well as its position within Chapter III speak in favor of treating it as an individual right; cf. also Anton Vedder and Laurens Naudts, ‘Accountability for the use of algorithms in a big data environment’ (2017) *International Review of Law, Computers & Technology*, 206, 213 et seq.; Moreover, Art. 11 of the Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, OJ L 119, 4.5.2016, p. 89–131, states that it is up to the Member States to determine whether automated individual decision-making should be prohibited. This could indicate that at least in the case of automated decision-making systems used by authorities for preventing, investigating, and detecting, and prosecuting criminal offences, Member States can prohibit the use of such systems “unless authorized by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller.” (Art. 11(1) Directive 2016/680). E.c. the use of automated decision-making systems is to be understood as a right, not a prohibition.

¹³⁷ Cf. Art. 15(1) Directive 95/46/EC which states: “Member States shall grant the right to every person not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.”

¹³⁸ Mendoza and Bygrave (above note 37), 83 with reference to Commission of the European Communities, Explanatory text for Proposal for a Council Directive concerning the protection of individuals in relation to the processing of personal data, COM (90) 314 final – SYN 287, 29.

¹³⁹ Cf. here Mendoza and Bygrave (above note 37), 84 with a historical analysis of how Art. 22 came to be. Cf. European Commission, ‘Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)’ COM(2012) 11 final (C7-0025/2012 – 2012/0011(COD)) and the proposed text of former Art. 20 (“Measures based on profiling”). Mendoza and Bygrave elaborate on how the Committee on Internal Market and Consumer Protection (IMCO) and the Committee on Industry, Research and Energy (ITRE)

processes, or understood as a separate activity.¹⁴⁰ The latter interpretation would not necessarily broaden the scope of Article 22, as the profiling activity nonetheless has to be followed by a decision producing either legal or other significant effects.¹⁴¹ Overall though, prohibiting the deployment of automated decision-making systems in sensitive sectors might become necessary if no responsibility for decisions made can be assigned in a satisfactory manner.

Irrespective of whether a data subject has to demand not to be subject to automated decision-making that significantly affect him or her, or whether such processing is prohibited *per se*, a shortcoming of the provision is the general focus on the data subject.¹⁴² This focus neglects the fact that the identifiability of a natural person is irrelevant for profiling activities.¹⁴³ In fact, profiling algorithms treat “people on the basis of group characteristics instead of on their own individual characteristics and merit.”¹⁴⁴ With profiling, we thus move away from the strict individual (personal) level towards the aggregated level.¹⁴⁵ On an individual level, data about an individual is recorded, and based on that recording, knowledge is created. On an aggregated level, statistical models determine correlations, patterns, and general rules between data points. In this sense, the goal of profiling is to infer knowledge about an individual that is not directly observed.¹⁴⁶

Combining two Approaches: Individual Rights and Due Processes

EU policymakers had the societal changes automated decision-making processes would bring along in mind when they drafted the GDPR. Importantly, policymakers wanted to ensure that individuals would remain in control over decision-making processes and that their position to influence such processes would not be weakened.¹⁴⁷ The idealization of individual control¹⁴⁸ is also why data protection law’s material scope includes a broad definition of personal data, which in turn led the law to be known as the “law of everything.”¹⁴⁹ Yet, the identifiability of a data subject reaches its limits when trying to regulate the side-effects of automated decision-making.

were more friendly to profiling than the Committee on Civil Liberties, Justice and Home Affairs (which had the lead role in negotiating the GDPR). The term profiling nonetheless received more attention, obtained an own definition (in Article 4(4) of the GDPR), and was included in several recitals, among others specifying the protection of user profiles of children (cf. Recital 38 of the GDPR).

¹⁴⁰ Cf. to Rec. 71 and Art. 4(4) of the GDPR; Art. 4(4) of the GDPR might indicate that profiling must be understood as a separate activity rather than an inherent element of automated decision-making. Yet, earlier drafts of the GDPR (see above 134) indicate that initially profiling was seen as the activity leading to automated decision-making.

