Reasoned A(I)dministration: explanation requirements in EU law and the automation of public administration

Fink, M.; Finck, M.

Reasoned A(I)dministration: Explanation Requirements in EU Law and the Automation of Public Administration

by

Melanie Fink and Michèle Finck

Reprinted from **European Law Review Issue 3 2022**

Sweet & Maxwell
5 Canada Square
Canary Wharf
London
E14 5AQ
(Law Publishers)

SWEET & MAXWELL

Analysis and Reflections

Reasoned A(I)dministration: Explanation Requirements in EU Law and the Automation of Public Administration

Melanie Fink*

Leiden University and Central European University

Michèle Finck**

University of Tübingen and Max Planck Institute for Innovation and Competition

Administrative decision-making; Artificial intelligence; Data protection; EU law; Reasons

Abstract

Mechanisms to control public power have been developed and shaped around human beings as decision-makers at the centre of the public administration. However, technology is radically changing how public administration is organised and reliance on Artificial Intelligence is on the rise across all sectors. While carrying the promise of an increasingly efficient administration, automating (parts of) administrative decision-making processes also poses a challenge to our human-centred systems of control of public power. This article focuses on one of these control mechanisms: the duty to give reasons under EU law, a pillar of administrative law designed to enable individuals to challenge decisions and courts to exercise their powers of review. First, it analyses whether the duty to give reasons can be meaningfully applied when EU bodies rely on AI systems to inform their decision-making. Secondly, it examines the added value of secondary law, in particular the data protection rules applicable to EU institutions and the draft EU Artificial Intelligence Act, in complementing and adapting the duty to give reasons to better fulfil its purpose in a (partially) automated administration. This article concludes that the duty to give reasons provides a useful starting point but leaves a number of aspects unclear. While providing important safeguards, neither EU data protection law nor the draft EU Artificial Intelligence Act currently fill these gaps.

Introduction

One of the primary functions of administrative law is the control of public power. It ensures that the administration stays within the bounds of the law and protects citizens from abuse of authority.[1] While

* Assistant Professor at Leiden University (Europa Institute) and APART-GSK Fellow of the Austrian Academy of Sciences at the Central European University (Department of Legal Studies).

** Professor of Law and Artificial Intelligence at the University of Tübingen and Affiliated Fellow, Max Planck Institute for Innovation and Competition. Member of the Machine Learning Cluster of Excellence, EXC number 2064/1 — Project number 390727645. The author gratefully acknowledges funding by the Carl Zeiss Foundation.

[1] C. Harlow, "Global Administrative Law: the Quest for Principles and Values" (2006) 17 E.J.I.L. 187, 190–191; see also S. Rose-Ackerman, P.L. Lindseth and B. Emerson, "Introduction" in S. Rose-Ackerman, P.L. Lindseth and B. Emerson (eds), *Comparative Administrative Law*, 2nd edn (Cheltenham: Edward Elgar, 2017), pp.1–3.

concrete duties and rights vary among different systems, they have at least one thing in common: control mechanisms are shaped around a human being as a decision-maker at the centre of the administration. These mechanisms protect against what we know to be the flaws of human decision-makers, such as arbitrariness or bias, through reasoning obligations, participation rights, or transparency. And they make use of the virtues of human decision-making, such as compassion or empathy, by creating room for human judgment in the form of administrative discretion.[2]

However, technology is radically changing how public administration is organised. Even though digitization and automation are most commonly associated with the private sector, also public sector decision-making is increasingly relying on Artificial Intelligence (AI).[3]AI systems are employed in a variety of areas, such as crime prevention, criminal justice, healthcare, immigration, or the administration of public benefits.[4] Reliance on these tools carries the promise of addressing some of the known flaws in human decision-making by improving accuracy, efficiency, and potentially bias.[5] At the same time, automating (parts of) administrative decision-making processes also poses a challenge to our human-centred systems of control of public power.

This article focuses on one of these control mechanisms: the duty to give reasons under EU law. This duty requires EU bodies to communicate the reasons for their decision to those affected by it. Disclosing the rationale of a decision is considered instrumental to enable individuals to challenge it and courts to exercise their powers of review. While humans can in principle articulate the reasons for their decision without major difficulties (whether these are the 'real' reasons is, of course, a different question),[6] this is not always the case when the decision is made by, based on, or informed by an AI system. The problem is that AI systems may be so complex that a human cannot understand how and why a system reached its conclusion (also referred to as the 'black box' problem).[7] A human decision-maker will therefore not be able to explain the concrete reasons that informed a decision significantly shaped by a 'black box'.

Against this background, this article explores the challenges posed by automation to the implementation of the duty to give reasons under EU law. It pursues a two-fold aim. First, it analyses whether this pillar of administrative law can be meaningfully applied when EU bodies rely on AI systems to inform their decision-making. Second, it examines the added value of secondary law, in particular data protection rules applicable to EU institutions and the draft EU Artificial Intelligence Act (AIA), in complementing and adapting the duty to give reasons to better fulfil its purpose in a (partially) automated administration.[8] Doing so, this contribution creates bridges between long-standing and deeply rooted explanation duties in administrative law and more recent debates about the explainability of AI that have largely centred on

[2] K. Yeung, "Why Worry about Decision-making by Machine?" in K. Yeung and M. Lodge (eds), *Algorithmic Regulation* (Oxford: Oxford University Press, 2019), p.29. More generally on the role of empathy see S. Ranchordas, "Empathy in the Digital Administrative State" (2022) 72 *Duke Law Journal* (forthcoming).

[3] M. Busuioc, "Accountable Artificial Intelligence: Holding Algorithms to Account" (2021) 81 *Public Administration Review* 825, 826; J. Cobbe, "Administrative Law and the Machines of Government: Judicial Review of Automated Public-sector Decision-making" (2019) 39 *Legal Studies* 636; K. Yeung and M. Lodge, "*Algorithmic Regulation*: an Introduction" in Yeung and Lodge (eds), Algorithmic Regulation (Oxford: Oxford University Press, 2019), pp.1–4.

[4] Busuioc, "Accountable Artificial Intelligence" (2021) 81 *Public Administration Review* 826.

[5] C.R. Sunstein, "Governing by Algorithm? No Noise and (Potentially) Less Bias" (2022) 71 *Duke Law Journal* 1175.

[6] Yeung, "Why Worry about Decision-making by Machine?" in *Algorithmic Regulation* (2019), pp.28–29.

[7] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, Massachusetts London, England: Harvard University Press, 2016).

[8] Regulation 2018/1725 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data [2018] OJ L295/39; "Proposal for a Regulation of the European Parliament and of the Council laying down the harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts" COM(2021) 206 final.

data protection law. It also seeks to inform the law-making process around the draft AIA by examining whether, and how, it can add value to existing requirements in EU law.

This article first provides a brief overview of the terms commonly used in context of AI and introduces concrete examples of how EU bodies may make use of these new technologies. It then maps and systematises the duty to give reasons in administrative law. Building on that, it assesses whether this duty is suitable to meet the unique challenges posed by AI-supported decision-making by the EU administration. Finally, it examines the added value of EU data protection law and the AIA in this respect. This article concludes that the duty to give reasons provides a useful starting point, but leaves a number of aspects unclear. Neither EU data protection law nor the AIA currently fill these gaps.

