An approach for a fundamental rights impact assessment to automated decision-making

Heleen L. Janssen*

Key Points

- Companies expect great and promising benefits from automated decision-making with personal data; however, scientific research indicates that legal uncertainty exists among private controllers with the interpretation of provisions relevant to automated decision-making under the General Data Protection Regulation (GDPR).
- Article 35 GDPR obliges private controllers to execute a Data Protection Impact Assessment (DPIA) prior to deploying automated decisions on humans. Assessing potential fundamental rights impacts is part of that DPIA.
- The objective of this article is to provide private controllers with a practical approach for a DPIA to automated decision-making to detect potential impacts on fundamental rights. The approach indicates levels of impacts and types of measures a controller should consider to achieve an appropriate risk management.
- The impact assessment is based on four benchmarks: (i) to identify fundamental rights potentially at risk; (ii) to identify risks occurring in their ADM systems at design stages and during operation; (iii) to balance fundamental rights risks and controller interests involved; and (iv) to establish to what extent data subjects exercise control over data processing.
- By responding to the benchmarks, controllers identify risk levels that indicate the type of measures that should be considered to achieve fundamental rights compliant ADM.
- This approach enables controllers to give account towards data subjects and supervisory authorities about envisaged risk management to potential impacts on fundamental rights. The proposed approach seeks to foster compliant, fair, and transparent automated decision-making.

Introduction

Automated decision-making (ADM) is already used across a variety of societal contexts,[1] from simplistic models that help online service providers to carry out operations on behalf of their users, for instance for billing purposes, or to create a better functioning social network[2]—to more complex profiling algorithms that filter systems for targeted advertisements, credit scoring, recommender systems, IoT-applications, insurance proposals, health care applications, or examination to enter education or training.[3]

Meanwhile, dynamics of ADM may collide with and exert pressure on fundamental rights.[4] For instance, Kramer and others described how Facebook

[*] Department of Computer Science and Technology, Compliant and Accountable Systems Research Group, University of Cambridge, Cambridge, UK. E-mail: heleen.janssen@cl.cam.ac.uk. The author acknowledges the support of Engineering and Physical Sciences Research Council, UK (Grant number EP/P024394/01). The author wishes to thank the editor of the journal, the reviewers, as well as John Morijn, Janneke Gerards and Jatinder Singh for their useful comments on this article.

[1] Information Commission Officer, 'What is Automated Individual Decision-making and Profiling' (2018) <https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling-1-1.pdf> accessed 18 February 2020.

[2] Hojung Kim, Joseph Giacomin and Robert Macredie, 'A Qualitative Study of Stakeholders' Perspectives on the Social Network Service Environment' (2012) 30 International Journal of Human-Computer Interaction 965.

[3] Mireille Hildebrandt and Serge Gutwirth, 'General Introduction and Overview' in Mireille Hildebrandt and Serge Gutwirth (eds), *Profiling the European Citizen* (Springer, Dordrecht 2008) 1.

[4] For a comprehensive overview of ADM impacts on fundamental rights, see eg Rinie van Est, Joost Gerritsen, Linda Kool, 'Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality', Expert report written for the Committee on Culture, Science, Education and Media of the

experimented with its algorithm that organizes news feeds of users to test how different fine-tunings in the application of the algorithm may impact behaviour and emotions of users.[5] Such use of ADM (determining on the basis of personal data that was targeted with emotionally charged messages) may impact one's privacy or freedom of thought and conscience.[6] Searches in Google's search engine with ADM for African-American names were more likely to show advertisements suggesting that the person had an arrest record.[7] Such use of personal data in ADM can lead to a discriminatory application of ADM. On group or societal scale, such applications of ADM may result in social stratification.[8]

These examples demonstrate the need to combat issues caused by ADM to fundamental rights. The application of ADM based on personal data comes with a number of risks which, if not properly addressed, can outweigh the benefits, and potentially erode public support for ADM-driven solutions.[9] O'Neill observed that there are countless tales of companies putting profit before individuals, including banks systematically misselling insurance policies in the knowledge that consumers will never be entitled to make claims on them, mortgage providers lending credits to people they know will almost certainly end up in foreclosure[10], or 'for-profit' education institutions targeting the more easily cheated underclasses in the USA.[11]

ADM may impact fundamental rights

Fundamental rights

To elaborate on what is meant by 'fundamental rights' in the context of this article, some demarcation is necessary. Fundamental rights—as rights of the highest rank in a legal system—are commonly referred to as 'fundamental' rights or 'human rights'. Under the law, fundamental rights commonly entail subjective rights which are accepted in a specific legal system, and approved by that legal system's statutes. 'Fundamental rights' in the context of this article refer to the rights enshrined in the European Union Charter of Fundamental Rights and Freedoms ('EU Charter') and in the European Convention of Human Rights and Fundamental Freedoms ('ECHR').[12]

Fundamental rights and human rights diverge at the point where fundamental rights are specific to and can be invoked in a particular legal system by those to whom the law applies (eg citizen rights such as voting rights of the EU Charter can only be invoked by EU

Parliamentary Assembly of the Council of Europe (PACE), The Hague: Rathenau Instituut 2017, <https://www.rathenau.nl/sites/default/files/2018-02/Human%20Rights%20in%20the%20Robot%20Age-Rathenau%20Instituut-2017.pdf> accessed 12 January 2020; Rob van den Hoven van Genderen, 'Privacy and Data Protection in the Age of Pervasive Technologies (2017) 3 AI and Robotics European Data Protection Law 338; Council of Europe Committee of Experts on Internet Intermediaries (MSI-NET), 'Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications' (2018) 45 <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5> accessed 12 January 2020; UN Human Rights Council, 'Report of the Office of the UN High Commissioner for Human Rights on the Right to Privacy in the Digital Age' (3 August 2018) UN Doc A/HRC/39/29, paras 1, 15; UN Human Rights Council, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression on A Human Rights Approach to Platform Content Regulation' (6 April 2018) UN Doc A/HRC/38/35; Janneke Gerards, Remco Nehmelman and Max Vetzo, *Algoritmes en Grondrechten* (Boom Juridische Uitgevers 2018, The Hague, in Dutch); Alessandro Mantelero, 'AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment' (2018) 34 Computer Law & Security Review 754; Filippo Raso and others, 'Artificial Intelligence & Human rights: Opportunities and Risks' (2018) Research Publication No 2018-6, 25 September 2018 <https://ssrn.com/abstract=3259344> accessed 11 January 2020; Lorna McGregor, Daragh Murray and Vivian Ng, 'International Human Rights Law as a Framework for Algorithmic Accountability' (2019) 68 International and Comparative Law Quarterly 309.

[5] Example borrowed from Adam Kramer, Jamie Guillory and Jeffrey Hancock, 'Experimental Evidence of Massive-scale Emotional Contagion Through Social Networks' (2014) 111 Proceedings of the National Academy of Sciences of the USA 8788.

[6] See respectively, arts 9 (1) and 8 (1) of the European Convention of Human Rights and Fundamental Freedoms (European Convention of Human Rights and Fundamental Freedoms of 3 September 1953, ETS 5 213 UNTS 221; 'ECHR'). The European Court of Human Rights (ECtHR) has repeatedly recognized 'mental health', or 'moral integrity' as part of the private sphere as enshrined in art 8 ECHR, see *X and Y v the Netherlands* App no 8978/80 (ECtHR, 26 March 1985) 8 EHRR 235, para 22; *Bensaid v United Kingdom* App no 44599/98 (ECtHR, 6 February 2001) 33 EHRR 10, para 47. The EU Charter of Fundamental Rights encompasses comparable rights in arts 10 and 8.

[7] Example borrowed from Solon Barocas and Andrew Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review 671, 682; see as regards the use of ADM to predict future criminals: Julia Angwin and others, 'Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks' (Pro Publica, New York) <https:www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 30 June 2019.

[8] Dennis Broeders, Erik Schrijvers and Ernst Hirsch Ballin, 'Big Data and Security Policies: Serving Security, Protecting Freedom' (2016) 6 The Netherlands Scientific Council for Government Policy (Policy Brief) 15.

[9] Cathy O'Neill, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Penguin Random House, London 2016) 4.

[10] Example borrowed from Steven Finlay, *Predictive Analytics, Data Mining and Big Data: Myths, Misconceptions and Methods* (Palgrave MacMillan, London, 2014) 101.

[11] O'Neill (n 9) 68, 71, where she explains that US federal state loans to students used to be the 'life-blood' of 'for-profit education factories', and that 'vulnerability [of students] is considered worth gold' since recruiters in such companies were directed to target 'welfare mom with kids, pregnant ladies, recent divorce, low self-esteem, low income jobs, experienced a recent death, physically/mentally abused, recent incarceration, drug rehabilitation, no future'.

[12] Charter of Fundamental Rights of the European Union [2000] OJ C 364/01; the Charter became legally binding when the Treaty of Lisbon entered into force on 1 December 2009; as regards the ECHR, see n 6. 'Fundamental rights' are often also enshrined in a national legal system's constitution.

citizens; the right to privacy can be invoked by any person residing on EU territory). By contrast, human rights have worldwide acceptance and belong to *all* human beings—irrespective of, for instance, their nationality, race, gender, or birth (eg the UN Charter of Human Rights), or of where they reside.[13] The term 'human rights' has its origins in international law.[14] While fundamental rights and human rights show large overlaps in substance (they both protect rights to not to be discriminated against, freedom of expression, access to an independent or impartial court, privacy, and so forth), the terms are of a different origin, and more importantly, raise different bifurcations as to their applicability. This article seeks to develop a practical approach to assess potential impacts on those rights due to a private organization's use of ADM. This approach has its foundations in the EU Charter of Fundamental Rights and in the EU General Data Protection Regulation (GDPR[15]). EU Member States are obliged to comply with the EU Charter whenever they implement EU law.[16] Whenever Member States do not implement EU law, fundamental rights are still covered by the legally binding ECHR, to which all EU Member States are a party, and by their national constitutions.[17]

Automated decision-making

Whenever a decision is reached solely by technological means, without human intervention, it qualifies under GDPR as a decision based solely on automated processing or 'ADM'.[18] ADM is constructed to process outputs automatically resulting in actions without direct human involvement. As many ADM systems contain some form of opacity in their data processing, its harm to individuals involved can unobtrusively be exacerbated.[19] Inappropriate or harmful outcomes may remain unnoticed, and potentially lead to biased outcomes which can result in discrimination. If these impacts remain unnoticed, they can effectively reduce one's right to obtain an effective remedy.[20] Exercising human control over ADM can become difficult, or even impossible. While ADM offers great opportunity for efficiency, and useful applications in many areas, its opacity can result in reduction of public and consumer trust.[21]

Private entities' insecurities over the application of the GDPR

The GDPR offers important mechanisms to protect the enjoyment of one's fundamental rights against intrusive ADM. Though data protection in the EU has been in place since 1995, including protection where ADM is involved, compliance with the GDPR has been perceived by many companies as significantly challenging. While many of the rights and principles of the GDPR are largely similar to those of the 1995 Data Protection Directive,[22] increased maximum fines, renewed media

[13] McGregor, Murray and Ng (n 4) 312.

[14] See United Nations Universal Declaration of Human Rights of 1948; UN General Assembly (Paris 1948). *Universal Declaration of Human Rights* (217 [III] A); or the International Covenant of Civil and Political Rights of 1966 (ICCPR); see International Covenant on Civil and Political Rights, adopted and opened for signature, ratification and accession by General Assembly resolution 2200A (XXI) of 16 December 1966 (entry into force 23 March 1976). Under international human rights law, states are required to put in place a framework that prevents human rights violations from taking place, that establishes monitoring and oversight mechanisms as safeguards, that holds those responsible to account, and that provides a remedy to individuals and groups who claim that their rights have been violated, see UN Human Rights Committee, 'General Comment No 31, The Nature of the Legal Obligation Imposed on States Parties to the Covenant' (26 May 2004) UN Doc CPR/C/21/Rev.1/Add.13, paras 3–8.

[15] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46 (General Data Protection Regulation) [2016] OJ L119/1.

[16] The EU Charter also applies to EU institutions in their relations with EU citizens (eg as regards voting rights) and those residing on its territory (eg freedom of expression, right to privacy, or right to data protection). The EU Charter does not apply to legal areas over which it has no competence. National security has explicitly been excluded from EU competence (art 4 (2) Treaty on the European Union).

[17] The next section on the legal framework elaborates how the EU Charter and the ECHR relate to each other.

[18] Art 22 (1) GDPR; Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (WP 251/rev01, 6 February 2018) 8; Michèle Finck, 'Smart Contracts as a Form of Solely Automated Processing under the GDPR' (2019) 2 International Data Privacy Law 78; in popular discussions these are often also termed 'AI' and may also be discussed by reference to 'algorithmic decision-making', see Jennifer Cobbe, 'Administrative Law and the Machines of Government: Judicial Review of Automated Public-Sector Decision-Making' (20 June 2019) 3 <https://ssrn.com/abstract=3226913> accessed 22 November 2019.

[19] Jenna Burrell, 'How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms' (2016) 3 Big Data & Society 1; Jatinder Singh and others, 'Responsibility & Machine Learning: Part of a Process' (28 October 2016) <https://ssrn.com/abstract=2860048> accessed 11 January 2020; Brent Mittelstadt and others, 'The Ethics of Algorithms: Mapping the Debate' (2016) 3 Big Data & Society 2, 6; Joshua Kroll and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633; but see Joshua Kroll, The Fallacy of Inscrutability, Philosophical Transactions A (15 October 2018), found at <http://dx.doi.org/10.1098/rsta.2018.0084> accessed 11 January 2020.

[20] Andrej Savin, 'Profiling in the Present and New EU Data Protection Frameworks' in Peter Nielsen, Peter Schmidt and Katja Dyppel Weber (eds), *Erhvervsretlige emne, Juridisk Institut CBS* (Copenhagen, Djøf Forlag 2015) 251.

[21] Recommendation CM/Rec(2010)13 adopted by the Committee of Ministers of the Council of Europe on 23 November 2010 and explanatory memorandum to Member States on the protection of individuals with regard to automatic processing of personal data in the context of profiling (Council of Europe Publishing) 28; see Damian Clifford, 'The Legal Limits of the Monetisation of Online Emotions' (DPhil thesis, University of Leuven 2019) 113.

[22] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ

attention, and high-profile data incidents have made data protection compliance a high-priority concern for private organizations. In the run-up and immediate aftermath of its introduction, the GDPR was the subject of a number of broad, large-scale quantitative surveys exploring readiness and attitudes to the GDPR.[23] At the time, most respondents did not expect to be compliant by the enforcement date.[24] A recent qualitative study demonstrated that insecurities in tech start-up companies exist with regard to interpretation and the framework of the GDPR more generally.[25] Their insecurities relate, among others, to the fact that the GDPR contains it is largely 'principles-based', which can be defined as high-level, broadly stated rules or principles to set standards by which regulated firms must conduct business.[26] In their eyes, these principles in the GDPR leave much room for interpretation, while in the event of non-compliance, high fines can be expected. This causes legal uncertainty over how legal compliance can be achieved.

This article seeks to build an approach for impact assessments to fundamental rights based on Article 35 GDPR. This right enshrines an organization's obligation to carry out a Data Protection Impact Assessment (DPIA). Article 22 GDPR in principle prohibits ADM, while laying down strict conditions under which ADM may be used. Both Articles entail open norms potentially creating legal uncertainty among business enterprises. The earlier mentioned qualitative study has shown that this leads to concerns among companies about the risk of unintentionally missing GDPR-compliance responsibilities.[27] Because of this legal uncertainty, some interviewees in that study suggested that many companies might simply 'wait and see' how the GDPR will be enforced, doing the minimum required—or even ignoring it altogether.[28] To prevent a situation in which open legal requirements such as Article 35 GDPR lead to legal uncertainty or even to non-application, this article explores an approach that seeks to offer a practical tool to companies to help them assess potential risks to fundamental rights in ADM.

Approach

The objective of this article is to help companies with a *practical* approach to a fundamental rights impact assessment at stages where they consider and explore the use of ADM. To achieve an early-stage analysis, such assessment should take place at the design stage of an intended ADM. Early-stage impact assessments are useful as the ADM can still be changed and tried out again until appropriate risk levels are achieved. The impact assessment may, once a particular ADM has started to deploy, serve as an appropriate standard to monitor and appreciate the ADM's functioning, and where necessary, lead to readjustment of the ADM. The approach to a fundamental rights impact assessment as proposed in this article is, as will be argued, part of the GDPR's DPIA.[29]

While few approaches to human rights impact assessments and ethical impact assessments and/or social impact assessments already exist, practical approaches are rare.[30] The Danish Institute for Human Rights has created a Human Rights Compliance Assessment tool[31] which can be voluntarily applied by private entities, of which only its 'Quick Check' is publicly available.[32] The

L281/31 (Data Protection Directive, hereafter: 'DPD'). See for automated decisions as regards art 15 of the DPD: Isak Mendoza and Lee Bygrave, 'The Right Not to be Subject to Automated Decisions Based on Profiling' (2017) University of Oslo, Legal Studies, Research Paper Series No. 2017-20, 3. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2964855> accessed 11 January 2020.

[23] IBM Institute for Business Value, 'The End of the Beginning: Unleashing the Transformational Power of GDPR' (2018) <https://www-01.ibm.com/common/ssi/cgibin/ssialias?htmlfid=86015886USEN> accessed 11 January 2020; Deloitte, 'The Time is Now. The Deloitte General Data Protection Regulation Benchmarking Survey' (2018) <https://www2.deloitte.com/content/dam/Deloitte/be/Documents/risk/emea-gdpr-benchmarkingsurvey. pdf> accessed 11 January 2020; SAS, 'Most Aren't Ready, Are You? Answers from a 2018 Survey of 183 Global, Cross-industry Business People Involved in Preparing for GDPR' (2018). <https://www.sas.com/content/dam/SAS/en_us/doc/infographic/gdpr-109649.pdf> accessed 11 January 2020; ISACA, 'The End of The Beginning' (2018) <http://www.isaca.org/Knowledge-Center/Documents/2018-GDPR-Readiness-Survey-Report.pdf> accessed 11 January 2020; Net App Survey, 'Gauging Global Awareness of Business Concerns' (2018) <https://www.netapp.com/us/media/netapp-gdpr-survey-findings.pdf> accessed 31 December 2018.

