Government by Algorithms at the *Light* of Freedom of Information Regimes

A Case-by-Case Approach on ADM Systems within Public Education Sector

MARÍA ESTRELLA GUTIÉRREZ DAVID*

ABSTRACT

What the Houston Court qualified as "mysterious 'black box' impervious to challenge" was in practice a sophisticated software of many layers of calculations, which rated teachers' effectiveness to make employment decisions. In the European Union, a system as such would fall under the Proposal for AI Regulation of 2021, which qualifies AI models in education and vocational training as "high-risk" systems. Automated decision-making systems (ADM systems), AI-driven or not, are being increasingly used by governments in public education for different purposes, such as handling applications for undergraduate admission or profiling students and teachers to assess their performance. Across cases and jurisdictions, there is growing evidence of how the use of ADM systems in the education sector is becoming quite problematic: arbitrary assignment of teaching posts in mobility procedures, undue barriers to access undergraduate studies, and frequent lack of transparency in their implementation and decisions. This Article discusses how Freedom of Information Act (FOIA) regimes may contribute to rendering governments' ADM systems (AI-driven or not)

* The author is Associate Professor of Information Law and Personal Data Protection at the Universidad Complutense de Madrid, Departmental Section of Constitutional Law at the Faculty of Communication Sciencies. This work is part of the research projects "The impact of artificial intelligence in public services: a legal analysis of its scope and consequences in healthcare" (PGC2018-098243-B-I00), and "Artificial Intelligence in the national health care system: solutions to specific legal problems" (PID2021-128621NB-100), directed by José Vida Fernández and founded by the Ministry of Science and Innovation of Spain (MCIN/AEI/10.13039/501100011033/) and by "FEDER: A way of making Europe."

Indiana Journal of Global Legal Studies Vol. 30 #1 (Winter 2023)

accountable. The analysis of the FOIA cases (Parcoursoup saga in France, MIUR in Italy, and Ofqual in the United Kingdom) shows to what extent decisions granting access to the source code, functional and technical specifications, or third-party audits allow public scrutiny of ADM systems, detection of their pathologies, and better understanding of their adverse impacts on rights and freedoms, individual or collective. This Article also addresses the constitutional value of the right of access to public records (Parcoursup), and the importance of proactive and mandatory public dissemination to ensure traceability, transparency, and accountability of the ADM systems for FOIA purposes. In this sense, some legal initiatives across jurisdictions (Canada, France, Spain, United States, European Union) enhancing transparency and accountability of algorithmic systems will be examined.

I. INTRODUCTION

Governments around the world have immersed themselves in the automation and algorithmization of their activities, as this is seen as "a way to improve, increase efficiency or lower costs of public services."[1] The so-called "Administration 4.0" is characterized by automation, interconnection, and artificial intelligence (AI), which is capable of performing complex calculation operations in a short time and emulating, to a certain and limited extent, the functioning of the human mind.[2]

In the public sector, AI techniques, such as machine learning and deep learning (ML and DL, respectively) have a very wide field of application: taxpayer profiling to predict the risk of fraud in relation to tax deductions or public aids applied;[3] predictive policing, crime mapping, and offender risk assessment;[4] identification of buildings that

1. ADA LOVELACE INSTITUTE, AI NOW INSTITUTE & OPEN GOVERNMENT PARTNERSHIP, ALGORITHMIC ACCOUNTABILITY FOR THE PUBLIC SECTOR, https://www.opengovpartnership.org/wp-content/uploads/2021/08/algorithmic-accountability-public-sector.pdf (2021).

2. Davide Ponte & Giulia Pernice, *L'intelligenza artificiale e l'algoritmo a contatto col diritto amministrativo: rischi e speranze* [Artificial Intelligence and the Algorithm in contact with Administrative law: Risks and Hopes], CONSIGLIO DI STATO, GIUSTIZIA AMMINISTRATIVA (2021), https://www.giustizia-amministrativa.it/web/guest/-/ponte-pernice-l-intelligenza-artificiale-e-l-algoritmo-a-contatto-col-diritto-amministrativo-rischi-e-speranze (It.).

3. Marlies Van Eck, *Algorithms in Public Administration*, BESTUURECHT & AI, (Jan 31, 2017), https://marliesvaneck.wordpress.com/2017/01/31/algorithms-in-public-administration/ (Neth.).

4. ALEXANDER BABUTA, MARION OSWALD & CHRISTINE RINIK, ROYAL UNITED SERVICES INSTI. & UNIV. OF WINCHESTER, MACHINE LEARNING ALGORITHMS AND POLICE DECISION-MAKING. LEGAL, ETHICAL, AND REGULATORY CHALLENGES 5–9 (2018), https://static.rusi.org/201809_whr_3-18_machine_learning_algorithms.pdf.pdf

should be subject to administrative inspection, or traffic light control to optimize traffic flow in cities;[5] prediction of vulnerabilities of homeless families in order to design social care policies providing for provisional shelters or permanent housing;[6] and implementation of "4P medicine"—personalized, preventive, predictive, and participatory medicine—for early detection of pathologies and adoption of tailored therapeutic strategies for each patient, or predisposition to certain diseases in order to deliver specific and timely prevention.[7]

On the one hand, many algorithmic systems—especially those based on ML and DL—are designed and deployed on the very same assumption: looking at the past to find patterns for making predictions or recommendations. On the other hand, this assumption seems to be quite sensitive when applied to individuals or collectives, because looking at their past behavior in a certain context (job, education, health, fulfilment of legal obligations) will give only an approximate indication of how they will behave in the future.[8]

In fact, in the ML realm, major learning algorithms (e.g., KNN, decision trees, or Bayesian networks) are universal in the sense that, given the appropriate data, they can learn anything. But learning from data requires making assumptions, and "different learners make different assumptions, which makes them good for some things but not others."[9]

Sometimes assumptions, data, learning models, and purposes are not only inappropriate for the intended use cases but also have adverse effects. Indeed, there is growing evidence demonstrating how "algorithmic systems in public service delivery can cause harm," while at the same time these systems are severely affected by "the frequent lack of transparency in their application, including opacity around decisions about whether and why to use them."[10]

---

5. Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 5, 1147, 1161 (2017).

6. CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY, 167, 181 (2016).

7. EUROPEAN COMMISSION, COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE, AND THE COMMITTEE OF THE REGIONS 3 (2020), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0066&from=ES.

8. *See* Sandra Wachter, Brent Mittelstadt & Chris Russell, *Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law*, 123 W. VA. L. REV. 735, 738 (2021).

9. PEDRO DOMINGOS, THE MASTER ALGORITHM. HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE YOUR WORLD 24 (2018).

10. ADA LOVELACE INSTITUTE, *supra* note 1, at 7.

A. The State of the Art of ADM Systems: The Human Rights Impact and the Black Box Problem

From a human rights approach, automation and algorithmization operated by public and private organizations are escalating the existing risks while creating new ones for rights and freedoms of citizens.[11]

Adverse impacts on human rights have been associated with the amplification of discrimination and social biases,[12] loss of privacy,[13] harmful consequences associated with facial recognition[14] and criminal

11. COUNCIL OF EUROPE, ALGORITHMS AND HUMAN RIGHTS: STUDY ON THE HUMAN RIGHTS DIMENSIONS OF AUTOMATED DATA PROCESSING TECHNIQUES AND POSSIBLE REGULATORY IMPLICATIONS 3–4 (2018), https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5 (explaining thoroughly why beyond the general concerns on opacity and unpredictability, there is an increasing awareness that specific human rights, such as fair trial and due process, privacy and data protection, freedom of expression, freedom of assembly and association, effective remedy, prohibition of discrimination, social rights and access to public services, or the right to free elections, are being particularly affected by algorithmic systems).

12. Wachter, Mittelstadt & Russell, *supra* note 8, at 741–44 (contending that most important categories of problematic bias in machine learning are "social bias" and "technical bias"; the ignorance of social bias such as historical inequality in society is very likely to lead to technical biases in the design of the automated system).

13. LORENZO COTINO, *Nuevo paradigma en las Garantías de los Derechos Fundamentales y una Nueva Protección de Datos frente al Impacto Social y Colectivo de la Inteligencia Artificial*, *in* DERECHO Y GARANTÍAS ANTE LA INTELIGENCIA ARTIFICIAL Y LAS DECISIONES 69–105 (2022) (referring to the so-called "paradox of privacy", and emphasizing, on the one hand, how citizens usually express concern about the way their personal data is processed and, on the other, their willingness to protect their privacy; however, their actual behavior do not match very often that willingness, as short-term rewards lead them to consent massive and harmful processing of their personal data in exchange for accessing digital services). *See also* EUROPEAN DATA PROTECTION BOARD & EUROPEAN DATA PROTECTION SUPERVISOR, JOINT OPINION 5/2021 ON THE PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONIZED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) 2–3 (June 18, 2021), https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf (widely welcoming the risk-based approach underpinning the EU Proposal for AI Regulation, but claiming its better alignment with the European General Data Protection Regulation ("GDPR") in some areas such as: the concept of "risk to fundamental rights"; the exclusion of international law enforcement cooperation from the scope of the Proposal; further requirements for controllers of high risk AI systems; or the lack of a general ban on any use of AI for automated recognition of human features in publicly accessible spaces to infer emotions or categorize individuals on grounds of ethnicity, gender, political or sexual orientation, or other grounds of discrimination).

14. INFORMATION COMMISSIONER'S OFFICE, ICO INVESTIGATION INTO HOW THE POLICE USE FACIAL RECOGNITION TECHNOLOGY IN PUBLIC PLACES 3, 31–32 (2019), https://ico.org.uk/media/about-the-ico/documents/2616185/live-frt-law-enforcement-report-20191031.pdf (highlighting that women and ethnic minorities are more prone to false positives, after having investigated the use of life facial recognition technology in two pilots undertaken by the Metropolitan Police Service first deployed at the Notting Hill

risk assessment,[15] or misinformation,[16] among others.[17]

These risks are even more exacerbated because the outcomes of many automated decision-making systems (ADM systems)—especially those based on AI models—are unintelligible, insofar as "the decision[-making] process is a black box."[18] The *black box* problem can be defined as "an inability to fully understand an AI's decision-making process and the inability to predict the AI's decisions or outputs."[19] And this is so, even for the human expert who designed the system.

Even though such systems can produce statistically reliable results, the end-user (e.g., public administrations) "will not necessarily be able to explain how these results have been generated or what particular features of a case have been important in reaching a final decision,"[20] thus raising "accountability concerns," especially in critical areas such

---

Carnival in August 2016 and South Wales Police's at the UEFA Champions League Final in June 2017). *See also* PATRICK J. GROTHER, MEI L. NGAN & KAYEE K. HANAOKA, NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, FACE RECOGNITION VENDOR TEST PART 3: DEMOGRAPHIC EFFECTS 2-3 (2019), https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf (concluding that, in domestic law enforcement images, the highest false positives were in American Indians, with elevated rates also in African American and Asian populations; being higher in women than in men, and even more elevated in the elderly and children).

15. Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (analyzing discriminatory bias on grounds of race incurred by software used across the United States, such as COMPAS, to predict future criminals). *See also* BABUTA ET AL., *supra* note 4, at 7–8, 21, 24 (describing the specific risks posed by ML models in policing context—unfairness and discriminatory bias, use of *black box* models or cost ratios of different types of error to predict individuals' proclivity for future crime; and urging caution when using proxy variables or historic data to produce forecasts based on new and unfamiliar data for policing purposes, giving careful consideration to the representativeness of the dataset used to train the algorithm).

16. HOUSE OF COMMONS DIGITAL, CULTURE, MEDIA AND SPORT COMMITTEE, DISINFORMATION AND 'FAKE NEWS': INTERIM REPORT 11, 18–21 (2018), https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/363/363.pdf (explaining the role of bots and algorithms in spreading fake news and their potential risks to the values and integrity of democratic systems).

17. MILES BRUNDAGE ET AL., TOWARD TRUSTWORTHY AI DEVELOPMENT: MECHANISMS FOR SUPPORTING VERIFIABLE CLAIMS 4 (2020), https://arxiv.org/abs/2004.07213 (summarizing all previous risks).

18. Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 3 (2017).

19. Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 2, 889, 905 (2018). *See generally* EUROPEAN PARLIAMENT RESOLUTION OF OCTOBER 20, 2020 WITH RECOMMENDATIONS TO THE COMMISSION ON A FRAMEWORK OF ETHICAL ASPECTS OF ARTIFICIAL INTELLIGENCE, ROBOTICS AND RELATED TECHNOLOGIES (2020/2012(INL)).