¹⁴¹ Cf. above section on Retrospective Transparency; Simon Henseler, ‘Was ist eine automatisierte Entscheidung? Zum Tatbestand von Art. 22 Abs. 1 DSGVO’ (2020) APARIUZ 21, 301, 308 et seq.; Differing opinion: Mendoza and Bygrave (above note 37), 90 et seq.

¹⁴² Cf. Kamarinou, Millard and Singh, (above note 123), 95 et seq. who point out the that original wording of Art. 22 GDPR by the European Commission’s proposal (see above 134) was broader including any “natural person” and not merely “data subjects.”

¹⁴³ Mittelstadt et al. (above note 19), 10 with reference to Schermer, (above note 53), 48 et seq.

¹⁴⁴ Lita van Wel and Lambèr Royakkers, ‘Ethical issues in web data mining’ (2004) Ethics and Information Technology 6, 129, 133.

¹⁴⁵ van Otterlo (above note 18), 43 et seq.

¹⁴⁶ van Otterlo (above note 18), 43 et seq.

¹⁴⁷ Mendoza and Bygrave (above note 37), 82 et seq. with further references.

¹⁴⁸ Cf. Woodrow Hartzog, ‘The Case Against Idealising Control’ (2018) European Data Protection Law Review 4, 423.

¹⁴⁹ Purtova (above note 61), 40.

Automated decision-making systems rely on data from multiple individuals that ultimately may not be affected by the decision-making process itself. Moreover, many remedies found within the GDPR focus on an individual rights-based approach, i.e., putting the onus of contesting automated decision-making practices in the hands of the data subject. A limitation of such an approach is that side-effects like discrimination are easier to perceive when not looking at an individual case but at the larger behavior of automated decision-making systems.¹⁵⁰ Merely studying individual remedies is unlikely to be the adequate strategy to address system-wide issues, such as those mentioned above. Policymakers seem to have been aware of this issue, as the ambiguously drafted Article 22(1) of the GDPR indicates; even if its wording includes the word “right,” Article 22 can be interpreted as a prohibition.¹⁵¹ In addition, too much focus on individual rights might lead to tilting the balance towards individual self-protection, rather than a focus on public notices or information to expert watchdogs. To address the unwanted side-effects of automated decision-making, we not only need to “expand the right to receive information regarding the prediction process to the entire public,”¹⁵² and establish “adequate public-facing or even expert-facing accountability,”¹⁵³ but also to obtain information about the underlying decision-making standards of automated decision-making processes.¹⁵⁴

Aside from the individual rights approach, the GDPR also includes due process measures, such as techniques that target the design of automated decision-making systems, like data protection impact assessments. Such assessments could provide greater accountability, but the fact that the GDPR does not mandate their publication (not even of a summary of the assessment) makes it questionable that such an effect is currently achieved.¹⁵⁵ Other collaborative governance instruments, i.e. those deploying private-public partnerships towards enhancing public governance goals, exist under the GDPR.¹⁵⁶ These instruments include formal co-regulatory tools, such as the delegation of the establishment of codes of conduct and certification schemes to the regulated data controllers,¹⁵⁷ as well as industry involvement, through a formal process initiated

¹⁵⁰ Kaminski (above note 88), 1558 with reference to Edwards and Veale (above note 113), 74 et seq.

¹⁵¹ Cf. above Section Forbidding Automated Decision-Making? with further references. Note that some authors have argued that a prohibition of certain automated decision-making processes are necessary. Cf. Mittelstadt et al. (above note 19), 14 with reference to Zarsky (above note 37) and Burrell (above note 42) stating that in the US the Credit Reporting Act, for instance prohibits machine-learning in credit scoring. In Europe restrictions on face recognition technologies were discussed, but ultimately the European Commission opted for a risks-based approach. Cf. Samuel Stolton, ‘LEAK: Commission considers facial recognition ban in AI “white paper”’ *Euractive* (17 January 2020) available at <<https://www.euractiv.com/section/digital/news/leak-commission-considers-facial-recognition-ban-in-ai-white-paper/>> (last accessed 11 December 2020); European Commission, ‘White Paper on Artificial Intelligence – A European approach to excellence and trust’ COM(2020) 65 final.

