

The questionable necessity of a new human right against being subject to automated decision-making

Elena Abrusci* and Richard Mackenzie-Gray Scott†

ABSTRACT

The development and interest in decision-making that is or can be automated have opened the doors of debate regarding the form and substance of related means of regulating its application. Part of this discourse involves proposals advocating for the creation of a new human right not to be subject to an automated decision. This article questions whether such a right is necessary in light of existing substantive rules under legal frameworks already applicable to automated decision-making, specifically data protection, non-discrimination and human rights. There are also procedural challenges requiring treatment if automated decision-making is to be adequately addressed by application of the law. Exploring these challenges helps appreciate the significance of ensuring that existing substantive law is better implemented for the purpose of protecting human beings in settings where automated decision-making poses risks to individuals and groups.

KEYWORDS: law and technology, automated decision-making, data protection, non-discrimination, human rights, procedural safeguards

INTRODUCTION

The shift from decisions being made by humans to automated systems has prompted a variety of responses. From the apparent cost-effectiveness of automated decision-making (ADM) depending on the domain, to its (in)appropriateness across sectors such as employment or immigration, there remains much to uncover in deciding how best to utilize ADM for the

* Lecturer, Brunel University London. E-mail: elena.abrusci@brunel.ac.uk.

† Postdoctoral Fellow, Bonavero Institute of Human Rights, and Fellow, St Antony's College, University of Oxford. E-mail: richard.mackenzie-grayscott@law.ox.ac.uk. Research funded by the British Academy (grant no. BAR00550-BA00.01). Many thanks to Halefom Abraha, John Croker, Katie Lines, Katie Pentney, Keri Grieman, Lillian Edwards, Martin Scheinin, Oliver Butler, Tanya Krupiy and Tarik Gherbaoui for their helpful comments on the original draft and contributions to the workshop *Algorithmic Discrimination and the Associated Challenges for Human Rights*, held at the Bonavero Institute on 1 November 2022, where this article was first presented. Many thanks also to the Information Law & Policy Centre for the opportunity to present some of this research at its Annual Conference on *Online Safety in a Connected World*, held at the Institute of Advanced Legal Studies on 17 and 18 November 2022, where discussions with other participants provided valuable insights regarding the content and scope of the article.

benefit of individuals, businesses and societies, whilst mitigating, if not eliminating, the associated risks of doing so, including those concerning respect for, and the protection of, human beings. This is not to imply that the idea of ADM is new, which has been attracting attention since at least the 1950s.¹ Nor should the contents herein be read to infer that ADM is necessarily inherently problematic, at least when considered in the abstract, and perhaps even with respect to concrete applications. ADM can ensure the safe management of energy grids, where decisions sometimes need to be made quickly in order to prevent system-wide crashes, such as by redirecting or shutting down supply when there are electrical surges that can exceed the capacity of generators in a particular area.² Another use of ADM that assists various sectors and has become almost ubiquitous is that based on the Global Positioning System (GPS), which allows people to receive information on how to navigate this world through devices equipped with digital cartography capabilities.³ ADM based on GPS and related data is also a helpful example to highlight the extents to which humans are and should (not) be deferential towards the options presented by such systems. Just because ADM produces some results recommending the ‘best’ options, does not mean they are the best for those people making the ultimate decision or should be relied upon in general when choosing between options (for example, some people prefer experience over efficiency—meaning a longer, scenic route is preferable to a shorter, dreary one). The insidious aspect of this dynamic is that should a GPS application be used to gather more individual data on users, then routing options may end up being geared towards presenting those that will ensure users traverse past locations where there are goods or services of interest to them.⁴ In other words, that GPS software may no longer only be offering possible routes to a destination, but routes that are likely to generate user expenditure. If poor Bobby’s GPS software has been fed plenty of data to indicate that he’s partial to a McDonald’s, then the application will recommend routes putting him on course to see the golden arches at every chance it gets. Great for the multinational companies involved—not so great for Bobby’s bank account, self-control, or general health.

Although results produced by ADM can be perceived as reliable, whether they actually are reliable is a key issue. ADM is a tool to be used or abused, underutilized or over-relied upon, purposeful or pointless—not so different from a hammer, with the potential to be constructive or destructive.⁵ The importance of regulating its application includes safeguarding the human rights with which there is interplay. The proliferation of ADM in recent years with corresponding commentary across various fields of research has prompted questions yet to be settled regarding how best to achieve regulatory effectivity, including with respect to providing people and groups of people protection under the law from the adverse impacts of ADM. This article engages with an aspect of the related debate, specifically that focused on the apparent necessity of a new human right not to be subject to ADM. It first briefly sets out some concerns regarding ADM that appear to have prompted the notion that there is a need for a new human right. Second, it is questioned whether such a right is actually necessary in light of the current law already applicable to ADM, in particular that under the legal frameworks governing data protection, non-discrimination and human rights. Analysis of these bodies of law highlights the current strengths and weaknesses of the rules in place, as well as the current gaps

¹ TH Davenport and JG Harris, ‘Automated Decision Making Comes of Age’ (*MIT Sloan Management Review*, 15 July 2005) <<https://sloanreview.mit.edu/article/automated-decision-making-comes-of-age/>>

² J Barron, ‘The Blackout of 2003: The Overview; Power Surge Blacks out Northeast, Hitting Cities in 8 States and Canada; Midday Shutdowns Disrupt Millions’ (*New York Times*, 15 August 2003) <<https://www.nytimes.com/2003/08/15/nyregion/blackout-2003-overview-power-surge-blacks-northeast-hitting-cities-8-states.html>>

³ Y Li and others, ‘Feedback and Direction Sources Influence Navigation Decision Making on Experienced Routes’ (2019) 10 *Frontiers in Psychology* 1.

⁴ See generally A Pentland, *Social Physics: How Social Networks Can Make Us Smarter* (Penguin 2015).

⁵ With thanks to Aislinn Kelly-Lyth, Halefom Abraha and Six Silberman for stimulating this point through an enjoyable discussion.

in their practical and theoretical application. Third, the article outlines procedural capacities and oversight mechanisms that require treatment if the implementation of legal rules is to be achievable, and an effective means of protecting people from the adverse impacts of ADM, whilst successfully regulating its application more generally. Lastly, some thoughts are offered on the economic incentives that drive the use or not of ADM, specifically the interdependence between regulation and innovation, and how further consideration of this dynamic could alter a root cause of problems concerning ADM. Bringing these strands together, the conclusion offers some takeaways, with an invitation to consider how best to ensure tangible protections for human beings whilst making the most of ADM, which are grounded in legal rules and readily enforceable across states through appropriate machinery in order to provide adequate oversight and due process guarantees.

CONCERNS PROMPTING THE ADVOCACY FOR A NEW HUMAN RIGHT

During the Cold War, a potentially catastrophic event could have transpired. The former Soviet Union's system for detecting missile attacks recommended to Lieutenant Colonel Stanislav Petrov that he act in response to what the system stated with 'high reliability' was a United States missile attack.⁶ A nuclear war could have begun if the data provided by the missile detection system alone had been relied upon. However, even in the absence of verifiable information, Stanislav and his colleagues decided that the system was wrong and did not act on the alert.⁷ The decision was the right one, as the system 'mistook the sun's reflection off clouds for a missile'.⁸ Consider the inaccurate information in this case being given more deference by the individuals involved in the decision ultimately reached, or no humans being part of that decision. Although technology has developed since, this story provides a stark illustration of what is at stake should decisions be made without the involvement of human beings. From the perspective of respecting and protecting human rights, although the stakes might be different from those in the abovementioned event, the deployment of ADM raises a number of concerns when people are subjected to its use.

One example is the rolling out of ADM as part of the judicial process, where sentencing recommendations are issued to judges based on software analysing past cases with similar fact patterns.⁹ In theory, although human judges can be rational agents, with discretion not to adopt a recommended sentence generated via ADM, in practice, research suggests that the existence of an externally created reference point anchors judges' rulings, which in turn 'can generate ridiculous outcomes'.¹⁰ This is but one of a number of issues that arises when ADM is used within the judiciary, prompting considerations regarding the right to a fair trial.¹¹ Another related example is the use of ADM to assess the probability of recidivism, which in order to determine whether an accurate conclusion has been reached requires, among other things, being able to

⁶ D Matthews, '36 Years Ago Today, One Man Saved Us from World-ending Nuclear War' (*Vox*, 26 September 2019) <<https://www.vox.com/2018/9/26/17905796/nuclear-war-1983-stanislav-petrov-soviet-union>>; See also D Hoffman, 'Cold-War Doctrines Refuse to Die' (*Washington Post*, 15 March 1998) <<https://www.washingtonpost.com/wp-srv/inatl/longterm/coldwar/shatter031598a.htm>>

⁷ *ibid.*

⁸ *ibid.*

⁹ RE Stern and others, 'Automating Fairness? Artificial Intelligence in the Chinese Courts' (2021) 59 *Columbia Journal of Transnational Law* 515; W Ji, 'The Change of Judicial Power in China in the Era of Artificial Intelligence' (2020) 7 *Asian Journal of Law & Society* 515.

¹⁰ JJ Rachlinski and AJ Wistrich, 'Judging the Judiciary by the Numbers: Empirical Research on Judges' (2017) 13 *Annual Review of Law & Social Science* 203, 215.

¹¹ S de Heer, 'Administrative Automated Decision-Making: What About the Right to an Effective Remedy?' (*Oxford Human Rights Hub*, 8 April 2021) <<https://ohrh.law.ox.ac.uk/administrative-automated-decision-making-what-about-the-right-to-an-effective-remedy/>>

access and understand the risk-scoring methodology of the system and the data informing its calculations.¹² Outside the courtroom context, ADM is used to assist in the policing of states' borders, including with respect to the processing of asylum applications, a policy that appears to be favoured by the European Union and the Organisation for Economic Co-operation and Development.¹³ Policing is also a function that shows how ADM can lead to (further) racial profiling by law enforcement.¹⁴ A separate example is the development of military technology incorporating ADM that is deployed in situations where choices regarding who to kill and how are commonplace (for example, armed drones selecting and attacking targets).¹⁵

Even if there is a human part of such decision-making processes, the influence of ADM on that person means they may not appropriately scrutinize the decision that was reached via automation, especially if the matter is time sensitive. The risk here is humans proceeding unquestionably in favour of 'recommendations' generated by ADM. Essentially, instead of providing appropriate oversight, people rubber stamp outputs of ADM systems. This problem highlights the issue of 'automation bias',¹⁶ whereby 'humans tend to imbue machines with outside authority',¹⁷ and can attempt to shift responsibility away from themselves by reducing their involvement in ADM.¹⁸ In addition, people can fail to search for and consider information that contradicts ADM, which can also be assumed to be 'more reliable and complete than their own' decision-making.¹⁹ Deference of this sort is dangerous, not to mention potentially unwarranted depending on the particular system of ADM.

A key feature of processes that utilize ADM is the data-driven approach underpinning the technology, which generates a veneer promising a combination of accuracy, neutrality and speed being injected into decision-making. It has been shown over decades' worth of research that humans are not that good at being objective or rational when making decisions, whether individually or collectively.²⁰ Groupthink occurs when overconfident individuals remain unchallenged in an assembly, even when such conformity is indefensible due to an inaccurate position being advanced.²¹ Individuals are also notoriously poor at making predictions and recognizing potential.²² People's decisions can become inconsistent and irrational depending on their feelings and mood, which are influenced by a variety of factors, ranging from the physiological (for example, hunger or thirst) to the psychological (for example, unconscious bias), to the sociological (for example, what people are present at the time when a decision is being made). The allure of ADM is that it has the potential to not be influenced by what are considered by some to be deficiencies in human judgment. A factor in support of this stance is that factual data are used to inform ADM, which, depending on provenance, can result in considerable deference being lent to the associated decisions because they are perceived as being based on

¹² A Završnik, 'Criminal Justice, Artificial Intelligence Systems, and Human Rights' (2020) 20 *Journal of the Academy of European Law* 567.

¹³ European Migration Network, *The Use of Digitalisation and Artificial Intelligence in Migration Management*, Joint EMN-OECD Inform (Brussels, February 2022).

¹⁴ D Moeckli, 'General Recommendation No. 36 (2020) on Preventing and Combating Racial Profiling by Law Enforcement Officials (C.E.R.D.)' (2021) 61 *International Legal Materials* 351.

¹⁵ UN Security Council, Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council (8 March 2021), UN Doc. S/2021/229, paras 63–64 and Annex 30 (p 148).

¹⁶ DK Citron, 'Technological Due Process' (2008) 85 *Washington University Law Review* 1249, 1271–72, defining 'automation bias' as the 'use of automation as a heuristic replacement for vigilant information seeking and processing'.

¹⁷ SK Glaberson, 'Coding Over the Cracks: Predictive Analytics and Child Protection' (2019) 46 *Fordham Urban Law Journal* 307, 355.

¹⁸ R Gsenger and T Strle, 'Trust, Automation Bias and Aversion: Algorithmic Decision-Making in the Context of Credit Scoring' (2021) 19 *Interdisciplinary Description of Complex Systems* 542, 547.

¹⁹ S Sacher, 'Risking Children: The Implications of Predictive Risk Analytics Across Child Protection and Policing for Vulnerable and Marginalized Children' (2022) 22 *Human Rights Law Review* 1, 6.

²⁰ D Kahneman, O Sibony and CR Sunstein, *Noise: A Flaw in Human Judgment* (Little, Brown Spark 2021); D Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux 2011).

²¹ CR Sunstein and R Hastie, *Wiser: Getting Beyond Groupthink to Make Groups Smarter* (Harvard Business School Press 2014).