20. THE ROYAL SOCIETY, MACHINE LEARNING: THE POWER AND PROMISE OF COMPUTERS THAT LEARN BY EXAMPLE 93 (2017).

as law enforcement, health, or education.[21]

In a context of growing automatization and algorithmization of public administrations, the so-called "algorithmic opacity" poses an undeniable "serious problem for judicial control and a risk of abandonment of the core principles governing public administration,"[22] and may lead to "dismant[ling] critical procedural safeguards at the foundation of administrative law." [23]

It is not by chance that the Italian State Council has asserted that "[t]he use of 'robotized' procedures cannot be a reason for circumventing the principles that shape our legal system and that govern the administrative activity."[24]

B. Discussion and Topics

Domingos has explained in a very wise manner that "[w]hen a new technology is as pervasive and game changing as machine learning, it's not wise to let it remain a black box."[25] Though the author refers to ML models, it is not unusual to see how courts dealing with government decisions made by ADM systems do often refer to them as "black boxes," regardless of whether the decisions reached rely on AI-models or not.[26]

---

21. DAVID FREEMAN ENGSTROM, ET AL., GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES 28 (2020), https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf

22. Manuel Fernández Salmerón, Address at the Universidad Carlos III & Indiana Journal of Global Legal Studies Conference on Digital Transformation of Government: The Risk of Government: "The Risk of a Government as a Big Brother" (June 23–24, 2022) (Spain).

23. Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1253 (2008).

24. Cons. Stato, Sez. VI, 8 April 2019, n. 2270/2019, §8.2; *see also* December 13, 2019, n. 8472/2019, §10.

25. DOMINGOS, *supra* note 9, at xvi.

26. An example of an statistical model qualified as "black box" can be seen at Raad van State 17 May 2017, n. 201600614/1/R2 & others, ECLI:NL:RVS:2017:1259, §14.3 (May 17, 2017) (Neth.) (ruling, *obiter dicta* and for the first time, on a semi-automated procedure, where the predictions of the algorithm were used to support the decisions of some municipalities for granting or dismissing licenses to livestock farms to make nitrogen depositions; finding that the software in question resulted in an "unequal procedural position of the parties", due to the "lack of insights of the choices made, as well as the data and assumptions used" by the algorithm; and thus concluding that the software could be regarded as "a black box" from the standpoint of the addressees who "cannot check on the basis of which a particular decision has been reached.")
Though from the facts of the case, it is not clear enough if the system at stake used DL or decision trees techniques, an example of an AI-driven model, also qualified as a "black box", can be found at Rechtbank Den Haag 6 March 2020, n. C-09-550982-HA ZA 18-388, ECLI:NL:RBDHA:2020:1878, §6.53, §§6.89-6.90 (Neth.) (considering that a technical

In line with Domingos' approach, our starting point is the fact highlighted by the Spanish authority on freedom of information, the State Council of Transparency and Good Governance (CTBG), that algorithmization of governments is resulting in "a growing public demand for transparency of the algorithms used by public administrations as an inexcusable condition to preserve accountability and oversight of the decisions made by public authorities and, ultimately, as an effective guarantee against arbitrariness or discriminatory biases in fully or partially automated decision making."[27]

There is enough evidence that ADM Systems, AI-driven or not, are being used in critical sectors, such as public education. Most relevant Freedom of Information Act (FOIA) cases analyzed in this Article show how governments are currently deploying ADM systems in the public education sector.

There are several reasons that justify addressing the potential adverse impacts of algorithmic systems on education. First, public service missions are one of the most important administrative activities, and public education guarantees the exercise of the human right to education. Second, ADM systems are being used in public education for different purposes, such as handling applications for undergraduate admission or profiling students and teachers to assess their performance. Third, there is growing evidence of quite problematic uses of ADM systems and algorithmic processing in public education because of the existence of discriminatory bias and adverse impacts: arbitrary assignment of teaching posts in mobility procedures, undue barriers to access undergraduate studies, non-renewal or termination of contracts, and frequent lack of transparency in their implementation and decisions.

In this Article, we will argue to which extent FOIA regimes may contribute to rendering government's ADM systems (AI-driven or not) accountable in two ways. Either by disclosing, at the request of any citizen seeking access, the source code, the algorithms and/or the ancillary documents explaining them (the right of access), or by making available to the public relevant information thereof, either proactively or under statutory obligations provided in FOIA or sectoral legislation

---

infrastructure called SyRI—used by the Netherlands Government to generate risk reports of individuals in order to prevent and combat fraud in the fields of social security, taxes, and labor—was inherently a *black box* type; emphasizing the inability of the Court to verify how a simple decision tree was generated by the system; and stressing the difficulties to comprehend how the affected person could be able to defend himself or herself "against the fact that a risk report has been submitted about him or her.")

27. Consejo de Transparencia y Buen Gobierno [CTBG] [Council of Transparency and Good Governance], May 5, 2021, Decision R/0058/2021, II(5) (Spain).

(public disclosure schemes).

Part II discusses the qualification of the source code and the algorithms as public records under FOIA regimes. In this sense, there is a growing consensus on their public records status, unless a statutory exemption is applied.

Part III analyzes how, in some civil law jurisdictions (Italy, Germany, or Spain), the legal status of computer programs and algorithms used by public administrations for decision-making has been long discussed by scholars, courts, and authorities. Recently, this debate has escalated even more due to the exercise of the right of access to the source code and algorithms held by public authorities. Some relevant cases on ADM systems discussed in Italy and in France under domestic FOIA regimes evidence the nature and extent of the arguments raised about this topic.

Some argue that FOIA regimes are not the appropriate instruments to guarantee adequate transparency of ADM systems. In this sense, Part IV is entirely devoted to presenting some counterarguments against the alleged futility of FOIA regimes to open the *black box*. The analysis of the FOIA cases (MIUR in Italy and Ofqual in the United Kingdom) shows how the decisions that grant access to public records—not only the source code or the algorithm, but also the functional and technical specifications, or third-party audits—allow public scrutiny of ADM systems, detection of their pathologies, and better understanding of their adverse impacts, individual or collective.

If Parts III and IV are focused on the facet of the right of access, Part V is entirely devoted to public disclosure schemes. In doing so, this part analyzes the constitutional value of the right of access to algorithms, and the importance of proactive or mandatory public dissemination to ensure traceability, transparency, and accountability of the ADM systems for the purposes of FOIA goals. On this occasion, some FOIA cases—as the Parcoursup saga in France—are again a pretext to discuss these topics. This Article will discuss various legal initiatives across jurisdictions (Canada, France, Spain, United States, European Union) to enhance transparency and accountability of algorithmic systems.

C. Terminology and Methodology

For the purposes of this Article, the FOIA cases analyzed and systematized deals with ADM systems used by public administrations that apply a wide range of algorithmic procedures. From the facts of the cases here documented, it is not always possible to discern (e.g., local and national algorithms of Parcoursup) whether AI techniques have

been implemented instead of more traditional methods.

This being so, it is worth noting that there is still little consensus on a general and universal definition of AI, neither within the scientific community nor across international and national organizations.[28] References to this concept usually encompass two meanings of AI: both as a science and a technology, according to the definition provided by the National Institute of Standards and Technology (NIST).[29]

In **Table 1** we include some common definitions of AI systems provided by regulators and organizations[30]:

---

28. DEPARTMENT FOR DIGITAL, CULTURE, MEDIA & SPORT, DEPARTMENT FOR BUSINESS, ENERGY & INDUSTRIAL STRATEGY, AND OFFICE FOR ARTIFICIAL INTELLIGENCE, ESTABLISHING A PRO-INNOVATION APPROACH TO REGULATING AI 12 (July 20, 2022), https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement.

29. NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, U.S. DEP'T COMMERCE, U.S. LEADERSHIP IN AI: A PLAN FOR FEDERAL ENGAGEMENT IN DEVELOPING TECHNICAL STANDARDS AND RELATED TOOLS 25 (Aug. 9, 2019), https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf (NIST has embraced the AI's twofold definition proposed by the American National Standard Dictionary of Information Technology (ANSI): "(1) A branch of computer science devoted to developing data processing systems that performs functions normally associated with human intelligence, such as reasoning, learning, and self-improvement. (2) The capability of a device to perform functions that are normally associated with human intelligence such as reasoning, learning, and self-improvement".)

30. OECD, RECOMMENDATION OF THE COUNCIL ON ARTIFICIAL INTELLIGENCE (May 25, 2019); INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, ISO/IEC 22989:2022 (EN), INFORMATION TECHNOLOGY—ARTIFICIAL INTELLIGENCE (2022); EUROPEAN COMMISSION, PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, COM(2021) 206 final (Apr. 24, 2021) (hereinafter *EU Proposal for AI Regulation*); HIS MAJESTY'S GOVERNMENT, INDUSTRIAL STRATEGY: BUILDING A BRITAIN FIT FOR THE FUTURE 45, 132 (2017), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664563/industrial-strategy-white-paper-web-ready-version.pdf; 15 U.S.C. § 9401 (3). For a wider definition, *see* also John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115–232, §238(g), 132 Stat. 1636 ("In this section, the term 'artificial intelligence' includes the following: (1) Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets. (2) An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action; (3) An artificial system designed to think or act like a human, including cognitive architectures and neural networks. (4) A set of techniques, including machine learning, that is designed to approximate a cognitive task. (5) An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting.")

Organization	Definition
Organization for Economic Co-operation and Development	An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.
International Organization for Standardization	An AI system is an engineered system that generates outputs such as content, forecasts, recommendations, or decisions for a given set of human-defined objectives. These systems can use various techniques and approaches related to AI to develop a model to represent data, knowledge, processes, etc., which can be used to conduct tasks. AI systems are designed to operate with varying levels of automation, which entails pertaining to a process or system that, under specified conditions, functions without human intervention.
European Commission	An AI system means software that is developed with one or more of the techniques and approaches listed in Annex I of the EU Proposal for AI Regulation 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. AI techniques and approaches listed in Annex I are: (a) machine learning approaches, including supervised, unsupervised, and reinforcement learning, using a wide variety of methods including deep learning; (b) logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines (symbolic), reasoning and expert systems; (c) statistical approaches, Bayesian estimation, search and optimization methods.
UK Government	Technologies with the ability to perform tasks that would otherwise require human

	intelligence, such as visual perception, speech recognition, and language translation.
United States Code	Machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems use machine and human-based inputs to: (a) perceive real and virtual environments; (b) abstract such perceptions into models through analysis in an automated manner; and (c) use model inference to formulate options for information or action.

It is important to note that Annex I of the EU Proposal for AI Regulation also includes "statistical approaches" among the list of AI techniques and approaches.[31] This is crucial because some of the automated systems that we will discuss in this Article are not based on learning algorithms but rather on statistical approaches (Ofqual).

But regardless of the technique implemented, the outcomes produced by AI or statistical models are always predictions based on prior assumptions, variables, and criteria that do not always respond to a causal relationship or, if they do, such causality is not self-evident from the results.[32] Moreover, in machine learning contexts, it is common to hear "correlation instead of causation."[33] And this is critical where an administrative decision in adjudication processes is at stake. In fact, what some of the FOIA requests reveal is precisely the lack of statistical accuracy of the outcomes (predictions) and its adverse individual and social impacts on the governed. This is the case of the Ofqual's algorithm that we will discuss later.

From the OECD and ISO definitions, it is clear that AI systems are usually designed to operate with varying levels of automation: to

31. EUROPEAN COMMISSION, ANNEXES TO THE PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, COM(2021) 206 final ANNEXES 1 to 9 (Apr. 21, 2021.)

32. NICOLAS DIAKOPOULOS, ALGORITHMIC ACCOUNTABILITY REPORTING: ON THE INVESTIGATION OF BLACKBOXES 18 (2013) (explaining that correlations between data found by algorithms "do[] not imply causation, nor intent on the part of the designer." *See also* INFORMATION COMMISSIONER'S OFFICE, BIG DATA, ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DATA PROTECTION ¶118 (2017) (highlighting the relevant distinction between correlation and causation, and urging organizations using machine learning algorithms to discover associations "to appropriately consider this distinction and the potential accuracy (or inaccuracy) of any resulting decisions.")

33. COUNCIL OF EUROPE, *supra* note 11, at 37.

support in the decision-making process or to make the decision. This ultimately explains why some national legislations have endorsed a broad notion of "automated decision-making system" (ADM systems) in the context of administrative decisions.