¹⁵² Zarsky (above note 46), 1543.

¹⁵³ Kaminski (above note 88), 1607.

¹⁵⁴ Wachter and Mittelstadt, (above note 6), 505.

¹⁵⁵ Tamò-Larrieux (above note 64), 182; cf. also Kaminski (above note 88), 1576; Martini (above note 94), 24.

¹⁵⁶ Kaminski (above note 88), 1601 et seqq. differentiating between formal collaborative governance methods (e.g., negotiating rulemaking, safe harbor agreements, adopting codes of conducts and technical standards) and less formal ones (e.g., audited self-regulation, delegation of interpretation of standards to the industry).

¹⁵⁷ Cf. Arts. 40, 41, and 42 GDPR.

by the European Commission, in the development of technical standards.¹⁵⁸ In addition, informal co-regulatory tools exist, including the aforementioned data protection impact assessments, the introduction of data protection officers,¹⁵⁹ and record-keeping to enable internal auditing.¹⁶⁰ The principle based approach of the GDPR furthermore allows for flexibility in that it relies on “state of the art” measures developed by the regulated industries.¹⁶¹

These collaborative governance approaches provide tools to (partially) address the side-effects of algorithmic decision-making. In particular, data protection impact assessments, which may be combined with more exhaustive social and ethical impact assessments,¹⁶² as well as the introduction of data protection officers, or more personnel in general, that take the broader impact of technologies into consideration, are means to shift managerial perspectives. Any measures that attempt to foster a more holistic understanding and organizational culture that considers the impact of automated decision-making should be welcomed. These measures could be extended by demanding public disclosure of failures of implemented automated decision-making systems, the criteria used in impact assessments and employed remedies, as well as strategies followed by data protection officers to ensure reasonable decision-making practices. Furthermore, the GDPR empowers regulated parties to establish codes of conduct and/or certificates but does not go as far as to require the involvement of third parties, such as NGOs, nor public transparency. If third parties are not able to obtain information about how data controllers determine the compliance of their products, services, and internal processes with data protection law, they cannot help data protection authorities with monitoring the activities of data controllers and the lodging of complaints for malpractice.¹⁶³

The analysis within this article shows that ‘the law of everything’ has different tools to regulate automated decision-making, that they interact with one another, and that such an interplay can be complementary. The tools within the GDPR constrain, enable, and level the playing field.¹⁶⁴ On the one hand, transparency requirements with respect to automated decision-making practices and individual rights can, for instance, be seen as an enabler or leveler. On the

¹⁵⁸ Irene Kamara, ‘Co-regulation in EU personal data protection: the case of technical standards and the privacy by design standardization ‘mandate’’ (2017) *European Journal of Law and Technology* 8(1); cf. also Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council Text with EEA relevance, OJ L 316, 14.11.2012.

¹⁵⁹ Arts. 37, 38, and 39 GDPR.

¹⁶⁰ Art. 30 GDPR; cf. Rec. 71 GDPR which can be interpreted as requiring auditing, cf. Kaminski (above note 88), 1606 with reference to Gianclaudio Malgieri and Giovanni Comandé, ‘Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation’ (2017) *International Data Privacy Law* 7(4), 243, 285.

¹⁶¹ Cf. overall on informal co-regulatory tools: Kaminski (above note 88) and Kamara (above note 158).

¹⁶² Cf. Alessandro Mantelero, ‘AI and Big Data: A blueprint for a human rights, social and ethical impact assessment’ (2018) *Computer Law & Security Review* 34(4), 754.

¹⁶³ Kaminski (above note 88), 1562 et seq.