Automated decision-making, algorithms, and artificial intelligence

Public administrations increasingly rely on various forms of computational analysis that enables them to render their operations more efficient and detect correlations in or across datasets that may be undetectable to human cognition. Automated decision-making (ADM) refers to the use of algorithms, i.e. a set of steps stated in mathematical or logical language, in a decision-making process.[9] A decision-making process can be fully or partially automated. In the latter case, the most common in the administrative context, ADM is used only for part of the decision-making process and a human takes the actual decision.[10]

ADM comprises various sets of techniques, including AI, a general-purpose technology that is powered by different kinds of algorithms. The current draft of the AIA defines AI systems as software that is developed with one or more of the techniques listed in the Act and which can, 'for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with'.[11] This article focuses specifically on AI-assisted decisions by EU institutions and authorities, that is to say situations in which an algorithmic output taking the form of a decision, recommendation, or prediction forms part of a human decision-making process.

Machine learning is the currently most widely used subset of AI that finds application in such processes. It essentially mines data to identify patterns that are transformed into predictive analysis that is in turn applied to new data.[12] Here, a machine learning algorithm is used to transform data into an output, which may take the form of an automated decision or form part of an automated decision-making process.[13] Machine learning is itself an umbrella term that encompasses different techniques, including deep learning. Deep learning is a subset of machine learning whose complex internal logic comprises different layers. Due to this complexity, it can be hard if not impossible for humans to determine *ex post* how the deep learning algorithm produced a given output on the basis of the data that it has processed. This phenomenon is conventionally referred to as the 'black box' problem pursuant to which algorithmic processes lack the transparency that would enable humans to understand how a given input led to a given output.[14] As a consequence, a public administration using such software may be unable to understand how the output was generated, as do citizens subjected to the relevant decision. There is a possibility that as artificial intelligence techniques become more sophisticated, this problem will proliferate.

[9] Busuioc, "Accountable Artificial Intelligence" (2021) 81 *Public Administration Review* 825, 828.

[10] L. Edwards and M. Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' is Probably not the Remedy you are Looking For" (2017) 16 *Duke Law & Technology Review* 18, 45.

[11] Draft EU Artificial Intelligence (AIA) Act art.3(1).

[12] See further Research Group on the Regulation of the Digital Economy, J. Drexl et al, "Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective" (2019) 19 *Max Planck Institute for Innovation & Competition Research Paper*.

[13] For a more detailed overview, see M. Finck, "Automated Decision-making and Administrative law" in P. Cane et al (eds), *The Oxford Handbook of Comparative Administrative Law* (Oxford: Oxford University Press, 2021).

[14] See further Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (2016).

The EU is investing heavily in the development of technologies, including AI, to be used in an increasing amount of policy areas by the EU administration itself and its Member States.[15] While current applications of AI-assisted ADM by EU bodies is limited, it is expected to increase significantly in the future.[16] A particularly illustrative example of the potential of these new technologies is the iBorderCtrl project. The aim of the project, funded under the European Commission's Horizon 2020 programme, was to speed up border control for third-country nationals crossing land borders of EU Member states by providing a decision support system for border control authorities.[17] The project, which included pilot tests at Hungarian, Greek, and Latvian land borders in 2019, experimented with a system that detects deception through facial recognition technology and the measurement of micro-expressions. If put into actual practice, travellers would be interviewed by a computer-animated border guard that identifies them as 'bona fide' or 'non-bona fide' based on what the team behind iBorderCtrl termed 'biomarkers of deceit', that is facial micro-expressions they claim humans produce when lying.[18] Depending on whether deception is detected, there will be more or less thorough follow-up checks by a human border guard, who will also take the final decision regarding the permission of entry.

Whereas it might seem compelling to believe that the use of algorithms such as these inevitably leads to an objective and verifiable truth, this is not actually the case. First, AI systems require design choices that add highly subjective or even incorrect elements. In particular, the way in which features are chosen and prioritised may enshrine the system designer's preferences and implicit biases.[19] Indeed, these systems necessarily embody 'contestable epistemic and normative assumptions'.[20] As a result, they can encode different forms of bias.[21] Secondly, machine learning relies on datasets to train its algorithms, the so-called 'training data'. This data informs the analysis that the trained algorithm performs on new datasets. Training data can be incomplete, contain incorrect data or fail to be representative and hence risk perpetuating biases. Even the use of a well-polished dataset may be unsuitable. For instance, an algorithm trained on employment data from the 1950 will conclude that women are unsuitable for most work outside of the home and apply that insight to contemporary data, including in recruitment processes. Some of these problems have also been observed in relation to the iBorderCtrl project. The system is based on a number of assumptions with regards to the relationship between lying and criminality or risk, the universality of facial expressions across persons of different ethnicity, gender, age, etc. that indicate deception, and the measurability of these expressions. These assumptions have either been proven incorrect or at least lack a solid basis in theoretical or empirical studies.[22] Since decisions based on algorithms thus cannot generally

[15] F. Chiusi, "Automating Society Report 2020" (Algorithm Watch, 2020) *Algorithmwatch.org*, *https://automatingsociety.algorithmwatch.org/*.

[16] For an overview see H.C. Hofmann, "An introduction to Automated Decision Making (ADM) and Cyber-Delegation in the Scope of EU Public Law" (2021) *University of Luxembourg Law Working Paper* No.8.

[17] J. Sánchez-Monedero and L. Dencik, "The Politics of Deceptive Borders: 'Biomarkers of Deceit' and the Case of iBorderCtrl" (2020) 25 *Information, Communication & Society* 1, 4.

[18] Sánchez-Monedero and Dencik, "The Politics of Deceptive Borders" (2020) 25 *Information, Communication & Society* 4–7.

[19] A. Danna and O.H. Gandy, "All that Glitters is not Gold: Digging beneath the Surface of Data Mining" (2002) 40 *Journal of Business Ethics* 373.

[20] R. Binns, "Algorithmic Accountability and Public Reason" (2018) 31 *Philosophy & Technology* 543, 546.

[21] L. Sweeney, "Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising" (2013) 11 *Queue* 10, 29.

[22] Sánchez-Monedero and Dencik, "The Politics of Deceptive Borders" (2020) 25 *Information, Communication & Society* 7–11. They also show that even if all the assumptions were to hold, the test conducted by the system used in the iBorderCtrl project will in practice not be able to accurately distinguish between 'bona fide' travellers and 'liars', See Sánchez-Monedero and Dencik, "The Politics of Deceptive Borders" (2020) 25 *Information, Communication & Society* 11–14.

be assumed to be objective and accurate, they need to be subject to control mechanisms, just like decisions taken by humans are.

The duty to give reasons in EU administrative law

The EU legal system enshrines a range of legal obligations that require authorities to explain the reasons behind a decision to those affected by it.[23] Sector-specific obligations aside, EU law sets out a general reasoning obligation that applies across the board, regardless of the policy area and EU body in question. This section sets out the parameters of this obligation, discussing first its legal basis and purpose and then the specific requirements a statement of reasons needs to fulfil.