For instance, in Canada, ADM systems encompass "any technology that either *assists or replaces the judgement of human decision-makers*. These systems draw from fields like statistics, linguistics, and computer science, and use techniques such as rules-based systems, regression, predictive analytics, machine learning, deep learning, and neural nets."[34]

In the United States, a recent bill sponsored by Representative Yvette D. Clarke—to require the impact assessments of automated decision systems and augmented critical decision processes—defines ADM systems as "any system, software, or process (including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques and excluding passive computing infrastructure) that uses computation, the result of which serves as a basis for a decision or judgment."[35]

To illustrate our core discussion and related topics, the Article will analyze and systematize a series of FOIA legislation and cases from different jurisdictions (Italy, France, United Kingdom, Spain, Germany, Netherlands, United States, Canada). Of particular interest is the doctrine emanated from the independent authorities upholding information rights, such as the Commission d'Accès aux Documents Administratifs in France (CADA) or the Information Commissioner's Office in the UK (ICO). Our comparative approach also resorts to constant references to prominent case law seeking to enrich discussion and topics.

II. SOURCE CODE AND ALGORITHMS AS PUBLIC RECORDS

The right of access to public records guarantees the ultimate goals of FOIA regimes—namely, strengthening "the principle of democracy and respect for fundamental rights"[36] by opening "agency action to the light

34. Directive on Automated Decision-Making, 2019 (Can.) [hereinafter *ADM Directive*] (emphasis added).

35. Algorithmic Accountability Act of 2022, H.R.6580, 117th Cong. § 2(2) (2021-2022) (as introduced in the House on April 3, 2022) (applied only to a person, partnership, or corporation under the jurisdiction of the Federal Trade Commission and automated decision systems or augmented critical decision process that impact on consumers).

36. C-28/08, Eur. Comm'n v. Bavarian Lager, 2010 E.C.R. I-06055 ¶ 14; *see also* Cases C-39/05 P and C-52/05 P, Sweden and Turco v. Council, [2008] ECR I-4723, ¶45–46; T-233/09, Access Info Europe v. Council [2011], ECR II-1073, ¶57, aff'd C-280/11 P, Council

of public scrutiny"[37] and ensuring the citizenry has the right to know: "what their government is up to";[38] "how decisions that affect them are taken, how public funds are managed or under which criteria our public institutions act";[39] and whether or not "administration acts with greater propriety, efficiency and responsibility vis-à-vis the citizens."[40]

Some argue that the right of access under FOIA regimes would be insufficient to ensure compliance with the principle of administrative transparency in the context of ADM systems,[41] as the knowledge of the source code cannot guarantee a full openness of the algorithmic process due to the inability of citizens—usually non-experts—to understand the language of the machines, especially in the case of AI-based systems.[42]

Conversely, some scholars are of the view that the right of access to the source code and the underlying algorithm can contribute to fostering algorithmic transparency,[43] insofar as such access would imply "some degree of public scrutiny" of the automated systems used by public authorities.[44]

Furthermore, in relation to the use of AI algorithms by governments, the Federal and State Information Commissioners in Germany have encouraged this approach in a joint statement:

---

of the European Union v. Access Info Europe, [2013] ECR I-000, ¶32 (emphasizing the liaison between the right to access and the democratic system).

37. U.S. Dep't of Justice v. Tax Analysts, 492 U.S. 136, 142 (1989).

38. Nat'l Archives & Rec. Admin. v. Favish, 541 U.S. 157, 171–72 (2004); U.S. Dep't of Just. v. Rep.'s Comm. For Freedom of the Press, 489 U.S. 749, 773 (1989).

39. *See* Law on Transparency, Access to Public Information, and good Governance (B.O.E. 2013, 295) (Spain).

40. T-211/00, Kuijer v. Council Eur. Union, 2002 E.C.R. II-485, ¶ 52.

41. *See* Andrés Boix Palop, *Los algoritmos son reglamentos: la necesidad de extender las garantías propias de las normas reglamentarias a los programas empleados por la administración para la adopción de decisiones*, 1 REVISTA DE DERECHO PÚBLICO: TEORÍA Y MÉTODO, 223, 242; Julián Valero Torrijos, *Las garantías jurídicas de la inteligencia artificial en la actividad administrativa desde la perspectiva de la buena administración*, 58 REVISTA CATALANA DE DRET PÚBLIC 82, 89 (2019) (both authors are of the 117pinión that current Spanish State Law 13/2019, on Transparency, is a very restrictive instrument to guarantee the effective transparency of algorithmic systems used by public administrations).

42. Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. Pa. L. Rev. 633, 638 (2017).

43. *See generally* NICOLAS DIAKOPOULOS, ALGORITHMIC ACCOUNTABILITY REPORTING: ON THE INVESTIGATION OF BLACKBOXES 12 (2013), http://www.nickdiakopoulos.com/wp-content/uploads/2011/07/Algorithmic-Accountability-Reporting_final.pdf.

44. JOSHUA NEW & DANIEL CASTRO, HOW POLICYMAKERS CAN FOSTER ALGORITHMIC ACCOUNTABILITY 8 (Center for Data Innovation, May 21 2018), https://www2.datainn ovation.org/2018-algorithmic-accountability.pdf; *see also* Angelo Giuseppe Orofino, *The implementation of the Transparency Principle in the Development of Electronic Administration*, 1 EUROPEAN REVIEW OF DIGITAL ADMINISTRATIVE & LAW 1–2, 127, 130 (2020) (who insists that, despite technical incomprehensibility to the average citizen, access to the code must be guaranteed in any case).

> In accordance with the *principles of freedom of information and administrative transparency, information essential to the Government about the algorithms and AI processes it uses must also be made available to the public* . . . . It makes sense to embed corresponding transparency regulations *in the respective freedom of information or transparency laws* or in the relevant specialized laws. Exceptions should be kept to a minimum.[45]

In fact, it can be noted that source code and algorithms held by governments should be deemed as public records subject to FOIA regimes and, thus, accessible information,[46] except when they fall under a statutory exception (e.g., national security, trade secrets, law enforcement).[47]

In the United States, the status of a computer program as "agency records" for the purposes of the 1966 Freedom of Information Act has been decided according to the "particular nature and functionality of the software at issue"; [48] and more specifically, whether the access to the software in question provides "any insight into agency decision making."[49] To put it simply, "The question is whether they are most properly regarded as vessels of information (like data), on the one hand,

---

45. BUNDESBEAUFTRAGTE FÜR DEN DATENSCHUTZ UND DIE INFORMATIONSFREIHEIT *ET AL.*, POSITIONSPAPIER IM RAHMEN DER 36. KONFERENZ DER INFORMATIONSFREIHEITSBEAUFTRAGTEN IN DEUTSCHLAND – 'TRANSPARENZ DER VERWALTUNG BEIM EINSATZ VON ALGORITHMEN FÜR GELEBTEN GRUNDRECHTSSCHUTZ UNABDINGBAR' (Oct. 16, 2018) [Transparency of the Administration in the Use of Algorithms to ensure the Indispensable Protection of Fundamental Rights] (Ger.), https://www.datenschutzzentrum.de/uploads/informationsfreiheit/2018_Positionspapier-Transparenz-von-Algorithmen.pdf) (supporting the joint statement, the Federal Commissioner of Data Protection and Freedom of Information, and the State Commissioners of Berlin, Bremen, Mecklenburg-Western Pomerania, Rhineland-Palatinate, Saxony-Anhalt, Schleswig-Holstein, Thuringia and Baden-Württemberg.) (emphasis added). *See* also, DATEN KOMMISIONEN, GUTACHTEN DER DATENETHIKKOMMISSION, 215 (Dec. 2019) (Ger.), https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/itdigitalpolitik/gutachtendatenethikkommission.pdf?__blob=publicationFile&v=6(welcoming the joint statement).

46. *See* Orofino, *supra* note 44, at 125, 127–29.

47. THE UNITED STATES DEPARTMENT OF JUSTICE, FOIA UPDATE: DEPARTMENT OF JUSTICE REPORT ON "ELECTRONIC RECORD" FOIA ISSUES, PART II, FOIA UPDATE, vol. XI, n. 3 (Jan.1, 1990), available at https://www.justice.gov/oip/blog/foia-update-department-justice-report-electronic-record-foia-issues-part-ii.

48. THE UNITED STATES DEPARTMENT OF JUSTICE, GUIDE TO THE FREEDOM OF INFORMATION ACT. PROCEDURAL REQUIREMENTS 11–12, (Aug. 20, 2021), https://www.justice.gov/oip/doj-guide-freedom-information-act-0.

49. Baizer v. U.S. Dep't of Air Force*,* 887 F. Supp. 225, 228 (N.D. Cal. 1995).

or as mere tools (like hardware), on the other."[50]

For example, the Northern District Court of California analyzed a decision dismissing access in relation to "CLERVER," a videoconferencing software developed by a contractor of the Department of Energy (DOE) and subject to a non-exclusive license. In its decision, the District Court concluded that the software could not be regarded as a public record "[e]ven if DOE actually owned and controlled CLERVER at the time of . . . FOIA request . . . because CLERVER does not illuminate the structure, operation, or decision-making structure of DOE."[51]

In contrast, the District Court of Columbia concluded that computer software programs associated to the agency report, the Philen Study, withheld by the Centers for Disease Control under FOIA Exemption 5 (predecisional internal communications) were agency records because "[u]nlike generic word processing or prefabricated software, Dr. Philen's programs are uniquely suited to its underlying database," and a consequence of such tailoring is "the software's design and ability to manipulate the data reflect[s] the Philen Study." As a result, the computer programs in question "preserve information and 'perpetuate knowledge' that are responsive to plaintiff's FOIA request because of their relation to the Philen Study."[52]

Much before the legislature did so, the French Commission d'Accès aux Documents Administratifs had already qualified the source code of algorithms as administrative documents—*documents administratifs*—in several cases where the independent authority had to review administrative decisions dismissing the access sought by citizens to the source code or the algorithms used by public entities.[53]

---

50. THE U.S. DEP'T OF JUST., *supra* note 47.

51. Gilmore v. U.S. Dep't of Energy, 4 F. Supp. 2d 912, 920–21 (ND Cal. 1998).

52. Cleary, Gottlieb, Steen & Hamilton v. Health & Hum. Servs., 844 F. Supp. 770, 781–82 (D.D.C. 1993) (where plaintiffs had filed a suit against the Department of Health and Human Services and Centers for Disease Control seeking the release of "various forms of information, including computer searches, *statistical analyses*, printouts, and *software* [emphasis added]" connected to a study referred as "Philen Study" which had reported a possible link between the ingestion of L-tryptophan and a rare syndrome).

53. *See* Commission d'Accès aux Documents Administratifs [CADA] [Commission for Access to Adminstrative Documents], Oct. 16, 2014, 20142953 (where the access to a software developed by a private company to build the Musée des Confluences de Lyon was granted, excluding some redacted parts which were affected by commercial secrecy); [CADA] Jan. 8, 2015, 20144578 (upholding the access sought by a researcher to the source code of a software developed by the General Directorate of Finance to simulate the calculation of the income tax; [CADA] June 23, 2016, 201611990 (qualifying as an administrative document the algorithm developed by the French Ministry of Education, known as APB –*Admission Post-Bac*– for processing the applications for admission to university degrees, and upholding its accessible character) (Fr.).

The doctrine produced by the CADA was eventually codified by the French legislature in the so-called *Loi Lemaire* of 2016.[54] Accordingly, the Article L300-2 of the Code of Relations between the Public and the Administration (CRPA)[55] qualifies the source code used by an administration as an "administrative document."

In view of the foregoing, the CADA has qualified as "administrative documents" not only the source code[56] or the algorithms implemented by an administration,[57] but also functional and technical specifications related to them.[58] For example, with regard to the source code of the Parcoursup platform to manage the pre-registration applications for undergraduate studies, the French Authority has upheld the right of access to the software specifications, "presented in a synthesized manner."[59]

---

54. *See* Loi 2016-1321 du 7 octobre 2016 pour une République Numérique [Law 2016-1321 of October 7, 2016 for a Digital Republic], JOURNAL OFFICIEL DE LA REPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE] n. 0235, Oct. 8, 2016.

55. *See* Ordonnance n. 2015-1341 du 23 octobre 2015 relative aux dispositions législatives du code des relations entre le public et l'administration [Ordinance No. 2015-1341 of October 23, 2015, relating to the legislative provisions of the code on relations between the public and the administration], JOURNAL OFFICIEL DE LA REPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE] n. 248, Oct. 25, 2015 (codifying the main provisions of previous legislation on the right to communication of administrative documents, the reasoning of administrative decisions, and the rights of citizens before public administrations, in particular, the Law n. 78-753 of 17 July 1978 concerning various measures to improve relations between the administration and the public and various provisions of an administrative, social and fiscal nature, the Law n. 79-587 of 11 July 1979 relating to the reasoning of administrative acts and to the improvement of relations between the administration and the public, and the Law n. 2000-321 of 12 April 2000 relating to the rights of citizens in their relations with the administrations.)