[24] Of every report listed, all reported a majority of respondents who did not expect to be GDPR compliant by 25 May 2018. These ranged from 54% to 85% of respondents.

[25] Chris Norval, Heleen L. Janssen, Jennifer Cobbe, Jatinder Singh and others, 'Data Protection and the Tech Start-ups: the Need for Attention, Support and Scrutiny' (6 June 2019) 4 SSRN <http://ssrn.com/abstract=3398204> accessed 11 January 2020.

[26] Julia Black, Martyn Hopper and Christa Band, 'Making a Success of Principles-Based Regulation' (2007) 1 Law and Financial Markets Review 191; ibid 3.

[27] Norval and others (n 25) 11.

[28] ibid 2.

[29] Arts 25, 35 GDPR.

[30] See eg Paul de Hert, 'A Human Rights Perspective on Privacy and Data Protection Impact Assessments' in Paul de Hert and David Wright (eds), *Privacy Impact Assessment* (Springer, Dordrecht, 2012) 33; Dariusz Klosza and others, 'Data Protection Impact Assessments in the European Union: Complementing the New Legal Framework Towards a More Robust Protection of Individuals' (1/2017) d.pia.lab Policy Brief 1-4; Mantelero (n 4) 754.

[31] See for a toolbox for companies, Nora Götzmann and others, 'Human Rights Impact Assessment Guidance and Toolbox' (2016) <https://www.humanrights.dk/sites/humanrights.dk/files/media/dokumenter/business/hria_toolbox/hria_guidance_and_toolbox_final_may22016.pdf_223795_1_1.pdf> accessed 11 January 2020.

[32] 'Quick check' found at <https://hrca2.humanrightsbusiness.org/docs/file/HRCA%20Quick%20Check_English.pdf> accessed 28 June 2019. Companies sign agreements vowing not to publish or share any element

Canadian Government recently developed an Algorithmic Impact Assessment tool (AIA).[33] AINOW has developed a practical framework for public agency accountability towards the public for the use of algorithmic decision-making.[34] These frameworks offer useful recommendations and guidance to public organizations, and call those to engage the public in their impact assessments. An important difference with our approach is that they cover public sector use, whereas the impact assessment in our approach focuses on the private sector. Assessment of fundamental rights impacts is not their specific focus. Others have issued or proposed recommendations for future reviews to human rights impact assessments,[35] while again others have developed ethical impact assessments for ethical boards, either in-company, or as an independent actor.[36] Yet, their proposals do not clarify how private organizations would be guided at more practical and concrete levels.

Fundamental rights have originally been a *public* legal affair. Yet, over the last two decades, the dichotomy between public and private application of fundamental rights in EU law has become somewhat blurred due to developments of ordinary EU legislation.[37] This can, for instance, be seen in the areas of equal treatment legislation or in data protection law.[38] These laws explicitly address private entities. This evolution in EU legislation has forced some Member States to reconsider their traditional view that fundamental rights should be binding and enforceable only against state authorities, and not against private bodies and individuals.[39]

For companies to detect potential fundamental rights impacts due to their use of ADM, four benchmarks are proposed in order to help them assess risks. Private entities should to that end (i) identify fundamental rights potentially at risk; (ii) identify risks occurring in their ADM systems at design stages and during operation; (iii) balance fundamental rights risks and interests involved; and (iv) establish to what extent data subjects exercise control over the processing of their data. Their responses to the benchmarks will result in an overall impact assessment. This overall assessment should result in risk factors (high, medium, or low risk). The risk factors should subsequently lead to finding an appropriate level of protection, by choosing and applying mitigating measures that correspond to the level of impact emerging from ADM. Private organizations should, when determining appropriate mitigating measures, avoid over- and under-inclusiveness of impacts on either the fundamental rights at stake, or on their own legitimate interests (such as computational or commercial interests).

Note that the benchmarks are presented as a starting point. In practice, more detailed questionnaires, guidance materials, standards, and best practices might work to both further develop the assessment processes, and perhaps the evolution of new benchmarks. This could be the topic of future research. As regards methodology, the approach is based on interpretation of the law, case law, relevant academic literature, and policy documents.

The Section 'Machine learning: concerns and legal framework' sets the scene for the technology. It introduces how ADMs commonly work, what the role of machine learning (ML) and algorithms in ADM is, and how they can adversely impact fundamental rights and pose challenges to the transparency of data processing. This Section outlines the relevant provisions surrounding the DPIAs (Article 35 GDPR), the legal framework for fundamental rights in the EU Charter of Fundamental Rights, and that of ADMs (Article 22 GDPR) that help shape the proposed approach.

The Section 'Benchmarks to assess fundamental rights impacts' explains and elaborates the four benchmarks which should assist private entities to achieve an appropriate fundamental rights impact assessment. Each benchmark contains examples, and a 'sliding scale' nature of the benchmarks and the risks will be presented in figures. The Section 'Finding appropriate levels of protection' merges the responses to the benchmarks into an

of the HRCA (Human Rights Compliance Assessment) reports they create. The tool incorporates the Universal Declaration of Human Rights and includes a database of 195 questions and 947 indicators.

[33] Government of Canada, 'Algorithmic Impact Assessment'; it applies to public entities and has taken effect as of 1 April 2019 and applies to any Automated Decision System developed or procured after 1 April 2020, see <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> accessed 20 November 2019. The AIA has become obligatory under the Canadian Directive on Automated Decision-Making, which has taken effect as of 1 April 2019.

[34] David Reisman and others, 'Algorithmic Impact Assessment: A Practical Framework for Public Agency Accountability (AINOW) (2018) <https://ainowinstitute.org/aiareport2018.pdf> accessed 16 November 2019.

[35] James Harrison and Mary-Ann Stephenson, 'Human Rights Impact Assessment: Review of Practice and Guidance for Future Assessments' (Scottish Human Rights Commission, 2010) <http://fian-ch.org/content/uploads/HRIA-Review-of-Practice-and-Guidance-for-Future-Assessments.pdf> accessed 1 July 2019.

[36] Mantelero (n 4) 754.

[37] Bruno de Witte, 'The Crumbling Public/Private Divide: Horizontality in European Anti-Discrimination law' (2009) 13 Citizenship Studies 1.

[38] See as regards equal treatment, eg Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services [2004] OJ L373/37.

[39] This is admittedly a simplification of a complex picture; the halting evolution towards greater 'horizontality' in fundamental rights law is examined in great detail by Andrew Clapham, *Human Rights Obligations of Non-State Actors* (OUP, Oxford 2006); for a critical view of ongoing efforts to give greater horizontal effect to international human rights, see John H Knox 'Horizontal Human Rights Law' (2008) 102 American Journal of International Law 1.

overall impact assessment, from which appropriate risk factors should emerge. Those risk factors will guide the controller towards certain types of risk mitigation measures. The final section presents concluding observations.

ML: concerns and legal framework

Automated decision-making and ML

ADM refers to decision-making by systems in which algorithmic processing is involved—often including ML—to automate human decision-making.[40] ML is the procedure in which a system trains itself to spot patterns and correlations in large data sets and to infer new information and make predictions based on those patterns and correlations, without being specifically programmed to do so.[41] 'Training data' is used to train ML systems, comprising large data sets often provided by the system designer.[42] In short, what can be derived *from* the data is determined by what is *in* the data, or what the ML system designers identify as the relevant factors for analysis.[43] The decision about what those 'relevant factors to be analysed' are, might involve ethical and fundamental rights considerations, to which a controller eventually has to be able to respond and account for.

Machine learning

As there are values and choices involved in ML, the processing that leads up ML has to be considered here. Whenever ML is used in the context of ADM, the designer gives the system the desired output from their analysis of the training data.[44] The desired output results in a so-called deterministic ML model. This is where ADM is distinguished from artificial intelligence (AI), which creates output without human instructions.

During training, the system passes the training data through its internal statistical model so as to produce a calculated output, and then automatically adjusts the internal values (or weightings) of the statistical model so that the output of the system moves slightly closer to the desired outcome.[45] The process of incrementally adjusting the weightings is repeated over and over again, until its outputs closely match the desired value.[46]

Values are involved at various stages when training ML. First, the designer determines (with the controller's ADM purposes in mind) the data feed that is used to train the ML. They decide which data/what type of data is used, and which data is left out. A designer's decisions about the data feed inherently involve choices (and values). As these choices have to be made, there may be bias and potentially unfairness and discrimination in data input, output, and decisions. The use of historical data is another potential source of human bias and discrimination.[47] Equally important are values involved in the determination of the weightings of input data in the statistical model, to achieve the desired outcome. Potential failures should be identified and dealt with by the algorithm designer, through dry runs of the system in experimental, but secured settings away from the public.

The (intended) outcome itself can also infringe fundamental rights. Facebook's ADM, which determined who was targeted with emotionally charged messages, infringed their user's privacy. The selection of persons receiving selected types of information was likely to be based on data about the user's online communication habits. Facebook's intended use of that information was not communicated to the users. Facebook manipulated user newsfeeds to assess, for instance, changes in user emotions with the introduction of 'feelings', in addition to the 'like' button, and to assess changes in the emotions of insecure youths with 'vulnerable' moods. Facebook also assessed changes in the effectiveness of their services via cameras on smartphones or laptops (for content delivery) and via image analyses of photos such as selfies (in order to dynamically generate emojis).[48]

Once trained, a system can be used to infer new information and/or to make predictions based on other data: this involves inputting that data to the system so that it runs through the model, which ultimately produces the desired output.[49] Article 22 (3) GDPR requires involvement of meaningful human intervention whenever ADM results in legal effects, or in similar, significant effects for humans. This applies where the processing of personal data in ADM is based on the

[40] Cobbe (n 18) 4.

[41] David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn About Machine Learning' (2017) 51 University of California Davis Law Review 653ff; Cobbe (n 18) 4.

[42] Cobbe (n 18) 4.

[43] Emre Bayamlıoğlu and Ronald Leenes, 'The 'Rule of Law' Implications of Data-Driven Decision-making: A Techno-Regulatory Perspective' (2018) 10 Law, Innovation and Technology 7.

[44] Cobbe (n 18) 4.

[45] ibid 4.

[46] ibid 4.

[47] Julia Powles and Helen Nissenbaum, 'The Seductive Diversion of 'Solving' Bias in Artificial Intelligence: Trying to 'fix' A.I. Distracts from the More Urgent Questions about the Technology' (*Medium*, 7 December 2019) <https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53> accessed 7 October 2019.

[48] Clifford (n 21) 111.

[49] Cobbe (n 18) 4.

explicit consent of the data subject, or where such processing is necessary for entering into, or the performance of, a contract between the data subject and a data controller. In the context of processing of personal data by private organizations under the GDPR, ADM functions at its best as a guide, or as one of several tools for a human decision-maker who ultimately brings their judgement to the final decision.[50]

Multiple issues such as bias, unfairness, or discrimination may arise during input (at the capture stage), in the analytics and outcome (the computation stage), and in any human intervention that is involved. It may arise in the final decision-making stage. It generally appears due to the absence of transparency and explicability,[51] which has been identified as problematic with ADM, and which may originate from the use of ML.[52] The 'Fairness, Accountability and Transparency Research Machine Learning Community' (FAT-ML community) has conducted particularly extensive research on fairness, transparency, and accountability so as to improve standards in ADM systems, and to bring them into compliance with these principles as required by data protection law.[53] While this research is extremely valuable as regards improving control, reliability, and trust in ML(including ADM), it is also necessary to involve and apply *legal* perspectives and objectives (from, eg, administrative law, contract law, criminal law, or fundamental rights law) whenever ADM is to be deployed in a context where that particular law also applies. Data protection law is an important auxiliary to protect fundamental rights in digital contexts; yet, data protection law is not the sole legal framework through which to make ADM fundamental rights compliant and trustworthy.

Concerns with the application of automated decision-making

There are important challenges from a fundamental rights perspective in the relationships between private controllers and individuals. A first concern with ADM is opacity, which may largely reduce meaningful human control.[54] Related to this is the inexplicability of the reasoning in ADMs, creating information and power asymmetries between private organizations and individuals affected by ADM.[55] These asymmetries derive not only from the opacity of ML, they also stem from the fact that individuals often lack knowledge or understanding of the practices of the private organizations holding and processing their data, the technical specifics of the private organization architecture that individuals are dealing with, the data-sharing arrangements of private organizations with third parties that individuals do not know the details of, and the associated (data, and often surveillance-driven) business models behind any particular ADM with which individuals might get involved.[56] Moreover, the analytics inherent in these practices mean that organizations can generate further insights into behaviour and interests, further amplifying these asymmetries. The information asymmetry in particular may result in unnoticeable consequences for individuals and groups, as they will simply not be able to detect discrimination—they lack knowledge about the processing. This potentially leads to the restricted exercise of procedural rights (the right to an effective remedy). Individuals may unknowingly make choices based on discriminatory or privacy-impacting algorithms. Hildebrandt explains that 'precisely because a person is not aware of the profiles that are applied to her, she may be seduced to act in ways she would not have chosen otherwise'.[57]

A second concern is that the probabilistic nature of profiles can lead to the inaccuracy of personal data and to the de-individualization of persons, impacting their right to privacy and personal autonomy.[58] Third, automated decision-making and profiling potentially have discriminatory effects.[59] ADM can perpetuate, amplify, and potentially solidify existing patterns of bias, leading to discrimination.[60] A fourth challenge is related to ethics, with which ML is trained in a commercial setting. It is not self-evident that organizations operating with a commercial objective consider and assess impacts on fundamental rights. Commercial drivers, and the

[50] ibid 5; see more generally with regard to human judgement in algorithmic loops and automated decision-making: Reuben Binns, 'Human Judgement in Algorithmic Loops; Individual Justice and Automated Decision-Making', LawArxiv (15 September 2019) <https://doi.org/10.31228/osf.io/kz4s2> accessed 22 September 2019.

[51] Kroll and others (n 19) 165.

[52] Barocas and Selbst (n 7) 671; Cobbe (n 18) 4.

[53] For an indicative reading list, see <https://www.fatml.org/resources/relevant-scholarship> accessed 4 March 2019. The requirements of fairness and transparency are laid down in art 5 (1)(a) GDPR.

[54] Burrell (n 19).

[55] Alessandro Mantelero, 'Social Control, Transparency and Participation in the Big Data World' (2014) Journal of Internet Law 23.

[56] Heleen L. Janssen, Jennifer Cobbe, Chris Norval, Jatinder Singh, 'Personal Data Stores and the GDPR's Lawful Grounds for Processing Personal Data' (2019) 2 <https://zenodo.org/record/3234902#.XdVDcy2ZOYU> accessed 20 November 2019.

[57] Mireille Hildebrandt, 'The Dawn of a Critical Transparency Right for the Profiling Era' in Jacques Bus and others (eds), *Digital Enlightenment Yearbook* (IOS Press, Amsterdam 2012) 49.

[58] Anton Vedder, 'Knowledge Discovery in Databases: The Challenge to Individualism' (2000) 1 Ethics and Information Technology 278.

[59] Mireille Hildebrandt, 'Defining Profiling: 'A New Type of Knowledge?' in Mireille Hildebrandt and Serge Gutwirth (eds), *Profiling the European Citizen* (Springer, Dordrecht 2010) 18.

[60] Raso and others (n 4) 18.

drivers of competitiveness might distract private organizations from other relevant rights and interests. This might perhaps originate from unawareness of the fact that both public and private organizations have to comply with the GDPR, or from negligence regarding knowledge of potential impacts. Fifth, the positive and negative impacts of ADM are not always evenly distributed across society.[61]

Controller interests in ADM

As innovation and commercial activities by private entities are crucial in the context of the EU internal market, fundamental rights will have to be calibrated with legitimate interests of private organizations, such as commercial or economic interests.[62] Article 16 of the EU Charter recognizes one's 'freedom to conduct a business in accordance with Union law and national laws and practices' as a fundamental right. Fundamental rights require, whenever they collide or compete for preponderance, an appropriate balancing. An under-inclusive balancing might entail risks for the fundamental rights of individuals, while an over-inclusive interpretation might lead to a violation of Article 16 of the EU Charter, potentially resulting in stifling technological innovation and competitiveness in the EU internal market.[63] This implies that a private entity has to prudently calibrate *all* relevant interests involved. Legal compliance (with the GDPR, fundamental rights) might be costly from a financial, organizational, or from a computational perspective. The benchmarks presented in this article seek to help private entities find an equitable calibration between their own interests and the rights of individuals involved in their ADM.

A tailor-made approach avoiding over- and under-inclusiveness ideally means that whenever risks to the fundamental rights of individuals are high, the mitigating measures should be appropriate. When the impacts of ADMs are assessed as mid-level, the level of protection might be instead adjusted in order to prevent over-inclusiveness and chilling effects on innovation. As many private organizations are uncertain about the interpretation of many of the open norms and principles in the GDPR, and also aware of the severe consequences of non-compliance with the GDPR, there is a need for practical guidance to assess fundamental rights impact assessments.[64]

The GDPR assigns 'data controllers' with the responsibility of designing ADM which is compliant with the GDPR. A data controller is defined in Article 4 (7) GDPR as the entity which determines the purposes and means of the processing of personal data to achieve those purposes. In private organizations, a person (or a group of persons) at board level act as data controllers. To be compliant with the GDPR, controllers considering the use of ADM are responsible for a range of activities, including undertaking risk assessments (Article 35 GDPR), the implementation of appropriate technical and organizational measures to ensure and to be able to demonstrate that their ADM is performed in accordance with the GDPR (Article 24).

From the perspective of the controller, there may well be legitimate reasons to use ADM. An often touted benefit of a well-deployed ADM is the speed with which decisions can be made.[65] A well-implemented ADM can, in certain contexts, make decisions that are more accurate than those made by humans.[66] Some claim that certain models of ADM are stable, as ADM based on well-trained ML may generate the same prediction when presented with different data, however, ADM models might adapt over time due to changes in weightings. In such cases, decisions may be barely, or not at all, reproducible. Whenever these models have impacts on fundamental rights, they may, depending on the gravity of the impact, become problematic if the decision-making process cannot be reproduced and accounted for. However, this may not be very different in humans—mood swings, troubles at home, being hungry, the time of day, and many other influences might lead to different conclusions and different decisions when they are repeated.[67]

Computational interests in ADM

A controller will usually choose optimal accuracy, speed, and computational resources to smoothly deliver

[61] ibid 17.

[62] Recital 4 GDPR.

[63] Recital 7 GDPR.