²² PE Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton University Press 2017).

scientifically and methodologically rigorous grounds, whilst providing a means of time-saving and cost-reduction.²³

The *Motherhood Plan* case serves as a helpful illustration.²⁴ Here the Treasury of the UK Government introduced the Self Employment Income Support Scheme in an attempt to assist those adversely affected by the economic impact of the COVID-19 pandemic and the associated measures adopted by the government.²⁵ The Treasury decided to set the payable amount based on a calculation accounting for 80% of the mean monthly profits of an applicant over the three years previous to 2020 (when the Scheme opened), which relied on ADM to predict how much revenue businesses would have otherwise generated in the ordinary course of events but for the pandemic and the related responses from government.²⁶ Past profitability was thus assumed to be an accurate proxy for profitability in the present. As the amount payable would be lower for women who had been on maternity leave anytime during those previous three years versus people who had not, the claimants argued that the Scheme discriminated against women that take maternity leave and thus infringed their rights under Article 14 when read with Article 1, Protocol 1 of the European Human Convention on Human Rights (ECHR).²⁷ The Court disagreed, holding that there had been no such discrimination.²⁸ The justification went as far as to express that even if the discrimination had been found, there were reasons that would have rendered it lawful, including that the Scheme was designed ‘by reference to data already held by HMRC’, meaning ‘claims could be automated, which achieved speed and cost savings.’²⁹ This reasoning has since been upheld on appeal.³⁰ The case is a cautionary tale of the deference that can be provided to decisions that are data-driven and cost-effective. There is a risk with equating these combined features with accuracy, regardless of potential (in)compatibility with data protection, non-discrimination and human rights laws.

Whatever decision is ultimately reached by ADM, there will be questions regarding what data is consumed by the particular system in order to reach the related conclusions. A key one is how a system establishes *relevant* predictors, meaning from which sources does it draw in order to create criteria that ultimately inform its data processing. Depending on these individual sources of information, when accumulated as part of an aggregate they can be over- or under-representative. As highlighted by the research of Veronika Fikfak, assessing whether any such dataset is (in)complete or (un)balanced is challenging.³¹ Whether ADM is indeed accurate and neutral therefore depends on the data underpinning the applicable system, raising considerations regarding what information is made available to a particular system, and what weight it is afforded. Those favouring ADM often laud it on the basis of the *quantity* of data that a particular system can process, which humans could not assess in the same amount of time, if at all.

However, the *quality* of the data is equally if not more important. Issues of bias and unfairness arise out of data mis-accumulation and subsequent mis-reading. In a world where the language of growth appears to be unstoppable, it is questionable whether the key drivers of change favouring the use of ADM (i.e. economic efficiency and optimization) favour quantitative data over

²³ R Mackenzie-Gray Scott, *Judicial Scrutiny of COVID-19 Regulations in the UK: Addressing Deference to Data-Driven Decision-Making in Human Rights Cases* (Bingham Centre for the Rule of Law 2021).

²⁴ *The Motherhood Plan v HM Treasury* [2021] EWHC 309 (Admin).

²⁵ *ibid* para 10.

²⁶ *ibid* paras 7–34.

²⁷ Convention for the Protection of Human Rights and Fundamental Freedoms (signed 4 November 1950, entered into force 3 September 1953), 213 UNTS 221.

²⁸ *The Motherhood Plan v HM Treasury* [2021] EWHC 309 (Admin), paras 62–68.

²⁹ *ibid* para 79; see also paras 80–85.

³⁰ *The Motherhood Plan v HM Treasury* [2021] EWCA Civ 1703; for commentary see A McColgan, ‘R (The Motherhood Plan & Anor) v HM Treasury’ (*Equality Law Blog*, 14 January 2022) <<https://equalitylawblog.com/2022/01/14/r-the-motherhood-plan-anor-v-her-majestys-treasury/>>

³¹ V Fikfak, ‘What Future for Human Rights? Decision-Making by Algorithm’ (*Strasbourg Observers*, 19 May 2021) <<https://strasbourgothers.com/2021/05/19/what-future-for-human-rights-decision-making-by-algorithm/>>

qualitative data, when it is arguably a combined balance of both that are necessary to produce utility-maximizing systems of ADM. Should the scales skew in favour of quantity, datasets may turn out to be faulty. And even if the data are not faulty, with the dataset being quantitatively and qualitatively robust, the manner in which it is interpreted by a system that involves ADM may be flawed. Long story short, data informing ADM can be flawed. As can it create inappropriate proxies and be inaccurately and unclearly interpreted. Yet making judgments about ADM and the data underpinning it rests on a further factor: explainability, which refers to whether and the extent to which a particular system of ADM is accessible and comprehensible to the person(s) assessing it.³²

Brent Mittelstadt, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter and Luciano Floridi have crafted a map highlighting the types of ethical concerns that are applicable to ADM.³³ These include those of an epistemic nature (inconclusive evidence, inscrutable evidence and misguided evidence), those raising normative considerations (unfair outcomes and transformative effects), and traceability, which is the common denominator across the other concerns.³⁴ Their research is instructive to the challenges that ADM poses to laws aiming to protect human beings, particularly those regarding data protection, non-discrimination and human rights. From the apportionment of responsibility across actors such as designers, companies and states for the consequences arising from ADM, to the levels of trust in the processes and reviews of related systems, it is understandable that there are discussions about creating new human rights to engage with the associated advancements in technology,³⁵ including a new human right not to be subject to ADM.³⁶

But what does this right entail? Dafna Dror-Shpoliansky and Yuval Shany set out their vision for such a right, which they believe can 'respond to wholly new threats or challenges that did not really exist before the digital age.'³⁷ The need for this new right, in their view, stems from a shift away from decisions being made by humans with moral intuitions and agency towards systems that lack the capability to incorporate these traits.³⁸ This is an appealing basis for a new right to be constructed upon, especially considering the significance of interpersonal interactions between people where their humanity can be 'mutually recognized.'³⁹ Having decisions made about an individual by other individuals is an important cornerstone of human rights, which, Dror-Shpoliansky and Shany highlight, dates back to the Magna Carta.⁴⁰ As ADM systems change this dynamic, it has been argued that a new right not to be subject to ADM is needed. This position may be supported by current practice—such as Article 22 of the EU General Data Protection Regulation (GDPR)—that could be geared towards the 'the emergence over time of a generally applicable new international digital human right not to be subject to an automated decision in decisions significantly affecting important areas of life.'⁴¹ According to Dror-Shpoliansky and Shany, what would justify the creation of this and other 'digital human rights' is the imperative of addressing 'structural causes for injustice and inequality.'⁴² On this view, human rights form part of movements towards these aims, because, by offering protection to autonomy, individuals are empowered to strive for better approaches to governance of particular social

³² U Franke, 'First- and Second-Level Bias in Automated Decision-Making' (2022) 35 *Philosophy & Technology* 1.

³³ BD Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) *Big Data & Society* 1, 4.

³⁴ *ibid.*; see also pp. 5–12.

³⁵ M Ienca and R Andorno, 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology' (2017) 13 *Life Sciences, Society & Policy* 1.

³⁶ D Dror-Shpoliansky and Y Shany, 'It's the End of the (Offline) World as We Know It: From Human Rights to Digital Human Rights—A Proposed Typology' (2021) 32 *European Journal of International Law* 1249, 1274–79.

³⁷ *ibid.* 1274.

³⁸ *ibid.* 1277.

³⁹ *ibid.*

⁴⁰ *ibid.* 1278; see also WS Holdsworth, *A History of English Law* (1956), 59, para 39.

⁴¹ Dror-Shpoliansky and Shany, 'It's the End of the (Offline) World as We Know It' (2021) 1279.

⁴² *ibid.*

groups.⁴³ ADM dehumanizes people, impacting a range of aspects concerning human life, including those specific to particular social groups. The argument of Dror-Shpoliansky and Shany appears to be along the lines of guarding against such impacts, in that having a new human right not to be subject to ADM will better protect the needs and interests of individuals that are boxed into ‘algorithmic categories’.⁴⁴ These commentators also believe that such a new right ‘would convey more clearly to technology companies the standards of conduct that they are expected to follow than would general standards derived from traditional human rights’, whilst asserting that implanting it in practice is ‘likely to improve compliance with international human rights law norms’.⁴⁵ This belief that a new human right would improve the practice across the continuum of public and private actors in their respect for and engagement with human rights is sanguine—a welcome part of a puzzle filled with dispiriting content.

Yet it is important not to fall into a trap. This trap is less about the concerns surrounding so-called rights inflation, and more about a general pretence that the substantive law alone enshrining human rights can adequately address problems posed by ADM.⁴⁶ While there are a number of criteria that can be created and referred to for the purpose of determining when any new proposed human right should be considered as such for the purposes of practice,⁴⁷ there is the related, but separate, overarching matter of whether that right is ultimately necessary from the perspective of what it would add to legal practice. One way of approaching this question is to scrutinize the law already in existence that is applicable to the subject matter of the proposed right and reflect on the related findings to evaluate what, if anything, it would add substantively to existing law. The next section, therefore, analyses existing law across data protection, non-discrimination and human rights that are applicable to ADM. In doing so, it is questioned why a new human right not to be subject to ADM is necessary from the perspective of legal practice. What also becomes apparent is the need to enhance the enforcement machinery and oversight mechanisms applicable to ADM, and alter the underlying causes of reliance on ADM that contribute to its use creating tensions with, and breaches of, existing legal rules.

THE NECESSITY QUESTION IN LIGHT OF CURRENT LAW AND ITS (POSSIBLE) INTERPRETATION

By analysing the applicability of data protection, non-discrimination and human rights law to ADM, it becomes clear that there is a considerable corpus of legal rules that can be drawn from in order to address the risks posed by ADM whilst assessing its legality. These rules also show that current legal frameworks can provide an adequate means of engaging with the problems posed by ADM without the need for a new ad hoc human right. What is concerning with respect to the applicable law at present are the issues of implementation and oversight, which require further attention and treatment through procedural developments.

Data protection

As data can be considered the backbone of the digital realm,⁴⁸ and the fuel of ADM systems, data protection laws have become a key source for regulating ADM and protecting individuals

⁴³ See C Beitz, ‘What Human Rights Mean’ (2003) 132 *Daedalus* 36.

⁴⁴ Dror-Shpoliansky and Shany, ‘It’s the End of the (Offline) World as We Know It’ (2021) 1280.

⁴⁵ *ibid* 1281.

⁴⁶ For insights on rights inflation see JT Theilen, ‘The Inflation of Human Rights: A Deconstruction’ (2021) 34 *Leiden Journal of International Law* 831.

⁴⁷ See, for example, P Alston, ‘Conjuring up New Human Rights: A Proposal for Quality Control’ (1984) 78 *American Journal of International Law* 607.

⁴⁸ See generally F Fabbrini, E Celeste and J Quinn (eds.), *Data Protection beyond Borders: Transatlantic Perspectives on Extraterritoriality and Sovereignty* (Bloomsbury 2021).

against possible problems produced by its use. Data protection at present is mostly a matter of domestic competence, and laws across states differ, sometimes significantly, with some states being comparatively more protective of data subjects' rights, and others being more accommodating to data controllers' agendas.⁴⁹ Within the picture of different data protection regimes, the GDPR stands out globally as a detailed and comprehensive instrument. The GDPR contains a dedicated human right not to be subject to ADM. Article 22(1) provides that:

the data subject shall have 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.⁵⁰

On first reading, this article appears to be, at least intuitively, what a human right not to be subject to ADM should look like. However, its formulation has proven controversial, particularly in relation to what constitutes a decision based 'solely' on automated processing.⁵¹ This condition can be interpreted in a manner whereby minimal human involvement in a process involving ADM, potentially with no feasible possibility to affect the related outcome, would make the subsequent decision fall outside the scope of this provision. That said, Rebecca Williams argues that, at least with respect to profiling, the 'solely' element may not be overly problematic in practice considering the European Data Protection Board has clarified that profiling involving automated processing will fall within the scope of article 22.⁵² Furthermore, should it be proven that human supervisors of ADM systems have succumbed to automation bias, then it could be argued that the particular system at issue was making decisions solely on automated processing, as the human in the loop was ineffectual in their role. Yet this could be a difficult argument to prove.

The overall formulation of this provision has been criticized elsewhere on a separate ground, relating to the requirement of producing a 'legal effect'. For example, it excludes practices such as price discrimination, where online offers are generally regarded as invitations to treat and do not produce legal effects in and of themselves.⁵³ In addition, the provision provides exceptions in paragraph 2, explaining that paragraph 1 does not apply if the decision:

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) is based on the data subject's explicit consent.

These exceptions are susceptible to various interpretations and possibly constitute something of a get-out-of-jail-free card for data controllers, depending on how sophisticated any legal

⁴⁹ See the study on Jordan, Lebanon, Palestine and Tunisia by Access Now in M Fatafa and D Samaro, 'Exposed and exploited: Data protection in the Middle East and North Africa', Access Now, 28 January 2021 and the report on the lack of an appropriate federal legislation in the United States; see also the progress made by the new data protection law in India and in Ecuador: <https://www.accessnow.org/data-protection-laws-in-2022/>; Data Protection Act 2018 (UK).

⁵⁰ GDPR, art 22.

⁵¹ See, among others, S Dreyer and W Schulz, 'The General Data Protection Regulation and Automated Decision-making: Will It Deliver?' (2019) Bertelsmann Stiftung 18–20; S Wachter and B Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2 Columbia Business Law Review 494.

⁵² R Williams, 'Rethinking Administrative Law for Algorithmic Decision Making' (2022) 42 Oxford Journal of Legal Studies 475; see also M Veale and L Edwards, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' (2017) 34 Computer Law & Security Review 398.