Legal basis and purpose of the duty to give reasons

EU administrative law contains three legal bases for the duty to give reasons. The first is art.296 Treaty on the Functioning of the EU (TFEU), which goes back to the 1957 Treaty of Rome and sets out that '[l]egal acts shall state the reasons on which they are based […]'. The second provision is art.41 Charter of Fundamental Rights (CFR), the right to good administration, which draws together a range of procedural rights that are to ensure procedural justice, the rule of law, and sound administrative decision-making.[24] Paragraph 2(c) of that provision is based on art.296 TFEU and explicitly sets out that the right to good administration includes 'the obligation of the administration to give reasons for its decisions'.[25] A third legal basis for the duty to state reasons can be found in the procedural standards and guarantees recognised on the basis of general principles of law.[26] The content of the duty to state reason as a general principle of law largely overlaps with art.41(2)(c) which in turn is not only inspired by art.296 TFEU, but also by the case law of the Court of Justice of the EU (CJEU or 'the Court') on the principles of sound administration.[27]

Guarantees of procedural justice are typically determined by two main rationales that exist alongside each other. On the one hand, following a set of procedural rules makes it more likely that decision-making is accurate and efficient. The ultimate goal of having them in place is accordingly achieving the best and most efficient implementation of policy. On the other hand, procedural fairness ensures that individual

[23] While there is no broad overview of these legal requirements, there is some literature discussing rules applicable to private bodies in EU law as well as literature outlining legal explainability requirements more generally. On the former, see P. Hacker and J. Passoth, "Varieties of AI Explanations under the Law: from the GDPR to the AIA, and Beyond" in A. Holzinger et al. (eds), *Lecture Notes on Artificial Intelligence 13200: xxAI - beyond explainable AI* (Springer, 2021). On the latter, see A. Bibal et al, "Legal Requirements on Explainability in Machine Learning" (2021) 29 *Artificial Intelligence and Law* 149; For an overview of different legal systems see H.P. Olsen, J.L. Slosser and T.T. Hildebrandt, "What's in the Box?: the Legal Requirement of Explainability in Computationally Aided Decision-making in Public Administration" in H. Micklitz et al (eds), *Constitutional Challenges in the Algorithmic Society* (Cambridge: Cambridge University Press, 2022).

[24] H.C. Hofmann and C. Mihaescu, "The Relation between the Charter's Fundamental Rights and the Unwritten General Principles of EU Law: Good Administration as the Test Case" (2013) 9 *European Constitutional Law Review* 73, 84; K. Kańska, "Towards Administrative Human Rights in the EU: Impact of the Charter of Fundamental Rights" (2004) 10 E.L.J. 296; H.P. Nehl, "Good Administration as Procedural Right and/or General Principle?" in H.C. Hofmann and A.H. Türk (eds), *Legal Challenges in EU Administrative Law: Towards an Integrated Administration* (Cheltenham: Edward Elgar, 2009); J. Mendes, "Good Administration in EU law and the European Code of Good Administrative Behaviour" (2009) *EUI Working Papers Law* No.2009/09.

[25] Explanations relating to the Charter of Fundamental Rights [2017] OJ C303/17 art.41.

[26] Nehl, "Good Administration as Procedural Right and/or General Principle" in *Legal Challenges in EU Administrative Law: Towards an Integrated Administration* (2009); Hofmann and Mihaescu, "The Relation between the Charter's Fundamental Rights and the Unwritten General Principles of EU Law" (2013) 9 *European Constitutional Law Review* 73, 84.

[27] Charter of Fundamental Rights [2017] OJ C303/17 art.41.

rights are effectively protected against abuse of public power.[28] The CJEU finds the duty to give reasons to serve two specific goals, both closely related to the latter rationale. First, the explicit communication of reasons enables competent courts to exercise their power to review the legality of a decision.[29] Secondly, where an individual decision is concerned, a person affected by it must receive enough information to be able to determine whether the decision is well-founded and challenge it if that is not the case.[30]

There is a particularly strong emphasis by the Court on the duty to state reasons as a right of defence, enabling individuals to initiate a review of legality.[31] In this vein, the Court frequently invokes art.47 CFR, the right to an effective remedy, to support the requirement of reason-giving.[32] On some occasions, it even deduced the right to a reasoned decision directly from art.47 CFR itself, with or without additionally relying on arts 296 TFEU or 41(2)(c) CFR.[33] In *R.N.N.S. and K.A. v Minister van Buitenlandse Zaken*, for instance, the Court held:

> "[I]t is settled case-law that if the judicial review guaranteed by Article 47 of the Charter is to be effective, the person concerned must be able to ascertain the reasons upon which the decision […] is based […] so as to make it possible for him or her to defend his or her rights in the best possible conditions and to decide, with full knowledge of the relevant facts, whether there is any point in applying to the court with jurisdiction."[34]

This exposes an instrumental understanding of the duty to state reasons in that it forms a gateway to ensure the enjoyment of other rights. However, in addition, reason-giving has an important role to play at a more abstract level. As Mashaw notes, authority without reason is dehumanising and thus 'fundamentally at war with the promise of democracy, which is, after all, self-government'.[35] In this understanding, the obligation to reason is essential to recognising individuals as subjects, rather than objects of the law and as such is fundamental to the moral and political legitimacy of a legal order.[36] This broader constitutional function of the duty to give reasons has also been recognised for the EU more specifically.[37]

[28] Nehl, "Good Administration as Procedural Right and/or General Principle?" in *Legal Challenges in EU Administrative Law: Towards an Integrated Administration* (2009), 343–346; L. Millett, "The Right to Good Administration in European Law" (2002) 47 *Public Law* 309, 312; M. Shapiro, "The Giving Reasons Requirement" (1992) *University of Chicago Legal Forum* 179, *https://chicagounbound.uchicago.edu/uclf/vol1992/iss1/8*.

[29] *France v Commission* (C-17/99) EU:C:2001:178 at [35]; *Club Hotel Loutraki AE v Commission* (C-131/15 P) EU:C:2016:989 at [46].

[30] *Elf Aquitaine SA v Commission* (C-521/09 P) EU:C:2011:620 at [148]; *Corus UK v Commission* (C-199/99) EU:C:2003:531 at [145].

[31] *PI v Landespolizeidirektion Tirol* (C-230/18) EU:C:2019:383; [2019] 3 C.M.L.R. 31 at [57]; See also J. Schwarze, "Judicial Review of European Administrative Procedure" (2004) 68 *Law and Contemporary Problems* 85, 91–94.

[32] *Club Hotel Loutraki* (C-131/15 P) EU:C:2016:989 at [49]; See also the Opinion of AG Kokott in *Housieaux* (C-186/04) EU:C:2005:70 at [32].

[33] *Unectef v George Heylens* (222/86) EU:C:1987:442; [1989] 1 C.M.L.R. 901 at [15]; *Kadi v Council of the European Union* (C-402/05 P and C-415/05 P) EU:C:2008:461; [2008] 3 C.M.L.R. 41 at [335]–[337]; *ZZ France v Secretary of State for the Home Department* (C-300/11) EU:C:2013:363; [2013] 3 C.M.L.R. 46 at [53]; *RNNS v Minister van Buitenlandse Zaken* (C-225/19 and C-226/19) EU:C:2020:951 at [43].