[64] Whenever controllers infringe the rights of data subjects under the GDPR, they may be subject to administrative fines up to € 20,000,000—or in the case of an undertaking, upto 4% of the total global annual turnover of the preceding financial year, whichever is higher, see art 83 (5) GDPR.

[65] Finlay (n 10) 7.

[66] ibid 8.

[67] See eg Dimitra Kamarinou, Christopher Millard and Jatinder Singh, 'Machine Learning with Personal Data' (7 November 2016) Queen Mary School of Law Legal Studies Research Paper No 247/2016, 5 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2865811> accessed 24 November 2019; Susan Dynarski, 'Why Talented Black and Hispanic Students Can Go Undiscovered' *The New York Times* (8 April 2016) <http://www.nytimes.com/2016/04/10/upshot/why-talented-black-and-hispanic-students-can-go-undiscovered.html?_r=0> accessed 12 January 2020; Proceedings of the National Academy of Sciences Paper 'I think it's time we broke for lunch...' *The Economist* (14 April 2011) <http://www.economist.com/node/18557594> accessed 12 January 2020. Some have suggested that in some contexts the use of ADM might lead to fairer decisions for individuals (Kamarinou, Millard and Singh, ibid 7).

services in a highly competitive market. They may have to balance computational interests with other interests, such as fundamental rights, or GDPR requirements such as a data subject's access to information.[68] Neural networks are considered opaque primarily because their mode of operation provides little insight into the processing and in the underlying model.[69] Decision trees are generally seen as more transparent, but they can be very complex as well.[70] Some ML techniques (neural networks, decision trees, or logistic regression) are more accurate and appropriate for solving problems in some domains than others. ML techniques to establish loan default predictions will be different compared to, for example, image recognition.[71] Models may also be combined so as to find optimal accuracy in ADM.[72] The systems designer commonly decides about the most appropriate type at the design stage. Controllers should thereby bear in mind that whatever the choice for a type of ML, the controller is accountable and responsible in the context of the GDPR towards data subjects and supervisory authorities regarding that choice.[73]

From a controller's perspective, ADM is primarily designed and deployed to perform accurately; a given ADM is considered better by a designer when its results show improvement according to agreed metrics.[74] These metrics assess aspects such as accuracy, speed, and computational resources. The extent to which these metrics align is not clear—or it may potentially collide with human interpretations.[75] Baumer stated that these interpretations are rarely, if ever, accounted for during the design and development of ADM training models. These interpretations often enter the picture once released in public—which can be explained (but not justified) from the viewpoint that the process of designing an ADM is often primarily driven by technical and financial constraints, timeliness, and competition and performance in the metrics.[76]

GDPR's DPIA

DPIAs seek to provide organizations with tools to manage data protection risks. While they initially emerged as voluntarily measure, DPIAs on ADMs in the GDPR[77] are, in certain circumstances, now a mandatory requirement when certain conditions are met.[78] This subsection explains the relevance of DPIAs for the proposed approach to fundamental rights impact assessments, as well as their potential shortcomings.

DPIAs and their relevance to fundamental rights impact assessments

Processing operations that may result in a high risk to the rights and freedoms of natural persons require the controller, to execute a DPIA to evaluate, in particular, the origin, nature, details, and severity of that risk [Article 35 (1) GDPR].[79] The EU legislator has recognized that potential risks to rights, such as discrimination or 'any other significant economic or social disadvantage', may herewith become part of that evaluation as well.[80] The Article 29 Working Party,[81] the European Data Protection Supervisor, and other experts[82] have suggested a broader impact assessment, including analysis of the societal and ethical

[68] Arts 13 (2)(f) and 14 (2)(g) GDPR refer to information that has to be provided to data subjects; art 15 (1)(h) GDPR refers to the obligation to provide data subjects with meaningful information about the logic involved and with the significance and the envisaged consequences of the processing.

[69] Singh and others (n 19) 4; Burrell (n 19) 5.

[70] Burrell (n 19) 5.

[71] ibid 5.

[72] ibid 5.

[73] Art 5 (2) and art 4 (7) GDPR.

[74] In this sense, see Eric Baumer, 'Toward Human-centered Algorithm Design' (2017) 4 (2) Big Data & Society 1; Berk Ustun and Cynthia Rudin, 'Methods and Models for Interpretable Linear Classification' (1 October 2014) https://arxiv.org/pdf/1405.4047.pdf> accessed 11 January 2020. Hildebrandt critically noted that sometimes 'profilers are not very interested in causes or reasons, [as] their interest lies in a reliable prediction, to allow [for] adequate decision-making', Hildebrandt (n 59) 18.

[75] The use case of how Facebook used metrics as mentioned in s 1.1 (fn 5) provides for an insightful example of how a controller's perspective on accurate metrics (eg mere commercial perspectives) can collide with human interpretations on accurate metrics (eg metrics reflecting respect for the right of privacy of the data subjects involved).

[76] Baumer (n 74) 2.

[77] Art 35 (1) reads: '[. . .] Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data.' art 35 (3)(a) GDPR reads: 'A data protection impact assessment referred to in section 1 shall in particular be required in the case of: [. . .] a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person; [. . .].'

[78] Reuben Binns, 'Data Protection Impact Assessments: a Meta-Regulatory Approach' (2017) 7 International Data Privacy Law 22; Finck (n 18) 90–91.

[79] Recitals 75, 84 GDPR; Article 29 Working Party, 'Guidelines on Data Protection Impact assessment (DPIA) and Determining Whether Processing is Likely to Result in a High Risk for the Purposes of Regulation 2016/679' (WP 248 rev.01 of 4 October 2017); for the evaluation impact assessments, see eg Binns (n 78) 22; Michael Veale, Reuben Binns and Jef Ausloos, 'When Data Protection by Design and Data Subject Rights Clash' (2018) 8 International Data Privacy Law 118; Alessandro Mantelero, 'Comment to Articles 35 and 36' in Mark Cole and Franziska Boehm (eds), *GDPR Commentary* (Edward Elgar Publishing, forthcoming) accessed 11 January 2020.

[80] Recital 75 GDPR.

[81] Article 29 Working Party (n 79).

[82] EDPS Ethics Advisory Group, 'Towards a Digital Ethics' (2018) <https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf> accessed 28 November 2019; see also Council of Europe Committee of Experts on Internet Intermediaries (MSI-NET) (n 4); see for an overview of initiatives Mantelero (n 4) 754.

consequences where personal data is used, and some have started to give shape to approaches to include broader assessments.[83]

DPIAs are under Article 35 (3)(a) GDPR particularly required in the case of a systematic and extensive evaluation of personal aspects relating to natural persons, where that evaluation is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person, or that similarly, significantly, affect the natural person. Whenever controllers aim at producing legally significant effects, or similar, by means of ADMs, they are usually obliged to assess whether their intended ADM entails high risks.[84]

A DPIA contains a systematic description of the purposes and the means through which these purposes should be realized, including the envisaged processing operations [Article 35(7)(a) GDPR]; an assessment of the necessity and proportionality of the processing operations in relation to the purposes [Article 35(7)(b) GDPR]; an assessment of the risks to the rights and freedoms of data subjects [Article 35(7)(c) GDPR]; and the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data, and to demonstrate compliance with the regulation, taking into account the rights and legitimate interests of data subjects and other persons concerned [Article 35(7)(d) GDPR]. 'Rights and freedoms' as mentioned in Articles 35(1) and 35 (7)(c) GDPR include *fundamental* rights (as enshrined in the EU Charter of Fundamental Rights), given that Recital 2 GDPR explicitly references to such, and that Recital 75 GDPR highlights that the use of personal data in the applications of new technologies can entail discrimination, physical, material or non-material damage, or other significant economic or social disadvantages.

Whenever a DPIA indicates that processing operations involve high risks that the controller cannot or does not mitigate using the appropriate measures in terms of available technology and costs of implementation, the supervisory authority must be consulted [Article 36(1) GDPR]. If it is of the opinion that the AI system would infringe the GDPR, it should provide written advice to the controller, and may also impose a ban on that controller's processing [Article 58 (2)(f) GDPR].

The GDPR, which currently offers the most advanced legal framework for data protection, does not explicitly include detailed approaches as to how or at what level of detail fundamental rights impacts should be assessed. It has been suggested that fundamental rights impact assessments should become mandatory to give teeth to the protection of these rights whenever the use of impactful technology is intended.[85] However, a voluntary application of a fundamental rights impact assessment as part of a DPIA might at this stage be preferred over a mandatory assessment, as more detailed guidance is currently missing, while a voluntary approach is more consistent with the GDPR, which itself focuses on risk assessment.[86] Once fundamental rights impact assessments have become practical and operable, they might start to function as a controller tool through which compliance can be demonstrated, whereby companies can distinguish their fundamental rights compliant services from other, non-compliant suppliers.

Interplay between DPIA and requirement of data protection by design and by default

Not only Article 35 GDPR urges controllers to seriously consider the need of a DPIA. The data protection by design requirement as enshrined in Article 25 GDPR naturally underpins the need for an early stage impact assessment of a controller's envisaged design and deployment of ADM. Data protection by design is now a qualified duty under Article 25 GDPR; it requires controllers to implement appropriate technical and organizational measures to comply with the GDPR at design stages where the means for processing are also determined during deployment.[87] These requirements not only apply to the technical processing, but also to business models *as such*.[88] The rationale for data protection by design and default is that it aims at controller compliance during the entire cycle of an ADM system, rather than fulfilling separate normative GDPR postulates.[89] The data protection by design and by default requirement therefore strongly supports

[83] See eg 23 Asilomar AI Principles of the Future of Life Institute <https://futureoflife.org/ai-principles/?cn-reloaded=1> accessed 28 June 2019.

[84] See also art 35 (1) GDPR; Article 29 Working Party (n 79) 7.

[85] David Wright, 'A Framework for the Ethical Impact Assessment of Information Technology (2011) 13 Ethics and Information Technology 199; Alessandro Mantelero, 'The Future of Consumer Data Protection in the E.U. Rethinking the 'Notice and Consent' Paradigm in the New Era of Predictive Analytics' (2014) 30 Computer Law and Security Review 643 [but see Mantelero (n 4) 756].

[86] See for references Mantelero (n 4) 756.

[87] Art 25 (1); art 25 (2) GDPR requires controllers to take into account the 'state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing'.

[88] Finck (n 18) 91 (emphasis added).

[89] Lee Bygrave, 'Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements' (2017) 4 Oslo Law Review 105; Finck (n 18) 91.

undertaking a DPIA including fundamental rights impact assessments.

It is currently not clear how the requirements of data protection by design and default should be applied in a DPIA.[90] Such a lack of clarity might seriously reduce DPIA effectiveness, and thus the effectiveness of the fundamental rights impact assessments embedded in the DPIA. To improve compliance, regulators should develop approved certification mechanisms, to enable controllers to demonstrate their compliance with the requirements of data protection by design and default. To date, however, no such mechanisms have been initiated. This is a rather urgent issue, given that regulators are, when considering fines, explicitly called on to give due regard to considering the degree of responsibility taken by the controller, thereby taking into account the technical and organizational measures implemented by controllers pursuant to Article 25.[91] Impact assessments at later stages, or omitting them altogether, will, given the explicit call to regulators in Article 83 (2)(d) GDPR, probably be critically judged if there are complaints about a company's ADM. Little space might be left for controllers to make appropriate adjustments at later stages, while consumer trust might already be damaged.

DPIAs: interaction between controller, data protection officer, and supervisory authority

An important accountability requirement is the designation of a data protection officer (DPO),[92] who monitors the organization's performance pursuant to, among others, the DPIA.[93] While a DPO might lack detailed fundamental rights knowledge to conduct the fundamental rights impact assessment, national human rights commissioners, NGOs or guidelines from expert groups such as those of the Council of Europe could provide elucidation and guidelines to DPOs.[94] DPOs fulfil a key position with regard to compliance with the GDPR, encompassing the design and the deployment of the ADM, while they also operate on behalf of the controller as a point of contact with supervisory authorities.[95]

The GDPR does not require a DPIA to be executed by an external independent body. From an awareness and a future effectiveness perspective, it is preferred that the impact assessment happens within the organization itself, rather than having it executed by a distant entity carrying out the assessment.[96] The responsibilities of controllers (and processors) include the obligation to support and maintain relevant expert DPO knowledge.[97] National Human Rights Commissioners and other knowledge bodies could assist DPOs in reducing knowledge gaps about fundamental rights interpretation and case law.

Challenges with DPIAs and the approach to fundamental rights impact assessments

While the DPIA entails a combination of self-assessment and control by supervisory authorities, the risk management as enshrined in the DPIA itself might, for various reasons, be less effective than foreseen.[98] First, DPIAs largely consist of *internal* assessments, the results of which are commonly not publicly available, which weakens their transparency. Secondly, where Article 29 Working Party encourages controllers to take an approach that includes involving the data subject, the consultation of data subjects is only required, according to Article 35 (9) GDPR, 'where appropriate', thereby allowing controllers to waive their involvement if their commercial interests or the security of processing operations come into play. This might effectively reduce *any* involvement of data subjects or other stakeholders.[99] Next, the Article 35 concept of "high risk" needs interpretation by controllers, who might prefer to underestimate the risks of their ADM so as to circumvent prior consultation with supervisory authorities.[100] Another, more practical issue is that the resources of supervisory authorities are commonly scarce,

[90] Veale, Binns and Ausloos (n 79) 118.

[91] Art 83(2)(d) GDPR.

[92] Art 37 (1)(b) GDPR. The DPO is an independent advisory person within a company, monitoring compliance with the GDPR and with the policies of the controller in relation to the protection of personal data, including the assignment of responsibilities, awareness raising, and training of staff involved in processing operations. See also Article 29 Data Protection Working Party, 'Guidelines on Data Protection Officer ('DPOs')' (WP 243 rev.01, 30 October 2017).

[93] Art 39 GDPR. The DPO is also a contact point for the supervisory authority on issues relating to processing, including its prior consultation with regard to the DPIA in art 36.

[94] At the European level, advisory committees [such as Council of Europe Committee of Experts on Human Rights Dimensions of automated data processing and different forms of artificial intelligence, see <https://www.coe.int/en/web/freedom-expression/msi-aut or the EU-European Group on Ethics in Science and New Technologies (EGE), see <https://ec.europa.eu/research/ege/index.cfm] develop guidelines regarding how to assess fundamental rights impacts in horizontal relations caused by the use of new technologies.

[95] Art 39 (1) (b) – (d) and 39 (1) (e) GDPR.

[96] A different approach is defended by Mantelero (n 4) 759, who argues that an impact assessment should provide a self-assessment tool, as 'there are cases though in which bridging the gap between the theory of rights and values and their concrete application is complicated by the nature of data use and complexity of the associated risks. In such cases, the assessments should be carried out by an ad hoc panel of experts, just as ethics committees apply general principles and guidelines to a specific case'.

[97] Art 38 (2) GDPR.

[98] Mantelero (n 79) 19.

[99] Veale, Binns and Ausloos (n 79) 118.

[100] Mantelero (n 79) 9.

which can result in delays to written advice by the supervisory authorities.

Despite potential shortcomings and uncertainties, DPIA remains a useful tool for controllers, as it brings societal benefits by forcing risks to be considered from the outset. DPIA is an *ex ante* form of risk assessment. Note that some risks may only be discovered once an ADM system is in operation, however, when a change of risk occurs and is identified, controllers should conduct another DPIA [Article 35 (11) GDPR]. Given the *ex ante* nature of DPIA, a controller designing ADM should consider how the rights and interests of natural persons will be safeguarded during deployment of their ADM as required by Articles 24 and 25 GDPR, as not all issues can be solved at design stages only.

It would be useful if (anonymized) controller experiences could at some point be shared with other controllers or relevant bodies, to offer them opportunities to learn from the way that relevant individual and controller interests were calibrated in actual cases. This would allow for the development of a body of best practices and of lessons to be learned from less successful impact assessments (eg where impacts were underestimated and mitigating measures absent). To keep track of the developments and effectiveness of DPIAs, including fundamental rights impact assessments—meanwhile respecting legitimate restrictions on private controllers not to disclose all details of their business models, IP rights-trusted organizations (perhaps regulators) could aggregate such DPIAs.[101] This would enable opportunities for inspection, comparison, research and, where necessary, new ways forward. A trusted organization might also be enabled to share the reports of impact assessments with the public, while respecting the legitimate restrictions on private controller interests.

Once the deployment of ADM has begun, controllers are obliged to actively inform users about data processing, and specific GDPR requirements apply in the case of ADM. To enhance the controller's transparency and accountability regarding obligations under Articles 5 (1)(a) and 5(2) GDPR, controllers have to provide data subjects with meaningful information about the logic involved, as well as the significance and the envisaged consequence of such processing for the data subject of ADM.[102] Data subjects can exercise the right to access their personal data against a controller, as well as their rights to rectification, erasure or objection against data processing (these are also called 'data subject access rights').[103] Data subject access rights seek to provide data subjects with an important mechanism to interrogate controllers about their intentions, reasons for data processing and their compliance with fundamental rights. Trade secrets or intellectual property rights (and particularly the copyright protecting the software) should not be adversely affected by data subject access rights, but controllers cannot use such arguments to entirely refuse access.[104]

Regulators might further strengthen these rights by requiring controllers to actively facilitate data subject access rights, for instance through technical architecture included in ADM, so that they become more accessible and effective. Such technical architecture could, if appropriately and comprehensively implemented, contribute to a controller's data protection by design obligation under the GDPR. These ideas might typically arise when conducting a DPIA to ADM.

Veale and others have suggested strengthening DPIA more generally through the use of other GDPR supported means at the operational stages of ADM.[105] The GDPR gives data subjects the right to mandate a not-for-profit-body to lodge a complaint on their behalf in order to exercise their rights.[106] Given that system infrastructures and technical architecture can be extremely complex and therefore incomprehensible to data subjects, there may be a need for organizations, such as not-for-profit bodies, who can understand, investigate, interrogate, or report potential data protection and fundamental rights issues related to conducting a DPIA, or who can publicly highlight the existence of state-of-the-art technologies that could assist controllers with better balancing between (fundamental) rights and controller interests in conducting the DPIA.[107] They might also assist users by developing templates to help them effectuate their data subject access rights.