⁵³ S Hanold, 'Profiling and Automated Decision-Making: Legal Implications and Shortcomings' in M Corrales, M Fenwick and N Forgo (eds), *Robotics, AI and the Future of Law* (Springer 2018), 123, 135.

arguments are in advancing a particular interpretation. This backdrop, therefore, benefits parties with the resources to hire legal teams that are experts in manipulating language for the benefit of their clients. Such asymmetry between data subjects and data controllers can be vast, making it a challenging feat for a data subject to enforce their rights enshrined in article 22(1). A related issue is the use of ADM by a public authority, meaning exception (b) of paragraph 2 can be evoked so long as the state in question is able to argue that 'suitable' safeguards are in place, which may not be suitable in the context—again meaning disputes on this ground will likely weigh in favour of states with substantial resources at their disposal to defend against any legal claims brought by data subjects with comparatively minimal resources. That said, article 22 and its related recitals require the data controller to establish safeguards even when invoking exceptions, including the addition of a human reviewer to address the data subject's concerns, who need not be ineffective in their role.⁵⁴

Although article 22 has limitations and offers legal loopholes for data controllers to play with,⁵⁵ the GDPR also regulates a wide range of elements both at the input and output stage of an ADM system. The GDPR takes a 'self-determination' approach,⁵⁶ with its starting point being the need to ensure that the data subject is in control of their data throughout the stages of ADM. This translates into the requirements, set by articles 12 and 13, for the data controller to obtain consent from the data subject for the collection or processing of the related data. The data subject should have full rights regarding their data, which means that they have the right to access (article 15), rectification (article 16) and erasure (article 17). They also have a right to withdraw their consent from data processing (article 21). All these provisions are, on paper at least, comprehensive, but raise significant challenges in their practical application. Thorough and strong implementation mechanisms are key, which could better protect individuals and groups against ADM-generated harms. But these appear to be lacking at present. For example, on the issue of consent, data controllers have repeatedly attempted to deceive and manipulate data subjects into providing their consent through 'I agree'-type check boxes that can be difficult to close, circumvent or reject, in addition to disclosing no or limited information on how data will be used. Transparency and access to information are two of the most important elements underpinning the GDPR, relevant throughout data lifecycles. If data subjects are to provide free, prior and informed consent, they need to receive adequate and intelligible information on how data are collected, stored and processed.

A related feature of access to information and explanations regarding ADM is being made aware of being subject to an automated decision, receiving adequate information on how the ADM processing data works, what data the applicable system has been 'trained' with, and what data have been used to profile people. The debate on whether or to what extent the GDPR contains a right to explanation is active.⁵⁷ The specific contours of this possible right are disputed. There are different ways of explaining ADM, depending on whether the explanation is *ex ante* (for example, of the system functionality) or *ex post* (for example, of the decisions reached by the system). According to Sandra Wachter and others, the GDPR does not contain a meaningful right to explanation, as accessibility appears to be limited to system functionality.⁵⁸ Gianclaudio Malgieri and Giovanni Comandé take a different view, arguing that if article 22 is read in conjunction with articles 13–15, the GDPR could provide a right to legibility, understood as the capability of individuals to independently

⁵⁴ See GDPR, art 22, paras 3 and 4 and commentary.

⁵⁵ Wachter and others, 'No Right to Explanation of Automated Decision-Making' (2017).

⁵⁶ BD Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) *Big Data & Society* 1, 14.

⁵⁷ AD Selbst and J Powles, 'Meaningful Information and the Right to Explanation' (2017) 7(4) *International Data Privacy Law* 233–42; Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) 16 *IEEE Security & Privacy* 3, 46–54.

⁵⁸ Wachter and others, 'No Right to Explanation of Automated Decision-Making' (2017), 87.

understand data and analytics algorithms.⁵⁹ Such a right would practically mean a right to know the existence of a given processing system and meaningful information about its logic, significance and consequences, both *ex ante*, as a preliminary information supply duty of data controllers,⁶⁰ and *ex post*, after specific requests by data subjects.⁶¹

The opacity of ADM, due to both lack of information available and/or information provided in unclear or obscure terms, also raises issues of trustworthiness. There is also the issue of responsabilization,⁶² requiring arguably absurd levels of awareness from data subjects. In order to exercise a right, people first need to be aware of its existence, such as in a legal instrument, the GDPR, say. Then they need to understand whatever right they think they have if they are going to be in a position to make any use of it (not that big of a problem: this only involves consuming and understanding pages of law and commentary ad nauseam or having access to someone that decided this type of work was a sensible career choice). People also need to realize how they are the subjects of a particular ADM system. Then they need to know how to proceed in making inquiries with respect to their data and its use in that system, and what they can do should data controllers be uncooperative. And potentially the most significant factor across all these steps (in addition to money, obviously)? Time. What people have the time to undertake this knowledge gathering, and then use it in a related inquest? There is a gap here between principle and practice, rendering any rights to explanation and not being subject to ADM ineffectual at scale. Ronan Hamon and others have shown that related problems may also lie in the technology at issue, and not just in the efforts and willingness of data controllers to be transparent.⁶³ Complex ADM systems, those that use 'deep-learning', for example, are often not able to make clear causal links between input data and models, thus failing to provide satisfactory, fair and transparent explanations when they are questioned. Additional data protection tools, such as algorithmic impact assessments and other algorithmic explanation models, thus require further exploration and operationalisation by bodies that have the resources to act on behalf of data subjects' interests.

The imprecise language of the GDPR may also leave components of ADM out of its scope. For instance, as Sandra Wachter and Brent Mittelstadt argue, inferences are 'economy class' personal data in the GDPR, as they do not have the same protection granted to other data by articles 13–21.⁶⁴ And ADM is based on inferences. Reduced protection on this category of data may jeopardize the protection of individuals and groups against the harm that ADM may produce. The GDPR offers protections around data collection and processing (as it focuses on data input), but does not offer such protection to data evaluation and data outputs. This created a regulatory gap, but a recent judgment by the Court of Justice of the European Union (CJEU) seems to have filled it. The Court recently ruled that article 9 GDPR should be interpreted as applied also to inferences.⁶⁵ This effectively extends the applicability of the GDPR to the use of data for processing and profiling. The GDPR also requires data controllers to assess the risks arising from profiling and ADM before their deployment. Article 35 of GDPR requires data controllers to evaluate the potential consequences of their data-processing activities via

⁵⁹ G Malgieri and G Comandé, 'Right to Legibility of Automated Decision-Making' (2017), 245.

⁶⁰ GDPR, arts 13–14.

⁶¹ *ibid* art 15.

⁶² J Pysiäinen, D Halpin and A Guilfoyle, 'Neoliberal Governance and 'Responsibilization' of Agents: Reassessing the Mechanisms of Responsibility-Shift in Neoliberal Discursive Environments' (2017) 18 *Journal of Social Theory* 215; N Rose, 'Governing "Advanced" Liberal Democracies' in A Barry, T Osborne and N Rose (eds), *Foucault and Political Reason: Liberalism, Neo-Liberalism, and Rationalities of Government* (University of Chicago Press 1996), 37.

⁶³ R Hamon and others, 'Bridging the Gap between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making' (2022) *IEEE Computational Intelligence Magazine* 73–85.

⁶⁴ S Wachter and B Mittelstadt, 'A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI' (2019) *Columbia Business Law Review* 2, 494–620, 499.

⁶⁵ Case C-184/20 [2022] *OT v Vyriausioji tarnybinės etikos komisija* (Chief Official Ethics Commission, Lithuania).

a data protection impact assessment and communicate these risks to the persons concerned. However, as pointed out elsewhere, these impact assessments can be limited and carried out as a ‘box ticking’ exercise.⁶⁶ Moreover, such evaluations focus exclusively on GDPR rules, ignoring potential impacts on human rights under other legal frameworks.⁶⁷ It is also unclear what legal is available if the impact assessment finds that there is a risk associated with an ADM system.

Overall, the GDPR contains a wide range of provisions that offer protection against possible harms generated by the use of ADM throughout its lifecycle, from data collection to delivery. However, the margin of interpretation of these provisions is wide enough to benefit data controllers with substantial resources.⁶⁸ Article 22, in particular, has been critiqued to the extent that there exist claims that it should be rewritten, because its current formulation, according to Paul De Hert and Guillermo Lazcoz, is too complex, ambiguous and lacks clarity, making ‘it difficult to apply.’⁶⁹ Reuben Binns and Michael Veale have commented that ‘[s]ome might conclude that Article 22 is thus conceptually beyond saving.’⁷⁰ The need for effective enforcement and monitoring is particularly noteworthy, especially considering the tangible risk of data controllers benefiting from merely superficial compliance with existing law. Shortcomings in the current data protection framework thus emphasize the significance of exploring other bodies of law that serve to protect against the problems associated with ADM weighing on data subjects.

Non-discrimination

In addition to data protection, non-discrimination laws are another crucial part of the toolbox of legal frameworks that can offer protection to individuals against the problems posed by ADM. Non-discrimination laws are established at a national level to avoid and sanction unequal treatment, including in situations and settings where ADM systems operate.⁷¹

However, the application of these laws to systems of ADM reveals specific pressure points. As Jeremias Adams-Prassl, Reuben Binns and Aislinn Kelly-Lyth observe, no ADM system is completely free of bias, so it is in a sense natural for them to discriminate.⁷² In particular, the practice of profiling increases the likelihood of discrimination to arise. ‘Machine learning’ systems, on which some ADM is based, have the goal of providing a seemingly rational basis upon which individuals and groups can be distinguished.⁷³ Discrimination is not just a side effect of such ADM but also a core feature of it, and, as such, it requires appropriate regulation in order to avoid arbitrary and/or unjustifiable discrimination. There are two types of discrimination under which the law provides guidance on compliance: direct and indirect. If an ADM system treats individuals less favourably on grounds of, or because of, a legally protected characteristic (such as race or sex), it will constitute direct discrimination. If this occurs, the deployment of ADM is unlawful, even if there was no intent to discriminate or harm people from those who provided or deployed the

⁶⁶ L Naudts, ‘How Machine Learning Generates Unfair Inequalities and How Data Protection Instruments May Help in Mitigating Them’ in R Leenes, R van Brakel, S Gutwirth and P De Hert (eds), *Data Protection and Privacy: The Internet of Bodies* (Bloomsbury 2018), 71, 72.

⁶⁷ ME Kaminski and G Malgieri, ‘Algorithmic Impact Assessment under the GDPR: Producing Multi-Layered Explanations’ (2020) 11 *International Data Privacy Law* 125.

⁶⁸ D Geradin, T Karanioti and D Katsifis, ‘GDPR Myopia: How a Well-Intended Regulation Ended Up Favouring Large Online Platforms—The Case of Ad Tech’ (2021) 17 *European Competition Journal* 47.

⁶⁹ PDHertandGLazcoz, ‘RadicalRewritingofArticle22GDPRonMachineDecisionsintheAIera’ (*EuropeanLawBlog*, 13 October 2021) <<https://europeanlawblog.eu/2021/10/13/radical-rewriting-of-article-22-gdpr-on-machine-decisions-in-the-ai-era/>>

⁷⁰ R Binns and M Veale, ‘Is that Your Final Decision? Multi-stage Profiling, Selective Effects, and Article 22 of the GDPR’ (2021) 11 *International Data Privacy Law* 319, 331.

⁷¹ R Allen and D Masters, ‘Artificial Intelligence: The Right to Protection from Discrimination Caused by Algorithms, Machine Learning and Automated Decision-making’ (2019) *ERA Forum* 20, 585–98.

⁷² J Adams-Prassl, R Binns and A Kelly-Lyth, ‘Directly Discriminatory Algorithms’ (2022) *Modern Law Review*.

⁷³ L Naudts, ‘How Machine Learning Generates Unfair Inequalities and How Data Protection Instruments May Help in Mitigating Them’ in R Leenes, R van Brakel, S Gutwirth and P De Hert (eds.), *Data Protection and Privacy: The Internet of Bodies* (Bloomsbury 2018), 71, 72.

ADM system. However, legally protected characteristics do not include poverty or socio-economic status at present, meaning ADM systems can lawfully discriminate against people that are poor or on the basis of their social class. An ADM system that is programmed to favour certain groups over others, such as has been the case with those utilizing algorithms to generate credit scores,⁷⁴ screen for benefit fraud⁷⁵ or process visa applications,⁷⁶ tends to fall within this category. But should such systems of ADM not discriminate on the basis of a protected characteristic, they will be considered lawful, even though they may discriminate on the basis of other characteristics not considered to merit protection under the law at this time.⁷⁷ Whether particular characteristics should become protected under equality law is therefore a pressing question in light of developments in ADM. Poverty in particular warrants attention, because data informing a number of ADM systems can be interpreted as a proxy for poverty, whether correctly or incorrectly, and the use of ADM in particular sectors can exploit and worsen the circumstances of poor people.⁷⁸ Whether acting as a gatekeeper for who gets welfare support or offering predatory loans to people in immediate need of money, ADM helps maintain occupancy on the lower rungs of societies' socio-economic ladders.⁷⁹

Unprotected characteristics aside, EU law and domestic law, such as the Equality Act in the UK, do not only address direct discrimination, but indirect discrimination as well, which is the result of an ostensible neutral treatment that nevertheless has a discriminatory effect on a legally protected group. A number of ADM systems fall into this category, such as those used in child protection and welfare services in the US,⁸⁰ and for creating and implementing 'no fly' lists at airports (also in the US).⁸¹ ADM gives the illusion of neutrality, but if the data used to 'train' the system is biased, or individuals are targeted based on profiling, then its application may produce significant discrimination. Non-discrimination law exists to protect discriminated groups in such cases, despite the burden of proof being a considerable hurdle depending on the case in question and the skills and knowledge of those representing the parties involved. Indirect discrimination is unlawful if the discriminatory actions are proven to be disproportionate to the legitimate aim sought to be achieved. This means that indirect discrimination can be justified, and, depending on the interpretation and understanding of the principle of proportionality, there can be (very) different conclusions on the lawfulness of the ADM system under scrutiny. It has been argued that this could generate mechanics of self-justifying feedback loops, where the same system that discriminates provides justification for its proportionality and legitimacy.⁸² One example is an ADM system used to identify areas with higher crime risk. If the system is 'trained' using biased data that would discriminate against specific groups, the

⁷⁴ See Decision no 216/2017 of the National Non-Discrimination and Equality Tribunal of Finland (21 March 2018), where factors including gender and age had been labelled as inputs in a credit scoring system.