[34] *RNNS* (C-225/19 and C-226/19) EU:C:2020:951 at [43].

[35] J.L. Mashawn, "Reasoned Administration: the European Union, the United States, and the Project of Democratic Governance" (2007) 76 *George Washington Law Review* 118.

[36] Mashawn, "Reasoned Administration" (2007) 76 *George Washington Law Review* 116–118.

[37] Kańska, "Towards Administrative Human Rights in the EU" (2004) 10 E.L.J. 299; See also J. Mendes, "The Foundations of the Duty to give Reasons and a Normative Reconstruction" in E. Fisher, J. King and A. Young (eds), *The Foundations and Future of Public Law: Essays in Honour of Paul Craig* (Oxford: Oxford University Press, 2020).

Requirements for the statement of reasons

The duty to give reasons requires that EU bodies disclose the reasoning followed to reach a decision in a clear and unequivocal fashion.[38] This means that the reasons must be logical and contain no internal inconsistency that would prevent a proper understanding of the reasons underlying the measure.[39] In terms of extent, the decision-making authority is required to set out the facts and the legal considerations having decisive importance and address all relevant counter-arguments.[40] This does not include being required to go into all points of law and fact, also because the statement of reasons is assessed not only in light of its wording but also its context and 'all the legal rules governing the matter in question'.[41]

Beyond these general requirements, the types of reasons to be given depend on the measure in question. According to the Court, the reasons must be appropriate, first, to the content of the measure and, second, the interests involved on the side of those concerned by the measure.[42] In relation to the latter, this includes that adverse decisions, especially when they have serious consequences, require more explanation than decisions in favour of the person concerned.[43]

As regards the former, two essential distinctions are made. The Court distinguishes 'routine' measures based on long-standing decision practice and well-established case law from 'exceptional' measures going beyond prior practice. Whereas routine measures can be reasoned in a summary manner, for instance referring to earlier decisions, exceptional measures need to provide a fuller account of their reasoning.[44] Another distinction is drawn between discretionary measures and non-discretionary measures. The former include discretionary powers regarding political or strategic policy choices (volitive discretion) and the large margin of appraisal administrative authorities enjoy regarding complex economic, technical, or scientific facts (cognitive discretion).[45] In these areas of factual complexity, where administrative authorities have special expertise, the Union judge may not substitute his or her own assessment for that of the (expert) administration. Judicial review of the substantive legality of such decisions is therefore limited to ascertaining whether the administration has committed a 'manifest error of appraisal'.[46] To counterbalance this limit to full review, the EU Courts consider adherence to procedural requirements, such as the duty to state reasons, fundamentally important and review these more strictly.[47]

[38] For many others see *Spain v Commission of the European Communities* (C-113/00) EU:C:2002:507; [2003] 1 C.M.L.R. 16 at [47].

[39] *Elf Aquitaine SA* (C-521/09 P) EU:C:2011:620 at [151].

[40] *Ryanair DAC v European Commission* (T-665/20) EU:T:2021:344 at [40]; *Elf Aquitaine SA* (C-521/09 P) EU:C:2011:620 at [159]–[169].

[41] *Landeskreditbank Baden-Wurttemberg - Forderbank v ECB* (T-122/15) EU:T:2017:337; at [124]–[125]; *Council v Bamba* (C-417/11 P) EU:C:2012:718; [2018] 1 C.M.L.R. 7 at [54].

[42] *Elf Aquitaine SA* (C-521/09 P) EU:C:2011:620 at [150]; *Deutsche Telekom AG v European Commission* (C-280/08 P) EU:C:2010:603; [2010] 5 C.M.L.R. 27 at [131]; *France v Commission* (C-17/99) EU:C:2001:178 at [36].

[43] *Vela Srl and Tecnagrind SL v Commission* (T-141/99, T-142/99, T-150/99 and T-151/99) EU:T:2002:270 at [168]; *Sgaravatti Mediterranea Srl v Commission (T-199/99) EU:T:2002:228* at [101]; *Elf Aquitaine SA* (C-521/09 P) EU:C:2011:620 at [152].

[44] *Société Francaise des Biscuits Delacre SA et al v Commission* (C-350/88) EU:C:1990:71 at [15]; *Elf Aquitaine SA* (C-521/09 P) (C-521/09 P) EU:C:2011:620 at [155].

[45] H.P. Nehl, "Judicial Review of Complex Socio-economic, Technical, and Scientific Assessments in the European Union" in J. Mendes (ed), *EU Executive Discretion and the Limits of Law* (Oxford: Oxford University Press, 2019), pp.168–169.

[46] Nehl, "Judicial Review of Complex Socio-economic, Technical, and Scientific Assessments in the European Union" in *EU Executive Discretion and the Limits of Law* (2019), pp.168–170, 175–176, 178–179.

[47] *Hauptzollamt München-Mitte v Technische Universität München* (C-269/90) EU:C:1991:438; [1994] 2 C.M.L.R. 187 at [13]–[14]; *France v Commission* (T-257/07) EU:T:2011:444 at [84]–[89]; *Pfizer v Council* (T-13/99) EU:T:2002:209 at [166]; Schwarze, "Judicial Review of European Administrative Procedure" (2004) 68 *Law and Contemporary Problems* 85, 94–96.

In short, thus, the legal reasoning must be disclosed in a concise and understandable manner. This means that it should consist of three elements: (1) the legal provisions applied, (2) the relevant facts, and (3) the decisive considerations in applying the law to the facts, such as specific interpretations of the law or the facts, discretionary choices made, or any other factor that carried weight in the assessment.[48] When the decision touches upon important interests of the recipient, is particularly consequential for that person, departs from prior decisions, or involves a high degree of discretion, e.g. due to high technical complexity, reasoning has to be more thorough and is more strictly checked by the court. In a context such as the iBorderCtrl project, this would mean disclosing to a traveller denied entry into the EU exactly what factors led to that decision, such as what fact they were found to have been untruthful about, what made the officer think this was a lie, what legal provision this triggered the application of, and why that required denial of entry.

Should an EU body have failed to give reasons that live up to these requirements, the measure in question can be annulled under art.263 TFEU.[49] The ground for annulment is the infringement of an 'essential procedural requirement' and consequently independent of the substantive legality of the measure, i.e. the question whether the reasoning is well-founded. A decision may well be substantively in accordance with the law, but nonetheless be annulled for lack of adequate reasons communicated to the addressee. The question whether the infringement of the obligation to state reasons may also render the Union liable is more contested, but has been summarily rejected by the Court in a long-standing line of case law.[50]

The duty to give reasons in the context of automation

The duty to give reasons has been carved out in relation to human decision-makers. This section places it in the context of AI, exploring the challenges to implement this duty in a (partially) automated decision-making process. It starts by discussing the opacity of AI as a challenge to the obligations arising for public authorities and then moves on to examine what the duty to give reasons requires from a public authority that uses an AI-system to support its decision-making.