Supervisory authorities (and courts) are always entitled to be adequately informed about trade secrets models, intellectual property rights or copyrights *ex post*, as regulators are called on by the GDPR to give due regard to the degree of responsibility taken by the controller's technical and organizational measures that were implemented.[108] Regulators should thereby also consider how the controller has facilitated the data subject access

[101] Chris Reed et al., 'Data Trusts: Legal and Governance Considerations' (April 2019) <https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf> accessed 11 January 2020.

[102] Arts 13 and 14 GDPR; regarding specific requirements relating to ADM, see arts 13 (2)(f) and 14 (2)(g) GDPR.

[103] Arts 15, 16, 17 and 21 GDPR.

[104] Recital 63 GDPR.

[105] Veale, Binns and Ausloos (n 79) 121.

[106] Art 80 (1) GDPR; art 80 (2) GDPR allows such bodies to exercise these rights, subject to Member State derogation without data subject mandate.

[107] Veale, Binns and Ausloos (n 79) 121, who envisages such a role for not-for-profit-bodies in the context of Data Protection by Design issues.

[108] Art 83 (2)(d) GDPR.

rights by way of their technical and organizational measures; however, at that stage, risks to fundamental rights might already have operationalized.

As DPIA might become a mandatory exercise for controllers of ADM systems, they will become one of the mechanisms for the governance of fundamental rights in ADM and in other emerging information technologies. DPIA, and in particular the notion of 'risks to the rights and freedoms of data subjects' which it seeks to protect, potentially embody a shift from classical legal practice to more risk-based approaches of these rights.[109] Merging risks and rights in the proposed fashion could change their meanings into something hardly predictable.[110] Future research might identify emerging gaps in the way DPIAs will be operationalized, and determine whether a risk-based approach to the fundamental rights might itself pose a risk to these rights, while detecting prospects for corrections and for lessons to be drawn from DPIA conducted by private controllers.[111]

GDPR and automated decision-making

Article 22 (1) GDPR reads: '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 affects him or her.'[112] Whenever this applies, a data controller cannot process personal data unless one of the following lawful grounds applies—if (i) the decision is necessary for entering into, or the performance of, a contract between the data controller and data subject [Article 22 (2) (a) GDPR]; if (ii) the decision is authorized by Union or Member State law to which the data controller is subject [Article 22 (2)(b) GDPR]; or if (iii) the decision is based on the data subject's *explicit* consent [Article 22 (2)(c) GDPR].[113] The GDPR mandates that where ADM is based on the first or third grounds of Article 22, the controller shall implement 'suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to human intervention on the part of the controller, to express his or her point of view and to context the decision'.[114]

Private controllers have to equip their ADM systems with human intervention mechanisms.[115] Working Party Article 29 has stipulated that human intervention must not just be nominal but rather take the form of a review carried out by someone who 'has the appropriate authority and capability to change the decision'.[116] This is important, as humans may tend to uncritically follow what was proposed by the ADM—which has the appearance of objectivity to many.[117] The moment of intervention—before the ADM is deployed, or after deployment, or at both stages—is not clearly prescribed by law.[118] The requirement that someone must have appropriate authority and ability to change the decision should contribute to the effectiveness of human intervention.[119]

The DPIA seeks to assist controllers in identifying potential issues with their obligations, and to help determine how data subjects can be informed in a compliant manner.[120] Being controller entails a range of other obligations under the GDPR, among which the obligation to satisfy the data subject's access rights (Articles 12–22 GDPR). A data subject's access rights include the right to obtain information about the existence of ADM: controllers have to provide meaningful information about the logic involved, as well as the significance and the envisaged consequences of the processing.[121]

[109] Niels van Dijk, Raphaël Gellert and Kjetil Rommetveit, 'A Risk to a Right? Beyond Data Protection Risk Assessments' (2016) 32 Computer Law & Security Review 286.

[110] ibid 286.

[111] In this sense, see ibid 286.

[112] Recitals 71 and 72 GDPR.

[113] Controllers cannot process personal data in the context of ADM if that processing is based on the controller's legitimate interests, see Art 6 (1)(f) GDPR.

[114] Art 22 (3) GDPR.

[115] Art 22 (2)(a) and (c) can be invoked by private controllers, whereas the authorization of ADM by law as envisaged by art 22 (2)(b) is typically a matter for public entities. In that case, human intervention is not obligatory (although legislators could still decide to include human intervention in a specific law).

[116] Article 29 Working Party (n 18) 27.

[117] Mary Cummings, 'Automation Bias in Intelligent Time Critical Decision Support Systems' (AIAA First Intelligence, Chicago, 20 September 2004) <https://arc.aiaa.org/doi/10.2514/6.2004-6313> accessed 1 July 2019.

[118] Finck (n 18) 84.

[119] There has been much debate about the scope of application of art 22 (1) GDPR. The DPD granted the right not to be subjected to automated decision-making to *everyone*. The GDPR has limited the right to the *data subject*. Kamarinou, Millard and Singh (n 67) 10–11 argue that art 22 GDPR may apply to group decisions, as 'the provision does not limit "profiling" as such to individual profiling but only requires that the decision based on such profiling is addressed to an individual, in a way that has legal or significant effects for him or her as an individual'. Extending applicability to group members would eventually result from equal applications of the right in concrete dispute settlements due to the legal principle of precedent.

[120] Controllers using ADM shall, as required by art 5 (1) GDPR, act in a lawful, fair, and transparent manner. Art 5 (2) holds controllers accountable and responsible as regards the data protection principles laid down in art 5 (1) (a)–(f).

[121] Arts 13 (2)(f) and 14 (2)(g) GDPR. The controller shall provide the data subject upon its request with access to information about the existence of ADM, referred to in art 22 (1) and (4) GDPR.

Relevant legal framework of fundamental rights

Fundamental rights reflect well-established basic rights in a society designed to secure the liberty of individuals, whereby these rights must be harmonized with the interests of society. These rights are technologically independent: they are not formulated so as to privilege a certain technology over another.[122] Most fundamental rights involve open norms, as they apply to all public sectors in society. States should not infringe these rights, unless such limitation is foreseeable and accessible, necessary in a democratic society, proportionate to the purposes, and laid down by law.[123] Private parties are generally allowed to act in society, *unless* the law prohibits them to do so.

Through their activity, fundamental rights increasingly apply to private parties as well, whereby this application is often achieved through 'indirect effect'. This happens, for instance, by judicial interpretation of civil lawsuits between private parties, or by enactment of national legislation (that will apply to private parties).[124] Other international legal frameworks have become more active in applying fundamental rights as norm-setting to business enterprises as well. As businesses are at the forefront of innovating and implementing ADM, the United Nations Guiding Principles on Business and Human Rights of 2011 are salient in ensuring that ADMs in horizontal relationships are deployed in a rights-respecting manner.[125]

If private parties, by their actions, violate fundamental rights of individuals as enshrined in the ECHR, the European Court of Human rights (ECtHR) commonly requires *states* to prevent, to combat, and to sanction violations of the rights.[126] The ECtHR has more than once urged states to take regulatory measures to effectively address and solve fundamental rights violations by private parties.[127] This more stringent approach particularly applies in criminal context, where the ECtHR has required states to enact legislation,[128] or to take effective preventive measures.[129] Yet, a state's obligation to act in relations between private parties is, under the ECHR, not unlimited. If effective legal redress exists,[130] if the complaint concerns a claim not encompassed by the ECHR,[131] or if limitations on a fundamental right appear relatively limited,[132] no such obligation might apply.

While indirect effect of fundamental rights to private parties has been confirmed, application and limitation of those rights largely remain a state responsibility. The determination of whether a private party's ADM lawfully limits fundamental rights as per the ECHR and, given its close ties to the ECHR, the EU Charter, may primarily have to be defined by regulators. Such demarcation of where limitations might be allowed will assist private parties in complying with the open norms of fundamental rights when conducting a DPIA. This appears an area for future research and further clarification and guidance by regulators.

Insofar as the EU Charter contains rights that correspond to the rights guaranteed by the ECHR, the meaning and scope of those rights in the EU Charter equal those in the ECHR.[133] The GDPR contains (implicit) references to fundamental rights in the context of ADM as well. Article 22 (1) GDPR stipulates 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

[122] The Dutch Constitution currently contains a technology-dependent right to communication secrecy, which is according to its letter currently solely covering communication via mail, telefax, and telephone, leaving, for example, e-mail or SMS and other means for communication unprotected. A constitutional legal proposal to adjust this is pending; see for references <https://www.eerstekamer.nl/wetsvoorstel/33989_verandering_in_de_grondwet> accessed 7 December 2019.

[123] See for limitations on fundamental rights as per the ECHR: Roza Pati, 'Rights and Their Limits: The Constitution for Europe in International and Comparative Legal Perspective' (2005) 23 Berkeley Journal of International Law 248; as regards limitations in relations between private parties, see Janneke Gerards, *General Principles of the European Convention on Human Rights* (CUP, Cambridge 2019) 136ff; there are only a few 'absolute' rights which do not allow for limitations, see, for example, the right not to be tortured, the right to think whatever one pleases (this is the so-called forum internum), the right not to be submitted to slavery, and the requirement to respect the principle of legality as enshrined in art 7 ECHR.

[124] Olha Cherednychenko and Norbert Reich, 'The Constitutionalisation of European Private Law: Gateways, Constraints and Challenges' (2015) 5 European Review of Private Law 797; Hans Micklitz, Constitutionalisation of European Private Law (OUP Oxford 2014); Justice Aharan Barak, 'Constitutional Human Rights and Private Law' (1996) Faculty Scholarship Series Paper 3698, 218, <http: digitalcommons.law.yale.edu="" fss_papers="" 3698=""> accessed 11 January 2019; fundamental rights norms might moreover permeate civil law norms whenever national courts interpret open civil norms (eg tort law) with fundamental rights norms. This applies, for instance, whenever the right to privacy or prohibition of discrimination applies in employer–employee relationships.

[125] UN principles Guiding Principles on Business and Human Rights of 2011 <https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf> accessed 12 January 2020.

[126] *Moreno Gómez v Spain* App no 4243/02 (ECtHR, 16 November 2004) paras 57, 61; Gerards (n 123) 136.

[127] *Siebenhaar v Germany* App no 18136/02 (ECtHR, 3 February 2011) para 38.

[128] *X and Y v the Netherlands* App no 8970/80 (ECtHR, 26 March 1985).

[129] *Osman v United Kingdom* App no 23452/94 (ECtHR, 28 October 1998, Grand Chamber) paras 115, 116.

[130] *Benderskiy v Ukraine* App no 22750/02 (ECtHR, 15 November 2007) para 60.

[131] *Smith v United Kingdom* App no 39658/05 (ECtHR, 4 January 2007, admissibility decision).

[132] *Appleby and others v United Kingdom* App no 44306/98 (ECtHR, 6 May 2003) paras 41, 47.

[133] Art 52 (3) EU Charter; this provision shall however not prevent Union law providing more extensive protection.

or her' …. Article 35 (3)(a) GDPR concerns high risks to rights that should be identified by a DPIA, thereby referring to 'high risks to the *rights and freedoms* of natural persons', and to '*legal effects*' (emphasis added). Legal effects and rights and freedoms also include fundamental rights. Controllers envisaging ADM in their data processing have to comply with a natural person's fundamental rights as well. The wording of Article 35(7)(c) GDPR endorses this approach.

Benchmarks to assess fundamental rights impacts

Building an approach to a fundamental rights impact assessment requires a selection of benchmarks, which can be applied in every private sector context where ADM is foreseen. A benchmark is a practical tool which allows a referential or calibration framework, rather than for the production of an exact outcome after its application.[134] In other words, the results from the benchmarks are not 'mathematicized', but rather leave space for a spectrum of impacts, within which low, medium, and high impacts on fundamental rights might emerge. Benchmarks can be recommended in a policy field where the impacts of innovation—typically including the impacts of ADM on fundamental rights—have not yet been stably, predictably applied.[135] In such a dynamic context, exact measurements might instead lead to false certainty, while benchmarks require explanations as to why a controller judges their own ADM as medium or high risk. Responding to the benchmarks enables controllers to assess the fundamental rights risk levels that their intended ADM might entail.

In this section, the benchmarks will be elaborated with examples. The sample responses to the benchmarks will be presented in schedules. After scrutiny of these benchmarks, an overall fundamental rights impact assessment emerges from which a controller should conclude a risk factor, which can be low, medium or high risk. These levels may either urge an adjustment in the technology and the broader organizational processes of the intended ADM, or to the adoption of appropriate mitigating measures so as to minimize the impact on fundamental rights.

Specifically, *Benchmark 1* requires the controller to identify fundamental rights that are potentially at risk due to the use of ADM. *Benchmark 2* requires the controller to identify potential elements in the design of their ADM system that may contribute to infringements, and to consider what appropriate technical or organizational, mitigating measures must be taken. *Benchmark 3* seeks a controller's explanation of how risks and interests are balanced, why ADM is necessary to achieve their commercial purposes, and why, given AI's risk to fundamental rights, other, less intrusive means cannot be used. *Benchmark 4* interrogates the controller as to the extent to which data subjects exercise control or mitigate the data processing in ADM.

Benchmark 1: what fundamental rights are impacted?

The identification of relevant fundamental rights as a first benchmark seeks to secure a private controller's focus on the legal framework within which the other three benchmarks are responded to. This subsection presents examples of ADM resulting in low, medium, and high impact on fundamental rights. The rights are discussed in clusters (in the context of the approach to a fundamental rights impact assessment: privacy rights, equal treatment rights, expressional rights, and procedural rights).[136] The assessments are based on the EU Charter, the ECHR, the case law of the ECtHR or on that of the EU Court of Justice.

Privacy rights

'Personal autonomy', as the capping principle of privacy rights, is an established right in the ECtHR case law.[137] It entails one's right to personal development and a right to establish and develop relationships with other human beings and the outside world, as well as a sphere within which individuals can freely pursue the development and fulfilment of their personality.[138] The Council

[134] Kenniscentrum Wetgeving en Juridische Zaken of the Ministry of Justice of the Netherlands, '"Factsheet Benchmarking" (2016) in Integraal Afwegingskader voor Beleid en Regelgeving' <https://www.kcwj.nl/sites/default/files/Factsheet_Benchmarking.pdf> accessed 19 November 2019.

[135] Michiel Bijlsma, Bastiaan Overvest and Bas Straathof, 'Meer Onzekerheid door ICT. Vroegtijdig Ingrijpen Niet Nodig. Marktordening bij Nieuwe ICT-toepassingen' (2016) Centraal PlanBureau Policy Brief 2016/09 <https://www.cpb.nl/publicatie/marktordening-bij-nieuwe-ict-toepassingen> accessed 11 January 2020.

[136] For the purpose of the approach to developing a fundamental rights impact assessment, it is considered useful to work with four 'clusters' of fundamental rights, without claiming that these clusters are complete. This is not an uncommon approach, see Gerards, Nehmelman and Vetzo (n 4). This approach should however not imply that the rights mentioned in the clusters are complete, as eg social, cultural, and economic fundamental rights are not explicitly included. Those rights may still form part of one of the clusters.

[137] *Johansen v Norway* App no 17383/90 (ECtHR, 7 August 1996) 23 EHRR 3; *Laskey, Jaggard and Brown v the United Kingdom* App nos 21627/93, 21826/93 and 21974/93 (ECtHR, 19 February 1997) 24 EHRR 39; *Pretty v United Kingdom* App no 2346/02 (ECtHR, 29 April 2002) 35 EHRR 1. The ECtHR referred explicitly to personal autonomy as a right in *Evans v the United Kingdom* App no 6339/05 (ECtHR, 7 March 2006) 46 EHRR 34, para 57.

[138] European Court of Human Rights, 'Guide on Article 8 of the European Convention on Human Rights Right to respect for private and family

of Europe Commissioner of Human Rights has held that '[...] personal autonomy is not about being able to do everything on your own, but about having control of your life and the ability to make decisions and have them respected by others'.[139] Personal autonomy is strongly linked with notions of human control and agency.

'Privacy rights' are laid down in Articles 7 and 8[140] EU Charter[141] and in Article 8 ECHR[142]. Privacy rights cover a broad range of human behaviours.[143] They cover a domain in societal life in which an individual enjoys a degree of autonomy, unhampered by interventions from governmental or other (private) institutions. A non-exhaustive list of examples of the private sphere is the right to relate to other people of one's own choice, the right to found a family, the sanctity of the home, freedom from unwanted physical or mental interference, and private correspondence.[144] The ECtHR has long recognized that the right to one's private life has no sharp boundaries, thereby allowing the case law to develop in line with social and technological developments.[145] The elaborate body of ECtHR case law might offer useful guidance to private controllers in their assessment of impacts on privacy rights.

For instance, a smart assistant in a domestic context could provide a controller with the automatic and continuous capture of sensor data around that assistant, through which the ADM conducts analytics, and automatically decides, based on those analytics, about the dwellers in the home (eg it senses noises to warn caregivers about the needs of their baby). This is particularly problematic for privacy rights of persons in the vicinity of that ADM who did not themselves consent to that system's data processing. Such data processing likely encroaches on privacy rights, such as one's right not to be subject to continuous business surveillance, or one's right to privacy of communication, and the sanctity of the home. Smart assistants may also have chilling effects on the dweller's freedom of speech or on their freedom of assembly.

The ECtHR has not yet established particular case law with regard to ADM. Its case law might, however, allow analogies to be built, which can be helpful in guiding controllers in detecting potential privacy breaches. In *Bărbulescu v Romania* the Court determined that an employers' continuous surveillance of employee behaviours using a full-time active camera had a strong impact on privacy rights and required the national government to take regulatory measures.[146] While the relationship between an employer and an employee is different from that of a smart assistant in a domestic context, the essence of the Court's reasoning—that the continuous surveillance of individuals by a private controller has a strong impact on privacy rights—may offer useful guidance in a situation in which the full-time data capture and storage of information in the domestic setting applies. There is certainly no hierarchy between the parties involved in the case of the smart assistant, while hierarchy exists in the relationship between employers and employees, and the purposes of the cases are different, but that does not by itself reduce the validity of the essence of the Court's point in the *domestic* context. In a domestic context, such infringements may result in more substantial infringements, given that behaviour and communication in domestic context are protected at higher levels than behaviour in the work space, which is of a more public nature.

life, home and correspondence' (updated 31 August 2018) <https://www.echr.coe.int/Documents/Guide_Art_8_ENG.pdf> accessed 5 November 2019; the EU Court of Justice publishes comparable compilations, eg 'Factsheet about the Protection of Personal Data' <https://curia.europa.eu/jcms/upload/docs/application/pdf/2018-10/fiche_thematique_-_donnees_personnelles_-_en.pdf> accessed 20 November 2019.