¹⁶⁴ Urs Gasser, ‘Perspectives on the Future of Digital Privacy’ (2015) *Rechtsfragen im digitalen Zeitalter Schweizerischer Juristentag ZSR Band 134 II*, 337; cf. also Tamò-Larrieux (above note 64), 193 et seq. relying on Gasser’s taxonomy for privacy by design measures.

other hand, the prohibition of automated decision-making that significantly affect individuals constrains automated decision-making practices. Technical tools complement the legal order, offering, for instance, new means to monitor the side-effects of automated decision-making systems.¹⁶⁵ While time alone will tell how the different approaches and tools will interplay with one another towards a governance of automated decision-making practices, scholars have proposed new paradigms to support the current data protection law approaches.

[5] New Paradigms to Address the Side-Effects of Automated Decision-Making

The shift from having machines supporting human decision-making to having machines make actual decisions has triggered a debate on how to regulate such practices. As shown above, much of the attention on this subject has focused on data protection law, which relies also on other statutes, such as anti-discrimination law and restricting trade secret laws, to achieve certain goals. Additionally, the discussion has also illustrated the shortcomings of said regulatory framework to address novel issues posed by automated decision-making practices and new proposals have emerged. These proposals take into account the difficulties of untangling the effects of automated decision-making on individuals, groups, and the society at large and reflect upon the complexity of decision-making in terms of involved actors, interdependencies among actors, and power dynamics.

Modular Guidelines for Reasonable and Legitimate Decisions by Machines

As a starting point, it will be key to acknowledge that the ‘law of everything’ even with its broad material and territorial scope does not address all the arising side-effects of automated decision-making. To certain extent the European Commission seems to have already acknowledged this fact by proposing the Digital Service Act (DSA) and Digital Market Act (DMA).¹⁶⁶ Former includes due diligence obligations and increased transparency requirements for providers of intermediary services with a focus on the removal of illegal content, an activity which often is executed by automated decision-making systems (see Section 1). Latter targets core platform services (so-called gatekeepers) and wants to ensure contestable and fair markets. While the DMA does not fully move away from a (personal) data-focused approach,¹⁶⁷ the established obligations for gatekeepers may open the door for more collective protection mechanisms, thereby targeting

¹⁶⁵ Schermer (above note 53), 50 et seq.; Peter Fule and John Roddick, ‘Detecting Privacy and Ethical Sensitivity in Data Mining Results’ (2004) Proceedings of the 27th Australasian Conference on Computer Science, 26, 159.

¹⁶⁶ Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, COM/2020/825 final; Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Market Act), COM/2020/842 final.

¹⁶⁷ Cf. e.g., Article 5(a) DMA which requires gatekeepers to refrain from combining personal data from services offered by the data controller.

side-effects arising at a societal level (see below From the Individual to the Collective: Towards Social Protection Mechanisms).

In order to develop technologies benefitting individuals as well as society, measures that focus on an earlier stage, namely the development of technical systems, are necessary. In its broadest understanding, targeting the development stage could lead to the establishment of ex ante legitimization criteria that prior to each development of a technical system must be fulfilled, thereby “achieving legitimacy in code.”¹⁶⁸ While current frameworks of what those criteria should be are broad and to a certain extent align with some of the ideals and objectives found within data protection law, more granular approaches tailored on automated decision-making systems employed in different contexts can be envisaged. Nonetheless, and as shown above (see Section 4) the fundamental principles of data protection law provide room for interpretation. For instance, understanding the principle of accuracy not as focused on the correctness of the input data but more broadly as the overall competency of an automated decision-making system could be a starting point to provide guidance to developers.