Explainability as a precondition and the 'black box' problem

A major challenge in relation to the control of AI systems is their opacity. This may be 'intentional opacity', where the AI system's workings remain undisclosed to protect intellectual property rights, 'illiterate opacity', where a system might be understandable, but only to those with specific expert knowledge on coding and computing, or 'intrinsic opacity', where a system is so complex that it is generally not understandable for humans.[51]

[48] Similarly see Olsen, Slosser and Hildebrandt, "What's in the Box?" in *Constitutional Challenges in the Algorithmic Society* (2022), pp.225–226.

[49] *Elf Aquitaine SA* (C-521/09 P) EU:C:2011:620 at [146]; *Commission v Sytraval and Brink's France* (C-367/95 P) EU:C:1998:154 at [67]; *France v Commission* (C-17/99) EU:C:2001:178 at [35].

[50] *Kind KG v European Economic Community* (106/81) EU:C:1982:291 at [14]; reiterated more recently in *Bank Refah Kargaran v Council of the European Union* (C-134/19 P) EU:C:2020:793; [2021] 2 C.M.L.R. 3 at [61]. However, this summary dismissal of the possibility of liability seems contradictory to the recognition that the right to a reasoned decision confers rights on individuals as an 'expression of specific rights' under the principle of sound administration (e.g. *Tillack v Commission of the European Communities* (T-193/04) EU:T:2006:292; [2007] 1 C.M.L.R. 5 at [127]) and the fact that other such 'expressions of specific rights' have since formed the basis of successful damages claims (e.g. *Staelen v European Ombudsman* (T-217/11) EU:T:2015:238 at [81]–[88]).

[51] See Cobbe, "Administrative Law and the Machines of Government" (2019) 39 *Legal Studies* 636, 638–639; J. Cobbe, M.S.A. Lee and J. Singh, "Reviewable Automated Decision-Making: a Framework for Accountable Algorithmic Systems" in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (2021), pp.699–700. They, in turn, rely on J. Burrell "How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms" (2016) 3 *Big Data & Society* 1.

The opacity of AI systems presents a particular challenge in light of the widely held view that transparency is a necessary condition to enable control over AI systems.[52] Sharing information on the algorithm broadly speaking, the data it relies on, or its internal logic,[53] may be difficult or outright impossible, if an AI system is opaque. Explanation requirements, such as the duty to give reasons, face a similar problem. As a specific form of transparency, they focus on the reasons or justification for an outcome. What is important is not *how* a decision was reached, but *why*.[54] Providing an explanation for a decision of an AI system would involve disclosing to a human observer how different factors were weighed by the system to reach a decision and what input was determinative.[55] When a decision is (partially) based on the output of an AI system, the possibility to give an explanation thus depends on the system itself being able to generate an outcome that is understandable by humans. In other words, some degree of 'explainability' (also 'interpretability') of an AI system is a precondition to produce an explanation (see Fig.1).[56]

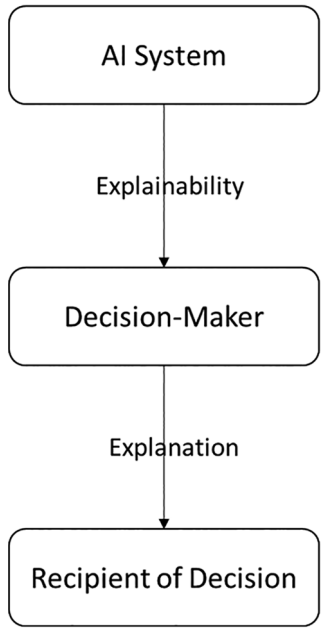

Fig.1

[52] Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (2016); see also M. Ananny and K. Crawford, "Seeing without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability" (2018) 20 *New Media & Society* 973, 977; Hacker and Passoth "Varieties of AI Explanations under the Law" in *Lecture Notes on Artificial Intelligence 13200: xxAI - beyond explainable AI* (2021), pp.1–2.

[53] B.D. Mittelstadt et al, "The Ethics of Algorithms: Mapping the Debate" (2016) 3 *Big Data & Society* 1–21, 4–7.

[54] Cobbe, "Administrative Law and the Machines of Government" (2019) 39 *Legal Studies* 636, 648.

[55] F. Doshi-Velez and M. Kortz, "Accountability of AI under the Law: the Role of Explanation" (2017) *Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper*, pp.2–3; Hacker and Passoth, "Varieties of AI Explanations under the Law" in *Lecture Notes on Artificial Intelligence 13200: xxAI - beyond explainable AI* (2021), p.2; S. Wachter, B. Mittelstadt and L. Floridi, "Why a Right to Explanation of Automated Decision-making does not Exist in the General Data Protection Regulation" (2017) 7 *International Data Privacy Law* 76, 78.

[56] It should be noted that for the purposes of this Article 'interpretability' and 'explainability' are treated as synonyms.

From the perspective of the duty to give reasons, 'illiterate opacity' presents a problem because the workforce of a public body necessarily consists of experts on the specific subject matter at the heart of that bodies' tasks, rather than AI experts. Thus, the public authority will not be able to substantially explain its own decision, if it uses an AI system's recommendation, the justification for which is interpretable only with expert knowledge.

Similarly challenging is 'intrinsic opacity', that is the 'black box' problem. As explained above, the image of the black box denotes that the inner workings of complex algorithmic systems, such as deep learning algorithms, can often not be reconstructed by humans. This means that the use of different variables to transform an input into an output cannot be retraced *ex post*. As a result, humans cannot understand how, exactly, these systems reach their conclusions.[57] In part, this stems from the fact that AI—at least for now—relies on correlations not causation.[58] Indeed, an actual causal relationship between input and output 'may simply not exist, no matter how intuitive such relationships might look on the surface'.[59] This also means they 'cannot support causal explanations of the kind that underlie the reasons traditionally offered to justify governmental action'.[60] In other words, a public authority cannot provide detailed reasons for a decision informed by an algorithm suffering from this type of opacity.

Think of a border guard who decides a person to be 'non-bona fide' on the basis of an interview and background checks and consequently rejects the application for entry. The border guard will be able to point to specific aspects in the person's documents or behaviour that led him or her to reach that conclusion and communicate these reasons to the person concerned. Regardless of whether those are the 'real' reasons, this provides the individual concerned with a basis upon which to challenge the decision. In that context, courts can assess whether the evidence presented—such as a transcription of the interview and the documents presented—supports the reasons given and whether the reasons themselves are capable of supporting the decision reached. If that border guard instead relies on the classification of a person as 'non-bona fide' by the iBorderCtrl project, his or her ability to justify the decision will depend on the system's ability to justify its recommendation. Should the AI system be too opaque for the border guard to understand what elements in the traveller's behaviour led to the 'non-bona fide' classification, the border guard's own reason-giving cannot amount to more than 'because the AI system said so'.

From the perspective of the duty to give reasons, this raises a fundamental question: Does a public authority comply with its reason-giving obligations if it simply defers to the recommendation provided by an AI system? Or does EU administrative law require the reasons underlying the AI system's recommendation to be disclosed? What exactly, in other words, is required from a public authority under the duty to give reasons, when it relies on a recommendation by an AI system?