[139] Commissioner for Human Rights, Human Rights and Disability, 'Equal Rights for All' (20 October 2008) Strasbourg CommDH/IssuePaper, 2, para 5.2. <https://rm.coe.int/16806dabe6> accessed 11 January 2020.

[140] A right to protection of personal data in the EU context originates from the right to privacy but with the Lisbon Treaty has evolved into a specific, independent fundamental EU right. The ECtHR has recognized data protection as one of the rights deriving from the right to privacy (art 8 ECHR). See for a historical perspective and the difference between the right to data protection and the right to privacy between the ECJ and the ECtHR, see Juliane Kokott and Christoph Sobotta, 'The Distinction Between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3 International Data Privacy Law 222.

[141] The right to hold a conviction or a belief (the forum internum), which is considered here a privacy right, is protected by art 10 EU Charter.

[142] The right to hold a conviction or a belief (the forum internum), which is considered here a privacy right, is protected by art 9 ECHR.

[143] For an ethical grounding of the protection of privacy and personal data, see Jeroen van den Hoven, 'Information Technology, Privacy, and the Protection of Personal Data' in Jeroen Van den Hoven and John Weckert (eds), *Information Technology and Moral Philosophy* (CUP, Cambridge 2008) 301; see for a typology of behaviours within the scope of the right to privacy, eg Bert-Jaap Koops and others, 'A Typology of Privacy' (2017) 38 University of Pennsylvania Journal of International Law 483 and Daniel Solove, 'A Taxonomy of Privacy' (2017) 154 University of Pennsylvania Law Review 477. For an overview of behaviours as recognized by the ECtHR within the scope of privacy of art 8 ECHR in the case law of the ECtHR, European Court of Human Rights, Guide on Article 8 of the European Convention on Human Rights: Right to respect for private and family life, home and correspondence (updated 31 August 2019) <https://www.echr.coe.int/Documents/Guide_Art_8_ENG.pdf> accessed 12 January.

[144] Privacy rights in the European legal context are recognized in art 8 ECHR and art 7 EU Charter.

[145] Council of Europe/European Court of Human Rights (n 136) 17.

[146] *Bărbulescu v Romania* App no 61496/08 (ECtHR, 5 September 2017, Grand Chamber judgment).

Equal treatment rights

Article 21 EU Charter comprises a general prohibition on discrimination.[147] At a lower regulatory level, various European Directives related to specific traits (eg ethnicity, gender) and domains (labour, pensions, or access to and the supply of goods and services) exist.[148] These directives prohibit *direct* (or intentional) discrimination. *Indirect* discrimination occurs when a controller's ADM which applies to everyone has the unintended *effect* that people with a certain protected characteristic are put at a disadvantage when compared with those who do not share it. There may occasionally be justifiable grounds for indirect discrimination.[149]

The effects of indirect discrimination in targeted advertising can, for instance, emerge where coaching institutions use ADM, which predominantly targets male viewers with recruitment proposals for high salary jobs. This might be due to the data feed (income data is a good proxy of gender), to the particular weightings attached to the data (eg higher values to current income, age, or grade) or to the feedback loop in the ML (which might amplify the effects of the data feed and the weightings). A discriminatory effect might also, however, be the (hidden) intention of the controller.[150] Direct discrimination on the basis of one's gender for job recruitment is generally prohibited. Controllers should realize that the European Court of Justice and the ECtHR do not leave room for justification when a person is discriminated against on the grounds of ethnicity or gender. Moreover, the European Court of Justice is, given the EU internal market freedom of labour, particularly critical of discrimination based on nationality. Indirect discrimination on other grounds, such as age or health status, might be justified under certain strict conditions.

Dry-runs of ADM to test the data feed, the weightings and the results of analytics during deployment may contribute to the identification and prevention of discriminatory ADM; however, some 'black boxes' seem immune to scrutiny, so that discriminatory practices cannot be detected—not even by the systems designer.[151] A fundamental legal problem with 'black boxes' is the production of evidence on the side of the individual. Producing evidence of intentional discrimination is particularly cumbersome, as a controller's intentional 'masking' of the intention can be very difficult for individuals to identify.[152]

A claim of indirect discrimination might in some cases be more helpful: an individual purporting that they are disadvantaged due to ADM might demonstrate that some types of ADM lead to a disproportionate number of disadvantageous decisions for a particular group. If the individual succeeds in demonstrating this, the controller has to evidence that it is reasonable to use that particular ADM. However, invoking indirect discrimination is only useful whenever the individual is able to evince a sufficient number of cases to demonstrate the discrimination. This entails a bold burden of proof on the side of the individual, thereby potentially reducing the usefulness of 'indirect' discrimination as a basis for a complaint.

A systems developer and a DPO acting on behalf of the controller should be aware that ADM based on ML might contribute to, or even result in, discriminatory decisions due to a biased data feed and/or biased algorithms at the analytics stage. Potential discriminatory effects in, for example, credit scoring, provide an

[147] Arts 20 and 23 EU Charter comprise general equality rights and equality of men and women respectively. Art 14 ECHR [as an 'accessory' right it can only be simultaneously invoked with (an) other provision(s) of the ECHR] and Protocol 12 (this right can be invoked independently of any other ECHR provision) ECHR aims at prohibiting discrimination.

[148] See eg the earlier mentioned Directive on the equal treatment of men and women in relation to access to and the supply of goods (n 38) and see Directive on equal treatment of persons on the basis of their racial or ethnic origin, and the Gender Directive of 2004 prohibiting discrimination in the access to and the supply of goods on the basis of one's sex, Council Directive 2000/43/EC of 29 June 2000 implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, OJ L180/22. As far as the EUDirectives do not cover a particular discrimination ground or a particular sector (labour, pensions, or goods and services), the EU Charter may fill such gaps. Art 14 ECHR and Protocol 12 of the ECHR offer protection to individuals and do not address a particular domain or ground.

[149] It must be borne in mind that the ECtHR and the Court of Justice of the European Union are very strict in the use of eg 'ethnic' data or data revealing one's race; indirect discrimination on these grounds will hardly ever pass a legal test. Other so-called 'suspect grounds' are nationality, gender, religion, and sexual preference. The ECtHR executes a 'very weighty reasons test' whenever a controller would defend their position—ie using those grounds to discriminate in the offering of goods and services. See for an overview of both Council of Europe and of EU legislation FRA: *Handbook on European Non-discrimination Law* <https://fra.europa.eu/en/publication/2018/handbook-european-law-non-discrimination> accessed 11 January 2020.

[150] Example based on example found in Amit Datta, Michael Carl Tschantz and Anupam Datta, 'Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination' (18 March 2015) 3 <https://arxiv.org/pdf/1408.6491.pdf> accessed 11 January 2020.

[151] Danielle Keats Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' (2014) 1 Washington Law Review 14; but see Kroll (n 19) who argues that 'algorithms are fundamentally understandable pieces of technology, and that system creators and operators always determine that the technologies they deploy are fit for certain uses, making no system wholly inscrutable'.

[152] This is problematic from the perspective of divergence in consequences as some national legal systems render intentional discrimination (eg based on ethnicity) a crime. In the context of the Netherlands, intentional discrimination is punished with higher fines and punishments compared to unintentional discrimination (arts 137 g and 429 quater of the Dutch Penal Code).

interesting example, as credit lending can play an important role in fostering the achievement and enjoyment of other fundamental rights. Access to credit (banking, lending) is an important means of promoting social and economic well-being. It helps disadvantaged individuals to better enjoy their economic, social, and cultural rights by providing them with the means to, for example, pursue higher education, access health care, purchase a property, or start a business. Prior to the introduction of ADM in assessing a borrower's risk of non-repayment, risk assessment consisted of compilations of information about an individual's personal affairs that were subject to the discretionary review of the lender.[153]

Data for credit scores based on ADM might now be based on a combination of an individual's payment history, such as the amount owed, the age of accounts, sources of credit and how much additional credit was borrowed recently.[154] In an ADM context, this set of data might appear objective at first instance; however, people without a long data record (a 'thin file'[155]), such as minorities, young adults, immigrants, or recently divorced women may be given a credit score that is not indicative of their true risk. It might even simply deny credit.[156] Even where the controller did not intend to be discriminatory, ADMs based on that information may result in the indirect discrimination of people based on their ethnic background, gender, age, or civil status.

Apart from unfair discrimination in ML due to a certain set of data, discrimination in ML can also result from deficiencies in the quality and quantity of the data available to test and train the algorithm.[157] An awareness of the impact of ML discrimination on individuals or groups in society is visibly growing in computer science communities.[158] Meanwhile, even if all relevant information is reported to a credit agency, there is no guarantee that the machine-learned credit scoring algorithm will consider *all relevant* information for an individual and make a just decision.[159] In order to consider that relevant information, information about the context of that individual must be taken into account, from which it might transpire that a person who did not get the credit based on ADM, may receive it after considering (other) relevant information. To date, it is not possible to embed context in ML. A person with a thin file who is a creditworthy person may thus be judged according to statistical but irrelevant information, a situation which should be avoided by the controller. Meaningful human intervention at this point is indispensable (and obligatory whenever EU law applies).[160]

A worrisome use of ADM-based credit scores *beyond* the lending context may further amplify discriminatory effects, as it is—in the USA—increasingly common for employers, landlords, and insurers to review an individual's credit score before offering them a job, renting them an apartment, or selling them insurance.[161] Critics note some employers who think that credit scores are a proxy for an applicant's integrity and responsibility, even though they have not been validated for that purpose.[162] The use of ADM based on a non-representative set of personal information demonstrates that concerns about this type of ADM use may have strong negative impacts which are unevenly distributed over the population.

Procedural rights

Where ADM underlies decisions that strongly impact the lives of humans, and those decisions can be challenged in court (or in front of another arbiter that can make legally binding decisions), it is crucial to guarantee that an essential rationale to procedural rights—solving conflicts through an objective, neutral institute that is not involved in the dispute—is fully respected.

[153] Sean Trainor, 'The Long Twisted History of Your Credit Score' *Time* (22 July 2015) <http://time.com/3961676/history-credit-scores/> accessed 11 January 2020.

[154] Example borrowed from Raso and others (n 4) 27. Their approach concerns the American situation.

[155] Raso and others (n 4) 27.

[156] Kenneth P Breevoort, Phillip Grimm and Michelle Kambara, 'Data Point: Credit Invisibles', Consumer Financial Protection Bureau Office of Research (May 2015) 6 <https://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf> accessed 29 November 2019.

[157] Kamarinou, Millard and Singh (n 67) 23.

[158] It is important to distinguish between the ML concept of discrimination (about the classification of information and bias, which lie at the heart of machine learning) and unfair discrimination in the legal/societal sense that leads to prejudicial treatment of individuals or groups of individuals. Awareness is growing in the computer science community that algorithmic decisions should not create discriminatory or unjust impacts when comparing across different demographics (eg ethnicity or sexual preference). See <http://www.fatml.org/resources/principles-for-accountable-algorithms> for references and for Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.

[159] Karlijn Kuijpers, Thomas Muntz and Tim Staal, 'Tientallen Grote Nederlandse Bedrijven Bespioneren Hun Personeel' (*Investico Platform*, 1 November 2018) <https://www.platform-investico.nl/artikel/tientallen-grote-nederlandse-bedrijven-bespioneren-hun-personeel/> 12 January 2020; Raso and others (n 4) 28.

[160] Art 22 (3) GDPR.

[161] This situation was found in the USA, see Raso and others (n 4) 28.

[162] Gary Rivlin, 'Employers Pull Applicants' Credit Reports' *New York Times* (11 May 2013) <https://www.nytimes.com/2013/05/12/business/employers-pull-applicants-credit-reports.html> accessed 30 June 2019.

Procedural rights include the right of access to a court,[163] the right to legal remedy and the right to a fair trial. Other, more specific rights have been derived from these rights and recognized by the ECtHR.[164]

The ECtHR has recognized some more specific rights which add to the effectiveness of the right to a *legal remedy*. Among these specific rights are the right to access effective legal redress,[165] including the redress of wrongful acts or damages, or access to means of evidence. The right to a legal remedy partially overlaps with the right of access to a court which includes the right to an independent,[166] impartial court[167] that can make binding decisions,[168] and 'real' access to a court—meaning that there should be a court, and that access should not be disproportionately (financial, procedural) burdensome for a litigant.[169]

Once a dispute has reached a court, the right to a *fair trial* entails specific procedural requirements that must be complied with in civil, criminal, or administrative procedures.[170] These requirements include a fair and open procedure which should guarantee the equality of arms[171] of the parties involved,[172] the requirement that proceedings should be adversarial,[173] the notion that parties should have equal opportunities to involve documents (including witness and expert reports),[174] and that burdens of evidence should be fairly distributed. Parties should have reasonable opportunities to access and process relevant documents,[175] although this does not entail full disclosure of all documents, as access to some documents may be restricted due to national security, intellectual property right, copyright or trade secrets interests.[176] Such restrictions have to be proportionate, and guarantees should be in place to provide the court with access to the documents to enable it to entirely and objectively judge the case at hand. Judgments must be well reasoned and made explicable to the parties involved.[177]

ADM entails transparency challenges, which might severely reduce the effectiveness of procedural rights. Many issues are caused by high levels of opacity in many ML systems, of which three forms have been identified. Burrell has described (i) intentional opacity where the system's workings are intentionally hidden to protect business secrets or intellectual property, (ii) illiterate opacity where a system is only comprehensible to those with the technical skills to write and understand code, and (iii) intrinsic opacity, where a system's complex decision-making process is difficult to understand or even not understandable at all for any human.[178] Opacity of any form is inherently problematic to, for instance, the right to an effective remedy. Whenever ADM is used to make decisions about humans by a company, informing the data subject in a meaningful way about the use of ADM is pivotal, as a person cannot, given opacity in ADM and in data processing by companies more generally, be fully aware of the incorrect use of their data in a given context, or of unfair

[163] *Golder v United Kingdom* App no 4451/70 (European Commission of Human Rights, 21 February 1975) para 36; *Stanev v Bulgaria* App no 36760/06 (ECtHR, 17 January 2012) para 230.

[164] EU Fundamental Rights Agency, *Handbook on European Law Relating to Access to Justice* (Publications Office of the European Union, Luxembourg 2016); European Court of Human Rights, *Guide on Article 6 of the European Convention of Human Rights, Right to Fair Trial* (civil limb) updated to 30 April 2019 <https://www.echr.coe.int/Documents/Guide_Art_6_ENG.pdf> accessed 2 September 2019; European Court of Human Rights, Guide on Article 6 of the European Convention of Human Rights, Right to Fair Trial (criminal limb) updated to 30 April 2019 <https://www.echr.coe.int/Documents/Guide_Art_6_criminal_ENG.pdf> accessed 29 June 2019.

[165] *Hornsby v Greece* App no 19953/1997 (ECtHR, 19 March 1997) paras 40, 45.

[166] *Zand v Austria* App no 7050/75 (European Commission of Human Rights, 12 October 1975) para 70 (objective criteria for appointment); *Sutter v Switzerland* (European Commission of Human Rights, 1 March 1979) paras 166–75 (term of appointment).

[167] *Kyprianou v Cyprus* App no 73797/01 (ECtHR, 15 December 2005) para 119 (objective and subjective impartiality).

[168] *Benthem v Netherlands* App no 8848/80 (European Commission of Human Rights, 23 October 1985) 102; *Van de Hurk v Netherlands* App no 16034/90 (ECtHR, 19 April 1994) para 45.

[169] *Steel and Morris v United Kingdom* App no 68416/01 (ECtHR, 15 June 2004) paras 61, 62 (legal aid to make access to a court effective); *Le Compte, Van Leuven and De Meyere v Belgium* App no 6878/75 and 7238/75 (European Commission of Human Rights, 23 June 1981) para 33 (access to a court does not imply that every instance must meet the conditions of art 6 ECHR; however, the final instance (and judgement) must meet art 6 ECRM requirements to qualify as a court).

[170] The right to a fair trial in criminal procedures entails extra guarantees whenever the prosecution falls within the scope of art 6 ECHR or art 47, first sentence, of the EU Charter.

[171] In criminal proceedings, see *Neumeister v Austria* App no 1936/63 (ECtHR, 28 June 1968); *Öcalan v Turkey* App no 46221/99 (ECtHR, 12 May 2005) para 140; in civil proceedings see *Dombo Beheer v Netherlands* App no 14448/88 (ECtHR, 27 October 1993) paras 34–35; in (expropriation) proceedings by the state see *Yvon v France* App no 44962/98 (ECtHR, 24 April 2003) paras 34, 36.

[172] *Kress v France* App no 39594/98 (ECtHR, 7 June 2001); Case C-199/11 *Otis* (EUCJ, 6 November 2012) ECLI:EU:C:2012:684.

[173] *Barberà, Meddegué and Jabardo v Spain* App nos 10558/83, 10589/83, 10/590/82 (ECtHR, 6 December 1988) para 78; *Brandstetter v Austria* App nos 11170/97, 12876/87 (ECtHR, 28 August 1991) paras 66–68 (both criminal proceedings); *Vermeulen v Belgium* App no 19075/91 (ECtHR, 20 February 1996) para 33 (civil proceedings); *Ruiz-Mateos v Spain* App no 12952/87 (ECtHR, 23 June 1993) para 63 (constitutional proceedings).

[174] *Mantovanelli v France* App no 21497/93 (ECtHR, 18 March 1997) paras 33, 36.

[175] *Schuler-Zgraggen v Switzerland* App no 14518/89 (ECtHR, 24 June 1993) para 52; *Augusto v France* App no 71665/01 (ECtHR, 11 January 2007) para 50.

[176] With regard to national security, see *Rowe and Davis v United Kingdom* App no 28901/95 (ECtHR, 16 February 2000) paras 61–65.

[177] In civil proceedings see *Hiro Balani v Spain* App no 18064/91 (ECtHR, 9 December 1994) paras 27, 28; in administrative proceedings see *Helle v Finland* App no 20772/92 (ECtHR, 19 December 1997) para 60; in criminal proceedings see *Vidal v Belgium* App no 12351/86 (ECtHR, 22 April 1992) para 34.