56. *See* Commission d'Accès aux Documents Administratifs [CADA] [Commission for Access to Administrative Documents], Jan. 16, 2020, 20191797; CADA, Sept. 6, 2018, 20182093; CADA, Sept. 6, 2018, 20182120; CADA, Sept. 6, 2018, 20182455; CADA, Sept. 6, 2018, 20182682; CADA, Apr.19, 2018, 20180276; CADA, June 23, 2016, 20161990; CADA, June 23, 2016, 20161989 (decisions granting the right of access to source code). *But see* CADA, Mar. 12, 2020, 20200496; CADA July 18, 2019, 20181891; CADA, Jan. 10, 2019, 20184400; CADA, May 31, 2018, 20180376 (dismissing access to the source code) (Fr.).

57. *See* Commission d'Accès aux Documents Administratifs [CADA] [Commission for Access to Administrative Documents], Sept. 6, 2018, 20182093; CADA, Sept. 6, 2018, 20182120; CADA, Sept. 6, 2018, 20182455; CADA, Nov. 30, 2017, 20173235; CADA, Oct. 6, 2016, 20163835; CADA, June 23, 2016, 20161990; CADA, June 23, 2016, 20161989 (upholding the right of access to algorithm held by Public Authorities). *But see* CADA, Sept. 10, 2020, 20201743; CADA, Jan. 10, 2019, 20184400 (Fr.).

58. *See* Commission d'Accès aux Documents Administratifs [CADA] [Commission for Access to Administrative Documents], Jan. 10, 2019, 20184400; CADA, Sept. 6, 2018, 20182120 and CADA, Sept. 6, 2018, 20182455 (Fr.).

59. Commission d'Accès aux Documents Administratifs [CADA] [Commission for Access to Administrative Documents], Sept. 6, 2018, 20182093 (Fr.).

Likewise, the source code of the computer program used by the National Family Allowance Fund (Caisse Nationale d'Allocations Familiales, or CNAF) for the full calculation of social financial assistance has also been qualified by the CADA as an "administrative document," along with the SQL files of the source code and the functional specifications used to calculate the different social benefits (e.g., housing and family allowances, solidarity income).[60]

In Spain, the State Council of Transparency (Consejo de Transparencia), and the Regional Authority in Catalonia (GAIP), reviewing decisions withholding the source code or underlying algorithms of applications used by public bodies, have also qualified them as "public information" for the purposes of FOIA legislation.[61]

Under the Freedom of Information Act 2000 (FOI), the UK authority, the Information Commissioner's Office has upheld the disclosure of the specifications and data dictionary associated with a software used by Student Loans Company (SLC)[62] for monitoring loan recovery data, as access to such information would allow the complainant "to understand exactly what queries are automated, what the system is for obtaining a data dump of the data, etc."[63]

Legal exemptions usually applied to prevent access to the source

---

60. Commission d'Accès aux Documents Administratifs [CADA] [Commission for Access to Administrative Documents], July 18, 2019, 20181891 (Fr.).

61. State Council of Transparency, Decisions 058/2021, §4 (May 5, 2021) (Spain) (in relation to an algorithm used by the Ministry of Social Security to calculate future pensions of public officials and employees); RT/0253/2021 (Nov. 11, 2021) (Spain) (in relation to the source code of an application used for the drawing of lots for members in boards associated with selective processes in matters of education in the Autonomous Community of Madrid). See also, GAIP, Decision of Sept. 21, 2016 §3, upholding joint claims 123/2016 and 124/2016; 200/2017 §3 (June 21, 2017) (Catalonia, Spain) (upholding access sought by the claimant to the source code used to randomly appoint the members of the boards for assessing official exams which give access to universities in Catalonia); 93/2019 §3 (Feb. 22, 2019) (Catalonia, Spain) (upholding access to the source code and algorithms used for the same purposes that of the previous decision).

62. *See About Us*, STUDENT LOAN COMPANY, https://www.gov.uk/government/organisations/student-loans-company/about (last visited Aug. 10, 2022) (The Student Loans Company is a public company whose activity is to provide loans and grants to university and vocational students and to collect the repayments of these loans. It is a body owned by the UK Government's Department for Education, the Scottish Government, the Welsh Government, the Northern Ireland Department for the Economy, the Revenue and Customs Authority and the University and Vocational Admissions Service)

63. ICO, FS50323800 (Dec. 9, 2010) (U.K.), ¶19, 25 (ruling that SLC incorrectly applied section 36(2)(c) of the FOIA when dealing with the information request of the complainant because it had not given enough evidence of having sought the opinion of a qualified person, insofar as the applied provision establishes that information can only be exempt if "in the reasonable opinion of a *qualified person* disclosure would, or would be likely to, lead to adverse consequences in relation to the effective conduct of public affairs" [emphasis added].)

code, the underlying algorithm or technical specifications are the protection of public security (including the security of the administration's own information systems);[64] the prevention, investigation, and punishment of criminal, administrative, or disciplinary offenses (including the risk of circumvention of law);[65] or the protection of intellectual property and commercial interests.[66]

For example, in some decisions the Information Commissioner's Office (ICO) has been reluctant to grant access to the source code of the software used by the administration, as it considered that the commercial interests of the administration or that of third parties outweighed the public interest in better knowing how automation of administrative procedures may have social and individual impacts.[67]

III. THEIR DISPUTED LEGAL STATUS

In some civil law systems, the legal status of computer programs, including their source code and the underlying algorithms, used by public authorities for the full or partial automation of decision-making, has been quite problematic. Scholars have qualified them differently, as administrative acts,[68] rules,[69] or internal administrative instructions,

64. *See* Commission d'Accès aux Documents Administratifs [CADA] [Commission for Access to Administrative Documents], Oct. 20, 2016, 20163619; and CADA, Mar. 12, 2020, 20200496 (Fr.).

65. Sheridan v. U.S. Office of Pers. Mgmt., 278 F. Supp. 3d 11, 20 (D.D.C. 2017).

66. ICO, FS 50630372 (July 18, 2019) (U.K.).

67. ICO, FS50371026 (Jan. 9, 2012), ¶27, 32-33. *See* also FS50459127 (Mar. 4, 2013), ¶27, 30-31 (UK) (dismissing in both cases the access to a software, called LiMA, used by the Department for Work and Pensions (DWP), in the creation of the IB85 Incapacity for Work Medical Report form, despite having acknowledged the existing public concerns about how the LiMA software worked and its "impact on the lives of many people." The Commissioner was mindful of the "amount of public concern and media attention the issue of medical assessments" had generated, and how better understanding the ways in which decisions were made by the software in question would lead to a "better informed debate and potentially increased confidence in the process." But, in balancing the public interest in disclosure against the public interest in maintaining the exemption, in both cases the Commissioner afforded significant weight to the fact that the DWP had a contract with a third party at the time of the request, and disclosure sought by the claimant would be likely to prejudice the authority's commercial interest and that of the licensee within the meaning of section 43(2) exemption of the FOIA).

68. Jean-Bernard Auby, *Algorithmes et Smart Cities: Données Juridiques*, REVUE GENERALE DU DROIT 29878, 21 (2018) (Fr.) (notice that the French term *actes administratifs* would be the equivalent to the *orders* or final dispositions in adjudication processes, as defined in 5 U.S.C. Subch II §551(6)–(7)).

69. Boix Palop, *supra* note 41, at 234–238 (notice that the term *reglamentos* used by the author would be equivalent to the *rules* as defined 5 U.S.C. Subch II §551(4)).

for example.[70]

The discussion on this topic has arisen precisely on occasion of the growing use of automated decision-making systems by public administrations, assisted or not by AI algorithms, and the subsequent growth in requests for access to these digital assets under FOIA regimes.

In Italy, the argument in favor of the software and the underlying algorithm being qualified as an "administrative act" has been upheld on judicial review to guarantee access to administrative documents[71] under

---

70. Elena Buoso, *Fully Automated Administrative Acts in the German Legal System*, 1 EUR. REV. OF DIGIT. ADMIN. & L., 2, 113, 121 (2020) (It.) (summarizing the approach of German scholars, according to which the algorithms used in automated administrative acts should be qualified as *Verwaltungsvorschrift* or administrative instructions, provided that such qualification would require the public disclosure of the algorithm in order to guarantee the traceability of the decision-making process; although, this qualification is notably limited as its scope of application would be restricted to deterministic algorithms, and not to AI algorithms.) Notice that the term *Verwaltungsvorschrift* or administrative instructions would be equivalent, to a greater or lesser extent, to non-legislative rules, such as "policy statements" as referred in 5 U.S.C. § 552(a)(1)(D). *Compare* Michael Asimov, *Nonlegislative Rulemaking and Regulatory Reform*, 2 DUKE LAW J., 381, 383 (1985) (explaining that a policy statement indicates how an agency intend "to exercise discretionary power in the course of performing some other administrative function", for instance, providing guidance on the factors to be considered and the goals to be achieved when agency conducts formal or informal adjudication), *with* Herman Pünder & Anika Klafki, *Administrative Law in Germany*, in: COMPARATIVE LAW. ADMINISTRATIVE LAW OF THE EUROPEAN UNION, ITS MEMBER STATES AND THE UNITED STATES 49, 70-71 (René Seerden, ed., 2018) (according to German Law, *Verwaltungsvorschriften* are internal regulations within an administrative organization issued by a higher-level administrative authority to its subordinate bodies or employees. For instance, an administrative instruction may regulate how to grant a specific social benefit. *Verwaltungsvorschriften* have no legal effects upon third parties unless the fundamental right to equal treatment is infringed due to the lack of adherence to the consolidated administrative practice established by the administrative instructions. This legal status of citizens before administrative instructions and their right to equal treatment can be enforced by Administrative Courts within judicial review.)
Compare also this notion of *Verwaltungsvorschriften*, *with* the *instrucciones u órdenes de servicio*, *i.e*, instructions or service orders regulated in Article 6 of the Spanish Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, B.O.E. n. 236, October 02, 2015 [Law 40/2015, of October 1, on the Legal Regime of Public Sector], according to which administrative authorities may direct the activities of their hierarchically dependent organs by means of instructions and service orders. Instructions shall be published in the corresponding official gazette when a specific norm provides so, or it is deemed appropriate in relation to the addressees affected by the instructions, notwithstanding to their publication in accordance with the provisions of the Law 19/2013, on Transparency. Failure to comply with instructions or service orders shall not affect the validity of acts issued by administrative organs, without prejudice to any disciplinary liability that may be incurred by public officials.) *See also*, S.T.C., Apr. 20, 1983 (B.O.E. n. 117, May 17, 1983), §2.

71. Orofino, *supra* note 44, at 124–25.

the current legislation, namely, the Law n. 241 of 1990.[72]

In a landmark decision, the Regional Administrative Court Lazio-Rome (T.A.R. Lazio) has qualified an algorithm as a "computer administrative act," within the meaning of the Article 22.1.d of the Law 241 of 1990.[73] This algorithm was used by the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) to assign vacant positions to teaching staff according to the interprovincial mobility call for the academic year 2016/2017.[74]

According to the facts of the case, the MIUR had used a third-party algorithm developed by contractors to fully automate the interprovincial mobility process to identify and assign the specific vacant positions to individual applicants, resulting in thousands of teachers being displaced hundreds of miles away from their home province, despite the fact that there were vacancies much closer. After the subsequent public outcry, the Federazione Nazionale Gilda Unams, a representative union in the Italian education sector, sought access to MIUR's algorithm, but the Ministry dismissed the request on grounds that the algorithm could not be deemed an administrative document and it was protected by intellectual property rights.[75]

So, the issue at stake was to discern whether the disputed software and its underlying algorithm was an accessible administrative document for the purposes of the Law n. 241 of 1990.

The T.A.R. Lazio reasoned that the algorithm in question supported the administrative procedure *itself* because, in practice, the identification and assignment of the specific position to the individual teacher within the mobility procedure was performed solely and exclusively by the algorithm. For this reason, "the algorithm becomes truly a *straightforward expression of the activity carried out by the public administration* which is, undeniably, an activity of public

72. Legge 7 agosto 1990, n.241, G.U. Aug. 18, 1990, n.192, Nuove Norme in Materia di Procedimento Amministrativo e di Diritto di Accesso ai Documenti Amministrativi [Law of August 7, 1990 on New rules on Administrative Procedure and Right of Access to Administrative Documents].

73. This provision defines the administrative document as "any graphic, photo-cinematographic, electromagnetic or any other kind of representation of the content of [administrative] acts, internal or otherwise, relating to a specific procedure, held by a public administration and concerning activities of public interest, regardless of the public or private nature of its substantial discipline."