These questions are explored in the following section. They are not only of relevance to public authorities but also the broader community working on making algorithms 'explainable'. Indeed, in the past years, a significant amount of work has gone into ensuring the explainability of AI systems at a technical level,[61] either by altering the algorithm itself, or adding an 'explanation algorithm' to the black box algorithm.[62] Investing in such approaches has been the predominant approach to solving the accountability problems

[57] C. Coglianese and D. Lehr, "Regulating by Robot: Administrative Decision Making in the Machine-learning Era" (2017) 105 *Georgetown Law Journal* 1147, 1159.

[58] B. Schölkopf, "Causality for Machine Learning" in H. Geffner, R. Dechter and J.Y. Halpern (eds), *Probabilistic and Causal Inference* (New York: ACM, 2022), pp.765–804.

[59] Coglianese and Lehr, "Regulating by Robot" (2017) 105 *Georgetown Law Journal* 1147, 1157.

[60] Coglianese and Lehr, "Regulating by Robot" (2017) 105 *Georgetown Law Journal* 1147, 1167.

[61] Cobbe, "Administrative Law and the Machines of Government" (2019) 39 *Legal Studies* 636, 648.

[62] Busuioc, "Accountable Artificial Intelligence" (2021) 81 *Public Administration Review* 825, 830; Doshi-Velez and Kortz, "Accountability of AI under the Law: the Role of Explanation" (2017) *Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper*, pp.6–9; see also *https://www.darpa.mil/program/explainable-artificial-intelligence*.

the lack of explainability creates.[63] However, different explanation requirements demand different degrees or types of explainability. For instance, ensuring the possibility of an explanation that is 'good' from a philosophical or psychological standpoint might have different implications than ensuring compliance with various legal explanation requirements.[64] Thus, finding technical solutions to make an AI system explainable requires knowing what kind of explanation it is supposed to be able to support.

Requirements for the statement of reasons in the AI context

Whether or not a public authority relies on AI does not in principle affect their obligation towards those affected by the decision to provide an explanation.[65] Yet, since the type and extent of reasoning required has been developed by the EU courts in relation to human decision-makers, the question arises how it applies when decision-making is (partially) automated.

It might seem intuitive to assume that in situations where public bodies rely on an AI system's recommendation, their reasoning obligations may be fulfilled by reference to that recommendation. The underlying idea is that what distinguishes this scenario from the one where the authority decides without AI support is added skill, knowledge, or expertise brought in by the system that makes the final decision more reliable and thus justifies a lowering of the usual reasoning requirement.

Apart from the fact that the underlying assumption that AI systems make 'better' decisions is not necessarily true, two considerations speak against this. First, EU courts have not adopted this particular approach in other areas where EU bodies rely on the findings of expert bodies, such as EU agencies or international institutions. In such scenarios, the decision-making institution nonetheless has to provide an explanation that specifically identifies the concrete reasons why a particular conclusion was reached.[66] This means that even if the use of AI in decision making were considered equivalent to relying on expert advice, providing a justification that simply refers to the fact that an AI system recommended the decision would not meet the requirements under the duty to give reasons.

Secondly, and more fundamentally, the perceived correctness of an AI system's output has no bearing on reasoning requirements. This has to do with the specific purpose of the reason-giving obligation. While it is certainly considered to promote accuracy by forcing an authority to lay out and thus ponder the reasons for a decision, its essential function lies in ensuring judicial control over administrative power and enabling private parties to challenge decisions. As explained above, the Court may thus well find that a decision is substantively in accordance with the law, but nonetheless annul it for lack of adequate reasons communicated to the addressee.[67] In other words, no matter how good a decision is likely to be from a substantive perspective, concrete reasons have to be provided so that the decision can be challenged and reviewed.

Thus, it is safe to say that 'the machine said so' will not be sufficient under the law. This also means that, at least where the AI recommendation is the only or main factor affecting the direction of a decision, the public authority that takes the decision needs to understand the reasons that determined the AI system's output itself, in order to comply with its reason-giving obligations. A degree of explainability of the AI

[63] See, however, Cobbe, Lee and Singh, "Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems" in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (2021), p.600.

[64] Suggesting to build more on insights from the social sciences in this respect see T. Miller, "Explanation in artificial intelligence: insights from the social sciences" (2019) 267 *Artificial Intelligence* 1.

[65] Cobbe "Administrative Law and the Machines of Government" (2019) 39 *Legal Studies* 636, 639–640.

[66] *European Commission v Kadi* (C-584/10 P, C-593/10 P and C-595/10) EU:C:2013:518; [2014] 1 C.M.L.R. 24 at [111]–[116]; implicitly also *Pfizer Animal Health v Council of the European Union* (T-13/99) EU:T:2002:209 at [505]–[514].

[67] *Commission v Sytraval and Brink's France* (C-367/95 P) EU:C:1998:154 at [65]–[67].

system used is therefore necessary to produce an explanation that complies with the requirements under EU administrative law.

Beyond this, the fact that AI is used may actually be a reason to *increase* the decision-maker's reasoning obligations. As explained above, in areas of high factual (economic, technical, or scientific) complexity, where the public administration enjoys a wider margin of appraisal, the EU courts have required them to more thoroughly reason their decisions vis-à-vis those affected. This is to counterbalance the fact that the Union judge may not, sometimes even cannot, fully review the substantive legality of decisions of the (expert) administration. Two considerations are relevant in this respect. First, AI offers the possibility to process vast amounts of information, a tool needed in areas of factual complexity. Thus, AI systems may often be used in areas already characterised by increased reasoning-obligations. Secondly, even when not actually used in an area of factual complexity, some AI systems, in particular those that suffer from the opacity described earlier, pose similar constraints to the Court's ability to conduct full substantive review. To balance this limit to full substantive review, the Court might consider increased scrutiny of reasoning obligations, similar to the approach adopted in areas of wide discretion and high complexity.

In addition, it can be argued that the Court should apply a higher explanation requirement when decision-making is automated to counter-act the risk of 'automation bias'.[68] Automation bias refers to the phenomenon that humans ascribe a certain authority to outcomes suggested by automated processes that leads them to neglect other available information or counter-indications. Common examples include car accidents that are caused by a human driver relying on a route suggested by a driving assistant even when real-life factors speak against it. Thus, even where an authority can give a justification, it may in substance have 'blindly' trusted the outcome suggested by the AI system. Preventing automation bias might necessitate additional safeguards, including, for instance, requiring the public authority that relies on AI to communicate how other available information or alternative outcomes were considered in reaching a decision.[69]

In sum, the duty to give reasons in EU administrative law provides a useful starting point to think about the rights individuals have to obtain an explanation in cases where public authorities rely on AI to reach their decisions.[70] It makes blind 'rubberstamping' of AI recommendations more difficult by requiring explanations that go beyond a simple 'the machine said so'. A case might even be made for more extensive reasoning obligations for public authorities when they use AI in their decision-making to make up for the possible limits this poses to substantive legality review and to counter-act the risk of automation bias.[71] In any case, at least where AI is the only or main factor affecting the direction of a decision, a degree of explainability would seem necessary to produce an explanation that complies with the requirements under EU administrative law.