[178] Burrel (n 19) 2.

decisions.[179] They may often even have no clue that ADM decisions have been reached.[180]

A practical use case in which parents with an Asian ethnic background receive targeted advertisements for homework assistance at a higher price compared to parents from different ethnic backgrounds demonstrates the potential consequences of opacity. Asian parents, it emerged from a controller's data mining, are on average willing to pay higher prices for homework assistance than, say, people of, for example, Caucasian ethnicity.[181] As such advertisements are individually targeted on the basis of personal traits and personal behaviour, it is very difficult for individuals—the Asian parents—to become aware of any discrimination. They are thus not in a position to effectively exercise their right to an effective remedy to combat ethnic discrimination.

The use of ADM in court proceedings can complicate one's right to access a court or to an arbiter that can make legally binding decisions. A practical example in this context is the Dutch e-Court (a private foundation), which based its 'decisions' on ADM. From 2011 to 2018, insurers and some other large enterprises agreed among themselves that dispute settlement would happen by arbitration or binding advice by e-Court.[182] To become legally binding, e-Court's 'decisions' had to be approved by appointed lower courts. The e-Court was involved in relatively simple types of dispute only. One of the appointed courts, however, could not substantially the test e-Courts' 'decisions', after which it was decided to refer the dispute, with legal questions about its legality and the court's problems, to interpret the 'decisions' to the Dutch Supreme Court.[183] The Supreme Court, however, did not reach a judgement, as the e-Court withdrew its 'decisions' from the Supreme Court.[184]

The right to a fair trial may equally suffer from opacity if ADM is used in court proceedings, as decisions will commonly not be accessible to complainants or accused persons: they may be extremely difficult (or impossible) for a human to explain. Such decisions affect the rights of equality of arms in court. These opacities lie at the foundations of the transparency issues noted above.[185] As long as ADMs prevent the effective exercise of procedural rights, particularly in criminal proceedings where the individual is in an inherently vulnerable position, ADMs should perhaps not be used for proportionate decision making until individuals can effectively exercise their procedural rights.

Freedom rights

Freedom rights entail one's freedom of expression, the right to gather information, or the right to peaceful assembly. The use of ADM entails three causes for concern with regard to freedom rights. First, freedom rights, and particularly the right to gather information, may be reduced due to a tendency to 'filter-bubbles' which in their turn may appear with the use of ADM. While the Internet was previously perceived as a motor of information diversity, powerful Internet service providers have developed into information providers (or gate keepers)—not by producing content themselves, but by using ADM to impart content produced by others to the public.[186] The content on the web is captured, analysed, and disseminated by these companies on the basis of commercial and marketing-driven incentives that are implemented in the weightings during the training of the ML process.[187] ADM based on ML techniques has now developed into 'gate keepers' of the freedom of expression and the freedom to impart information.[188] The ADMs of private controllers now determine at an unprecedented level which information is provided to citizens from the Internet. The involvement of commercially driven ADM at a global level in particular entails risks of reducing individual access to pluriform digital information.

[179] Janssen and others (n 56).

[180] Mireille Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Cheltenham UK: Edward Elgar Publishing 2015) 101.

[181] Research showed that the company (Princeton Review) used price differentiation practice which led to higher prices for people with an Asian background: 'Customers in areas with a high density of Asian residents were 1.8 times as likely to be offered higher prices, regardless of income', Julia Angwin and Jeff Larson, 'The Tiger Mom Tax: Asians are Nearly Twice as Likely to Get a Higher Price from Princeton Review' (*ProPublica*, New York, 1 September 2015) <https://www.propublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review> accessed 7 July 2019.

[182] Article 1020 of the Dutch Code of Civil Procedure (arbitration); Article 7:900 Dutch Civil Code (binding advice); see Henriëtte Nakad-Weststrate, Ton Jongbloed, Jaap van den Herik and Abdel-Badeeh M. Salem, 'Digitally Produced Judgments in Modern Court Proceedings' (2015) 4 International Journal of Digital Society 1102.

[183] *Zilveren Kruis Insurance v X* (30 January 2018) ECLI:NL:RBAMS:2018:419 (Court Amsterdam).

[184] One lower court rejected one of the 'decisions' of the e-Court in an ordinary civil dispute, as procedural mistakes were made by the e-Court. e-Court is no longer operable.

[185] Mantelero (n 55); Janssen and others (n 56).

[186] See, about recommender systems, Jennifer Cobbe and Jatinder Singh, 'Regulating Recommending: Motivations, Considerations, and Principles' (SSRN, 15 April 2019) <http://dx.doi.org/10.2139/ssrn.3371830> accessed 20 September 2019.

[187] Cobbe (n 18) 6ff.

[188] Vaira Vīķe-Freiberga and others, 'A Free and Pluralistic Media to Sustain European Democracy' (2013) Report of the EU High Level Group on Media Freedom and Pluralism, 27 <http://ec.europa.eu/information_society/media_taskforce/doc/pluralism/hlg/hlg_final_report.pdf> accessed 12 January 2020.

A second, related concern is that ADM causes the algorithmic censorship of digital expressions. The automated deletion of 'illegal' content by private entities has already had a severe impact on these rights. The deletion of content, including that of women breast feeding their children, and content delivered by the Dutch organization 'Women on Waves' about abortion on board a boat on international territory, has been blocked by some governments—and thus also by private providers.[189] Others, like photographers, have experienced the same restrictions by Facebook on their freedom of expression and artistic freedom.[190] Controllers should pay particular attention to artistic or 'opining' expressions, as the ECtHR has firmly established that the freedom of expression also protects expressions that 'shock, offend or disturb'. This norm is equally applicable in the digital environment.[191]

Private controller ADM can contribute to the serious reduction of expression rights. This might appear if controllers act on the basis of their own policies[192], or if they act according to national law, whereby that law itself can be contrary to fundamental rights as enshrined in the ECHR and in the EU Charter. As regards their own policies, and when acting under EU or ECHR jurisdiction, private controllers might instead take the EU Charter and the ECHR as legal standard in the design of their ADM.[193] Fundamental rights in the EU Charter and those in a national constitution might be applied differently (eg whenever freedom rights are applied in a more restricted manner at national level); however, adjusting ADM to correct fundamental rights protection levels in a national context might be complicated for private controllers. Guidance and clarification as to how fundamental rights would apply to private controllers by regulators would be helpful here.

A third concern is that ADM may have chilling effects, as individuals might adjust their expressions and behaviour whenever a particular artistic expression or opinion, or presence at a particular assembly or demonstration entails negative consequences for them within their online and offline communities. Individuals might instead choose to stay on the safe side, as determined by the policies of the private entities. Such adjustment in behaviour could, for instance, emerge in the context of the domestic IoT, where smart assistants may capture every word from every dweller in a home.

The ECtHR has developed rather detailed rules about restrictions on the freedom of expression that should be respected - including by private entities. Expressions that call for violence or hate speech may legitimately be refused by online private web administrators, but the ECtHR warned in *Delfi v Estonia* that filter mechanisms can ultimately lead to censorship.[194] In *Tierfabriken v Switzerland* the ECtHR found that the postal service cannot refuse to send particular advertisements because the company does not agree with its content.[195] Such an obligation is, however, not unlimited. In *UPC Telekabel Wien*, the Court of Justice determined that Internet service providers cannot be expected to prevent copyright harms through the use of filtering techniques in every situation if that reduces access to information for others.[196] The values in the data feed and the values of the weightings entail, in a European context, restrictions that can violate rights of expression at the stage of data capture and data analytics, and subsequently at the stage of the decision to disseminate particular content to particular individuals, potentially leading to violations of expressional freedoms.

Figure 1 provides examples of potential fundamental rights impacts, and a first impression of where the risks of certain contexts in which ADM is used reside. Potentially low impacts are on the left side of the spectrum, while higher impacts are on the right side. To achieve a proper establishment of impacts, a controller has, in order to be compliant with Article 35 (7) GDPR, to describe the use case in which the ADM should serve to achieve its purpose with as much detail as possible. The purposes must also be specified. Controllers should thereby bear in mind that high level descriptions might result in conclusions that there is no (high) risk. It is

[189] See the example of breast feeding being banned by Facebook (2017) <https://www.womenonwaves.org/en/page/4599/women-on-web-website-is-blocked> accessed 11 January 2020.

[190] See about the picture of the naked Vietnamese girl covered with napalm <https://en.wikipedia.org/wiki/Phan_Thi_Kim_Phuc> providing an example of a picture that was deleted by Facebook <https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo> accessed 11 January 2020.

[191] *Sunday Times v United Kingdom* App no 6538/74 (ECtHR, 26 April 1979) 2 EHRR 245 para 65 and *Delfi AS v Estonia* App no 64569/09 (ECtHR, 16 June 2015, Grand Chamber judgment) [2015] ECHR 586.

[192] Jack Balkin, 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation' (2017) 51 University of California, Davis 1182 <https://lawreview.law.ucdavis.edu/issues/51/3/Essays/51-3_Balkin.pdf> accessed 11 January 2020.

[193] Note however that the ECHR might not have direct effect in every jurisdiction; see eg the UK or Germany, where national law prevails over international law such as the ECHR.

[194] *Delfi AS v. Estonia* App no 64569/09 (GC) (ECtHR, 16 June 2015); *Sunday Times v United Kingdom* App no 6538/74 (ECtHR, 26 April 1979) 2 EHRR 245 para 65. The Court however acknowledges that such filter mechanisms may sometimes be the only solution to protect individuals and groups against hate speech or terrorist content. As regards the freedom to impart information, there is no such clear case law at the moment.

[195] *Verein gegen Tierfabriken v Switzerland* App no 48703/08 (20 September 2011, admissibility decicion) 34 EHRR 4.

[196] Case C-314/12 *UPC Telekabel Wien* [2014] ECLI:EU:C:2014:192.

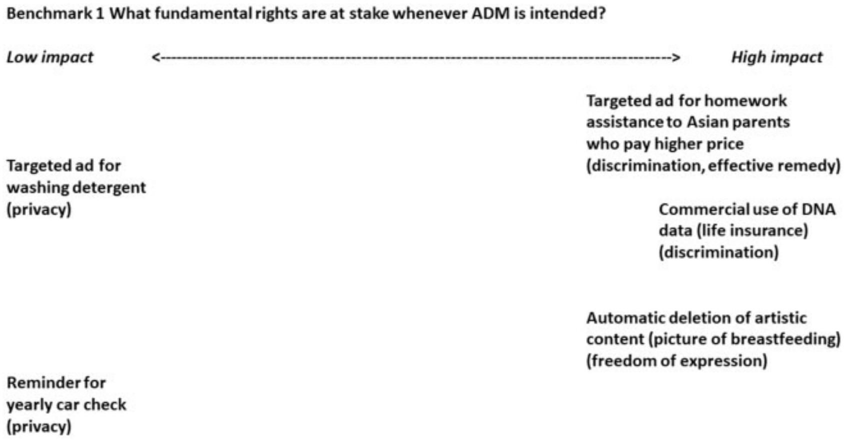

Benchmark 1 What fundamental rights are at stake whenever ADM is intended?
Low impact
<--->
High impact
Targeted ad for homework assistance to Asian parents who pay higher price (discrimination, effective remedy)
Targeted ad for washing detergent (privacy)
Commercial use of DNA data (life insurance) (discrimination)
Automatic deletion of artistic content (picture of breastfeeding) (freedom of expression)
Reminder for yearly car check (privacy)

Figure 1: Benchmark 1: What fundamental rights are at stake whenever ADM is intended?

important that potential impacts are considered at both granular, individual levels, and at group levels.[197]

While little case law exists where the uses of ADM were judged as regards fundamental rights compliance, analogies might be found or be developed from existing case law (where the use of certain technologies was said to impact a person's rights). Such analogies have to be explained in the impact assessment. Supervisory authorities, National Human Rights Commissioners, or non-governmental organizations (NGOs) might assist private controllers by identifying (and by building a body of) relevant case law, and provide other useful guidance and explanatory materials to inform and guide private controllers and DPOs in their fundamental rights impact assessments whenever they envisageADM in a particular context. The consultation of these bodies might be particularly appropriate where ADM relates to 'public character' goods and services (eg if it impacts access to health care, credit lending, access to housing, or access to education).

Benchmark 2: risks at design stages and during operation

The design, development, and implementation of ADM must not unduly hinder a person's fundamental rights. However, the risks of perpetuating bias, decision-making patterns impacting one's enjoyment of fundamental rights, and challenges related to interpretability and complexity of systems, are very difficult to address in ADM. Beyond complexity, the validity of a system itself must be considered, as correlations might be strong. For an ADM to function appropriately, Article 5 (1)(b) requires controllers to formulate robust purpose specifications, while 5(1)(d) GDPR requires controllers to consider the accuracy of data at stages of design, development, and deployment. If the data used in an ADM is inaccurate, any resultant decision or profile will be flawed. Inaccuracies could potentially lead to inappropriate predictions or decisions about, for example, someone's health, credit, or insurance risk. Even if raw data is recorded accurately, the data set may not be fully representative, or the analytics may contain hidden bias. To prevent this, controllers should introduce robust measures to verify and continually ensure that data that is reused, or obtained indirectly, is accurate and up to date.[198]

Impact assessments apply at design and *operation stages*

The second benchmark should involve assessing and testing the design of the algorithms used, the data feeds, and the weightings attached to the data. 'Dry-runs' may be required, and aspects of the system adjusted to make the processing GDPR and fundamental rights compliant.[199] Eventual risks during dry-runs should be mitigated with appropriate technical or organizational measures; controllers are obliged to do so under Articles 24 and 25. Article 32 GDPR requires them to take measures to safeguard the security of personal data used. The risks of ADM may be dynamic, however: ADM changes as it processes data. *Ex ante* controls built in at design stages should therefore be complemented with the continuous monitoring of risks and with an appropriate risk mitigation strategy. For example, in job vacancy scoring, any risk of discrimination should be reassessed with each

[197] See Benchmark 2.

[198] For instance, in a credit lending situation, a Machine Learning model might learn that a borrower's quality of clothing correlates with their income and thus with their creditworthiness. However, a credit model based on the borrower's clothing (or based on any other irrelevant aspect such as civil status or ethnic background) should be rejected during ex ante evaluation of the system (at dry-running stages), example borrowed from Kroll (n 19) 3.

[199] In this sense Kramer, Guillory and Hancock (n 5) 8788.

set of new applications, as the demographics may have changed. Designers should be aware that ADM involves dynamic processing; it should therefore be constantly reviewed [Article 35(11) GDPR]. This should be done through rigorous, continuous logs and audits of the processing in the ADM system. Various other technical measures, such as that around data and decision provenance, can help controllers to determine whether the use of ADM is—and remains—acceptable.[200] 'Decision provenance' entails using provenance methods to provide information exposing decision pipelines: it presents chains of inputs to, the nature of, and the flow-on effects from, the decisions and actions taken (at design and run-time stages) throughout ADM.[201]

Special category data

The GDPR discerns 'special categories' of data [Article 9 (1) GDPR] from 'personal data' [Article 4 (1) GDPR]. Special category data comprise data that reveal one's ethnic origin, genetic data, or one's sexual orientation, which, given the high risk of discriminatory application,[202] merit higher legal protection.[203] Processing special categories data is in principle prohibited, unless appropriate safety measures are in place. Their use can, in the context of processing by private controllers, be based on explicit consent only.[204] The automatic identification of special category data, however, has its challenges. While some data will clearly fall within a special category, there will often be situations where seemingly innocuous data can reveal information that falls into special categories. For instance, when data acts as a proxy (eg where a postcode might imply race[205]); or where certain benign data can be combined with other data to become particularly revealing.[206] It has thus been argued that in practice all personal data might, in effect, become a special category.[207] This is relevant in an ADM context where data will be captured automatically, which would mean that a prudent private controller using ADM should treat all data captured as a special category. Indeed, ADM needs a great wealth of personal data; where capture and analytics happen without the private controller necessarily 'knowing' the nature of all data and the associated contexts involved.[208]

The establishment of the weightings given to data and the inferences from analytics in ADM may pose risks to the fundamental rights of individuals, such as the right not to be discriminated against (targeted advertisements based on one's ethnicity), to privacy rights (whenever ADM is used in sensing technologies in the intimate domestic sphere and targets, for example, voice sensing in the bedroom) or in expressional rights (where, for example, artistic expressions with a naked breast are being denied based on controller policy). The weightings that are adjusted by the systems designer so as to lead to the desired outcomes from the data analytics represent a controller's values. The weightings are tested in manifold feed-back loops. Failures should be noted and be dealt with in a timely manner by the systems designer, and could entail, for example, adjustment of the weightings or to adjustment of the desired outcome to prevent fundamental rights impacts.[209]

Accuracy of inferences

The accuracy of inferences is, besides involving the accuracy of data, a crucial consideration at this stage, as consequences can have a strong impact on one's enjoyment of fundamental rights. Even where a system can derive information with 95 per cent accuracy, this means that at least 5 per cent of every 100 decisions will involve inferred or predicted inaccurate information on which the decision may, in part, be based.[210] A decision-making system that is claimed to be 95 per cent accurate may still have a false positive rate of over one-third.[211] Where inferences are

[200] Jatinder Singh, Jennifer Cobbe and Chris Norval, 'Decision Provenance: Harnessing Data Flow for Accountable Systems' (2019) 7 IEEE Access 6562ff.

[201] ibid 6562.

[202] Recital 71 GDPR.

[203] Art 9 (1) GDPR: 'Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.' See for its relationship with fundamental rights Recital 51, explicating 'special categories' of personal data: 'Personal data which are, by their nature, particularly sensitive *in relation to fundamental rights and freedoms* merit specific protection as the context of their processing could create significant risks to the fundamental rights and freedoms' (emphasis added).

[204] Art 9 (2)(a) GDPR. Another legitimate ground relevant in horizontal context might be art 9 (2) (e) allowing personal data which are manifestly made public by the data subject.

[205] A Datta and others, 'Proxy Discrimination in Data Driven Systems: Theory and Experiments with Machine Learnt Programs' (25 July 2017) <https://arxiv.org/pdf/1707.08120.pdf> accessed 11 January 2020.

[206] See, for example, Müge Fazlioglu, 'Beyond the Nature of Data: Obstacles to Protecting Sensitive Information in the European Union and the United States' (2019) 46 Fordham Urban Law Journal 271.

[207] ibid 295.

[208] Recital 71 GDPR contains high-level references to the use of appropriate mathematical or statistical procedures for profiling, and to technical and organizational measures appropriate to ensure that inaccuracies are corrected and the risk of errors is minimized, but it does not prevent the collection of data that produces such proxies.