⁷⁵ Tax authorities in the Netherlands used an algorithm to attempt the detection of benefit fraud, which included nationality as a risk factor. See Amnesty International, 'Dutch Childcare Benefit Scandal an Urgent Wake-Up Call to Ban Racist Algorithms' (25 October 2021) <<https://www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal/>>

⁷⁶ BBC News, 'Home Office Drops "Racist" Algorithm from Visa Decisions' (4 August 2020) <<https://www.bbc.com/news/technology-53650758>>

⁷⁷ But see the potential of realizing article 26 of the ICCPR in practice considering its language and the possibility of extending protection to groups based on other identity factors: 'All persons are equal before the law and are entitled without any discrimination to the equal protection of the law. In this respect, the law shall prohibit any discrimination and guarantee to all persons equal and effective protection against discrimination on any ground such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status.'

⁷⁸ V Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (Macmillan 2018).

⁷⁹ M Gilman, *Poverty Algorithms: A Poverty Lawyer's Guide to Fighting Automated Decision-Making Harms on Low-Income Communities* (Data & Society 2020).

⁸⁰ SK Glaberson, 'Coding Over the Cracks: Predictive Analytics and Child Protection' (2019) 46 *Fordham Urban Law Journal* 307.

⁸¹ DK Citron, 'Technological Due Process' (2008) 85 *Washington University Law Review* 1249, 1256–57.

⁸² Adams-Prassl, Binns and Kelly-Lyth, 'Directly Discriminatory Algorithms' (2022).

fact that more police patrols will take place in the areas where those groups are more present, will likely increase arrest rates (even if the same arrest rate could have been achieved by deploying the same amount of police patrols in another area with a different group composition), thus superficially justifying the use and results of the particular ADM system.

Discrimination caused by ADM poses additional challenges to the application of the existing non-discrimination legal framework, especially in relation to the evidentiary requirements for bringing a claim under non-discrimination law. In particular, there are issues concerning the awareness of discrimination, the access to information for judicial review, and the attribution of liability. To take the latter issue first, discrimination caused by ADM raises issues of attribution of liability to the actor that deploys the ADM system, and to the providers of that system. At present neither the EU nor UK legal frameworks on non-discrimination currently appears to address this matter, meaning the assessment is left to judges on a case-by-case analysis.⁸³ However, further support could be provided by rigorous impact assessments and oversight mechanisms.⁸⁴ A broader issue under non-discrimination law is that individuals or groups need to experience inequality or a disadvantageous treatment to be able to trigger the application of the relevant procedures. However, on numerous occasions, when ADM discriminates, it does so in a subtle and intangible manner, which is difficult to detect or experience, even for the targeted individual. Compared to other discriminatory practices, it is difficult for an individual to know of discrimination generated by ADM.⁸⁵ For instance, how could someone know if they have been discriminated against by an ADM system when purchasing something online *without* any reference point to distinguish price variations, or when *not* seeing certain job advertisements? As Sandra Wachter, Brent Mittelstadt and Chris Russell reflect, in addition to this unseeability being an issue in and of itself, as it decreases the likelihood of people bringing forward discrimination claims, it also complicates evidence submissions to courts and tribunals, as well as judges' task of providing comprehensive rulings based on available evidence.⁸⁶ When deciding on allegedly biased ADM systems, judges appear to heavily rely on contextual elements and intuition on a case-by-case basis, and applicants struggle to provide evidence of the discrimination they felt they had suffered.⁸⁷

To assess whether a targeted group has experienced a disadvantage, alleged discriminatory treatment requires a comparison with the 'neutral' treatment, or the treatment of the general group, at issue. In online settings where groups can be treated differently because of their data profile, it is difficult if not an insurmountable hurdle to identify such a 'gold standard'.⁸⁸ These elements significantly decrease the actual protection that the law can offer to individuals and groups who suffer discrimination by way of ADM. However, there may be several corrective measures that could be implemented to deliver on the promises of non-discrimination laws when applied to ADM systems. For example, there are proposals that aim to support the judicial review of discrimination cases involving ADM.⁸⁹ Such guidance, which should not replace judicial intuition, but be used in conjunction with it, shows that the limits of the current legal protections against ADM do not so much lie in the legal framework of non-discrimination, but in its implementation and oversight.

⁸³ J Buyers, *Artificial Intelligence: The Practical Legal Issues* (Law Brief Publishing 2018), 67–70.

⁸⁴ See L Edwards, 'Regulating AI in Europe: Four Problems and Four Solutions' (*Ada Lovelace Institute*, 31 March 2022) <<https://www.adalovlaceinstitute.org/report/regulating-ai-in-europe/>>

⁸⁵ R Xenidis and L Senden, 'EU Non-discrimination Law in the Era of Artificial Intelligence: Mapping the Challenges of Algorithmic Discrimination' in U. Bernitz et al. (eds), *General Principles of EU law and the EU Digital Order* (Kluwer Law International 2020), 151–82.

⁸⁶ S Wachter, B Mittelstadt and C Russell, 'Why Fairness Cannot Be Automated: Bridging the Gap between EU Non-Discrimination Law and AI' (2021) 41 *Computer Law & Security Review* 6.

⁸⁷ *ibid.*

⁸⁸ A Tischbirek, 'Artificial Intelligence and Discrimination: Discriminating against Discriminatory Systems' in T Wischmeyer and T Rademacher (eds), *Regulating Artificial Intelligence* (Springer 2020), 103, 121.

⁸⁹ Wachter, Mittelstadt and Russell, 'Why Fairness Cannot Be Automated' (2021).

Human rights

The third occupant in the toolbox of legal frameworks that can regulate ADM is human rights law. Although nowhere in international human rights law, nor in the regional or domestic implementation of human rights, is there a specific right not to be subject to ADM, the broad and comprehensive human rights framework offers protection against the possible adverse impacts of ADM. If combined, existing legal provisions under this body of law arguably offer protection to all human rights that could be negatively impacted by the use of ADM, to which a new ad hoc right would need to refer. From a human rights perspective, assessing whether and how a specific ADM system should be limited or prohibited requires scrutiny based on the challenges it poses to specific human rights. It follows that what is required is not a new human right on ADM, but enabling existing human rights to be effectively understood, interpreted, applied, and enforced in contexts where ADM is provided and deployed.

Whether adopting a general natural law standpoint or a position of intersubjective constructivism, human rights exist because of their attachment to human beings, meaning their existence and applicability does not depend on the domain where they may be called upon.⁹⁰ Human rights law can apply wherever humans are present (online, offline, Earth, space, the moon) and is thus applicable to ADM.⁹¹ As Roger Brownsword and Morag Goodwin argue, human rights can be considered as key boundary-marking concepts of regulatory legitimacy of new and emerging technologies, including ADM.⁹² Likewise, human rights law can be used as a framework to help guide the design and deployment of ADM, whilst providing a measure of accountability when such systems produce problems that impact individuals.⁹³ That said, if it is true in theory that human rights law applies to systems of ADM, its enforcement in practice poses challenges.

Several specific human rights are called into question by the use of ADM systems, which also have the potential to protect against possible harms. There are two groups of rights in this respect: (1) those that affect the functioning of ADM and may prevent harm from the outset if accounted for properly, and (2) those that protect against the harm as a consequence of the negative impact of an ADM system. The first category includes human dignity and autonomy, the right to privacy and non-discrimination, while the second comprises a plethora of rights depending on the context where the ADM is deployed, including the right to life and to education, and freedom of expression, thought, religion, assembly and association.

It is helpful to begin discussions on ADM and human rights within a framing of human dignity. Human dignity, closely linked to human autonomy and agency, can be considered at the core of the general relationship between human rights and so-called artificial intelligence, as it upholds the fundamental components of humanity. However, this cornerstone of human rights has been stretched and twisted by commentators, legal practitioners and policymakers to serve and justify many actions, omissions and policies.⁹⁴ Human dignity is a blurry concept. Immanuel Kant reflected that '[h]umanity itself is a dignity; for a human being

⁹⁰ J Gardner, 'Simply in Virtue of Being Human: The Whos and Whys of Human Rights (2008) 2 Journal of Ethics & Social Philosophy 1; J Griffin, 'Discrepancies Between the Best Philosophical Account of Human Rights and the International Law of Human Rights' (2001) 101 Proceedings of the Aristotelian Society 1.

⁹¹ L McGregor and others, 'The Universal Declaration at 70: Putting Human Rights at the Heart of the Design, Development, and Deployment of Artificial Intelligence' (*Human Rights, Big Data and Technology Report* 2018), 1–8.

⁹² R Brownsword and M Goodwin, *Law and Technologies of the Twenty-First Century* (Cambridge University Press 2012), 188–91; see also C Beitz, *The Idea of Human Rights* (Oxford University Press 2009), 44.

⁹³ L McGregor, D Murray and V Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019) 68 International and Comparative Law Quarterly 309.

⁹⁴ See a full discussion of this matter in A Sharkey, 'Autonomous Weapons Systems, Killer Robots and Human Dignity' (2018) 21 Ethics and Information Technology 75, 79–82.

cannot be used merely as a means by any human being [...] but must always be used at the same time as an end'.⁹⁵ This position assumes that humans are moral agents who are entitled to determine preferences and assign value to them.⁹⁶ From this understanding, human dignity protects the autonomy of the individual and their agency to freely exercise their rights. International human rights law recognizes the foundational nature of human dignity in article 1 of the Universal Declaration of Human Rights, which states that 'All human beings are born free and equal in dignity and rights'. While untrue in practice, it is difficult to argue with the principle.

Human dignity is considered a catch-all principle to which everyone can subscribe. When applied in the context of ADM, human dignity challenges the use of ADM that risks hindering autonomy and agency,⁹⁷ with exceptions being made where there has been consent. Human dignity requires some individual control over the ADM system at issue. What this means for the purposes of practice is the realization of meaningful transparency (knowledge about being subject to ADM), explainability (access to comprehensible information about the function and impact of the particular system) and express consent over its deployment. This human rights approach bolsters related requirements under data protection and non-discrimination laws. An ADM system that would operate without the consent of the individual concerned, negatively affect core aspects of their life, and limit their agency, should not be provided or deployed when considered from this perspective of human dignity.

For example, an ADM system that takes the ultimate decision on whether an individual should be subject to certain medical treatments, and does so without the express consent of the patient or their loved ones, would run contrary to human dignity and would thus be prohibited on this view. Likewise, a defence of human dignity should question the use of ADM in armed conflict for deciding the target of a deadly attack, or in criminal justice systems across states, especially in states where ADM is likely to aggravate existing prejudices. To respect human dignity, the requirement of transparency and access to intelligible information for free, prior and informed consent should not be merely nominal, but able to allow the individual in question to operate freely and come to a reasoned decision about their own life. Situations where the individual does not have any other choice but to 'consent', or where the information provided is minimal and/or obscure, would ultimately fail an assessment based on human dignity. Studies conducted by Hauke Behrendt and Wulf Loh show how discrimination ties into such assessments, in particular, because vulnerable groups are structurally more pressured to provide their data and 'consent' to data collection and processing, which can end up further increasing discrimination against them.⁹⁸

In addition to human dignity, the right to privacy, enshrined in article 17 of the International Covenant on Civil and Political Rights (ICCPR), acts as a gatekeeper right for ADM systems.⁹⁹ The interests and values the provision aims to protect are linked and overlap with data protection law. Yet its international appearance confers it further scope of applicability, also in the absence of a data protection framework. The intertwined nature with other human rights also makes the right to privacy more comprehensive than under data protection law. The right to privacy is considered necessary to protect the 'autonomous zone' within each person's life, where individuals

⁹⁵ I Kant, *The Metaphysics of Morals*, trans. and ed. Mary Gregor (Cambridge University Press 1996), 209.

⁹⁶ Brownsword and Goodwin (2012), 193.

⁹⁷ S Wachter and B Mittelstadt, 'A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI' (2019) 2 Columbia Business Law Review 494, 497.

⁹⁸ H Behrendt and W Loh, 'Informed Consent and Algorithmic Discrimination—Is Giving Away Your Data the New Vulnerable?' (2022) 80 Review of Social Economy 58.

⁹⁹ Lorna McGregor and others, 'Putting Human Rights at the Heart of the Design, Development, and Deployment of Artificial Intelligence' (2018) 11.

should be able to make decisions without interference.¹⁰⁰ As such, in the context of ADM, the right to privacy supplements human dignity and establishes related requirements that can be used to guide its design, deployment and development. As echoed by the Office of the High Commissioner for Human Rights, ADM, the collection and processing of personal data and the inferences that lead to profiling, significantly threaten the right to privacy and to self-determination of the individual.¹⁰¹ The amalgamation of scattered information about an individual can be processed so as to create detailed—and potentially inaccurate—insights about who that person is and what they like and dislike. This practice can be undertaken for the purpose of attempting to predict what an individual may do in the future, and also to attempt to change their thoughts and conduct towards pre-determined outcomes.¹⁰² This practice risks violating the right to privacy, the protection of which is crucial to preventing the breach of other rights (both offline and online), which are contextually applicable to ADM.¹⁰³ As explained earlier with respect to human dignity, subject consent is crucial and should be free, prior and informed, echoing the data protection requirements of explainability and transparency applicable to ADM systems for which data are collected and processed.¹⁰⁴ Yet it should be recalled that the imprecision of these international requirements and, more problematically, the lack of adequate enforcement and monitoring machinery, can make them somewhat ineffective for the purposes of legal practice, even though they do carry weight in other respects, such as politically.