Given that EU administrative law is applicable across the board, including when AI systems are used, secondary law that establishes specific explanation rights for automated technologies is not absolutely necessary. However, secondary law could play an important role in specifying the exact extent to which

[68] N. Carr, *The Glass Cage: Where Automation is Taking us* (London: Bodley Head, 2014); L.J. Skitka, K.L. Mosier and M. Burdick, "Does Automation bias Decision-making?" (1999) 51 *International Journal of Human-Computer Studies* 991; S. Alon-Barkat and M. Busuioc, "Human-AI Interactions in Public Sector Decision-making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice" (2022) *Journal of Public Administration Research and Theory* (forthcoming).

[69] Other possible safeguards may include introducing a 'four eyes principle' where a second official has to approve a decision and the reasoning behind it before it enters into force.

[70] See also M. Oswald, "Algorithm-assisted Decision-making in the Public Sector: Framing the Issues using Administrative Law Rules Governing Discretionary Power" (2018) 376 *Philosophical transactions Series A Mathematical, physical, and engineering sciences* 3; Olsen, Slosser and Hildebrandt, "What's in the Box?" in *Constitutional Challenges in the Algorithmic Society* (2022), p.223.

[71] There are examples of legal systems that do contain higher explanation threshold when the decision-making process is automated. Noting this for the French legal system see Bibal et al, "Legal Requirements on Explainability in Machine Learning" (2021) 29 *Artificial Intelligence and Law* 149, 154–155.

the public authority's reasoning may rely on an AI system's recommendation and what that means for the degree of explainability required from the system. In addition, since the CJEU is not known for its easy accessibility to private litigants, more accessible fora to enforce explanation obligations would also provide important added value. The following section explores two important areas of secondary legislation—data protection law and the currently negotiated EU Artificial Intelligence Act—with a view to assessing their potential to complement the duty to reason in EU administrative law.

Explanation obligations for automated technologies in secondary legislation

Data protection law and the 'right to an explanation'

Over the past years, a vivid debate has unfolded about whether the General Data Protection Regulation ('GDPR') creates a 'right to an explanation'. Article 22 GDPR sets out a qualified prohibition on the use of 'solely automated decision-making' (which can include AI). In some cases, this prohibition does not apply, namely where such processing is necessary for a contract between the data subject and controller, where it is based on law, or where the data subject has consented explicitly. Article 22(3) foresees that in such circumstances, a non-exhaustive list of 'suitable' measures to safeguard the data subject's rights and freedoms apply, including a right to obtain human intervention. Confusingly, Recital 71 GDPR adds that in those situations the data subject additionally has a 'right to an explanation'. While discussions around this alleged right have focused on the GDPR, the latter does not apply to European institutions. Instead, Regulation 2018/1725 (the 'EUDPR') governs the processing of personal data by EU institutions, bodies, offices and agencies.[72] Importantly, both regulations largely overlap from a substantive perspective and their provisions in relation to solely automated decision-making are identical.[73]Article 24 EUDPR restates the GDPR's qualified prohibition of the solely automated processing of personal data where this produces legal effects or similarly significantly affects the data subject. Just like Recital 71 GDPR, Recital 43 EUDPR suggests that where one of the exceptions to this qualified prohibition applies, the data subject has the right 'to obtain an explanation'.

The bifurcation between the legislative texts and recitals is due to the fact that originally, a right to an explanation was considered for the legislative text but dropped in the course of trilogue negotiations.[74] This has been the source of ample debates in recent years. On the one hand, it is doubtful whether a legal right to an explanation actually exists, given that it is mentioned only in a non-legally binding recital but not the legislative text. On the other hand, it remains unclear what such a right would practically entail, if it existed. Whereas some consider that data protection law requires 'human-intelligible explanations of algorithmic decision-making'[75] others only understand it as an *ex-ante* right to information (about the use of AI) rather than an *ex-post* right to learn how the output was generated.[76] Tellingly, in light of its post-Brexit reform of data protection law the United Kingdom is planning to do away with this provision as the 'current efficacy and operation of Article 22 is subject to uncertainty'.[77]

In its (non-legally binding) guidelines on automated decision-making, the art.29 Working Party, the predecessor of the European Data Protection Board, stressed the importance of the 'right to an explanation' in relation to the broader principle of transparency, arguing that data subjects 'will only be able to challenge

[72] For a definition, see Regulation 2018/1725 art.3(10).

[73] Regulation 2018/1725 art.2(5).

[74] Edwards and Veale, "Slave to the Algorithm?" (2017) 16 *Duke Law & Technology Review* 50.

[75] B. Goodman and S. Flaxman, "European Union Regulations on Algorithmic Decision-making and a 'Right to Explanation'" (2017) 38 *AI Magazine* 50.

[76] Wachter, Mittelstadt and Floridi, "Why a Right to Explanation of Automated Decision-making does not Exist in the General Data Protection Regulation" (2017) 7 *International Data Privacy Law* 76.

[77] Department for Digital, Culture, Media & Support, *Data: A New Direction* (2021), p.38.

a decision or express their view if they understand how it has been made and on what basis'.[78] These guidelines also clarified that the information transmitted to the data subject should be sufficiently comprehensive to 'understand the reasons for the decision' but that this does 'not necessarily [require] a complex explanation of the algorithms used or disclosure of the full algorithm'. In this understanding, the 'right to an explanation' under the EUDPR resembles the duty to give reasons in EU administrative law, both in terms of the rationale for the existence of the obligation as well as the substance of the explanation required. That conclusion is also reinforced by the fact that the reference to an explanation in Recital 71 GDPR is followed by a reference to the fact that the data subject also has a right to 'challenge the decision', just as in administrative law one of the rationales of the right to a reasoned decision is to enable such challenges. However, due to its non-binding nature, it is unclear to what extent the art.29 Working Party's guidance will influence future judicial interpretations.

Despite the uncertainty and contradictions surrounding the provisions on solely automated processing in data protection law, the case can thus be made that any possible right to an explanation is very similar to the right to a reasoned decision in administrative law. In essence, the aim is to make decisions understandable so that they can be challenged and thus be subject to some form of control. This would also correspond to one of the core objectives of data protection law: to address power asymmetries between data subjects and data controllers. Importantly, however, even in this form, a 'right to an explanation' under data protection law would be more limited than its equivalent in EU administrative law, given that it is applicable only to fully automated decisions that legally or otherwise significantly affect the data subject.

In this light, it is paradoxical that discussions around the explainability of AI have focused almost exclusively on data protection law, neglecting not only obligations in administrative law, but also other areas of EU law where similar obligations exist, such as public procurement law, consumer protection law, and financial regulation.[79] The acknowledgement that explanation obligations already exist in other areas of EU law is important more generally, especially in the context of claims that EU data protection law should 'introduce' explanation requirements.[80] Only then can the interplay between general and sectoral requirements, as well as the advantages of one versus the other, be properly evaluated.

In any event, this area is riddled with controversy and uncertainty and future case law needs to clarify the existence and boundaries of a possible 'right to an explanation' in data protection law. If it is to have added benefit in the specific context of EU bodies' use of AI in fully or partially automated decision-making processes, it would be important to be aware where exactly the duty to give reasons in administrative law falls short so as to usefully complement it. In light of the previous section, aspects to focus on should include determining the exact requirements an explanation has to fulfil in the context of automation, specifying what this in turn means for the level of explainability required from AI systems, and creating pathways for individuals to enforce the 'right to an explanation'.