[209] Cobbe (n 18) 5.

[210] ibid 26.

[211] D Colquhoun, 'An Investigation of the False Discovery Rate and the Misinterpretation of P-values' (2014) 1 Royal Society Open Science 140216, 1 <https://doi.org/10.1098/rsos.140216> accessed 12 January 2020.

based on personal data, controllers have to ensure that the data (in this case inferences drawn in the process of ADM) is accurate.[212]

Article 35(7)(d) GDPR requires controllers to scrutinize potential risks at various levels, calling for fundamental rights compliant processing in ADM. To achieve appropriate risk assessments, this benchmark provides private controllers with an opportunity to test and distinguish the impacts of their AI system at various levels of granularity. Impacts can be considered at higher, more generalized levels (at societal level), or at more granular, individual levels (at individual level). This 'appropriate level', however, needs to be defined. The response to a benchmark might perhaps establish an 'acceptable risk' assessment for a particular group of individuals, while the same assessment may oversee unacceptable risks to other individuals. Whenever a credit scoring system does not affect, for example, 98 per cent of potential credit applicants, but outlaws 2 per cent of the people with a so-called 'thin file', this would not be considered a fair ADM. Data processing for ADMs, however, has to be compliant with fundamental rights standards of *all* individuals who could potentially be affected by the ADM.

Whenever private controllers are using ADMs in a context that concerns access to or the supply of goods and services with a public character, the need to include granular testing levels becomes even more urgent, as each of these types of goods entails a broad range of fundamental rights contributing to better chances in society that should be accessible for all. As ADM systems are sociotechnical in nature, the second benchmark offers scope for considering whether challenges or mitigations relate to the technology, workflows, business processes, or business models.

Tools to achieve better compliance

A useful approach to test and dry-run design might be the use of a 'regulatory sandbox', a tool to assist controllers in developing their new ADM systems in a safe and more informed manner, while enabling regulators (supervisory authorities) to remain up-to-date with technological trends. Some national governments have already created such tools for fintech applications, allowing companies to test software projects during a limited period of time, in real-life situations in a closed environment (designed to safe experiments), and to be able to anticipate licensing new products.[213] The Information Commissioner in the UK has carried out a public survey for evidence and to collect initial views on creating a regulatory sandbox.[214]

National Human Rights commissioners might, for instance, contribute to the awareness of controllers by interrogating contexts or aspects in data processing where potential fundamental rights risks might apply. The use of a regulatory sandbox may be strongly recommended whenever the purpose of the ADM encompasses a good or a service with a 'public character', for example, whenever access to health care service is at stake, credit lending is involved or access to education is an envisaged purpose of ADM. Future research should contribute to the establishment of how risks will manifest themselves at early stages, and how effective audits or other mechanisms can be developed that help to early detect risks during the deployment of the models.

Benchmark 3: balancing risks and interests

Article 5 (1)(a) GDPR requires private controllers to process personal data in a fair and transparent manner, and to inform data subjects about their objectives and means[215], which should contribute to a more equitable use of their personal data in ADM; however, the share of the 'digital dividend' between private controllers and data subjects has become more uneven than ever. [216] Dominant platforms are able to discriminate by combining knowledge they extract from data as a monopoly, and by vertical integration in the markets. Unfair and deceptive practices have arisen.[217] As noted earlier, these asymmetries have become systemic in digital ecosystems, so that compliance with fairness and transparency requirements is relatively limited.[218]

Article 35 (7)(b) interrogates a private controller as to why ADM is necessary to achieve their economic and commercial purposes, and why, if risks to fundamental rights may emerge, other, less intrusive means cannot be used to achieve their purposes. Although commercial and economic interests are legitimate interests, these purposes should not override the fundamental rights of

[212] Art 5 (1)(d) GDPR.

[213] For the UK example, see Financial Conduct Authority, 'Regulatory Sandbox' (November 2015) <https://www.fca.org.uk/publication/research/regulatory-sandbox.pdf> accessed 12 January 2020.

[214] Information Commissioner's Office, 'Technology Strategy for 2018-2021', (2018) 8 <https://ico.org.uk/media/about-the-ico/documents/2258299/ico-technology-strategy-2018-2021.pdf> accessed 11 January 2020; 'ICO Call for Views on Creating a Regulatory Sandbox' <https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-call-for-views-on-creating-aregulatory-sandbox/>; 'ICO's Summary of Responses and ICO Comment' <https://ico.org.uk/media/about-the-ico/consultations/2260322/201811-sandbox-call-for-viewsanalysis.pdf> accessed 10 August 2019.

[215] Arts 13, 14 GDPR.

[216] European Data Protection Board (EDPB), 'Opinion 8/2016 on Coherent Enforcement of Fundamental Rights in the Age of Big Data' (23 September 2016) 15.

[217] ibid 13.

[218] Janssen and others (n 56) 2-3.

data subjects and other persons concerned, unless a private controller can justify why this is proportionate.

Fairness is not explicated in the Recitals to the GDPR. Its foundations can be found in the *travaux préparatoires* to the EU Directive 95/46, stating that 'a fair and lawful processing of personal data presupposes that the data subject makes his decision whether or not to disclose data related to him to the collector on a reliable factual basis as regards the purpose of the processing, the identity of the controller and the question whether he is under legal obligation to disclose the data or whether disclosure is voluntary'.[219] Controllers should thus take due account of the rights and reasonable expectations of data subjects, and should not relentlessly and unobtrusively override them to serve their business model. The principle of fairness in the context of GDPR implies that several interests, such as controller and data subject interests, compete for preponderance, the position of which should be determined in a fair balance. Under the GDPR, the controller is responsible for striking a fair balance between the interests involved.

In ADM, the concept of 'balancing' can, from a data subject perspective, be perceived as problematic, as controller subjectivity in a 'balancing' exercise might result from the unfair definition, and the weights and inequality of the interests might further amplify unfair outcomes. The weighing of interests is rarely only two-dimensional; rather, a multi-level dimensional weighing of interests might apply, involving a more complex calibration of interests such as, in addition to fundamental rights and commercial interests, market position interests, monetizing interests (by sharing personal data with third parties, see below) or computational interests. The interests at stake may be both abstract interests and concrete interests, which potentially leads to a difference in weight and abstraction, and subsequently to challenges in comparability.[220] Private controllers should be aware of these risk factors when calibrating interests at stake. More detailed questions in this benchmark may help controllers to balance the interests at stake, and contribute to achieving more equitable and fair data processing in an ADM context.

A practical example explains how the assessment risks and benefits of ADM can operationalize. A doctor in a private hospital processes a patient's ethnic data to establish a diagnosis and subsequent treatment for the benefit of a patient. Some diseases strongly or almost exclusively correlate with ethnicity, or with DNA data.[221] The use of ADM in this context is likely to appoint the data subject as a beneficiary of the data processing. Whenever data subjects benefit, this might imply a low impact on their fundamental rights, and improve a person's access to health care. In other words, the use of ethnic data, which is perceived as sensitive under Article 9 (1)(a) GDPR and more generally in relation to equal treatment rights, might in this specific context be less problematic, as the data subject is beneficiary.

This will change whenever a patient's ethnic data is processed in ADM for purely commercial purposes; the use of this type of data is likely to result in a high risk to that person's right not to be discriminated against. Controllers might argue that individuals use their services for free, and that individuals 'pay' for the use of AI-based services by sharing their data. Those controllers might however forget that values *other* than only the financial value of their personal data may be at stake. From a fundamental rights value perspective, a history of browsing data or geolocation data can be of less value in one context, but highly valuable in another. For instance, the use of browsing data to predict one's future health in an insurance context would likely be burdensome, while pattern detection based on that data to achieve better billing methods from a service provider might perhaps be less intrusive.

Sharing the results of ADM with third parties

Sharing data used in, or the results of ADM with third parties is often part of private controller business models.[222] Given that ADM usually involves rich data sets,

[219] COM(90) 314 final COM(90) 314 final – SYN 287, 13 September 1990, 27 at <https://resources.law.cam.ac.uk/cipil/travaux/data_protection/4%2013%20September%201990%20Proposal.pdf> accessed 11 January 2020; the legal literature about fairness is currently growing, see eg Damian Clifford and Jef Ausloos, 'Data Protection and the Role of Fairness' (2018) 37 Yearbook of European Law 130; Peggy Valcke, Inge Graef and Damian Clifford, 'Fairness – Constructing fairness in IT (and other areas of) law through intra- and interdisciplinarity' (2018) 34 Computer Law & Security Review 707; Friso Bostoen, 'Neutrality, Fairness or freedom? Principles for Platform Regulation' (2018) 7 Internet Policy Review 1.

[220] See about the concept of balancing eg Janneke Gerards, *Belangenafweging bij Rechterlijke Toetsing Aan Fundamentele Rechten* (Alphen aan den Rijn, Kluwer 2006). As regards the concept of 'balancing' in case law of the EU Court of Justice, see Bart van der Sloot, 'The Practical and Theoretical Problems with 'Balancing': *Delfi*, *Coty* And the Redundancy of the Human Rights Framework' (2016) 23 Maastricht Journal of European and Comparative Law 439.

[221] This is the case with, for instance, sickle cell disease, most often (but not exclusively) found in people from African or African-Caribbean background (see <https://www.nice.org.uk/guidance/qs58/chapter/introduction> accessed 11 January 2020). People with a Mediterranean, South Asian, Southeast Asian, and Middle Eastern background are more often diagnosed with Thalassaemia than others—and may need diagnoses and treatment based on their ethnicity (<https://www.nhs.uk/conditions/thalassaemia/> accessed 12 July 2018).

[222] Mantelero (n 55) 24.

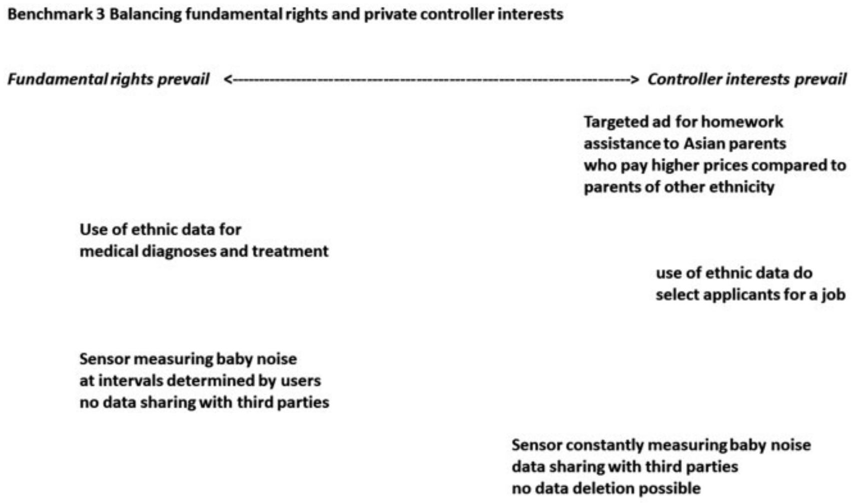

Benchmark 3 Balancing fundamental rights and private controller interests
Fundamental rights prevail <--> Controller interests prevail
Targeted ad for homework assistance to Asian parents who pay higher prices compared to parents of other ethnicity
Use of ethnic data for medical diagnoses and treatment
use of ethnic data do select applicants for a job
Sensor measuring baby noise at intervals determined by users no data sharing with third parties
Sensor constantly measuring baby noise data sharing with third parties no data deletion possible

Figure 2: Benchmark 3: Balancing fundamental rights and controller interests.

private controllers may have an interest in monetizing those data sets as well as the results from ADM itself. Third parties may be interested in buying these results, as computations at scale give them valuable insights they can use to refine their own business models.[223] New inferences based on ADM might, however, create new threats to the fundamental rights of the data subjects involved, and may require new impact assessments under Article 35 (11) GDPR. Risks may particularly arise from the fact that once data moves beyond a private controller's control (when it moves to a third party) then its usage is essentially out of sight for the data subjects (and of the controllers who shared the data). Such risks could, for example, include unexpected or undesired inferences and profiling which could subsequently result in discriminatory uses, or where the data could be used to influence, nudge or manipulate data subjects.[224] Private controllers sharing personal data and/or the results of ADM with third parties are likely to be involved with high risks to the fundamental rights of data subjects. Where such sharing applies, controllers will have to explicitly involve this in their balancing of risks and interests.

Figure 2 demonstrates that in a commercial context one party is commonly assigned as sole beneficiary in a particular context, while middle positions may, depending on the controller's motivation, be better placed. Whenever the benefit is relatively fairly distributed between controller and individual, the use of personal data in that ADM might, depending on the controller's motivation and the measures taken to protect individuals against risk, become a (more) acceptable option. Again, controllers have to be alert, as it is not unusual that weightings have to be adjusted during the life cycle of the processing, as the feed-back loops may affect the weightings over time. As a consequence, the establishment of who is benefitting from ADM should be logged and audited, and where necessary be regularly adjusted. Benchmark 3 encourages controllers to both consider and explain why their use of AI constitutes a fair trade-off in relation to a person's enjoyment of fundamental rights, or a restriction thereof.

Benchmark 4: control and agency over the data processing

Data subjects generally have little control over what, how, why, and by whom data is captured, analysed, transferred, stored, or otherwise used.[225] This benchmark, which interrogates controllers about who is in control of the data processing in ADM, seeks to identify the extent to which a data subject can mitigate and therefore exert influence over data processing in ADM. If a data subject can intervene or stop the continuous

[223] Aantti Eskola and others, 'MyData, A Nordic Model for Human Centred Personal Data Management' (2015) 4 <http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf?sequence=1&isAllowed=y> accessed 23 November 2019.

[224] Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) Columbia Business Law Review 1.

[225] Peter Tolmie and others, 'This Has To Be the Cats - Personal Data Legibility In Networked Sensing Systems' (2016) in Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing 490 <http://www.cs.nott.ac.uk/~pszaxc/work/CSCW_1_2016.pdf> accessed 11 January 2020; Tim Berners Lee, 'One Small Step for the Web. . .' open letter by Tim Berners Lee (23 October 2018) <https://inrupt.com/blog/one-small-step-for-the-web> accessed 20 October 2019.

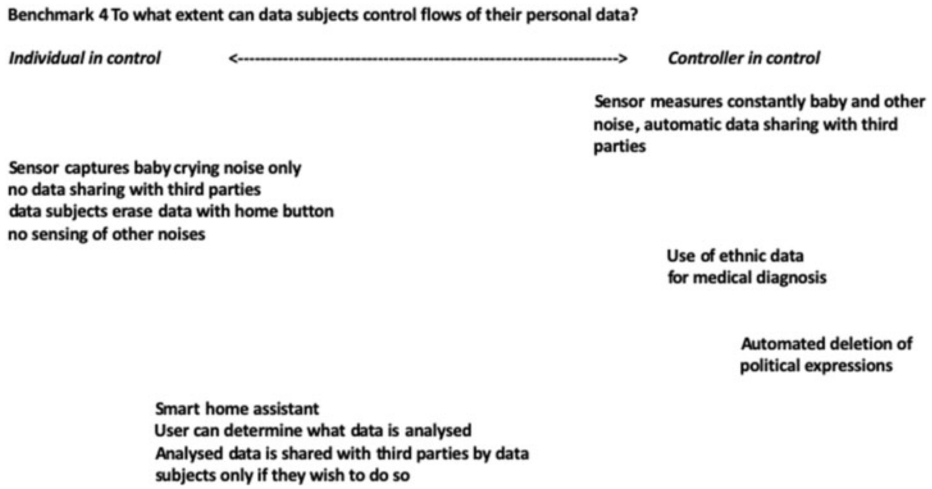

Benchmark 4 To what extent can data subjects control flows of their personal data?
Individual in control
<--->
Controller in control
Sensor measures constantly baby and other noise, automatic data sharing with third parties
Sensor captures baby crying noise only
no data sharing with third parties
data subjects erase data with home button
no sensing of other noises
Use of ethnic data for medical diagnosis
Automated deletion of political expressions
Smart home assistant
User can determine what data is analysed
Analysed data is shared with third parties by data subjects only if they wish to do so

Figure 3: Benchmark 4: Who controls the data flow?

data capture or sharing of data with third parties, then such interference might indicate a lower risk level compared to a situation in which no such interference is possible.

Some controllers might themselves consider processing personal data on a more incidental basis, but they might also give data subjects opportunities to adjust the data capture or data sharing. This might apply whenever a device, for instance a baby monitor, captures sensor data, thereby solely recognizing and capturing baby noise. The device might even allow caregivers to delete all the data at a caregiver's press of the home button of that device. The sensor might be adjusted so as to not capture personal data from other dwellers in the house. If no data is shared with third parties, the use case might become classified in the more fundamental rights-friendly side of the spectrum of this benchmark. On the other side of the spectrum, controllers behind smart assistants may commonly use extremely frequent or even continuous personal data capture for their ADMs, with no control for the caregiver over data erasure or sharing with third parties.

Figure 3 presents various situations, thereby distinguishing between situations in which individuals are not exercising any control over the processing of their personal data on the left, and situations in which the controller is (partially) in control of the processing. As this article presents an approach at this stage, rather than a detailed questionnaire, work needs to be done to enable controllers to document their impact assessment on this point. These questions must be consistent with the approach developed in this study. Detailing the benchmarks should be the topic of future research.