With a more tailored focus on automated decision-making systems, a key shift would be to implement what has been called a “right to reasonable inferences.”¹⁶⁹ This right requires a prior discussion of what “reasonable” inferences or reasonable automated decision-making entail and establishment of guidelines on how to determine said reasonableness.¹⁷⁰ In order to ensure strong public support, such a discussion should include various stakeholders. Norms on what ‘ought’ to be considered a reasonable decision is context dependent,¹⁷¹ and guidelines should address the domains in which automated decision-making systems are employed (e.g., content personalization, medical assistants), the effects of such systems to various stakeholders, and the measures of oversight to be set in place. In this sense, modular guidelines would guide developers through how to decide which outputs are or are not reasonable in a given context; this task requires not only knowledge of why and when decision-making is reasonable, but also why and when an alternate decision is not made.¹⁷² Clarke proposes that such guidelines focus on the life-cycle of data and include “do’s” and “don’ts” for each phase of its processing,¹⁷³ thereby not only providing guidance on how to determine the reasonableness of certain operations, but also stating what decisions considered unreasonable in any given phase. Such an approach would likewise

¹⁶⁸ Laurence Diver, ‘*Digisprudence: the design of legitimate code*’ (2020) *Law, Innovation & Technology*, 13(2), accepted for publication (pre-print available < <https://osf.io/preprints/lawarxiv/nechu> > last accessed 11 December 2020).

¹⁶⁹ Wachter and Mittelstadt (above note 6); Sartor and Lagioia (above note 6), 41.

¹⁷⁰ Cf. Roger Clarke, ‘Guidelines for the responsible application of data analytics’ (2018) *Computer Law & Security Review* 34, 467, 471 et seq.; Cf. also Sartor and Lagioia (above note 6).

¹⁷¹ Bayamlioglu (above note 32), 433 et seq.

¹⁷² Bayamlioglu (above note 32), 443 et seq. stating that “The conditions under which the decision would not have been made is a type of information necessary to construct counterarguments. (...) [c]ontrastive explanations (counterfactuals) as to the differences between two decisions enable a meaningful distinction between several decision alternatives: Counterfactuals are the building blocks of moral behavior as well as scientific thought.” Cf. also Wachter, Mittelstadt and Russell (above note 116).

¹⁷³ Clarke (above note 170), 471 et seq.; cf. for other life cycle-based approaches Tamò-Larrieux (above note 64), 6 et seq.; Wieringa (above note 95), 7.

move away from the regulation of the collection of data, as currently established within data protection law with a focus on personal data and provide guidance on how to analyze and make use of such data analysis. However, as the creation of rigorous guidelines covering every potential algorithmic decision-making system is unrealistic, a modular approach is necessary that, following a life cycle approach, provides design recommendations depending on the context of development/employment of a system as well as its effects.¹⁷⁴

Guidelines for reasonable automated decision-making would align with the literature in arguing that the focus of regulation should rest more on the way's companies employ data to achieve specific outcomes.¹⁷⁵ Thereby, moving away from the focus on input data (e.g., personal data) towards classification parameters, and more importantly the resulting decision-making processes. Taking a step into this direction is the Regulation 2019/1150,¹⁷⁶ which aims to promote fairness and transparency in the context of online intermediation services,¹⁷⁷ and requires service providers to “set out in their terms and conditions the main parameters determining ranking and the reasons for the relative importance of those main parameters as opposed to other parameters.”¹⁷⁸

Overall, the establishment of guidelines are also key in order to be able to implement principles of duty of care. Duty of care principles are a key tool to promote responsibility and, in the end, allocate accountability for certain actions. Developers and companies employing automated decision-making systems must take responsibility for their creation and implementations. An individual or collective duty of care provision could strengthen their commitment to responsibility instead of shifting the blame to the algorithms themselves. In fact, especially during the Covid-19 pandemic where we have more strongly relied on automation, we have seen how the public has started to blame the algorithms for its decisions (e.g., “f*** the algorithm”-protests have emerged in the UK after the grading scandal (see Section 1) or in the USA when the Stanford hospital relied on automated decision-making systems to determine who to vaccinate first). These protests and the blaming of automated decision-making system is not surprising considering the value-laden understanding of such systems (see Section 2). Yet comparing automated decision-making systems and artificial intelligence to humans and thereby (unconsciously) shifting the blame to algorithms instead of the humans relying on the algorithm is dangerous (see below Reframing the Discussion)

¹⁷⁴ Wieringa (above note 95) on modular accountability framework, 7.