74. T.A.R. Lazio-Rome, Sez. III Bis, 22 Marzo 2017, n. 3769 2–4 (It.) (pursuant to the Article 3.1 of the Order 241/2016 issued by the MIUR, applications for the interprovincial mobility call during the 2016/2017 school year had to be submitted by the teaching staff at pre-school, primary and secondary school through the POLIS Portal, *Presentazione On Line delle Istanze*).

75. See *id*. at 8.

interest."[76] In consequence, all the interim activity of gathering the relevant data for moving forward with administrative procedure, including the issuance of the final administrative decision assigning the post to the individual teacher,

> do converge and are *exhausted in the mere operation of the algorithm* resulting in the *assimilation of the algorithm in question to the administrative act . . . or rather . . . the recognition of the direct attribution of the software that governs the algorithm to the category of the so-called computer administrative acts* referred to in letter d) of art. 22 of law n. 241 of 1990."[77]

Although the software had been developed by a private contractor, it was directly attributable to the administration, insofar as the software in question was designed in accordance with the criteria and purposes of an administrative nature on the basis of precise indications given by the MIUR. Accordingly, the Court concluded that the software materialized the final will of the public administration: "it is with the software that, ultimately, *the administration constitutes, modifies or extinguishes individual legal situations."*[78]

For its part, the Italian State Council has not taken a clear position on the issue. While in a first judgment, the Council agrees with the opinion of the T.A.R. Lazio in qualifying the MIUR's algorithm as a "computer administrative act";[79] in another later decision, it is more

76. *See id*. at 9 (emphasis added).

77. *See id*. at 9–10 (emphasis added).

78. *See id*. at 14–15 (emphasis added).

79. Cons. Stato, Sez. VI, 8 aprile 2019, n. 2270, 8.2 ("The technical rule governing each algorithm remains a general administrative rule, designed by a human and not by a machine, to be then applied (even exclusively) by the machine." Accordingly, the Council first contends that this "algorithmic rule" has a full administrative value, even if it is expressed in a mathematical manner; and, therefore, it must be subject to the general principles of the administrative activity, such as those of publicity and transparency (Art. 1 l. 241/90), reasonableness and proportionality. In the second place, the algorithm shall not give rise to discretionary applications (not admissible in the field of programming), but it shall reasonably provide for a well-defined solution for all possible cases, even the most improbable (as this feature distinguishes the algorithm from many general administrative rules). In addition, the Administration is to play an *ex ante* role by constantly testing, updating and adjusting the algorithm (especially, in the cases of machine learning and deep learning). In the fourth place, the algorithmic rule shall provide for an adequate judicial control, given that the "robotized decision requires the judge to assess the correctness of the automated process in all its components." Finally, the Council comes to the conclusion that "*the algorithm, namely, the software, shall be deemed to all legal effects as a computerized administrative act* [emphasis added]." However, it should be noted that

inclined to consider the disputed algorithm as a tool at the service of administrative activity.[80]

Scholars have echoed the different approaches taken by Italian courts. On the one hand, some scholars have qualified the computer programs used by the administration as "instruments of administrative action," mere technical tools usually designed by contractors of the administration, who merely execute the instructions given by the contracting authority. On the other hand, computer programs to automate the decision-making process are considered as true "administrative acts," insofar as they would express the will of the administration which would be conditioned to the occurrence of the factual premise previously identified and defined by the program in question.[81]

For example, Orofino is of the view that computer programs cannot be considered administrative acts but rather they, at best, constitute the object of an administrative will: "the will of making of one's own, upon performing the functions, the decisions made by the machine." Moreover, if administrative acts are declarations of the will dictated by the authority, the declaration, to be such, must be communicated in a way that allows the addressees to understand its meaning. In that case, being the software, a set of signs written in a programming language, usually unintelligible for a layperson, it cannot be considered an administrative act. For this reason, the software is rather an "instrument of administrative action."[82]

In contrast, for Cavallaro and Smorto, the approach taken by the T.A.R. Lazio on the (technical) role of the algorithm in the allocation of the specific teaching position, due to its very characterization—a

---

the arguments posed by the Council are quite confusing because it first seems to identify the algorithm with a "technical rule" (*regola tecnica*)—and, more specifically with a "general administrative rule" (*regola amministrativa generale*) and, then qualifies it as "computer administrative act." Furthermore, the ruling equates the algorithm to the software, which is technically inaccurate.)

80. Cons. Stato, Sez. VI, 13 dicembre 2019, n. 8472, 10 (ruling that the use of the algorithm must be properly framed "in terms of an organizational module." a "procedural tool" *strumento procedimentale ed istruttorio* "subject to the verifications typical of any administrative procedure, which remains the modus operandi of the authority's decision, to be carried out on the basis of the law bestowing the power on the public body, holder of the power, and the [public] ends defined according to that law.")

81. Giorgio Mancosu, *Les algorithmes publics déterministes au prisme du cas italien de la mobilité des enseignants* [Deterministic Public Algorithms through the Prism of the Italian Case of Teacher Mobility], 1 RIVISTA ITALIANA DI INFORMATICA E DIRITTO 75, 79 (2019) (Fr.) (summarizing the Italian doctrine on the issue).

82. Angelo Giuseppe Orofino, *The implementation of the Transparency Principle in the Development of Electronic Administration*, 1 EUR. REV. OF DIGIT. ADMIN. L. 1–2, 123, 126 (2020).

sequence of interim acts that lead to the final decision—rather evokes the definition of the administrative procedure. However, what the authors consider the most relevant finding of the T.A.R.'s judgment is the innovative scope of the automated administrative decision: "it may happen that the algorithm, conceived as a technical rule, will finally assume a role that goes beyond the prerequisites on which the decision is based, being able to constitute a system for the formation of the procedural will itself." In this sense, from the facts established in the MIUR's judgment, the authors are of the opinion that it is difficult to discern the extent to which the final administrative decision on teacher's mobility was the result of an assessment made by the algorithm itself and assumed entirely by the authority, or whether it was rather an assessment made by the authority that was based on the outcome of the algorithm.[83]

The French Commission on Access to Administrative Documents (CADA) has taken a different approach to the matter. In a further step, the French Commission has recently upheld the right of access to the full source code of the Parcoursup platform, to automatically process the national pre-enrolment procedure in the first year of public university education. In this regard, the French Authority has underlined that

> public algorithm is an *administrative procedure, fully or partially automated*, involved in a decision-making process for citizens. The source code is the computer translation of this algorithm. *It explains the method of administrative decision-making, allows to control the interpretation and application of the law implemented by the public authorities* and reinforces the confidence of the users in the system.[84]

On the contrary, a Spanish Administrative Court expressly denied the status of administrative act or rule of the source code of an application called BOSCO, used by the Ministry of Ecological Transition, to verify whether the applicant complies with the legal requirements to be considered a vulnerable consumer to receive social benefits consisting in discounts on electricity supply bill. The civil association, CIVIO, had sought access to the source code in production,

---

83. *See generally* Maria Cristina Cavallaro & Guido Smorto, *Decisione pubblica e responsabilità dell'amministrazione nella società dell'algoritmo* [Public Decision and Responsibility of Administration in the Society of the Algorithm], FEDERALISMI. RIVISTA DE DIRITTO PUBBLICO ITALIANO, COMPARATO, EUROPEO 16 (2019).

84. Commission d'Accès aux Documents Administratifs [CADA] [Commission for Access to Administrative Documents], Jan. 13, 2022, 20213847 (Fr.) (emphasis added).

the technical specifications, the results of the tests made to verify the compliance of the application with functional specifications, and any deliverable explaining how the application worked. CIVIO alleged that the application had mistakenly and systematically denied eligibility for aid to applicants who met legal and regulatory requirements to be beneficiaries, so the information on the contested application was sought to verify the correctness of BOSCO's design and the existence of possible errors. The Court finally dismissed the access sought by CIVIO to the source code of the disputed application on grounds of public security of the information systems and intellectual property rights of the Ministry.[85]

IV. HIDDEN PATHOLOGIES IN ADM SYSTEMS: THE PUBLIC INTEREST IN FAVOR OF DISCLOSURE

"Opacity opens the door to error and misuse."[86] This is particularly true in the case of ADM systems, AI-driven or not.

To a greater or lesser extent, FOIA cases show that access to the code, the underlying algorithm, and other relevant documentation (functional and technical specifications, description and characterization of the dataset, validation metrics, or third-party audits) can contribute to better understanding and public scrutiny of: the automated systems deployed and implemented by public administrations; their explicit, implicit, or unintended purposes; and their individual and collective impacts.

Most importantly, access to technical information may contribute to a better understanding of some of the pathologies (errors, misuse, or unintended purpose) afflicting the algorithmic systems and unlock the door to judicial review of automated administrative decisions, whatever their level of automation may be.

This has been the case for the MIUR's algorithm. The illogic consequences of the decisions produced by the algorithm at least suggested the existence of pathologies in the decision-making process: thousands of teachers were transferred hundreds of kilometers from their homes; other teachers with lower scores who were assigned to a

85. Juz. Cont. Adm., Feb. 18, 2019, (R.J. No. 0701, p. 2018) (Spain) (granting access to technical specifications and tests undergone to verify the functionality of the application, but denying access to the source code on grounds of intellectual property, aff'd. Sentencia 143/2021 del Juzgado Central de lo Contencioso Administrativo n. 8 (Dec. 30, 2021), https://www.consejodetransparencia.es/dam/jcr:80688e50-c994-4850-8197-4f19dc46a6ad/R128_S143-2021_CIVIO.pdf (upholding the dismissal also on grounds of public security).

86. DOMINGOS, *supra* note 9, at xvi.

position in the same province where they lived. As a result, there were numerous claims and appeals.[87]

As the Italian State Council pointed out when deciding on appeal against the teaching positions assigned by the MIUR's algorithm, "the algorithmic rule must not only be *cognoscible in itself*, but also be subject to the full knowledge and review of the administrative judge."[88] In the view of the Council, this requirement responds to the undeniable need to review how the power has been exercised, appearing ultimately as a direct expression of the right of defense of the citizen, who cannot be prevented from knowing the modalities (even if they are automated) with which a decision affecting his legal sphere has been taken. In this sense, the automated decision requires the court to first assess the correctness of the computer process in all its components, in order to "ensure that such process, at the administrative level, *takes place transparently, by knowing the data entered and the algorithm itself*."[89] Second, the court must be able to review the very logic and the reasonableness of the robotized administrative decision, that is, the "rule" that governs the algorithm.[90]

The impossibility of understanding how the MIUR's algorithm made those decisions, especially those incurred in manifest arbitrariness or having harmful effects on teachers, led the Court to render such decisions null and void: "the inability to understand the manner in which the algorithm in question assigned the vacant positions constitutes in itself such a defect as to invalidate the whole procedure."[91]

---

87. BIAGIO ARAGONA, ALGORITHM AUDIT: WHY, WHAT, AND HOW? xi (2022) (reporting that teachers from Puglia and Calabria had to move to the province of Milan, when they should have been assigned to their respective regions).

88. Cons. Stato, Sez. VI, 8 aprile 2019, n. 2270, § 8.

89. *See id.* at 8 (emphasis added).

90. *Id*. at 8.

91. *Id.* at §9. (The Council finally upheld the appeal due to the breach of the principles of impartiality, publicity and transparency, "since it cannot be understood that the legitimate expectations of the persons placed in a certain position on the ranking list have been defrauded … Not only that, but the results of the procedure do appear to be characterized by the illogic and irrationality claimed by the appellants, as paradoxical situations have arisen, where some teachers with many years of duty have been assigned to territorial areas that they had never applied for and located hundreds of kilometers from their city of residence; while other ones, with less qualification and seniority, have obtained the same positions they applied for.")

A. Errors in Programming and Flawed Models: the MIUR and Ofqual's Algorithms

The development and deployment of ADM legal systems means that the law usually expressed in natural language needs to be reshaped into a formal representation by means of programming language to be understood by computers. This process implies a "transformation of legal sources."[92]

Yet it may happen that the transformation of legal norms into code language incur in an incorrect translation—desired or not—resulting in legal consequences not envisaged in the legal norm. Moreover, such consequences could have *extra legem* or *contra legem* effects.[93] The right of access may provide an opportunity to oversee the correctness of this process.

Again, the MIUR's algorithm is a clear example of the further consequences resulting from the disclosure of source code ordered by T.A.R. Lazio's judgment far beyond the access itself by the representative union. Whether or not the programming language is understandable to the layperson or to the judge themself, the right of access allows verification of the correctness of the automated administrative decision, if not directly by the addressee of the decision, then by an expert.[94] The alleged lack of expertise of citizens to understand the language of the algorithms can no longer be the main argument justifying the futility of the right of access.[95]

---

92. Dag Wiese Schartum, *Law and Algorithms in the Public Domain*, 1 ETIKK I PRAKSIS. NORDIC JOURNAL OF APPLIED ETHICS, 16 (2016), https://www.ntnu.no/ojs/index.php/etikk_i_praksis/article/view/1973/1984.