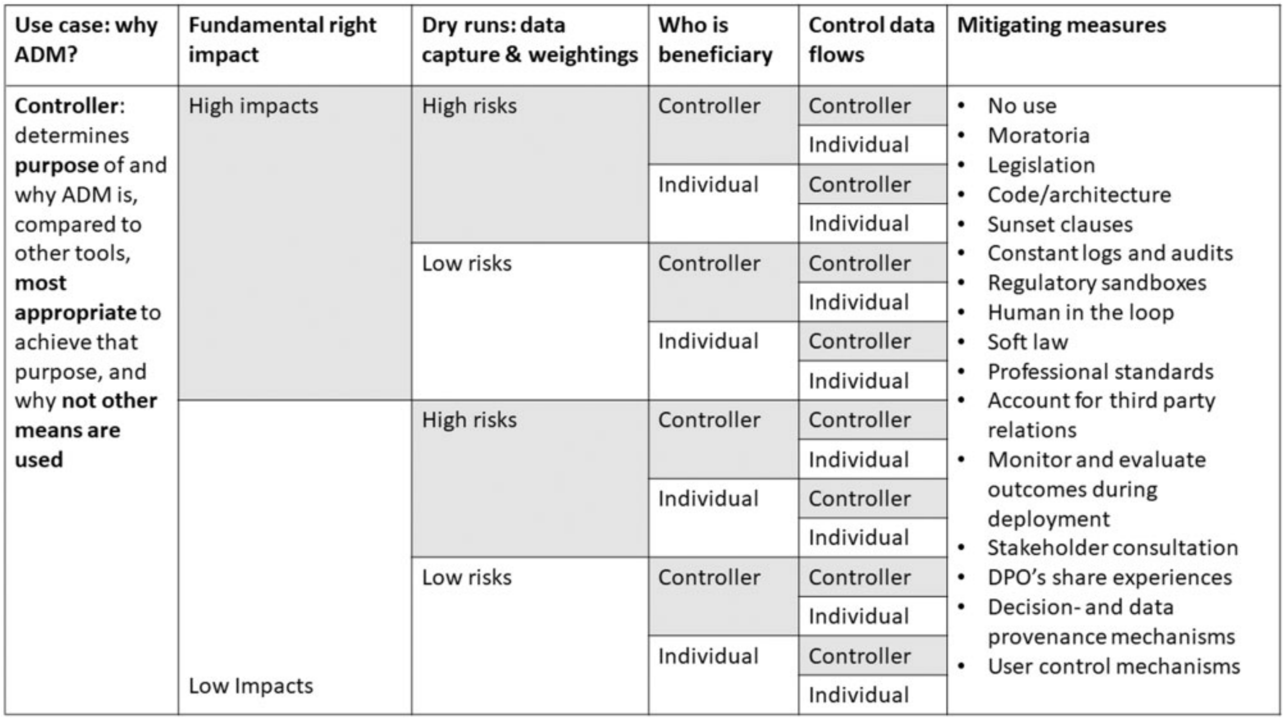
Finding appropriate levels of protection

The main objective of the proposed approach is to help companies with a practical fundamental rights impact assessment which is grounded in the DPIA as enshrined in Article 35 (1), (3) and (7) (a)–(d) GDPR, and which is based on European fundamental rights legislation (EU Charter and ECHR) at stages where they consider and explore the use of ADM. Four benchmarks were presented to achieve such impact assessment. To repeat, these benchmarks interrogate controllers about: (i) the fundamental rights that were considered impacted; (ii) to identify risks occurring in their ADM systems at design stages and during operation; (iii) how the risks and interests were balanced; and (iv) whether data subjects can exercise any control over the data processing in the ADM.

If a controller responds to the four benchmarks, an overall risk assessment of fundamental rights impacts will emerge. That risk assessment will indicate certain risk levels. As stated earlier, the responses will not produce a mathematical outcome, but instead indicate where a particular ADM will sit in a spectrum of risk levels between low risk and (very) high risk. These risk levels indicate technical and organizational measures that a controller should take in order to mitigate the identified risks.

Integrating the benchmarks in a comprehensive impact assessment

In order to complete the approach of the fundamental rights impact assessment, the final responses to the four benchmarks in an individual ADM are now merged and presented in a comprehensive impact assessment (Figure 4):



Use case: why ADM?	Fundamental right impact	Dry runs: data capture & weightings	Who is beneficiary	Control data flows	Mitigating measures
Controller: determines **purpose** of and why ADM is, compared to other tools, **most appropriate** to achieve that purpose, and why **not other means are used**	High impacts	High risks	Controller	Controller	• No use • Moratoria • Legislation • Code/architecture • Sunset clauses • Constant logs and audits • Regulatory sandboxes • Human in the loop • Soft law • Professional standards • Account for third party relations • Monitor and evaluate outcomes during deployment • Stakeholder consultation • DPO's share experiences • Decision- and data provenance mechanisms • User control mechanisms
				Individual	
			Individual	Controller	
				Individual	
		Low risks	Controller	Controller	
				Individual	
			Individual	Controller	
				Individual	
	Low Impacts	High risks	Controller	Controller	
				Individual	
			Individual	Controller	
				Individual	
		Low risks	Controller	Controller	
				Individual	
			Individual	Controller	
				Individual	

Figure 4: Integration of benchmarks in the overall approach.

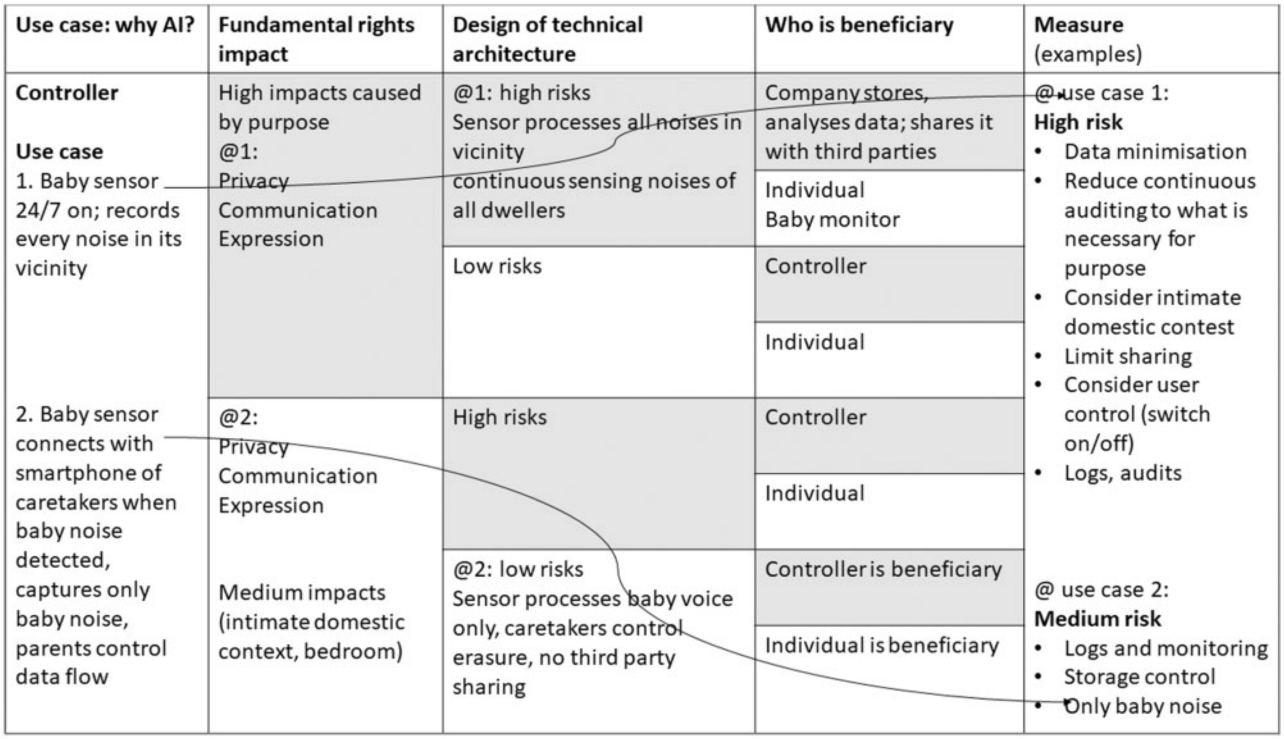
Controllers should respond to the benchmarks as follows:

- In the first column, the controller systematically describes the specific purpose(s) of the processing, including the legitimate interests pursued by the controller [Article 35(7)(a) GDPR] and why the use of ADM cannot be achieved by other, less intrusive means (proportionality requirement, Article 35(7)(b) GDPR).
- In the second column (Benchmark 1), the controller indicates whether the intended ADM may entails any impact on fundamental rights [as required by Article 35(7)(c) GDPR].
- In the third column (Benchmark 2), the controller describes the envisaged processing operations and the specific context in which they are used [Article 35(7)(a) GDPR]. This includes, at least, a complete description of the specific purpose for the use of personal data in the ADM system, a specification of the types of data, the weightings attached to that data, whether and how feedback loops were adjusted during dry runs to comply with the GDPR and fundamental rights; the accuracy achieved and whether that accuracy is proportional to potential fundamental rights risks; whether there is meaningful human intervention [as required by Article 22 (3) GDPR], and whether personal data or the results of analytics are shared with third parties.
- In Benchmark 3, which is presented in the fourth column, the controller describes how the interests were weighed and calibrated, and how the use of ADM is proportionate in relation to (potential) impacts on a person's—or a group of persons'—fundamental rights.
- In the fifth column (Benchmark 4), the controller describes the extent to which an individual has options to control data processing. Individuals might be enabled to opt for a method of data processing that is less invasive of their fundamental rights (eg to opt-out from data sharing with third parties, to choose fewer points of data capture, options to erase data feeds after services have been delivered, or to opt-out from data-sharing practices with third parties).

Determining identified risks as risk levels

Once controllers have responded to the benchmarks, their ADM should be positioned in the spectrum between no risk and high risk. In the context of this approach, risks have not been translated into numbers or other, sharp delineations. Future research might however consider whether a reliable, sharp delineation between risk levels can be realized.

Figure 4 indicates that where controller responses to the benchmarks indicate high risk, the overall risk will probably be assessed as high risk as well. Conversely,



Use case: why AI?	Fundamental rights impact	Design of technical architecture	Who is beneficiary	Measure (examples)
Controller **Use case** 1. Baby sensor 24/7 on; records every noise in its vicinity	High impacts caused by purpose @1: Privacy Communication Expression	@1: high risks Sensor processes all noises in vicinity continuous sensing noises of all dwellers	Company stores, analyses data; shares it with third parties	@ use case 1: **High risk** • Data minimisation • Reduce continuous auditing to what is necessary for purpose • Consider intimate domestic contest • Limit sharing • Consider user control (switch on/off) • Logs, audits
			Individual Baby monitor	
		Low risks	Controller	
			Individual	
2. Baby sensor connects with smartphone of caretakers when baby noise detected, captures only baby noise, parents control data flow	@2: Privacy Communication Expression	High risks	Controller	
			Individual	
	Medium impacts (intimate domestic context, bedroom)	@2: low risks Sensor processes baby voice only, caretakers control erasure, no third party sharing	Controller is beneficiary	@ use case 2: **Medium risk** • Logs and monitoring • Storage control • Only baby noise
			Individual is beneficiary	

Figure 5: Examples of two practical use cases in the integrated overall approach.

lower risk impacts might instead indicate the lower overall risk of that ADM. It might however be that a controller assesses their ADM overall as high risk, except for one benchmark, which scores low risk. This might apply in the earlier mentioned situation, where a doctor uses ethnic patient data to be able to diagnose (or to exclude) particular diseases. While the use of ethnic data would likely score very high risk in the first, second and fourth benchmark, it might remain at low level in the third benchmark (the balancing of rights and interests). The comprehensive risk assessment might, depending on the controller's compliance with other GDPR requirements, instead result in lower to mid-level fundamental rights impacts.

Application of the benchmarks in a scenario

The way that risk levels may be inferred from controller responses to benchmarks, based on this research, might be clarified by taking two practical use cases. These are presented in Figure 5.

Figure 5 presents a scenario where a baby monitor senses baby noise to warn its caregivers in a domestic context. We consider two situations. In situation (1), the full-time functioning monitor captures every noise in the home, including that of other dwellers. The data controller provides the monitors for free, while the controller also transfers and stores the data from the domestic environment to their own servers, uses it for data analytics to improve their service, and shares it with third parties. In situation (2), caregivers buy the monitor from Company A, and pay Company B for the monitoring. The monitor works to solely sense and process baby noise (which occurs on-device, without transmission). Caregivers can operate the home button to stop the device 'listening' once warnings are received.

Regarding *Benchmark 1*, given that the monitor sensor captures every noise in its vicinity in situation (1), it probably has a high impact on a number of privacy rights, including that of intimacy of the home, communication privacy and chilling effects on the freedom of speech of (other) dwellers in the home. This seems different in situation (2), where the monitor only registers baby noise, which is detected by sensing sound and then performing local (on-device) processing to determine whether it appears to be a baby; the same rights may be impacted but these impacts potentially appear at lower risk levels. Situation (1) implies high-level impacts, while (2) entails low-risk impacts on rights.

In *Benchmark 2*, the controller in situation (1) captures data (sounds) from the environment, including sounds of other dwellers. Its ADM seeks to identify and predict potential stress levels around the baby which might cause it to start making noises. The caregivers are automatically notified even before the baby starts to cry, which might prevent high-stress levels in the baby. The data from other dwellers is therefore regarded important, as it helps to indicate the situation of the baby. Although the controller might claim that outcomes of their ADM are accurate, they do not state whether they find the accuracy in balance with the identified fundamental rights risks. Other than accuracy testing, there is no human intervention or involvement. The data from the household (a domestic setting and therefore including personal data) is sent to the controller, and the data and results of analytics are shared with third parties.

Controller (2) has set up a different system. The device, although capturing domestic sound, uses on-device ADM only to identify baby noise, and where appropriate, it raises alerts. Beyond the very short period of time for analysis (probably seconds), no other noises are stored on-device, as they are deleted immediately after notification to the caregivers. No transfer of data occurs. Caregivers can stop the data processing by pressing the home button on their device. Data is not shared with third parties. Controller (2) is aware of potential divergences in its ADM system once it is utilised. It follows that this approach would be considered low risk under Benchmark 2.

In *Benchmark 3*, the controller in situation (1) will gain revenue from the processing of data that has an impact on rights, while caregivers profit from free services. The controller does not, however, appear to have balanced the necessity of the processing against rights, nor how their commercial interest overrides the fundamental rights of the caregivers and other persons involved. This would qualify their ADM processing as very high risk. It is rather unlikely that the supervisory authority will allow their system. The ADM-driven system of controller (2), although also commercial, is more respectful of the fundamental rights of the caregivers, as less invasive operations are involved in their system, and mitigating measures (such as on-device processing, minimal data storage) were taken. Their system has risks to fundamental rights, while the interests involved seem better balanced.

In *Benchmark 4*, the system of controller (1) does not allow caregivers to affect the data processing in their system at any stage. If caregivers wanted to reduce or stop data processing, they would probably have to withdraw their explicit consent, so that no use can be made of the services of controller (1). The absence of any control over the data processing probably categorizes an ADM as high risk. Controller (2) allows caregivers to adjust the data processing: caregivers can stop data processing after notification with a simple press of the home button on the device. Such a measure strengthens the caregivers fundamental right to personal autonomy, thereby reducing the risk of infringements to privacy and expression rights.

Connecting assessments with mitigating measures

Examples of risk management

Once the risks involved in a particular ADM at DPIA stages have been identified, controllers must take appropriate technical and organizational measures to mitigate these risks.[226] Such measures might vary from highly stringent (eg no use of ADM in certain sectors, moratoria for certain applications until appropriate legislation or technology is in place) to less severe types of measures (eg incidental logs and audits). This spectrum of measures is presented in Figure 4 (in the last column). Determinations of which measure(s) should be chosen to mitigate ADM risks are always context specific. This implies that more general approaches to how risk levels correspond to a particular stringency in mitigating measures may be difficult to capture in one-size-fits-all approaches.

Still, some general reflections that apply to all risks can be safely shared, as some measures will nearly always be useful in ADM. Logs, audits, and evaluations should be in place, as they will help controllers to identify (unexpected) risks once ADM is deployed, and they may help to determine where adjustments are needed during deployment. The same applies to measures that assist in decision and data provenance. Such measures should be baked into the technical architecture of the ADM.[227]

How regulators should further promote fundamental rights impact assessments

The GDPR has endowed supervisory authorities with strong enforcement powers, including carrying out investigations, imposing bans on controller ADM and imposing high fines. They might adopt standards or issue certificates, and they regularly publish guidance. They might also consider professional standards development for ADM systems designers, and best practices in cooperation with other oversight bodies or NGOs.

[226] Art 24 GDPR.

[227] See eg the development of personal data stores, which seek to offer data subjects more control over the processing of their personal data, thereby aiming at strengthening user privacy and data protection (eg Andy Crabtree and others, 'Building Accountability Into the Internet of Things: the IoT Databox Model' (2018) 4 Journal of Reliable Intelligent Environments 39, 51; for considerations as regards regulatory tools with regard to the development of robots see Ronald Leenes and others, 'Regulatory Challenges of Robotics: Some Guidelines for Addressing Legal and Ethical Issues' (2017) 9 Law, Innovation and Technology 1 <https://www.tandfonline.com/doi/full/10.1080/17579961.2017.1304921> accessed 11 October 2019.

Regulators might define regulatory ceilings (eg produce guidance for contexts where price discrimination is acceptable), moratoria (no use of ADM in certain sectors until safe ADM has been developed) or even ban ADM in specific contexts (eg facial recognition in the supply of commercial health care services). Sector-specific legislation might be necessary to protect fundamental rights, for instance where ADM reduces equal access to utilities that are important to the public at large, such as credit, education, or health insurance. Such legislation might be accompanied by sunset clauses, meaning that a law automatically ceases to exist unless the legislator explicitly decides to evaluate and review the law. As stated earlier, there is scope for regulatory sandboxes, and for cooperation between oversight bodies to share and exchange experiences. Regulators might establish trusted organizations to offer opportunities for scientific research and to keep track of developments and the effectiveness of DPIA, including fundamental rights impact assessments.

Concluding observations

Fundamental rights, an important body of law that fortifies individual freedom in society, are a relatively new set of rules for private controllers. This article works towards a practical approach for private controllers to conduct a DPIA as required by the GDPR, and to engage them in a productive discussion about the implementation of technical and organizational measures to mitigate fundamental rights impacts at individual and societal levels by ADM at early design and development stages. It seeks to interrogate controllers by way of four benchmarks about potential fundamental rights risks due to the use of ADM systems, whether their design is appropriate, and how they can correct inappropriate functioning of their ADM system. The approach will help controllers to achieve better compliance with fundamental rights, while it also offers them opportunities to explain and account for their use of ADM. Meanwhile, regulators should assist private controllers with appropriate regulatory guidance, and, where necessary, with regulatory measures as to how fundamental rights are engrained in DPIA (eg by developing practical questionnaires). Interdisciplinary scientific research may help regulators forward in this space. Though an initial, practical concept, this approach to a fundamental rights impact assessment presages a way forward to ensure that private controller ADM systems of the future are better lined up with fundamental rights, including personal freedoms and human dignity.

doi:10.1093/idpl/ipz028
Advance Access Publication 6 March 2020