Another component of the human rights framework that is constantly at stake when ADM systems operate and that should inform their conception, design, development and deployment is the principle of equality and the right to non-discrimination. Set forth by the Universal Declaration in article 2 and by the ICCPR in articles 2 and 26, the right to non-discrimination shares the underlying premise of domestic non-discrimination laws. As in the case of the right to privacy, an international provision has the merit of a wider and cross-border application with some sense of commonality in its language and application. However, internationalization does not appear to have thus far provided tangible solutions for addressing the complexities and shortcomings of domestic non-discrimination laws, which remains an obstacle to meaningful protection for individuals subject to ADM, because these are the rules by which individuals can enforce their rights under domestic legal machinery.¹⁰⁵

Non-discrimination also links to privacy with respect to personal data and profiling based on protected characteristics, which combine in settings where ADM is utilized and can bring further human rights into play depending on the applicable domain(s). For instance, using ADM in judicial systems for the ostensible purpose of supporting judges' deliberations, or to decide on bail and parole, is controversial also from a right to fair trial standpoint, in addition to privacy and non-discrimination.¹⁰⁶ The right to a fair trial under human rights law contains specific safeguards such as presumption of innocence and being tried before a competent, impartial and independent court or tribunal,¹⁰⁷ which when used as a yardstick highlights some warnings

¹⁰⁰ *Niemietz v Germany*, (1993) 16 EHRR 97, para 29.

¹⁰¹ UN High Commissioner for Human Rights, Report on 'The Right to Privacy in the Digital Age', A/HRC/39/29, 3 August 2018, para 30.

¹⁰² See generally S Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Profile Books 2019).

¹⁰³ UN Human Rights Council, Resolution on 'The Right to Privacy in the Digital Age', UN Doc. A/HRC/34/7 (19 December 2016); UN General Assembly, Resolution on 'The Right to Privacy in the Digital Age', UN Doc. 68/167 (18 December 2013); D Kaye, UN Special Rapporteur on Freedom of Expression, Annual report 2015, UN Doc. A/HRC/29/32 (22 May 2015).

¹⁰⁴ UN High Commissioner for Human Rights, Report on 'The Right to Privacy in the Digital Age', A/HRC/48/31 (13 September 2021).

¹⁰⁵ FJ Zuiderveen Borgesius, 'Strengthening Legal Protection Against Discrimination by Algorithms and Artificial Intelligence' (2020) 24 *International Journal of Human Rights* 1572, 1576–78.

¹⁰⁶ M Langford, 'Taming the Digital Leviathan: Automated Decision-Making and International Human Rights' (2020) 114 *AJIL Unbound* 141.

¹⁰⁷ See generally A Clooney and P Webb, *The Right to a Fair Trial in International Law* (Oxford University Press 2021).

when applied to the ADM systems used in these settings.¹⁰⁸ Susie Alegre provides examples, including defendants being considered ‘guilty’ from their facial expressions when analysed by pseudoscientific software, or judges’ reasoning being hampered by ‘advice’ generated by ADM that may be inaccurate, biased and unexplainable.¹⁰⁹ Fair trial rights of this sort link to privacy (for example, surveillance, extraction and accumulation of sensitive data), freedom of thought (particularly not being forced to reveal thoughts)¹¹⁰ and go hand in hand with the prohibition of discrimination, which ADM in criminal justice systems continuously feed and exploit.¹¹¹

In addition to the courtroom context, ADM systems are also prevalent elsewhere, prompting considerations of yet further sets of human rights. A key area here is the management of online content. This includes content filtering and curation, whether deciding what content should appear at the top of social media newsfeeds, or be visible (or invisible) to select users, and content moderation, including the removal of unlawful or potentially harmful content. These practices can restrict the right of users to receive adequate and varied information, as well as to freely explore and express their ideas and opinions.¹¹² In turn, freedom of thought can be negatively impacted, including by influencing opinions to the extent where people unwittingly become a means to an end, such as commercially or politically, with their minds at risk of being manipulated in various ways. Although it is necessary to appreciate the limits of human rights law in providing meaningful and tangible measures to address such issues involving ADM,¹¹³ working human rights considerations into related operational structures is possible, and can also inform monitoring and oversight bodies.¹¹⁴ Other systems of ADM have the potential to negatively impact various other human rights, such as that to life and education, as well as freedom of religion, assembly and association.¹¹⁵ This includes situations where the use of an ADM system produces a ‘chilling effect’ on human rights, meaning that such systems may not contravene human rights *law* directly, but deter and discourage the enjoyment of human rights for fears of possible negative consequences, including where ADM is utilized by states, especially those of an authoritarian nature.

The combination of law across human rights, data protection and non-discrimination set thresholds for what conduct involving ADM is legally permissible, even though these thresholds are not always clear. There are thus numerous ways in which the design or deployment of ADM can be legally impermissible from the perspective of these three bodies of law. What is hopefully apparent at this stage is that there is a toolbox of *substantive* legal rules that are already applicable to ADM. And this set of tools is numerically and functionally significant. When considering the processes and impacts of ADM systems, it is worth considering what current rules *cannot* cover that a new right not to be subject to ADM could. It is difficult to envisage how the

¹⁰⁸ JM van Gyseghem, ‘Fundamental Rights and the Use of Artificial Intelligence in Court’ in LA DiMatteo et al. (eds.) *The Cambridge Handbook of Lawyering in the Digital Age* (Cambridge University Press 2021), 257–71.

¹⁰⁹ See generally S Alegre, *Freedom to Think: The Long Struggle to Liberate Our Minds* (Atlantic Books 2022).

¹¹⁰ S Alegre, ‘Rethinking Freedom of Thought for the 21st Century’ (2017) 3 *European Human Rights Law Review* 221, 225; UNGA, Freedom of thought, Interim report of the Special Rapporteur on freedom of religion or belief, A/76/380 (5 October 2021), para 25.

¹¹¹ H Behrendt and W Loh, ‘Informed Consent and Algorithmic Discrimination—Is Giving Away Your Data the New Vulnerable?’ (2022) 80 *Review of Social Economy* 58.

¹¹² D Kaye, *Speech Police: The Global Struggle to Govern the Internet* (Columbia Global Reports 2019); D Kaye, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/73/348, 29 August 2018; I Khan, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/HRC/47/25, 13 April 2021.

¹¹³ H Hannum, *Rescuing Human Rights: A Radically Moderate Approach* (Cambridge University Press 2019); J Tasioulas, ‘Saving Human Rights from Human Rights Law’ (2019) 52 *Vanderbilt Journal of Transnational Law* 1167.

¹¹⁴ R Mackenzie-Gray Scott, ‘Managing Misinformation on Social Media: Targeted Newsfeed Interventions and Freedom of Thought’ (2023) 21 *Northwestern Journal of Human Rights* 109; B Sander, ‘Democratic Disruption in the Age of Social Media: Between Marketized and Structural Conceptions of Human Rights Law’ (2021) 32 *European Journal of International Law* 159; B Sander, ‘Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation’ (2020) 43 *Fordham International Law Journal* 4.

¹¹⁵ Access Now, *Human Rights in the Age of Artificial Intelligence* (November 2018).

application of ADM has affected a human being without some form of domain contextualization. And with that context will likely come substantive legal rules that are already applicable to the situation. In other words, there appears to be no feasible scenario in which a human right not to be subject to ADM would not already be addressed by another rule under at least data protection, non-discrimination and human rights law. What use of ADM could *not* be addressed by rules under these bodies of law that are already applicable to such systems?

ADM use in a courtroom: rules on the right to a fair trial, and freedom of thought. ADM use for the local police force: rules on non-discrimination, and privacy. ADM on your favourite social media platform: rules on data protection, and freedom of expression. ADM in child welfare and protection services: rules on information privacy and social security. The list can go on and on. And it is worth recalling the other areas of law that have not been examined here, which are also applicable to ADM (for example, competition, consumer, public, labour and tort),¹¹⁶ and can inform whilst (in certain instances) provide recourse to people that have been negatively impacted by ADM,¹¹⁷ as well as guiding the design, development and deployment of such systems.¹¹⁸ To briefly give a more specific example, the duty of candour under public law requires that public authorities justify their conduct through the transparent production of relevant information, and, according to the Court of Appeal, in judicial review cases there is a 'very high duty on central government to assist the court with full and accurate explanations of all the facts relevant to the issue that the court must decide'.¹¹⁹ Such reasoning can be applied to the providers and deployers of ADM.¹²⁰ The same can be said for the test of reasonableness under public law that is applied to decision-making, which can extend to ADM.¹²¹ All this is to say that there is already considerable substantive law that applies to practices involving ADM across the continuum of public and private actors. It is a challenge to envisage what a new human right not to be subject to ADM would bring to the table in terms of offering legal protections that are not already provided by other existing legal rules. The combination of the three bodies of law examined here offers substantive protections that span the apparent gamut of possible uses of ADM. With respect to the applicability of human rights law in particular, the use of ADM in a specific domain would inform the existing rules that could be brought to bear on the specific situation. While there is some political value that comes with the introduction of a new human right not to be subject to ADM, which brings different sets of questions to be answered, the value of such a rule for the purposes of legal practice is at the very least questionable.

Intersectionality via extrapolation: A saving grace?

However, there is a defence for the existence of a new human right not to be subject to ADM that lies beyond its value in terms of political advocacy and the related pressure for change that could be generated across states via socialization. There may be something to be gained in legal practice from such a standalone human right that lies in the nature of intersectionality.¹²² There

¹¹⁶ With many thanks to Bethany Shiner for prompting this point.

¹¹⁷ See Williams, 'Rethinking Administrative Law for Algorithmic Decision Making' (2022).

¹¹⁸ L Edwards, R Williams and R Binns, *Legal and Regulatory Frameworks Governing the Use of Automated Decision Making and Assisted Decision Making by Public Sector Bodies* (Legal Education Foundation 2021); A Drake and others, 'Legal Contestation of Artificial Intelligence-related Decision-Making in the United Kingdom: Reflections for Policy' (2022) 36 *International Review of Law, Computers & Technology* 251; B Cartwright, 'Regulating the Robot: A Toolkit for Public Sector Automated Decision-Making' (2021) 10 *Oxford University Undergraduate Law Journal* 23.

¹¹⁹ *R (Quark Fishing Ltd) v Secretary of State for Foreign and Commonwealth Affairs (No. 1)* [2002] EWCA Civ 1409 [50].

¹²⁰ See also *Graham v Police Service Commission* [2011] UKPC 46 [18] where it was made clear that the duty of candour requires the disclosure of 'materials which are reasonably required for the court to arrive at an accurate decision'.

¹²¹ *Associated Provincial Picture Houses v Wednesbury Corporation* [1948] 1 KB 223 (CA); *Ex p Barry* [1997] AC 584; *Tandy* [1998] AC 714; *R (Quila) v Home Sec* [2011] UKSC 45, [2012] 1 AC 621; *R (Wandsworth v Schools Adjudicator)* [2003] EWHC 2969, [2004] ELR 274; *R (Law Society v Legal Services Commission)* [2010] EWHC 2550, [2011] ACD 16; *A v Home Secretary* [2004] UKHL 56, [2005] 2 AC 68.

¹²² With many thanks to Katie Pentney for raising this consideration and for her insightful comments on the matter.

is value in having a new rule that encompasses elements from separate but related rules, because this approach can avoid shoehorning dated rules into contemporary problems. Lawmaking occurs in order to create such new rules, including at the international level,¹²³ where the method of extrapolation is used to create a new rule that is a mix of components of existing rules in order to present a better option for addressing a particular problem.¹²⁴ When reflecting on the legal frameworks of data protection, non-discrimination and human rights, a new human right against being subject to ADM could account for the interactions between existing legal rules, including those that exist beyond these three bodies of law, meaning potential breaches would not need to be determined individually under separate bodies of law. A downside of creating new rules in this way is that:

If the content of a rule stems from different, albeit related, rules, then, because of the dissimilar content used to create the whole, there can be diverse opinions on it, which can create inconsistencies in how the rule is interpreted and applied.¹²⁵

Yet the upsides of having a general rule that encompasses components of specific rules from different bodies of law are that it improves integration in legal practice on a particular issue, combats fragmentation and can develop and prescribe concrete lines of conduct on that issue for interested actors.¹²⁶ Providers and deployers of ADM could therefore be better informed on what they should and should not do if a new human right against being subject to ADM existed that manages to capture key elements from different legal frameworks and distil them in a clear and coherent way. Cohesion and commonality to the particular issue of ADM could therefore be achieved through the introduction of a new rule, instead of fragmented approaches to issues involving ADM through distinct bodies of law. And what better way to ground such a rule than in a human rights framing?

But regardless of whatever substantive law ends up being created or is already applicable to ADM, the toolbox of procedural capacities to implement and oversee these substantive rules in practice is comparatively deficient. The significance of legal rules lies, at least in part, in their effective application in practice. This is not the same as conflating the existence of law with its enforcement, but instead highlights that without tangible implementation of legal rules, their content can become meaningless for those hoping to rely on the law as a means of protection. Enhancing the related enforcement and oversight machinery surrounding legal rules is therefore crucial, which in the case of ADM provides opportunities to better regulate it. Although there are intrinsic limitations in this regard, bolstering procedural capacities relating to the existing substantive rules currently applicable to ADM arguably serves a practical purpose for the protection of human beings by application of the law.

BOLSTERING APPLICABLE PROCEDURAL CAPACITIES FOR BETTER REGULATION OF ADM

It is worth reaffirming that the advocacy behind creating a new human right not to be subject to ADM is important. Efforts of this sort are helpful, because they form part of pushes for change

¹²³ The Articles on Responsibility of States for Internationally Wrongful Acts created by the International Law Commission is one example (Articles on Responsibility of States for Internationally Wrongful Acts, UN Doc. A/56/10, 43, UN Doc. A/RES/56/83 (2001), Annex, UN Doc. A/CN.4/L.602/Rev 1, GAOR 56th Session Supp 10, 43).