93. *See generally* Citron, *supra* note 23, at 1254–55 (2008); Danièle Bourcier & Primavera de Filippi, *Les algorithmes sont-ils devenus le langage ordinaire de l'administration?*, GENEVIEVE KOUBI, LUCIE CLUZEL-METAYER & WAFA TAMZINI, LECTURES CRITIQUES DU CODE DES RELATIONS PUBLIC ET ADMINISTRATION, 200, 207 (2018).

94. T.A.R. Lazio-Rome, Sez. III Bis, 22 Marzo 2017, n. 3769 15–16 (It.) ("[T]he circumstance that the software is compiled by means of programming languages that are usually incomprehensible not only to the official who utilizes it to draft the final decision of the administrative procedure but also to the private addressee of the act itself does not seem to be diriment; since, on the one hand, the aforementioned circumstance is a consequence of the discretionary choice of the Administration to resort to an innovative tool, such as computer programming, for conducting a procedure of its own prerogative and authority; and, on the other hand, the *private addressee of the* [administrative] *act may legitimately resort to the professional services of a competent computer scientist for the purposes of comprehension and verification of* [the] *correctness* [of the administrative decision] [emphasis added]."

95. *See* DATENETHIKKOMMISSION, GUTACHTEN DER DATENETHIKKOMMISSION 170 (2019) (Ger.) ("However, the ubiquitous complexity cannot refute the goal of making algorithmic systems transparent or justify opacity).

Precisely, one of the immediate consequences of the T.A.R. Lazio judgment in this case is that, by granting the plaintiff-union the access to the source code, it made possible the expert analysis of the controversial code.

The subsequent audit carried out on the code by computer experts from the Universities of Tor Vergata and La Sapienza in Rome shows the importance of this issue and to what extent the right of access to source code guarantees some of the ultimate goals of FOIA regimes.

This being so, the audit revealed that the MIUR's algorithm had been designed and developed by using two different programming languages: "COBOL language"—an obsolete programming language—for phase A of the algorithm, and "C language" for phases B, C and D. This duality was considered a malpractice, taking into account that the different phases of the algorithm should interact and share data and outcomes with each other. The audit also highlighted that the excessive price paid to the contractors was not justified at all by the needs of rationalization and efficiency of public expenditure.[96]

The audit also determined the existence of relevant omissions in the information provided by the MIUR, which prevented adequate scrutiny of the disputed software. In particular, it was observed that in phases B, C, and D of the algorithm, the functions made use of a database. However, the documentation relating to the structure and format of the database had not been provided by the administration to the experts, so a correct and complete verification of the program in question was not possible. In this sense, the audit considered that the possible shortcomings of the algorithm could be attributed not only to errors in the design of the source code but also to an inappropriate preparation and management of the input data being processed, which irremediably would have affected the final outcome produced by the algorithm. Finally, the audit found that that certain files were blocked, thus restricting the possibility of a direct and automatic verification by means of specific software and compilers that would have accelerated the review of the logical and syntactic correctness of the program. Therefore, the only way to proceed was by manually copying the lines of code, a challenging task. In phase A alone, this meant to re-writes up to 29,600 lines.

The conclusions of the audit could not have been more devastating:

> It is obvious that . . . lack of the claimed lines of code, the database, the lines used by the software to read and

96. GILDA DEGLI INSEGNANTI ORISTANO, PERIZIA TECNICA SUL CONTESTATO ALGORITMO (Jun. 15, 2017), http://gildaoristano.blogspot.com/2017/06/perizia-tecnica-sul-contestato.html.

> write the data . . . along with the technical specifications results in a *non-transparent conduct* [of the MIUR] . . . . *Such omissions irreversibly preclude any possibility of a complete review of how the algorithm works and thus how teacher's mobility has been determined across the country*.[97]

In this context, it is worth recalling that the European Court of Human Rights has outlined that the "obstinate reluctance" or "dilatory" attitude of administrative bodies in providing access to public information, in breach of decisions of the supervisory authorities or courts upholding the applicant's right of access, must be considered as an "arbitrary restriction" contrary to the principle of legality and an arbitrary interference with Article 10 of the European Convention on Human Rights (ECHR).[98]

Once again in the field of education, the Ofqual's algorithm in the UK is another example of how defectively designed algorithmic models can have individual and social adverse impacts.

In this case, the right of access to the source code or to the algorithm used by the administration was not in dispute, since the technical documentation, including the explanation of the algorithmic model, had been released by the authority. Here, the issue at stake was the public interest in the access to certain disaggregated results of the algorithm that had not been previously made public to achieve a better understanding of the consequences resulting from the contested predictive model and, therefore, a broad public scrutiny of the public decisions based or adopted on that model.

Due to the COVID-19 pandemic, the official A-Level examinations in the UK, which give access to university studies, were suspended and replaced by an algorithmic model developed by Ofqual, the regulatory body for official examinations and qualifications. The aim of the algorithmic model was to predict the grade that students would have achieved had exams taken place.

Following the release of the grades, there were numerous

---

97. Alessandro Salvucci, et al., *Perizia tecnica preliminare sull'analisi dell'algoritmo che gestisce Il software della mobilità docenti per l'a.s. 2016/201*, GILDA VENEZIA 12–17 (June 4, 2017), https://www.gildavenezia.it/wp-content/uploads/2017/06/Perizia-tecnica-preliminare2017.pdf (emphasis added).

98. *See* Kenedi v. Hungary, App. No. 31475/05, Eur. Ct. H.R. (2009), ¶45; Youth Initiative for Human Rights v. Serbia, App. No. 48135/06, Eur. Ct. H.R. (2013), ¶24-26 (where the Court found in both cases that the persistent obstructive maneuvers of the State authorities precluding the access to the information sought by the applicants led to a violation of the human right to receive information without interference by public authority, enshrined in Article 10.1 of the ECHR).

complaints across the country about the process and the results. Nearly 40 percent of the 700,000 ratings submitted by the centers had been revised downward by at least one level, and 3.5 percent had been downgraded by two or more levels.[99] The Information Commissioner's Office even echoed the deviations of the algorithm from the estimates made by the centers and the "widespread criticism within the mainstream media."[100]

One of the most recurrent criticisms was that the downgrades made by the algorithm had particularly affected public schools, which usually have larger numbers of students and are financially weaker than private schools.[101]

In this regard, private schools saw the proportion of the A-level grades awarded raised by more than double than that of public schools. For independent schools, with small student numbers, the results for A or A* level grades grew by 4.7 percentage points, from 43.9 percent in 2019 to 48.6 percent in 2020; however, for public schools, the results only varied two points, from 19.8 percent in 2019 to 21.8 percent in 2020.[102]

But how did the controversial algorithm actually work? To determine the grades (predictions), the relevant regulatory body, Ofqual, designed an algorithm whose purpose was not to directly predict individual students' A-level grades, but rather the percentage of students in a given school "j" who should receive a grade "k" within the possible ranges A*, A, B, C, D, E, and U. The algorithm in question was more of a heuristic type[103] and, despite some opinions, did not implement machine learning or deep learning models.[104]

---

99. Richard Adams et al., *A-level results: almost 40% of teacher assessments in England downgraded*, THE GUARDIAN, (Aug. 13, 2020) https://www.theguardian.com/education/2020/aug/13/almost-40-of-english-students-have-a-level-results-downgraded.

100. ICO, IC-70514-H7K5 (Aug. 5, 2021), ¶54.

101. Julian Faraway, *An Alternative to the Ofqual Algorithm*, Aug. 28, 2020, https://julianfaraway.github.io/post/an-alternative-to-the-ofqual-algorithm; *see also A-levels and GCSEs: How did the exam algorithm work?* BBC, Aug. 20, 2020, https://www.bbc.com/news/explainers-53807730.

102. David Hughes, *What is the A-level algorithm? How the Ofqual's grade calculation worked - and its effect on 2020 results explained*, INEWS, Aug. 17, 2020, https://inews.co.uk/news/education/a-level-algorithm-what-ofqual-grades-how-work-results-2020-explained-581250.

103. Sophie Bennett, *On A Levels, Ofqual and Algorithms,* Aug. 20, 2020, https://www.sophieheloisebennett.com/posts/a-levels-2020/.

104. Tim Paulden, *A cutting re-mark*, 17 SIGNIFICANCE 5, 4–5 (2020). https://doi.org/10.1111/1740-9713.01436. *Cf.* Yannique Hetch, *UK's Failed Attempt to Grade Students by an Algorithm. Why engineering alone isn't enough to fix broken social systems*, TOWARDS AI, Sept. 4, 2020, https://pub.towardsai.net/ofqual-algorithm-5ecbe950c264; Selin Akgun & Christine Greenhow, *Artificial intelligence in education:*

To carry out this process, Ofqual asked the centers to submit, for each student and each subject they were entered for, two pieces of information: first, the grade estimated by the teachers that students would have most likely achieved if they had taken their exams, called "Centre Assessment Grade" (CAG); second, a ranking of each student in each grade range, ordered from best to worst, compared to the rest of the students in the same center with the same CAG.[105]

This was theoretically because in practice Ofqual did not apply the CAG in all cases. In particular, the CAG was the only or main reference for the assessment of private candidates and centers with a small number of enrollments in subjects considering the 2020 academic year and the historical series. The rationale behind this was the smaller the number of students, the weaker the statistical evidence. In contrast, in the case of centers with larger numbers of enrolled students, the standardization process applied by the algorithm was based on an approach that gave greater weight to the statistical historical evidence of center performance (given the prior attainment of students) than the submitted CAGs.[106]

In making its predictions, the algorithm took into account the following sources of information. First, the algorithm considered historical distribution of grades in schools for each subject in the last three academic years (2017–2019).[107] Second, the student rankings produced by each center were moderated at the national level by imputing a "proxy" grade. For this purpose, the algorithmically generated distribution of grades was subjected to a standardization process by transforming the ordinal grades (A*, A, B, etc.) into pseudo-numerical scores assigned from the order established by the ranking.[108] The result was a mark scale with notional cut-scores that determined

---

Addressing ethical challenges in K-12 settings, AI ETHICS 2, 431, 436 (2021), https://link.springer.com/content/pdf/10.1007/s43681-021-00096-7.pdf (identifying Ofqual's algorithm with machine learning techniques).

105. OFQUAL, RESEARCH AND ANALYSIS. AWARDING GCSE, AS, A LEVEL, ADVANCED EXTENSION AWARDS AND EXTENDED PROJECT QUALIFICATIONS IN SUMMER 2020: INTERIM REPORT, 97–102 (2020) (U.K), https://www.gov.uk/government/publications/awarding-gcse-as-a-levels-in-summer-2020-interim-report (click "Awarding GCSE, AS, A level []: interim report".).

106. *See id.* at 11, 92–93, 95–97.

107. This historical distribution was subjected to a double adjustment process. First, a correction was added for the differences between the academic results of the 2020 cohort and the previous results of the 2017-2019 cohorts. Second, it was taken into account the proportion of students who, in a given center, could be matched with the historical student cohort based on their previous attainment.

108. For example, if there were 3 students ranked as 1, 2 and 3 with grade A in a given center, the scores might be as follows: student 1: 600 ('high' A); student 2: 550 ('medium' A); student 3: 500 ('low' A).

the final grade boundaries at a national level.[109] The final grades for each center were assigned by distributing students across the range of grades available for each pseudo-numerical grade.

According to experts, the CAG was systematically ignored for large centers, and this fact and the standardization process introduced unfairness into the grading system. For example, if the CAG estimate for a student in a particular school was an A* level in mathematics, the algorithm would have reduced that estimate to an A level, or even a B or C level, depending on the school's historical cohort and the pseudo-numerical cut-score imputed from the ranking.[110]

After several days of public outcry and media pressure, Ofqual ignored the "synthetic grades" and replaced them by the grades originally set by centers. The "mutant algorithm"—as it was coined by Boris Johnson—was withdrawn not for statistical reasons but for political ones.[111]

The preceding context explains the decision of the ICO in response to a request of information that had been previously dismissed by Ofqual. The information requested by the claimant precisely concerned the disaggregated results of the algorithm and, in particular, the "center name; % grades up 2 grades at that center; % grades up 1 grade; % the same grade; % -1 grade; %-2 grades; % -3 grades."[112]

In the age of big date analytics, it is said that there is usually a "myopic focus on input data" to the detriment of the output data.[113] Rather, the terms of the complaint and the Commissioner's decision upholding the claimant's request show how the outputs of the algorithmic processing may also give insightful information about the risks of processing on individuals and a better understanding of the individual and collective impacts.