¹⁷⁵ George, Reutimann and Tamò-Larrieux (above 87), 297; Hirsch (above note 7), 463 with reference to Fred Cate, Peter Cullen and Victor Mayer-Schönberger, ‘Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines’ (2014) available at <https://www.oii.ox.ac.uk/archive/downloads/publications/Data_Protection_Principles_for_the_21st_Century.pdf> (last accessed 29 May 2020), 8.

¹⁷⁶ Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (Text with EEA relevance), PE/56/2019/REV/1, OJ L 186, 11.7.2019, 57–79.

¹⁷⁷ I.e., online services that enable businesses to offer products and services to consumer, cf. Art. 2(2) Regulation 2019/1150.

¹⁷⁸ Art. 5(1) Regulation 2019/1150.

From the Individual to the Collective: Towards Social Protection Mechanisms

As highlighted above, the effects of automated decision-making processes occur not only on the individual or group level but have a systemic impact on society and democracy (see Section 3). Similar to addressing the health concerns during a pandemic, where the focus cannot merely rest on ‘own’ or ‘individual’ responsibility but instead should focus on a ‘joint’ or ‘communal’ perspective (i.e., doing what is best for the community), the effects of decision-making by machines necessitate a collective redress mechanism.

While the ideal of privacy does combine an individual and social perspective, current data protection law is still to a large extent focused on control and the individual. It necessary to move towards a more social or collective data protection frameworks. Different ways to shift towards social protection mechanisms have been proposed. Mayer-Schönberger, for instance, argues that audits and investigations by data protection authorities and public disclosure thereof, direct regulatory enforcement by data protection authorities, and the establishment of a network of information privacy professionals should support the individual rights based approach in Europe.¹⁷⁹ Similarly, Hirsch highlights the need for social protection and argues that data protection authorities must be empowered to decide which automated decision-making practices are “safe for individuals and consistent with social values, and which are not.”¹⁸⁰ He does not, however, argue for a complete paradigm shift towards the sole control of public authorities, but rather states that as soon as individuals cannot, due to the complexity of automated decision-making practices, meaningfully choose whether to consent to certain practices or not, the individual control paradigm fails, and should therefore be replaced by a social control system.¹⁸¹ Such calls align with literature on accountability (of algorithmic systems) that distinguishes, among other aspects, the audience to whom the account is directed to, and states that account needs to be given to an audience that is able to “engage with the material in a critical way.”¹⁸²

While the Directive 95/46/EC already contained provisions strengthening social control, the GDPR has not only strengthened the powers of data protection authorities, but also includes the duty, under certain conditions, to designate a privacy professional. Moreover, Art. 80 of the GDPR enables data subjects to mandate NGOs or other organizations to lodge complaints on their behalf with data protection authorities.¹⁸³ Member States may enable these organizations to “independently of the data subjects mandate” lodge complaints.¹⁸⁴ Thus, while the GDPR at least takes into account representative and collective protection mechanisms, the landscape within the

¹⁷⁹ Mayer-Schönberger (above note 7), 1883; Hirsch (above note 7), 439 et seq.; Hartzog (above note 148), 423 et seq.

¹⁸⁰ Hirsch (above note 7), 462; cf. also Hartzog, (above note 148), 423 et seq.

¹⁸¹ Hirsch (above note 7), 462 et seq.

¹⁸² Wieringa (above note 95), 4 with reference to Jakko Kemper and Daan Kolkman, ‘Transparent to whom? No algorithmic accountability without a critical audience’ (2018) *Information Communication and Society* 1.

¹⁸³ These NGOs or organization must follow various requirements determined by Member States; Art. 80(1) GDPR.

¹⁸⁴ Art. 80(2) GDPR.