¹²⁴ See S Sivakumaran, 'Techniques in International Law-Making: Extrapolation, Analogy, Form and the Emergence of an International Law of Disaster Relief' (2017) 28 *European Journal of International Law* 1097.

¹²⁵ R Mackenzie-Gray Scott, 'Due Diligence as a Secondary Rule of General International Law' (2021) 34 *Leiden Journal of International Law* 343, 350.

¹²⁶ See *ibid* 354 and 360.

and the hope of better forms of governance. Although it is difficult to envisage what such a new human right would add in terms of substance to the legal frameworks examined above, save perhaps with respect to providing an intersectional approach to issues involving ADM, should such a rule continue to be advocated for *and* entail the premise that no human should *ever* be subject to its use, then such a stance would help push back against the normalization and legitimization of ADM, even if the realization of this position is perhaps unrealistic in a world today where ADM is already used in a variety of sectors.¹²⁷ On the flip side, if such advocacy declares subjecting humans to ADM is acceptable, but that the processes of doing so must be limited by parameters constructed by a proposed new human right, then such a rule would serve to further legitimize the use of ADM, even if seeking to curb that use. The position of Dror-Shpoliansky and Shany seems to be on this path:

Regulating through human rights norms the division of labour between human and algorithmic decision-makers would also make an important contribution to the human right-friendly development of AI, big data and other digital technologies applied in cyberspace.¹²⁸

The focus on so-called artificial intelligence from various fields of research is trendy. There is considerable hype regarding the subject. Whether this is warranted will not be addressed here, including because it has been addressed elsewhere,¹²⁹ but suffice to say that every bit of further work undertaken by researchers ostensibly concerned about the problems that ADM poses, further legitimizes the development and use of this technology unless that research proposes outright bans. The role of any proposed new human right not to be subject to ADM requires careful consideration in this respect.

Yet regardless of the precise contours of such a right, the significance of related advocacy is nonetheless not legal in nature. The significance instead extends to the signalling effect it provides for social practice and political purposes. Whether during a session of an intergovernmental organization or at a local protest, advocating for the creation of a new human right not to be subject to ADM serves different purposes. But it is important not to conflate these purposes and related processes with the application of the law for the purpose of regulatory guidance and determining responsibility. And should conflation not occur, it has been shown that attempting to apply such a new human right for the purposes of legal practice would add little, if anything, in terms of substance to the law that already exists across data protection, non-discrimination and human rights, save perhaps in the nature of intersectionality.

However, although these bodies of law can address problems posed by ADM, the extent to which they can do so *effectively* depends on related procedural machinery. While there have been calls for new sector-specific rules to enable the functionality of substantive legal rules, so that the latter may actually impact the use of ADM systems,¹³⁰ it is clear there are matters requiring treatment with respect to implementation. This section, therefore, seeks to map out a few areas of exploration related to the procedural capacities associated with the substantive rules across the bodies of law examined above. Procedural elements of governance frameworks play a significant role in effective regulation, including with respect to ADM. Ari Ezra Waldman points out that a common denominator of procedural machinery applicable to such systems is that it provides ‘guardrails to prophylactically

¹²⁷ T Hooley, ‘A War against the Robots? Career Guidance, Automation and Neoliberalism’ in T Hooley, R Sultana and R Thomsen (eds), *Career Guidance for Social Justice: Contesting Neoliberalism* (Routledge 2018).

¹²⁸ Dror-Shpoliansky and Shany, ‘It’s the End of the (Offline) World as We Know It’ (2021), 1280.

¹²⁹ See F Kalthéuner (ed), *Fake AI* (Meatspace Press, 2021).

¹³⁰ FJ Zuiderveen Borgesius, ‘Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence’ (2020) 24 *International Journal of Human Rights* 1572.

shoehorn algorithmic decision-making into an accountability regime'.¹³¹ Ensuring due process is dependent on the practical implementation of procedure, and has been described as the 'essence' of human rights applicable to ADM.¹³² This is not a difficult position to support.

It is all hunky-dory having a substantive rule stipulating y must do or not do x , but if there are no accompanying steps at the procedural level, x can become ineffectual for the purposes of legal practice. Helmut Aust understands this frustration well: 'There is no better recipe for disappointment than an immense development of international law in substantive terms which then falls foul of a lack of accompanying steps on the procedural level'.¹³³ In order to guide decision-making on the use of ADM, and assess the apportionment of responsibility for when these systems malfunction, governance regimes require means of gathering and scrutinizing necessary information. Without evidence, there is no way of understanding what occurs during the use of a particular ADM system, what steps were taken for the purposes of mitigating possible harms, and what are reasonable expectations under the circumstances of its deployment, which in turn forms part of determining whether due diligence has been exercised or not by those that utilize ADM systems.¹³⁴ As aptly summarized:

Procedural requirements like algorithmic impact assessments, source code transparency, explanations of either the result or the logic behind it, and a human in the loop who can hear someone's appeal move opaque automated systems closer to more familiar, and more accountable, decision-making regimes.¹³⁵

Oversight and impact assessments

Real-time governance of ADM is crucial, with human oversight of such systems being of particular importance. But being in the literal position to provide for appropriate oversight of ADM is a capacity-building and resource allocation issue. Not every operator or regulator of ADM systems has the same capabilities and resources with respect to providing adequate oversight. Aside from the assessment of substantive legal rules in light of such circumstances, in that they can create differentiated legal obligations across actors that create and utilize ADM,¹³⁶ there is the overarching problem of educating people to oversee ADM systems. While there is a need for a human in the loop, such a tick box requirement is practically useless if the human in question does not understand how the related ADM system functions.¹³⁷ Familiarity with the particular system, and with the context in which it operates, is key. For this reason, in addition to deployers of ADM systems hiring new employees that are first trained and then tasked with oversight duties, these deployers could also re-train employees to undertake oversight duties, and/or re-hire those that have been laid off, perhaps due to the very introduction of ADM.¹³⁸ The added benefit of this latter approach is that previous employees will be familiar with the sector where the ADM operates. And that contextual, sector-specific knowledge is valuable for ensuring that

¹³¹ AE Waldman, 'Power, Process, and Automated Decision-Making' (2019) 88 *Fordham Law Review* 613, 624.

¹³² DK Citron, 'Technological Due Process' (2008) 85 *Washington University Law Review* 1249, 1255.

¹³³ HP Aust, *Complicity and the Law of State Responsibility* (Cambridge University Press 2011) 427.

¹³⁴ A Coco and T Dias, 'Cyber Due Diligence': A Patchwork of Protective Obligations in International Law' (2021) 32 *European Journal of International Law* 771.

¹³⁵ Waldman, 'Power, Process, and Automated Decision-Making' (2019), 624.

¹³⁶ B Griffey, 'The "Reasonableness" Test: Assessing Violations of State Obligations under the Optional Protocol to the International Covenant on Economic, Social and Cultural Rights' (2011) 11 *Human Rights Law Review* 275.

¹³⁷ McGregor, Murray and Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019).

¹³⁸ HJ Holzer, 'Understanding the Impact of Automation on Workers, Jobs, and Wages' (*Brookings Institution*, 19 January 2022) <<https://www.brookings.edu/blog/up-front/2022/01/19/understanding-the-impact-of-automation-on-workers-jobs-and-wages/>>

the system is monitored properly. It also cuts the cost of recruitment for employers and alleviates the financial burdens of redundancy for previous workers.

People overseeing ADM need to be well-equipped to identify issues and take corrective action when needed. Being properly trained could help counteract automation bias, especially if employers create mechanisms whereby the responsibility of employees is renounced and instead lies with the provider and/or deployer. As alluded to above, people fear involving themselves with ADM in case something goes wrong, so as to pre-emptively relieve responsibility for involvement in the process. Creating positive incentives for those who oversee ADM is therefore crucial, so that they can work without worrying that they will be considered responsible should the ADM system they are overseeing produce harmful outcomes. Such an approach to apportioning and assessing responsibility gives preference to collective, over individual, responsibility.

Such oversight also includes the use of impact assessments. Information gathering about ADM processes will help providers and deployers of ADM see the true value of the particular system under scrutiny. This is particularly important considering the prevalent assumption that ADM is a means of cost savings, because it might not be. For example, imagine a situation where a particular system is poorly designed by a provider, the data informing it is not qualitatively and quantitatively robust, and the deployer using it does not become aware of its problems until potentially both actors are ultimately (successfully) sued before a court or tribunal. Here, that provider and deployer will have suffered costs associated with litigation, including those concerning rulings handed down or settlements reached, and the costs of the ADM system being suspended pending the outcome of proceedings, or declared unlawful afterwards and ordered prohibited.¹³⁹ As Robin Allen and Dee Masters emphasize, auditing of these systems helps ensure they are defensible whilst providing proof for providers and deployers of compliance with applicable rules.¹⁴⁰ If proper impact assessments are undertaken during the design and use of ADM, then people overseeing a particular system can identify and address any issues pre-emptively, reducing the risks of being sued and incurring the associated costs. In addition, those providing ADM systems would be better informed to take into consideration the possible harms that could be incurred because of their designs. A difficulty however is that when the benefits of compliance with the law are outweighed by the benefits of non-compliance, then legal proceedings will likely favour providers and deployers of ADM that have enough financial resources to sustain (potentially multiple) lawsuits over those that have been negatively impacted by ADM.¹⁴¹ One way of helping address this issue is the use of adverse inferences and shifting the burden of proof to providers and deployers of ADM to show how their systems comply with applicable laws, instead of claimants having to show how these systems have resulted in infringements. At present, heavy burdens are placed on claimants to first be in a position to try and prove they have been negatively impacted by ADM, and then to go about doing so by application of the law. If nothing else, the current asymmetry underscores the significance of collective action.

Nonetheless, impact assessments serve to benefit providers and deployers of ADM, those subjected to its use, and those looking to improve methods of regulation. In this latter respect, there appears to be increasing attention towards the idea of introducing human rights considerations

¹³⁹ J Henley and R Booth, 'Welfare Surveillance System Violates Human Rights, Dutch Court Rules' (*The Guardian*, 5 February 2020) <<https://www.theguardian.com/technology/2020/feb/05/welfare-surveillance-system-violates-human-rights-dutch-court-rules>>

¹⁴⁰ R Allen and D Masters, 'Government Automated-Decision Making' (*AI Law*, 30 March 2020) <<https://ai-lawhub.com/2020/03/30/government-automated-decision-making/>>

¹⁴¹ See generally K Pistor, *The Code of Capital: How the Law Creates Wealth and Inequality* (Princeton University Press 2019).

into impact assessments of ADM.¹⁴² As provided by the EU Digital Services Act,¹⁴³ human rights considerations have the potential to be better accounted for at all stages of the development and deployment of any system that incorporates ADM, including what safeguards should be put in place in order to respect and protect the human rights that are applicable to the particular system.¹⁴⁴ Related measures at the design phase could help weed out ADM systems that are likely to result in human rights infringements upon their use. Impact assessments serve as an important informational resource that helps individuals and those that represent them understand if human rights have been infringed, which can also aid investigations, and inform what remedies are appropriate in the circumstances.

Although time-consuming and potentially resource-intensive (depending on the provider or deployer), impact assessments should not focus exclusively on human rights considerations. Critics of taking a human rights approach to ADM, without taking the necessary practical steps to attain implementation, argue that this body of law is not easily incorporated into practical guidance or benchmarks, because the law remains, on the whole, too abstract and vague to provide meaningful input into governance measures, such as ADM impact assessments.¹⁴⁵ Barrie Sander reflects on the complexity of translating human rights language into rules understood by ADM systems, and the pitfalls of using the vocabulary of human rights law without structural changes enabling their practical effectivity.¹⁴⁶ Even if an impact assessment raises flags for a particular ADM system with respect to its impact on specified human rights, how such concerns can be appropriately addressed may not necessarily be clear to those attempting to amend the system, save by decommissioning it. Lilian Edwards and Michael Veale suggest moving away from an exclusive focus on individual rights, towards algorithmic transparency and the creation of ‘better systems, with less opacity, clearer audit trails, [and] well and holistically trained designers.’¹⁴⁷

Combining these improvements would assist public oversight bodies in their scrutiny of ADM systems. Impact assessments being readily available would mean any complaints lodged regarding a particular ADM system could be dealt with in a systematic manner. Feedback could also then be given to providers and deployers on how to develop their impact assessments, including by dealing with data protection, non-discrimination and human rights considerations. Legislatures across states are in a position to ensure such assessments are mandatory for providers and deployers of ADM.¹⁴⁸ These public bodies also have a duty to engage with both rights-holders and stakeholders so as to ensure adequate public participation in oversight mechanisms. These relationships, informed by the findings of impact assessments, perhaps serve as an avenue to reach a consensus between providers, deployers and data subjects on whether justifications for ADM are adequate or not in the circumstances.

Explainability to data subjects can be collectivized—granting each person more power than if they were to attempt influencing ADM processes individually. Public oversight bodies can be placed in a position to assess how well an operator of ADM has undertaken impact assessments,

¹⁴² See, among others, M Latonero, ‘Governing Artificial Intelligence: Upholding Human Rights and Dignity’ (*Data & Society* 2018); McGregor, Murray and Ng, ‘International Human Rights Law as a Framework for Algorithmic Accountability’ (2019).

¹⁴³ EU Digital Services Act, 2022.

¹⁴⁴ Amnesty International, ‘European Union: Digital Services Act Agreement a ‘Watershed Moment’ for Internet Regulation’ (23 April 2022) <<https://www.amnesty.org/en/latest/news/2022/04/european-union-digital-services-act-agreement-a-watershed-moment-for-internet-regulation/>>

¹⁴⁵ See NA Smuha, ‘Beyond Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea’ (2021) 34 *Philosophy & Technology* 91.