Ofqual had found section 36(2)(c) of Freedom Information Act

---

109. Theoretically, the application of these notional cut-scores tried to avoid grade inflation from one year to the next.

110. BENNETT, *supra* note 103 (explaining that the standardization process at national level used by Ofqual really meant that students' grades could be "shifted downwards or upwards depending on where their pseudo-score place[d] them relative to the rest of the student cohort").

111. Timandra Harkness, *How Ofqual Failed the Algorithm Test*, UNHERD, Aug. 18, 2020, https://unherd.com/2020/08/how-ofqual-failed-the-algorithm-test/.

112. ICO, IC-70514-H7K5 (Aug. 5, 2021) (UK), ¶17.

113. Sandra Wachter & Brent Mittelstadt, *A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI*, COLUM. BUS. L. REV. 2, 494, 514 (2019), https://doi.org/10.7916/cblr.v2019i2.3424 (contending that European data protection law fails to address properly the outputs of processing —e.g., inferred and derived data, profiles, and decisions—as it provides only a few mechanisms that are much weaker than those for input processing).

(FOI)[114] applicable to justify the dismissal of the request in question, arguing that disclosure of such information would "prejudice or be likely to prejudice the effective conduct of public affairs." The regulator argued that the decision not to publish educational data at center level for 2020 was taken to ensure that the teachers could produce their CAGs without fear of judgment to ensure the robustness of the grading process. In particular, the regulator considered that centers had a legitimate expectation that their center level performance would not be made publicly available. Consequently, disclosure would undermine "current government policy also" and the regulator's ability to perform its functions properly in its relationship with stakeholders, particularly teachers, schools, and their representatives.[115] Finally, Ofqual argued that disclosure could lead to comparison, scrutiny, and judgements made on individual centers based on the variance in CAGs and adjustments made by the algorithm; this would likely lead to an unfair perception or potential criticism of specific centers as being less reliable, more demanding, or more lenient than others. Such a situation would require then a diversion of resources for managing the adverse publicity that disclosure could cause, especially to those centers at the extremes of the variation.[116]

Although the Commissioner acknowledged that section 36 of the FOI had been correctly applied by Ofqual, she found that the regulator had failed to weigh the public interest in disclosure.[117] In the first place, the disclosure would provide a bigger picture of the disputed assessment process, thus holding centers accountable for any discrepancies or misapplication in relation to the CAGs awarded. In the second place, any potential adverse effect that the scrutiny of public opinion could have on the centers would be outweighed by the fact that disclosure would prompt students, or their parents, to engage in complaints procedures.[118]

But was the algorithm flawed? Put simply, the design of the algorithm gave rise to relevant "technical biases" that were not "proactively identified and corrected."[119] And this was so despite the fact that the Royal Statistical Society had offered its help, which it finally had to withdraw due to the restrictive confidentiality agreement that Ofqual intended to impose.[120]

---

114. Freedom of Information Act, (2000) §36(2)(c) (UK.), UK ST 2000 c. 36 Pt II s. 36.
115. *See supra* note 112, at ¶ 35–36.
116. *Id.* at ¶19, 32, 34–36, 42 and 44.
117. *Id.* at ¶28.
118. *Id.* at ¶61–62, 65–67.
119. Wachter, Mittelstadt & Russell, *supra* note 8, at 739.
120. Faraway, *supra* note 101.

Some of the shortcomings of the algorithm were identified by the experts after analyzing the report published by Ofqual with the details of the model.[121]

First, the algorithm had not been applied uniformly across centers. For those with five or fewer students, only the grade set by the centers was considered, thus discarding the application of any standardization process; for small centers, Ofqual had used a combination between the CAG and a simplified version of the algorithm. This immediately introduced a first point of unfairness in the rating system, as centers with fewer students were more likely to be private schools, and the ratings produced by these centers were generally higher than those generated by the algorithm.

Second, the standardization process at the national level by attributing pseudo-numerical cut-scores resulted in higher grades for students from smaller schools lowering the grades of students from larger schools, which already had their grades downgraded by the algorithm. While schools with a higher number of students enrolled were more likely to be public schools, those with a lower number were more likely to be private schools. This fact further increased the unfairness and disparity in the grades predicted for large schools.

Third, in the process of matching historical distributions of grades with their corresponding cohort, the model did not take into account the variability of grades from one year to the next, which especially affected public schools located in more deprived areas.

Fourth, the decisive influence of the rankings together with the standardization process at the national level ultimately generated paradoxical situations in which two students with almost identical performance in the same center and with the same CAG had been assigned different A-Level grades by the algorithm.

B. Validation of Algorithmic Models and Technical Bias: Ofqual, Once Again

As Zlotnic points out, both errors and their potential impacts (individual or social) of AI systems must always be assumable by organizations implementing those systems.[122] This is true not only for AI systems but also for any predictive model, such as the one developed by Ofqual.

---

121. *See generally* BENNETT, *supra* note 103; Harkness, *supra* note 111; Paulden, *supra* note 104; George Constantinides, *A-Levels and GCSEs in 2020*, THINKING, Aug. 15, 2020, https://constantinides.net/2020/08/15/a-levels-and-gcses-in-2020/.

122. *See generally* Alexander Zlotnik, *Artificial Intelligence in Public Administrations: Definitions, Project Feasibility assessment and Application Areas*, 84 BOLETIC (2019).

In fact, one of the constraints detected by the experts in Ofqual's model was precisely the lack of quantification of uncertainty, in the sense that the A-Level grades attributed by the algorithm (predictions), in practice, did not take into account the relevance of the error in the corresponding estimates. This was particularly problematic for two reasons. First, the overall accuracy of the model was far from good. Second, it was largely ignored that most statistical models tend to be limited in their ability to accurately predict outcomes for individual subjects (as opposed to population samples).[123]

Bearing in mind that accuracy in both statistical and AI models refers to how often the model gets the correct answers measured against correctly labeled test data,[124] how had Ofqual proceeded to validate the model and determine its accuracy? The white paper published by the regulator explained that the algorithm in question had been validated for different subjects using different models, including linear regression and logistic regression.

The method used by the regulator was described as "flawed" by experts.[125] Indeed, to assess the accuracy of the model, Ofqual had used historical data from the 2019 cohort. However, there were no rankings produced by the centers for this historical cohort, as was the case for the 2020 cohort. This created circularity in the model validation process as, in practice, Ofqual used the actual 2019 grade data—instead of the rankings, as was done for the 2020 cohort—to predict the 2019 grades. This flawed method for assessing model accuracy ultimately led to an overestimation of the actual accuracy of the model used to predict 2020 grades. In sum, the accuracy of the final model was not reliable.[126]

---

123. BENNETT, *supra* note 103.

124. INFORMATION COMMISSIONER'S OFFICE, GUIDANCE ON AI AND DATA PROTECTION, 38–40 (Oct. 14, 2020), https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection-0-0.pdf; EUROPEAN UNION AGENCY FOR CYBERSECURITY, AI CYBERSECURITY CHALLENGES, 19 (Dec. 15, 2020), https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges. (In AI models, statistical accuracy is about how closely an AI system's predictions match the correct labels as defined in the test data. This means comparing the performance of the model's outputs to some "ground truth". For instance, a medical diagnostic tool designed to detect malignant tumors could be evaluated against test data, containing true cases of malignant and benignant tumors of known patients.)

125. A detailed analysis of the main flaws in the validation of Ofqual's algorithm can be found at BENNETT, *supra* note 103 and Paulden, *supra* note 104.

126. *See generally* OFQUAL, *supra* note 105, at 52–54. (*There* was significant variation (approximately 40%–75%) depending on the subjects and the validation model considered by Ofqual. For example, biology, with the validation approach NO. 8 based on a linear regression model, the accuracy achieved was less than 0.4 was obtained. And, with the validation approach No. 3—based on a variant of the model used— the accuracy was below 0.7.)

In AI-driven automated decision-making systems, expressions such as "automation bias" or "automation-induced complacency" have been coined to describe to what extent human users routinely rely on the output generated by the system and stop questioning whether it might be wrong, unfair, or even harmful.[127]

Likewise, the automation of decisions and, particularly, the implementation of AI systems can amplify existing discriminatory biases and social inequalities or even distort the purpose of public policies.[128] As a consequence, for example, of a model trained, tested, or validated with incomplete or unrepresentative data or a selection of data from biased sources.[129]

With this in mind, Ofqual assessed whether the algorithm unfairly biased certain groups of individuals. However, it only checked what happened when all students and schools received the grades predicted by the algorithm, but not whether some schools—predominantly the wealthiest—had unduly obtained higher grades when using the CAG rather than the algorithm. Moreover, the impact of the algorithm on equality was only analyzed in a small subset of the models tested.

Because of the shortcomings in the design and validation of the model, high-performing students in high-performing schools received higher grades, while high-performing students in low-performing schools saw their grades lowered compared to their peers. In practice, this trend most disproportionately affected Black, Asian, and Minority Ethnic (BAME) students.[130]

In this regard, it is not surprising that the ICO took into account these aspects in its decision when weighing the public interest in the access to the disaggregated information withheld by Ofqual:

> There were concerns that the algorithm itself was unlawful, not only breaching anti-discrimination standards but also Article 22 of the GDPR which outlines the right not to be subject to fully automated decision-making that significantly affects individuals. The complainant has made this request based on

127. REUBEN BINNS & VALERIA GALLO, AUTOMATED DECISION MAKING: THE ROLE OF MEANINGFUL HUMAN REVIEWS (Apr. 12, 2019), https://ico.org.uk/about-the-ico/media-centre/ai-blog-automated-decision-making-the-role-of-meaningful-human-reviews/.

128. Danielle Citron & Ryan Calo, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 4, 797, 805, 816 (2021).

129. Danièle Bourcier; Primavera de Filippi, *La transparence des algorithmes face à l'Open Data: Quel statut pour les données d'apprentissage?*, 167 REVUE FRANÇAISE D'ADMINISTRATION PUBLIQUE 7, 534–536 (2018).

130. BENNETT, *supra* note 103; Wachter, Mittelstadt & Russell, *supra* note 8, at 738–39.

concerns that students attending lower.[131]

In light of the foregoing, it seems crucial to determine and verify the level of accuracy of AI models in relation to the task, its purpose, and the context of its use. In many cases, users of AI systems emphasize model error metrics while omitting a corresponding evaluation of the potential impacts of errors. For instance, a very low probability of error (e.g., 0.1% of false negatives), but with potential adverse impacts arising from this error (e.g., death of a patient), may not be assumable by the organization.[132] In addition, trade-offs between precision and recall must be carefully addressed as differences between them may affect the fairness of the model or may lead to adverse impacts.[133]

This is why in the context of AI systems, it is very welcomed that the 2021 EU Proposal for AI Regulation has stressed in Recital (44) that training, validation, and testing data sets should be sufficiently relevant, representative, complete, and free of errors in the light of the intended purpose of the system, and should also have the appropriate statistical properties, taking into account their intended purpose, the features, characteristics, or elements that are particular to the specific geographical, behavioral, functional setting, or context within which the AI system is intended to be used.[134]

131. ICO, *supra* note 100, ¶55.

132. Zlotnik, *supra* note 122, at 27–28.

133. ICO, *supra* note 112, at 40.

134. In measuring the statistical accuracy of the model, often, the available dataset is randomly split into: (1) a *training set*, data used for setting the internal model's parameters¾e.g., weights¾in order to minimize the difference between inferred outcomes and the desired result; (2) a *validation set*, a sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters in order to find the optimal values that will give the best possible performance; and (3) a *test set*, data used to assess the performance of the final model to ensure that it can generalize well to new and unseen data, comparing the testing accuracy against the training accuracy in order to avoid overfitting the model. Measuring the statistical accuracy should reflect the balance between two different kinds of errors: false positives, where the model incorrectly labels as positive, and false negatives, where the model incorrectly labels as negative. Another way to measure these types of errors is by including *precision*, which is the percentage of cases identified as positive that are in fact positive, or *recall*¾or *sensitivity*¾the percentage of all cases that are in fact positive having being identified as such. A low precision may indicate a large number of false positives, while a low recall may reveal a large number of false negatives.

C. Challenging Statistical Accuracy in Houston: The Problem of Relying on Proprietary Algorithms

The use of complex and sophisticated algorithms, of the sort being contested in Houston[135] to evaluate teacher performance and make employment decisions (tenure, salary, merits, or termination of the employment) in state-run schools across the United States,[136] shows again the relevance of the statistic accuracy, external audits, and appropriate use of error metrics as previously described for its European counterpart in the MIUR's and Ofqual's algorithms.