The EU Artificial Intelligence Act: ensuring explainability through law

The draft EU Artificial Intelligence Act (AIA) is currently being negotiated under the ordinary legislative procedure. If adopted, it creates a horizontal risk-based regime containing various kinds of legal obligations

[78] Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 [2018] WP251rev.01, p.27.

[79] Directive 2014/24 on public procurement [2014] OJ L094; Directive 2019/2161 amending Council Directive 93/13 and Directives 98/6, 2005/29 and 2011/83 as regards the better enforcement and modernisation of Union consumer protection rules [2019] OJ L328/7; Regulation (EU, Euratom) 2018/1046 on the financial rules applicable to the general budget of the Union [2018] OJ L193/1.

[80] Wachter, Mittelstadt and Floridi, "Why a Right to Explanation of Automated Decision-making does not Exist in the General Data Protection Regulation" (2017) 7 *International Data Privacy Law* 76, 90.

in relation to AI systems.[81] The AIA will also apply to European institutions, offices, bodies and agencies when acting as a provider or user of an AI system except where systems are developed or used for military purposes and their use falls under the exclusive remit of the Common Foreign and Security Policy.[82] The European Data Protection Supervisor will act as the competent authority for the supervision of EU institutions, agencies and bodies.[83]

The draft AIA creates different levels of legal obligations that apply to the creation and use of AI applications on the basis of their perceived risk. High-risk AI systems (HRAIS) are the riskiest category of AI and consequently subject to the most stringent legal obligations. This category includes applications of AI that can be used by public administration, such as systems used for access to and enjoyment of essential private and public services and benefits, law enforcement, systems used in migration, asylum and border control management (such as iBorderCtrl), or systems used for the administration of justice and democratic processes.

Two provisions of the draft AIA are of particular relevance here: art.52 and art.13. Article 52 creates a general transparency obligation that applies to AI systems 'intended to interact with natural persons' notwithstanding their specific risk level. In this context, natural persons must be informed that they are interacting with an AI system unless this is obvious from the circumstances.[84] This is for instance the case where an individual is subjected to an emotion recognition or biometric categorisation system.[85] This provision will not apply where AI is used as one element of a decision-making process where the citizen's point of contact remains a natural person. The obligation further does not apply 'to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence'.[86] Except for this very limited right to be informed, however, and in line with its overall product liability approach, the AIA creates no new rights and remedies for individuals that are subject to an AI-assisted decision by an administration.[87]

Article 13 AIA applies to high-risk AI systems and requires that these be 'designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately'.[84] Users must moreover have access to instructions for use containing information about 'the characteristics, capabilities and limitations of performance', including the level of accuracy, robustness and cybersecurity, risks to health, safety or fundamental rights, specifications for the input data, expected lifetime of the AI system and necessary maintenance measures.[85] Two aspects of art.13 are particularly noteworthy for our purposes.

First, the obligation applies to *providers* (i.e. those designing) of AI systems that must ensure that *users* can interpret the output. As such it does not create any new rights or remedies for those subject to AI-enabled administrative decisions. Secondly, art.13 does not create a right to an explanation (for neither the user nor those subject to AI-supported decisions), but rather an obligation of explainability. Stated otherwise, it requires that AI systems be built in a manner that allows for them to be explainable. While thus not creating an additional right to an explanation, the AIA facilitates the exercise of existing rights to an explanation by requiring high-risk applications of AI to be explainable by design and by default. This may have the additional benefit of facilitating public procurement processes as administrations no longer need to independently verify whether high-risk AI systems are interpretable but can assume this is so in light of this new legal obligation.

[81] An AI system currently being defined as 'artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

[82] Draft EU Artificial Intelligence Act Recital 12.

[83] EU Artificial Intelligence Act art.59(8).

[84] Draft EU Artificial Intelligence Act art.13(1).

[85] Draft EU Artificial Intelligence Act arts 13(2)–(3).

While much of the details are still unclear at this stage, in its current form, the AIA provides a useful complement to the duty to give reasons, but does not comprehensively address the gaps identified in the previous section. It can be anticipated that future case law will need to deal with questions such as what the requirement of interpretability implies from a technical perspective, whether providers can pre-suppose a certain level of technological literacy on behalf of users, and—relatedly—whether administrations will have to provide training to personnel using such systems.

Conclusion

The efficiency with which AI can process large amounts of data and its capacity to help public authorities to make informed decisions can play an important role in making the administration more efficient, fast, and fair. Yet, assumptions underlying AI systems may be generally flawed and biases may disadvantage particular groups. It is thus essential that decisions taken on the basis of suggestions by AI systems, just like any other decision, can be challenged and reviewed. This is only possible when the recipient of the decision, but also courts or other bodies tasked with ensuring legality, know and understand the reasoning behind the decision, which, in turn, requires some degree of explainability of the AI system used.

The duty to give reasons is a long-standing pillar of EU administrative law, requiring EU bodies to explain why they reached a certain decision. While developed outside the context of automation, it provides a useful starting point to determine the obligations EU bodies have when they rely on AI to support their decision-making, also because regulation adopted specifically for automated technologies has so far failed to guarantee such rights. Blind 'rubberstamping' of AI recommendations would certainly seem to not live up to the requirements under EU administrative law. A case might even be made for more extensive reasoning obligations, at least when AI systems substantially pre-determine a decision to make up for the limits this might pose to substantive legality review and to counter-act the risk of automation bias.

Having said this, the specific challenges the application of the duty to give reasons raises when public bodies rely on AI systems do justify additional safeguards. Given the discussion that has evolved around a possible 'right to an explanation' in data protection law, the existence of any such right in the context of public bodies relying on AI to support their decision-making should be made explicit for the sake of legal certainty. In addition, guidance on the benchmarks used to assess compliance with the right to a reasoned decision in the AI context would be helpful for both public authorities faced with obligation and for those affected by a decision. Such guidance should revolve around two related questions. What does the right to a reasoned decision actually require in terms of the nature and depth of the communication of reasons by the public authority that relied on an AI system? And what does that in turn require from the AI system's design in terms of its explainability? Finally, since access to the CJEU is notoriously difficult for individuals, alternative mechanisms to enforce explanation rights in the context of automation could go a long way in ensuring access to justice.

These aspects are central to an individual's possibility to challenge AI-supported decisions. For that reason, they should not be left to be worked out through litigation. It is in this respect that explanation obligations specific to automated technologies in secondary legislation could have real added value. However, neither data protection law nor the draft AIA do—in their current form—fully close this gap. Data protection might guarantee a 'right to an explanation', but its precise content remains highly uncertain and would in any case only apply to fully automated decision-making legally or otherwise significantly impacting the data subject, thus not covering the majority of use-cases of AI in the public sector. The proposed AIA, in turn, does not include any obligations for AI users to explain or justify their decisions towards those affected by them, even less a corresponding right on the part of individuals to demand that. It does helpfully set a clear obligation for producers to make high-risk AI understandable by humans, which facilitates the implementation of explanation rights between users of AI systems and those affected

by their decisions. Thus, while both frameworks do provide useful safeguards, they fall short of closing the specific gaps identified in this article.