¹⁴⁶ B Sanders, ‘Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation’ (2020) 43 *Fordham International Law Journal* 1005.

¹⁴⁷ L Edwards and M Veale, ‘Slave to the Algorithm? Why a “Right to an Explanation” Is Probably Not the Remedy You Are Looking For’ (2017) 16 *Duke Law & Technology Review* 18, 82.

¹⁴⁸ EL Kernell and CB Veiberg, *Guidance on Human Rights Impact Assessment of Digital Activities* (Danish Institute for Human Rights 2020), 17–21.

and make recommendations for development of the related system being scrutinized, including on the methodology underpinning the assessment of that system.¹⁴⁹ If such bodies do not exist within states, then states can create them, lest they fail to provide suitable avenues for individuals and groups to protect their rights outside of courtrooms. Philip Sales, for example, suggests establishing an algorithm commission in the UK, which would be accountable to Parliament.¹⁵⁰ This approach towards the interactions between providers and deployers of ADM and data subjects as a collective, with oversight bodies acting as an intermediary, dilutes the issue of feasibility. Too much responsibility is currently placed on data subjects to know their rights and how to effectively enforce them under current governance frameworks applicable to ADM. Collectivization helps remedy this situation to an extent where those with the power to represent the interests of data subjects, actually do so without data subjects having to know where to go to exercise rules that can protect them.¹⁵¹ Such efforts can give people adversely affected by ADM the opportunity to challenge the (ir)relevance and (in)accuracy of data about them, and how it is used. Public oversight bodies need to be equipped with both the competence and independence to judge systems of ADM and have their decisions implemented by providers and deployers. Perhaps the further this recommendation centred on transparency is realized, the more public trust can be built in ADM.¹⁵²

Guarantor institutions specifically tasked to oversee the use of ADM in practice would serve to safeguard the rights of data subjects,¹⁵³ whilst providing forums for individual concerns to be recorded and responded to appropriately. The outcomes of these processes could also help embed human rights considerations into ADM design, which could in turn allow for compliance monitoring to be more straightforward, including with respect to setting clear purpose limitations on data use. However, there are issues associated with ADM that extend beyond oversight and impact assessments, calling for the enhancement of related complaints procedures and better prospects for accessing remedies, thus going to the very essence of due process guarantees.

Complaints procedures and access to remedy

When conduct, such as that contained in an impact assessment, is considered to have infringed rules under data protection, non-discrimination and/or human rights law, then people require outlets to lodge complaints and seek remedies. Courts and tribunals may be the first port of call for certain parties. However, the costs associated with accessing judicial bodies can be considerable, and due to factors such as docket backlogs, decisions may not be reached in a manner that is timely enough to be meaningful for those that initiate proceedings.¹⁵⁴ As alluded to above, time and money are obstacles that are too often overlooked by those who consider judiciaries as some sort of bastion for protecting rights. Litigation costs are one of a number of reasons why researchers consider poor people to have no effective rights, whether they are those enshrined in domestic, regional or international law.¹⁵⁵ The elephant in the room with respect to access to judicial remedies is the assumption that everyone has the resources to access legal fora in the

¹⁴⁹ A Tutt, 'An FDA for Algorithms' (2017) 69 *Administrative Law Review* 83.

¹⁵⁰ P Sales, 'Algorithms, Artificial Intelligence and the Law' (2020) 25 *Judicial Review* 46.

¹⁵¹ For example: <https://www.facebookclaim.co.uk/>; see also A Hsu, 'In a Stunning Victory, Amazon Workers on Staten Island Vote for a Union' (NPR, 1 April 2022) <<https://www.npr.org/2022/04/01/1089318684/amazon-labor-union-staten-island-election-bessemer-alabama-warehouse-workers>>

¹⁵² Ada Lovelace Institute, 'The Rule of Trust: Findings from Citizens' Juries on the Good Governance of Data in Pandemics' (28 July 2022) <<https://www.adalovelaceinstitute.org/report/trust-data-governance-pandemics/>>

¹⁵³ See generally T Khaitan, 'Guarantor Institutions' (2021) 16 *Asian Journal of Comparative Law* 40.

¹⁵⁴ Docket backlogs are also a reason why ADM is being proposed in such settings. See L Scialabba and others, *Government Backlog Reduction: Five Ways Government Agencies Can Improve Services and Mission Delivery* (Deloitte 2019).

¹⁵⁵ See, for example, KM Bridges, *The Poverty of Privacy Rights* (Stanford University Press 2017); J Linarelli, ME Salomon and M Sornarajah, *The Misery of International Law: Confrontations with Injustice in the Global Economy* (Oxford University Press 2018); T Pogge, 'Recognized and Violated by International Law: The Human Rights of the Global Poor' (2005) 19 *Leiden Journal of International Law* 717.

first place. Many people, if not the large majority, do not. This issue is especially acute when it comes to gathering and marshalling evidence towards proving a violation stemming from ADM.¹⁵⁶

Thankfully, there are at least two more options in addition to the route of seeking judicial remedies. First, public oversight bodies can take collective action against providers and deployers of ADM on behalf of a class of individuals that have had their rights infringed, or individuals can petition oversight bodies to take action against the provider and/or deployer in question. Both approaches under this option are likely to be time-consuming, especially if the oversight body in question is under-resourced, because they require analysing individual cases on their merits and then deciding whether to take complaints forward against providers and/or deployers of ADM. Nonetheless, the advantage of collective action is that it collectivizes accountability in a way in which censuring can have considerable and lasting effects. If large-scale complaints get traction in the public arena, then the particular ADM provider and/or deployer at issue may be pressured into changing their mode(s) of operation that resulted in the infringements of the law(s) at issue. This pressure is likely to intensify should such a movement have a noticeable impact on the provider or deployer, whether financially or politically.

Loss of financial value or supporters helps move the needle of accountability. For example, Spotify lost ~2.1 billion USD in market value over three days after Neil Young removed his work from the platform in protest to Joe Rogan's podcast which was accused of spreading misinformation.¹⁵⁷ For ADM operators that are governments, the loss of supporters serves to affect change in similar ways as losing financial value does for corporations. Governments want to stay in power. They tend to do whatever the political winds indicate should be done in order to attain more power, such as disbanding a particular ADM system.¹⁵⁸ For example, Joe Biden has proposed a new labour rule in the US that would make it more difficult for companies to treat workers as independent contractors, requiring that they instead be considered employees, entitled to more benefits and legal protections.¹⁵⁹ This change came about in part because of knowledge becoming more widespread about the appalling conditions of workers that sustain ADM systems—without which these systems could not work.¹⁶⁰ If an ADM system creates cost savings, it is partly because people providing the resources to maintain it are likely being exploited. Pay such workers a fair wage and ensure decent working conditions and legal protections, and the costs of many ADM systems will rise. Those that are quick to praise ADM for its apparent cost-reduction potential need to take a breath before doing so and reflect on *why* the particular system in question may be a cost-effective alternative.

But individual claims brought against a provider and/or deployer of ADM may not garner the same levels of public attention and support that a large-scale collective action spearheaded by a public oversight body might. The overarching takeaway is that providing complaints procedures through public machinery aimed at regulating ADM can catalyse change. Oversight and activism can unite when providers or deployers of ADM overstep. While perhaps not immediately recalled when the term remedy is used, such processes and outcomes can provide remedies

¹⁵⁶ With many thanks to Katie Pentney for making this link.

¹⁵⁷ T Spangler, 'Spotify Lost More Than \$2 Billion in Market Value After Neil Young Pulled His Music Over Joe Rogan's Podcast' (*Variety*, 29 January 2022) <<https://variety.com/2022/digital/news/spotify-2-billion-market-cap-neil-young-joe-rogan-1235166798/>>; A Blake, 'The Coronavirus Misinformation on Joe Rogan's Show, Explained' (*Washington Post*, 2 February 2022) <<https://www.washingtonpost.com/politics/2022/02/02/actual-joe-rogan-coronavirus-misinformation/>>

¹⁵⁸ J Niklas, 'Poland: Government to Scrap Controversial Unemployment Scoring System' (*Algorithm Watch*, 16 April 2019) <<https://algorithmwatch.org/en/poland-government-to-scrap-controversial-unemployment-scoring-system/>>

¹⁵⁹ D Wiessner, N Bose and D Shepardson, 'Biden Labor Proposal Shakes Up Gig Economy that Relies on Contractors' (*Reuters*, 12 October 2022) <<https://www.reuters.com/world/us/new-biden-labor-rule-would-make-contractors-into-employees-2022-10-11/>>

¹⁶⁰ A Williams, M Miceli and T Gebru, 'The Exploited Labor Behind Artificial Intelligence' (*Noema Magazine*, 13 October 2022) <<https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/>>

to those that have been adversely affected by ADM. In this sense, the remedy referred to here is that with respect to a particular situation or outcome, so that it does not occur again in the future. These remedies may not be legal in nature, or provide financial compensation to which claimants may be entitled, but they can serve the same role of accounting for something that went wrong in the past, and attempt to right it by learning from past mistakes. They may also do more to affect change in policymaking, which parties bringing complaints may value more highly than the prospect of receiving financial compensation. This last point is particularly pertinent with respect to the comparison with litigation on human rights grounds. As uncovered by research from Luke Moffett, Cheryl Lawther, Sunneva Gilmore and Ebba Lekvall:

most victims of such [human rights] violations do not receive reparations. While states are increasingly being brought before international and regional bodies to fulfil their duties and satisfy victims' rights, they have often struggled or been reluctant to implement their international obligations.¹⁶¹

The second alternative to complaints procedures intended for the pursuit of judicial remedies is complaints procedures being provided by providers and deployers of ADM systems. One example, while perhaps not a good one,¹⁶² is the Oversight Board of Facebook.¹⁶³ This model can inform the creation of others. What is fundamental to providers and deployers of ADM making available complaints procedures and redress mechanisms is creating them as part of ADM systems at the outset, not merely setting them up as a reactive response to accusations of wrongdoing. By having these measures in place, providers and deployers of ADM systems have a means of being transparent and explaining their systems to those affected by them. This provides a meaningful form of accountability, namely that of answerability.¹⁶⁴ In situations where ADM operators 'flout relationship-defining demands', then by already having in place complaints procedures, they are choosing 'to put themselves in a position to be susceptible to being held to account' for any wrongdoing.¹⁶⁵

Developing complaint procedures to accompany ADM systems and managing them properly has the potential to help safeguard the rights enshrined across the laws examined above. Questions of due process and dignity become interlinked here. Should complaints procedures be readily accessible to any subject of ADM, not merely those with subject-specific knowledge of the laws applicable to such systems, there will be opportunities for people to be heard. The power of having another human actively listen to a grievance should not be

¹⁶¹ L Moffett and others, 'The Limits of the Law: Putting Reparations into Practice' (*EJIL: Talk!* 2 July 2019) <<https://www.ejiltalk.org/the-limits-of-the-law-putting-reparations-into-practice/>>

¹⁶² T E Kadri, 'Juridical Discourse for Platforms' (2022) 136 *Harvard Law Review* 163; J Cows, P Darius, D Santistevan and M Schramm, 'Constitutional metaphors: Facebook's "supreme court" and the legitimization of platform governance' (2022) *New Media & Society*.

¹⁶³ Oversight Board: <https://www.oversightboard.com/>; for commentary see E Douek, 'The Facebook Oversight Board's First Decisions: Ambitious, and Perhaps Impractical' (*Lawfare*, 28 January 2021) <<https://www.lawfaremedia.org/article/facebook-oversight-boards-first-decisions-ambitious-and-perhaps-impractical>>; D Morar, 'Trump Deplatforming Decision Highlights the Impotence of Facebook's Oversight Board' (*Brookings Institution*, 7 May 2021) <<https://www.brookings.edu/blog/techtank/2021/05/07/trump-deplatforming-decision-highlights-the-impotence-of-facebooks-oversight-board/>>; P Olson, 'Don't Dismiss Facebook's Oversight Board. It's Making Some Progress' (*The Washington Post*, 25 October 2021) <https://www.washingtonpost.com/business/dont-dismiss-facebooks-oversight-board-its-making-some-progress/2021/10/25/fe4bedd4-3561-11ec-9662-399cfa75efee_story.html>; S Levy, 'Oversight Board to Facebook: We're Not Going to Do Your Dirty Work' (*Wired*, 5 May 2021) <<https://www.wired.com/story/oversight-board-to-facebook-not-going-to-do-your-dirty-work/>>; J Cook, 'MPs Urge Facebook to Reveal Pay of Alan Rusbridger and Other Members of Its "Supreme Court"' (*The Telegraph*, 7 May 2020) <<https://www.telegraph.co.uk/technology/2020/05/07/mps-urge-facebook-reveal-pay-alan-rusbridger-members-supreme/>>

¹⁶⁴ CV Blatz, 'Accountability and Answerability' (1972) 2 *Journal for the Theory of Social Behaviour* 101; AM Smith, 'Responsibility as Answerability' (2015) 58 *Inquiry* 99.

¹⁶⁵ R Mackenzie-Gray Scott, *State Responsibility for Non-State Actors: Past, Present and Prospects for the Future* (Bloomsbury 2022), 134; D Shoemaker, 'Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility' (2011) 121 *Ethics* 602, 623.

understated.¹⁶⁶ Responding appropriately, and, if necessary, taking further steps to address complaints, requires human-to-human interaction at some point. Investing in such processes and not relying solely on ADM in an attempt to deal with complaints swiftly takes seriously the *dignity of being heard*. Humans can provide rational and compassionate ex-ante and ex-post explanations of ADM—an algorithm cannot.¹⁶⁷ Access to a human that can actually help alleviate concerns and right wrongs forms part of the bedrock of any functional system of governance. It is one reason why judges play such a vital role. Yet the judges assessing the compatibility of ADM systems with applicable laws need not be confined to those in the judiciary. It is worth reflecting on a holistic understanding of remedy, one in which being heard is fully appreciated. The process of having a dispute dealt with clearly and fairly can be more important to the interested parties than the ultimate result. Setting up effective complaint procedures to complement ADM systems may help people subject to such systems feel like their concerns are being taken seriously and acted upon. Opening functional lines of communication between potential adversaries can dissipate disputes before they escalate. It is thus in the interests of providers, deployers and subjects of ADM to have complaints procedures engrained within the fabric of these systems. The trickier question is whether providers and deployers of ADM should supply them or official oversight bodies. The benefits of the latter may well concern independence and impartiality, and the benefits of the former may well concern costs and capacity, especially when the providers and deployers are private companies with healthy revenue streams.