Member States is likely to remain scattered.¹⁸⁵ In fact, further going suggestions, such as establishing Union-wide legal collective action mechanisms (class action rights),¹⁸⁶ and enabling consumer associations to “obtain the right to bring a representative action focused on anti-discrimination, data protection and other fundamental rights,”¹⁸⁷ however, have not been codified. This is not to say that the individual and due process norms found in data protection law do not offer social protection mechanisms at all. In fact, better individual (legal) tools could play a key role in enabling such legal collective action mechanisms. For instance, better, real-time access and portability (which interestingly is being proposed in the DMA) to data and the logic of automated decision-making system through simple interfaces could enable organizations, collectives, and researchers to better understand the side-effects of such systems. In a way such individual tools could enable more crowdsourced approaches to detect issues, raise awareness, with the aim to change the public and policy making discourse.

Increased institutionalized oversight likewise calls for audits and the creation of automated decision-making systems that enable “after-the-fact oversight.”¹⁸⁸ As most legal norms will typically lack clear specifications, developers must deal with the lack of precision by considering after-the-fact oversight solutions.¹⁸⁹ Likewise audits of algorithms either by the data processors themselves, external regulators, or even researchers have been proposed.¹⁹⁰ Policymakers have picked up on the idea of auditing in particular providers of intermediary services and gatekeepers in the proposed DSA and DMA. It remains to be seen how these proposals will evolve into regulation and impact the deployment of automated decision-making systems (e.g., for the removal of illegal content, recommendations). Nonetheless, it is necessary to establish criteria to determine when automated decision-making systems by companies that exist outside of the scope of those new regulatory proposal should be audited. These criteria

¹⁸⁵ Michael Frenzel, ‘DS-GVO Art. 80 Vertretung von betroffenen Personen’ in Boris P Paal and Daniel Pauly (eds), *Datenschutz-Grundverordnung* (3rd edn, beck-online, Munich, 2021) Rn 15 states that the provision is the result of a political compromise.

¹⁸⁶ Lilian Edwards and Michael Veale, ‘Enslaving the Algorithm: From a “Right to an Explanation” to a “Right to Better Decisions”?’ (2018) *IEEE Security & Privacy* 16(3) 46, 47 et seq.; Cf. also Proposal for a Directive of the European Parliament and of the Council on representative actions for the protection of the collective interests of consumers, and repealing Directive 2009/22/EC, COM/2018/0184 final - 2018/089 (COD) and the Outcome of Proceedings of the General Secretariat of the Council of 30 June 2020 available at <<https://www.consilium.europa.eu/media/44766/st09223-en20.pdf>> (last accessed 11 December 2020).

¹⁸⁷ Martini (above note 94), 35.

¹⁸⁸ Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu, ‘Accountable Algorithms’ (2017) *University of Pennsylvania Law Review* 165, 663, 696.

¹⁸⁹ Kroll (above note 188), 696 et seq. The reason for the lack of precision is that social norms and public opinions are volatile and imprecise. This is reflected in legislation as well as case law, which often interprets norms taking a teleological perspective. As policymakers are unable to account for every change and might be uncertain how to address the side-effects of automated decision-making, they leave room for future flexible interpretations of their legislation.

¹⁹⁰ Zarsky (above note 37), 125 et seq.; Kitchin (above note 41), 22 et seq.; Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort, ‘Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platform’ (2014) Paper presented to “Data and Discrimination: Converting Critical Concerns into Productive Inquiry” a preconference at the 64th Annual Meeting of the International Communication Association available at <<http://www-personal.umich.edu/~csandvig/research/Auditing%20Algorithms%20--%20Sandvig%20--%20ICA%202014%20Data%20and%20Discrimination%20Preconference.pdf>> (last accessed 29 May 2020); Philip Adler, Casey Falk, Sorelle A. Friedler, Tionney Nix, Gabriel Rybeck, Carlos Scheidegger, Brandon Smith, and Suresh Venkatasubramanian, ‘Auditing black-box models for indirect influence’ (2018) *Knowledge and Information Systems* 54 95; Martini (above note 94), 25 et seq. and 36 et seq.

should consider the lasting side-effects of such decisions on individuals as well as society at large. Audits should then be conducted by an independent auditor group with individuals that have different background and ask broader questions, such as: “Is the algorithm suitably transparent to end-users? Is it likely to be used in a socially acceptable way? Might it produce undesirable psychological effects or inadvertently exploit natural human frailties? Is the algorithm being used for a deceptive purpose? Is there evidence of internal bias or incompetence in its design? Is it adequately reporting how it arrives at its recommendations and indicating its level of confidence?”¹⁹¹

Reframing the Discussion

The mimicking and augmenting of human intelligence via algorithmic decision-making might project human traits of how we think, judge, and learn to algorithms. This projection might load automated decision-making with values, thereby complicating the relationship among humans and such systems. Indeed, individuals and society might pose unrealistically high expectations to automated decision-making systems. In the end, a decision by an algorithm is an elegant set of mathematical functions and not a judgment formed by parents, educators, and peers. When under the law we decide who is capable of judging and acting, as in creating legal rights and obligations, we apply established case law or legal norms that break down what “capable of acting” means. For instance, we determine that given a certain age and mental ability one is able to take legal actions that impact other individuals and society (e.g., get married and divorced, buy and drive a car). At this point, we do not have the same criteria or filter for automated decision-making systems, nor the possibility to make an algorithm subject to rights and obligations. The closest filter we have to describe the capacities of automated decision-making systems are automation levels (i.e., from zero to fully autonomous systems). These levels, however, are not helpful in taking a multi-dimensional perspective on the capacities and resulting impacts of automated decision-making systems. It is thus unsurprising that we are witnessing a complicated relationship with respect to assigning responsibility. While the “f*** the algorithm”-protests seem to blame the algorithms, it is clear that the anger is directed towards the effect of the algorithms and in the end the authorities in charge of these effects. Nonetheless, these chants might frame the public discussion in a too narrow way. Instead, we should focus on the natural and legal persons designing the systems, deciding whether and when to rely on them, implementing them, and determining when to alter them. Modular guidelines and collective protection mechanisms addressing these stages should not be static but enable a constant feedback and updates according to new developments. But such legal tools are only one aspect within the

¹⁹¹ James Guszcza, Iyad Rahwan, Manuel Cebrian and Vic Katyal, ‘Why We Need to Audit Algorithms’ (28 November 2018) Harvard Business Review available at <<https://hbr.org/2018/11/why-we-need-to-audit-algorithms>> (last accessed 29 May 2020).

much broader discourse and enabling a more educated discourse as well as the development of technical tools will be key to ensure that future automated decision-making systems serve the needs of individuals and society at large.

[6] Conclusion

The delegation of decision-making to machines has unsurprisingly triggered a debate on how to regulate the side-effects of such decision-making. A great deal of attention has been centred around data protection law, which combined an individual's rights approach with due processes and relies upon further statutes. The GDPR, which was shaped by policymakers concerns about decision-making capacities of machines, provides regulatory tools to address some of the raised side-effects. Nevertheless, sometimes the scope of those tools needs to be broadened in order to provide a redress mechanism. Depending on the extent of the broadening, conflicts with the principle of the rule of law arise and thus lead to unwanted legal uncertainty. Even if the current regulatory framework (including the statutes discussed within this articles that data protection law refers to) is interpreted broadly, shortcomings in addressing all the identified side-effects of automated decision-making systems still prevail. These shortcomings have led scholars to argue for further regulatory action. Yet, so far, novel regulatory approaches to address the issues remain more conceptual and abstract, and more granular guidance will be needed in the future. As the deployment of automated decision-making systems continues, their adaptation in each given context and resulting impact should be closely monitored in order to provide further guidance for policymakers on how to prevent or mitigate unwanted side-effects.