ECONOMIC INCENTIVES: ALTERING A ROOT CAUSE OF PROBLEMS CONCERNING ADM

The final matter in need of mention is the economic incentives that apply to the use or not of ADM. It is helpful to signpost some problems on this topic, in particular one assumption that requires challenging, because it concerns the application of the rules across the areas of law examined above. The assumption in question is that from an economic standpoint, regulation hinders innovation. What this means in the case of ADM is that if further procedural requirements are placed on providers and deployers of these systems, for the purpose of ensuring the effectual operation of substantive rules across data protection, non-discrimination and human rights law, these actors will be prevented from making the most of ADM, including in ways that may benefit societies. The basis for this assumption may not lie on strong foundations. While grounded in a pillar supporting theories on neoliberal governance, namely that state regulation should be avoided,¹⁶⁸ the position that regulation of ADM is bad for innovation oversimplifies the interdependence between the two variables into an inverse correlation.

Care is needed here. Questions of regulation and innovation are not all-or-nothing choices. They are a question of quality. Hammering home this point to policymakers fanatical about the innovation potential of ADM is particularly important right now when governments trying to get a handle on their economies are looking for a leg up in challenging times by trying to attract business.¹⁶⁹ Further procedural requirements being placed on providers and deployers of ADM might therefore be assumed to exist as a barrier to, or at least a check on, innovation, which in turn may drive businesses away to more favourable states where there are no (or at least less)

¹⁶⁶ S Turkle, *Reclaiming Conversation: The Power of Talk in a Digital Age* (Penguin 2016).

¹⁶⁷ S Wachter and others, 'No Right to Explanation of Automated Decision-Making' (2017), 91.

¹⁶⁸ K Vallier, 'Neoliberalism' in EN Zalta and U Nodelman (eds), *The Stanford Encyclopedia of Philosophy* (2021) <<https://plato.stanford.edu/entries/neoliberalism/>>

¹⁶⁹ For example, see UK Government, *National AI Strategy—AI Action Plan* (18 July 2022) <<https://www.gov.uk/government/publications/national-ai-strategy-ai-action-plan/national-ai-strategy-ai-action-plan>>

regulations, making states with more regulations on ADM less competitive in a global market of actors seeking growth. This understanding presents a superficial, albeit logically coherent, picture that there exists an economic incentive for states not to further regulate ADM in order to foster investment and the innovation that is promised to accompany it.¹⁷⁰

However, as Rory Gillis argues, unregulated (or poorly regulated) ADM can ‘exacerbate larger negative trends that harm business’.¹⁷¹ Moreover, innovation is measured on a variety of metrics concerning, for example, productivity, or patenting.¹⁷² Innovations can be high-value or low-value to providers and deployers of ADM (in terms of their contribution to economic growth). Philippe Aghion, Antonin Bergeaud and John Van Reenen find that ‘regulation discourages low-value innovation, but has no impact on high-value innovation’.¹⁷³ They go on to show that ‘the chilling effect of regulation on innovation is not an issue for high-value patents and is instead confined to lower-value patents’.¹⁷⁴ This research provides a springboard to suggest that so long as regulations on ADM only discourage, if any, low-value innovations of providers and deployers, then those that are high in economic value can still be produced, and may even be encouraged depending on how well related regulations have been designed. Although there are compliance costs for ADM providers and deployers that come with not being completely free to conduct themselves however they please, well-made regulations can generate additional incentives for other innovative activities, including those that could be of high value, whether in the short or long term.

Knud Blind makes an observation that is also pertinent to enhancing the procedural machinery accompanying the substantive law applicable to ADM:

If liability rules are too strict, innovators do not introduce new products and services in the market – especially radical innovations, because the risks are high, the expected revenues decrease, and the users of the products reduce their self-protection efforts, leading to more accidents.¹⁷⁵

There are at least three factors regarding the development of procedural frameworks applicable to ADM that are noteworthy in light of this point. First, by putting in place competent oversight bodies, states can assist ADM providers and deployers to navigate the substantive rules applicable to their systems, helping them to undertake conduct that is profit-maximizing *and* legally compliant. By focusing on the former before or instead of the latter, providers and deployers risk incurring costs further into the future, meaning from a financial standpoint effective oversight mechanisms are in the interests of providers and deployers. Second, should ADM providers and deployers conduct proper impact assessments of their ADM systems that account for data protection, non-discrimination and human rights considerations, then the results of these will further guide compliance questions, where amendments can be made to ADM systems so as to reduce the likelihood of non-compliance. Third, whether provided by an oversight body and/or

¹⁷⁰ K Jones, ‘AI governance and human rights: Resetting the relationship’ (Chatham House, January 2023) <<https://www.chathamhouse.org/sites/default/files/2023-01/2023-01-10-AI-governance-human-rights-jones.pdf>> For some history behind why economic considerations have become predominant in matters of policy see RH Thaler, *Misbehaving: The Making of Behavioral Economics* (Norton 2016).

¹⁷¹ R Gillis, ‘Pro-Innovation and Pro-Ethics? Threading the Needle in UK AI Policy’ (*The Oxford University Politics Blog*, 6 December 2022) <<https://blog.politics.ox.ac.uk/pro-innovation-and-pro-ethics-threading-the-needle-in-uk-ai-policy/>>

¹⁷² See K Blind, ‘The Impact of Standardization and Standards on Innovation’, Nesta Working Paper 13/15 (November 2013).

¹⁷³ P Aghion, A Bergeaud and JV Reenen, ‘The Impact of Regulation on Innovation’, *London School of Economics, Centre for Economic Performance*, Discussion Paper No. 1744 (January 2021), 3; see also R Griffith and G Macartney, ‘Employment Protection Legislation, Multinational Firms, and Innovation’ (2014) 96 *Review of Economics and Statistics* 135.

¹⁷⁴ Aghion, Bergeaud and Reenen, ‘The Impact of Regulation on Innovation’ (2021), 34.

¹⁷⁵ K Blind, ‘The Influence of Regulations on Innovation: A Quantitative Assessment for OECD Countries’ (2012) 41 *Research Policy* 391, 394.

the provider and deployer in question, should adequate complaints procedures be made readily available for those subject to ADM systems, so that they can voice any concerns that arise, people satisfied with the processes and outcomes of such apparatus will likely be less inclined to initiate legal proceedings. Combined, these three factors have the potential to reduce liability risk for ADM providers and deployers. Standardization can therefore reduce the costs associated with ADM systems failing to function in accordance with data protection, non-discrimination and human rights law.

Innovation can be reconciled with regulation in the case of ADM, including when dealing with difficult concepts such as explainability and transparency, as has been argued by Sophie Stalla-Bourdillon and Alison Knight.¹⁷⁶ The more interesting question is whether further regulation of ADM systems *decreases* provider and/or deployer diversity in a market, creating a condition that helps monopolies to emerge (or remain). Consider for instance that the GDPR, so far at least, seems to have had a negative impact on small businesses, by placing a burden on them that demands changes be implemented in order to fulfil compliance requirements, which is significantly heavier than the burden placed on large firms with the capacity and resources to easily adapt.¹⁷⁷ Whether intentionally or not, creating conditions for a reduction in competition can come about from lopsided regulatory impacts. If regulations create excessive burdens on some businesses but not others, in severe instances to the extent that potential competitors are pushed out of a market, then those regulations may benefit the very actors whose power they should be limiting. It is understandable that there have been calls for modifications to how the GDPR applies, ‘taking into account that the regulation has put smaller companies at a disadvantage’, especially considering that in contrast ‘there is no evidence that large technology companies, such as Facebook and Google, experienced any reductions in either sales or profits.’¹⁷⁸ Interestingly, at the time of writing, in a nod towards favouring regulatory flexibility that promotes market competition, in discussions on the implementation of their data protection framework inspired by the GDPR, the Brazilian national data protection authority is currently debating whether entity size should be part of determining whether and to what extent certain provisions of the law apply,¹⁷⁹ and waived the requirement of having a data protection officer for small businesses and start-ups.¹⁸⁰

This tidbit aside, the general takeaway is that by creating positive economic incentives for providers and deployers of ADM systems through better regulation that enhances procedural capacities, a root cause of problems stemming from ADM could be extinguished, in that innovation and compliance with substantive rules are not at odds, but become complementary. While this approach will require budget increases across states and providers and deployers of ADM in order to be realized, such initial costs are worth viewing as an investment in the potential for long-term benefits to both subjects and providers and deployers of ADM, ensuring such systems can be utilized, whilst at least mitigating (if not eliminating) the issues that come with their use. There need not be further damage done from nearsighted and reactive approaches to policymaking.

¹⁷⁶ See S Stalla-Bourdillon and A Knight, ‘Data Analytics and the GDPR: Friends or Foes? A Call for a Dynamic Approach to Data Protection Law’ in R Leenes, R van Brakel, S Gutwirth and P De Hert (eds), *Data Protection and Privacy: The Internet of Bodies* (Bloomsbury 2018), 249.

¹⁷⁷ See CB Frey and G Presidente, ‘The GDPR Effect: How Data Privacy Regulation Shaped Firm Performance Globally’ (CEPR, 10 March 2022) <<https://cepr.org/voxeu/columns/gdpr-effect-how-data-privacy-regulation-shaped-firm-performance-globally>>; J Jia, G Jin and L Wagman, ‘The Short-run Effects of GDPR on Technology Venture Investment’ (CEPR, 7 January 2019) <<https://cepr.org/voxeu/columns/short-run-effects-gdpr-technology-venture-investment>>; A Acquisti, C Taylor and L Wagman, ‘The Economics of Privacy’ (2016) 54 *Journal of Economic Literature* 442.

¹⁷⁸ Frey and Presidente, ‘The GDPR Effect’ (2022). See also Irish Council for Civil Liberties, ‘5 years: GDPR’s crisis point’ (2023) <<https://www.iccl.ie/digital-data/iccl-2023-gdpr-report/>>

¹⁷⁹ Other criteria are also being discussed, such as volume of data processing operations.

¹⁸⁰ Diário Oficial da União, Resolução CD/ANPD Nº 2, de 27 de Janeiro de 2022 <<https://www.in.gov.br/en/web/dou/-/resolucao-cd/anpd-n-2-de-27-de-janeiro-de-2022-376562019#wrapper>>

While challenging, the above approach is intended to work within neoliberal governance structures, and, for better or worse, does not attempt to change them.¹⁸¹ Although Margaret Thatcher was well off-base with her famous TINA slogan of ‘there is no alternative’, here is not the place to set forth an agenda aiming to overhaul (broken) governance systems rooted in neoliberalism, from which all the noise celebrating ADM has grown. Instead, it is recalled that:

We are not locked in an AI race that can only be won by those countries with the least regulation and the most investment. Instead, the real advances in AI pose both old and new challenges that can only be tamed if we see AI for what it is. Namely, a powerful technology that at present is produced by only a handful of companies with workforces that are not representative of those who are disproportionately affected by its risks and harms.¹⁸²

CONCLUSION

It is understandable that some advocate for a new human right not to be subject to ADM. The fanfare surrounding this technology is loud. But too much is being made of its supposed benefits, and more needs to be learnt about its actual benefits when factoring in the interests of states, providers and deployers of ADM, and groups and individuals subject to its use—including the respective costs to each of these actors. Those favouring the use of ADM technology also appear to frequently conflate *can* with *should*, in that because the capabilities exist to do something, so it should be done. Fewer notions could be further from the truth. Perhaps this conflation is being imposed on providers and deployers of ADM systems by the prevalent governance approach that demands wealth extraction. But why this technology is considered to serve the common good needs further scrutiny from a range of perspectives. Proposing a new human right helps attract attention to these issues, as well as push back against the normalization and, potentially, the legitimization of ADM. However, it is hard to grasp what this right would add in terms of substance to the legal frameworks already applicable to ADM, save perhaps with respect to providing an intersectional, general rule that encompasses elements from different, specific rules. In terms of legal practice being used as means to protect groups and individuals, what is currently imperative is developing comprehensive and robust procedural capacities for the effective enforcement and implementation of existing legal frameworks. Although concerning human rights, it is not the role of human rights *law* to fix these current problems regarding accountability, management and oversight of ADM. Expecting too much from this body of law, and trying to expand its principles wherever possible, is a reason why the overall subject takes a bashing for being purportedly ineffective, when it is not. In addressing issues associated with ADM, there will also be tensions between bodies of law that may not be capable of being relieved. A noteworthy example that speaks to the issue of explainability is that those providing and deploying ADM systems have a right to exploit their intellectual property for economic gain. Proprietary code used in the design of an algorithm is protected from disclosure. This example illustrates how the law can also hinder the regulation of ADM. Yet an equally if not more powerful stimulus than the law for guiding the use of ADM is economic incentives. If these are of the positive variety and can be nourished by regulations that concretize data protection, non-discrimination and human rights law, perhaps the future will involve less friction between providers and deployers of ADM, and those subject to its use.

¹⁸¹ For a variety of reasons why neoliberal governance requires changing, examined from the perspective of ADM, see Waldman, ‘Power, Process, and Automated Decision-Making’ (2019), 624–27.

¹⁸² Kaltheuner, *Fake AI* (2021), 15.