In the MIUR's and Ofqual's algorithms, the access to the source code and/or the relevant documentation (e.g., audit by experts, explanatory documentation of the algorithm) revealed the inaccuracy and the existence of errors that invalidated the model and their results. In contrast, a repeated denial of FOIA requests seeking access to such information was evidence of government malpractice leading the Houston Court to overturn the model on grounds of the violation of due process rights without the necessity of opening the black box. Yet another difference must be highlighted. Whereas in the MIUR and Ofqual cases the algorithms challenged were in-house developments, in Houston the algorithm whose accuracy was contested was proprietary.

Whatever the grounds may be, in all the cases referred, the verification of the (in)accuracy and the detection of errors (or the impossibility to do it) highlight the issue at stake: the use by governments of automated models of low statistical confidence with adverse impacts on the governed.

In Houston, the Southern District Court of Texas had to deal with the problem of the validation of a value-added model (VAM) used by the Houston Independent School District (HISD) during the 2011–2015 school years to rate teacher effectiveness.

The evaluations were applied to make decisions of termination for poor performance. Plaintiffs sought a declaratory judgment and permanent injunction against the use of Educational Value–Added Assessment System (EVAAS) scores in termination or nonrenewal of teacher contracts.[137]

---

135. *See* Hous. Fed'n of Tchrs., Loc. 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1171, 1177 (S.D. Tex. 2017).

136. Mark Paige, Audrey A. Beardsley & Kevin Close, *Tennessee's National Impact on Teacher Evaluation Law & Policy: An Assessment of Value-Added Model Litigation*, 13 TENN. J.L. & POL'Y, 523, 527–528 (2019), https://ir.law.utk.edu/tjlp/vol13/iss2/3 (noting that these models usually fail to take into account the complexity of teaching and the impact of relevant variables, e.g., individual motivation of students, so in high-stakes employment decisions are an "invitation for legal action").

137. Hous. Fed'n of Tchrs., 251 F. Supp. 3d. at 1174.

The VAM was a proprietary algorithmic model based on three components: (1) instructional practice; (2) professional expectations; and (3) student performance. The weight assigned to each component varied over the years. The focus of this litigation was on the criterion of student performance, which was calculated by means of a value-added model, the EVAAS. The model assessed teacher effectiveness by attempting to track the teacher's impact on the student test scores over time.[138]

From the background of the case, it was clear that HISD had repeatedly denied discovery and FOIA requests of the plaintiff-union to the source code, computer algorithms, and underlying data of VAM ratings necessary to verify "the accuracy of their scores and, in particular, any error that may exist" on grounds that it was vendors proprietary information and this required "the production of proprietary, trade secret information not in the custody, control, or possession of the District."[139]

Plaintiffs argued that these procedures were constitutionally inadequate for teachers threatened with termination on the basis of low value-added scores because they were "denied access to the computer algorithms and data necessary to verify the accuracy of their scores."[140]

An interesting point of the judgment is the Court's own definition of "accuracy" for the purposes of the litigation: "'accuracy' simply means that the EVAAS score is correctly calculated according to the vendor's own algorithms, using the right data (e.g., correct test scores for the teacher's own students as well as all other students with whom they are compared) and executed by properly performing software that has been suitably tested and maintained according to appropriate quality control measures."[141]

According to the judgment, the HISD had conceded that the scores had been generated by very complex algorithms, employing "sophisticated software and many layers of calculations,"[142] and

---

138. *Id.* at 1171–72 (explaining that teacher's EVAAS score was based on comparing the average test score growth of students taught by the teacher with the statewide average for students in that grade or course).

139. *Id.* at 1177 (emphasis added).

140. *Id.* at 1176.

141. *Id.* at 1176–77, n. 25.

142. *Id.* at 1177. Such description of and further reference to the algorithm as "a mysterious 'black box'" suggest that the model used to produce the EVAAS scores would probably rely on neural networks or any ensembled models (e.g., random forests). *Id.* at 1179. These are the type of models that are described genuinely as *black boxes*. *See generally* INFORMATION COMMISSIONER'S OFFICE & THE ALAN TURING INSTITUTE, EXPLAINING DECISIONS MADE WITH ARTIFICIAL INTELLIGENCE, https://ico.org.uk/for-

admitted that lack of audit procedures of the EVAAS scores. And what was more problematic: "any effort by teachers to replicate their own scores, with the limited information available to them, would necessarily fail." In the same way, any independent verification of a negative EVAAS score would be impossible at all. The Court emphasized that "[a]ccording to the unrebutted testimony of plaintiffs' expert, without access to vendor's proprietary information—the value-added equations, computer source codes, decision rules, and assumptions—EVAAS scores will remain a mysterious 'black box,' impervious to challenge."[143]

The impossibility of replicating the scores and examining the algorithm to challenge its accuracy led the Court to infer that the EVAAS score might have been erroneously calculated for any number of reasons, ranging from data-entry mistakes to flaws in the source code itself. The Court continued, "[a]lgorithms are human creations, and subject to error like any other human endeavor." HISD has acknowledged that mistakes can occur in calculating a teacher's EVAAS score; moreover, even when a mistake is found in a particular teacher's score, it will not be promptly corrected." But one of the most remarkable things conceded by the HISD was that any attempt to re-analyze at the system level seeking to overview, and if necessary, correct an error in only one teacher score, would imply the "potential to change all other teachers' reports." In what the Court qualified as a "house-of-cards fragility of the EVAAS system," it concluded that

> [t]his interconnectivity [of teacher evaluations] means that the accuracy of one score hinges upon the accuracy of all. Thus, without access to data supporting all teacher scores, any teacher facing discharge for a low value-added score will necessarily be unable to verify that her own score is error-free.[144]

The Court agreed with the defendant in that the Due Process Clause did not empower plaintiffs to put the vendor out of business by requiring disclosure of its trade secrets. But, by the same token, the vendor's trade secrets "[did] not empower, much less compel, HISD to violate the constitutional rights of its employees." Thus, "[w]hen a public agency adopts a policy of making high stakes employment decisions based on secret algorithms incompatible with minimum due

---

organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/annexe-2-algorithmic-techniques/.

143. Hous. Fed'n of Tchrs., 251 F. Supp. 3d at 1179 (emphasis added).

144. *Id*. at 1177–78.

process, the proper remedy is to overturn the policy, while leaving the trade secrets intact."[145]

By denying access to the source code and the underlying data, the *Houston* Court concluded that teachers could not protect against the government's deprivation of their property right to employment. "HISD teachers have no meaningful way to ensure correct calculation of their EVAAS scores, and as a result are unfairly subject to mistaken deprivation of constitutionally protected property interests in their jobs."[146]

Many times, codes and algorithms behind ADM systems are not in-house solutions but custom software designed and developed by contractors. In the context of FOIA requests, administrations and courts are usually very reluctant to grant access to the source code or algorithms on grounds of trade secrecy or other confidential privileges by government contractors. But even if the software is an in-house development, public administrations still may assert proprietary rights in many FOIA regimes (such as the Spanish one).[147]

Yet in some jurisdictions, supervisory authorities for FOIA rights and courts are moving toward a different approach, taking into account the public interest at stake. For example, in weighing the interest protecting the intellectual property rights of the MIUR and the contractor against the public interest in access, the T.A.R. Lazio granted qualified access to the plaintiff-union, while precluding a "general access" by the public.[148]

The MUIR merely stated that the software in question was an intellectual work, but it never made clear whether the exclusive rights had been transferred or licensed under the agreement between the administration and the contractor. For that reason, the T.A.R. Lazio presumed that the contractor would have transferred to the Ministry all the exclusive rights on the software or, at least, that no exclusive right would have been retained by the contractor.[149]

The status of the intellectual work of administrative documents is not a ground for exempting the right to access according to Law 241/1990. In particular, the Court emphasized the different interests protected by intellectual property rights and the right to access administrative documents: ensuring economic interests of the author or owner of the intellectual work on the one hand, and on the other, effecting "widespread forms of control over the institutional functions

145. *Id*. at 1179.
146. *Id*. at 1180.
147. Juz. Cont. Adm., Feb. 18, 2019, (R.J. No. 0701, p. 2018) (Spain).
148. T.A.R. Lazio-Rome, Sez. III Bis, 22 Marzo 2017, n. 3769 23–24 (It.).
149. *Id.* at 20.

and the use of public resources, and … promot[ing] participation in public debate."[150]

As the access intended by the plaintiff-union had no economic exploitation purposes, the T.A.R. concluded that no exclusive right could be infringed. In consequence, the Court reasoned that access should be granted in the manner requested by the plaintiff (i.e., by displaying and obtaining a copy of the software in question). However, the information obtained in that way had to be restricted to a proper use—that is, to a use solely functional to the applicant's interest, which, according to the request made, was the protection of the rights of its affiliated members. Therefore, the access granted was solely and exclusively for such purpose, resulting in the subsequent liability before the owner for any use of the data obtained for purposes rather than those of the FOIA request.[151]

The Court conceded that the access intended was particularly pervasive insofar as it was directed precisely to the source code and the algorithm. Nevertheless, it reasoned that the public interest underlying the request (the assessment of the functionality of the algorithm and the existence of possible errors) could not be satisfied by a mere description of the algorithm in a memorandum, being necessary the inspection of the information sought.[152]

In France, the CADA seems to be inclined to apply the intellectual property exemption only in cases where the source code has been developed by a third party and the public entity is not the owner of exclusive rights. In this regard, the French Authority has regretted that the intellectual property rights of a third party may constitute an "obstacle" to the access, when the source code has been developed with "public funds" and in the framework of the "public service missions" carried out by the administration; thus, urging public institutions to review the terms in license agreements in this respect.[153]

V. AUTOMATED STATES IN THE SUNSHINE

It is worth revisiting Dehausse's well-known statement and rephrasing it like this: "'Government in the Sunshine' is a standard

150. *Id.* at 21, 23.

151. *Id.* at 21. *See also* GAIP, *supra* note 61 (applying the same a similar procedural solution—"conditional access"—restricting applicant's access on condition of using the source code displayed according to the FOIA request, *i.e.*, to verify the fairness of the results produced by the algorithm).

152. See *supra* note 148, at 22.

153. Comission d'accès aux documents administratif [CADA] [commission for access to administrative documents], May 31, 2018, 20180376.

problem of contemporary [algorithmic] governance."[154]

Scholars have regretted the lack of a clear "mapping of the [current] uses of AI in the public sector"[155] or "any 'roadmap' showing which systems a given public authority is planning, procuring, or deploying."[156] Thus, in the context of ADM systems, algorithmic opacity is not only the inability to understand why the algorithm produced a specific outcome but also the inability to know under which circumstances governments are using algorithmic systems (use cases), why (purposes), and how (correctness of the entire model).

In his description of the "Black Box Society," Pascal pointed out that transparency is not only an end in itself but also "an interim step on the road to intelligibility."[157] Applying this statement to algorithmic decision-making within public administrations, we could say that by way of ensuring a public scrutiny of ADM systems (AI-driven or not), FOIA regimes may contribute to lifting the veil of algorithmic opacity —in the sense described herein— and facilitate a better understanding of what, why, and how.

Therefore, the alleged futility of FOIA regimes to deal with the algorithmic opacity of ADM systems is not such.

Significantly, the Council of Europe has stressed that "transparency enhancement measures" on algorithms may facilitate scrutiny not only by the public but also independent experts or specialized agencies.[158]

---

154. Renaud Dehousse, *European Institutional Architecture after Amsterdam: Parliamentary System or Regulatory Structure*, 5 COMMON MKT. L. REV. 3, 595, 615 (1998).

155. Lorenzo Cotino, *SyRI, ¿A Quién Sanciono? Garantías frente al Uso de Inteligencia Artificial y Decisiones Automatizadas en el Sector Público y la Sentencia Holandesa de Febrero de 2020* [SyRI, Who shall I Sanction? Safeguards against the Use of Artificial Intelligence and Automated Decisions in the Public Sector and the Dutch Judgment of February 2020], 4 LA LEY PRIVACIDAD (2020) (Spain), https://www.researchgate.net/publication/349494176_ SyRI_a_quien_sanciono%27_Garantias_frente_al_uso_ de_inteligencia_artificial_y_decisiones_automatizadas_en_ el_sector_publico_y_la_sentencia_holandesa_de_febrero_de_2020.

156. ANSGAR KOENE ET AL., A GOVERNANCE FRAMEWORK FOR ALGORITHMIC ACCOUNTABILITY AND TRANSPARENCY 56 (2019), https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf.

157. FRANK PASQUALE, THE BLACK BOX SOCIETY. THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 8 (2015).

158. COUNCIL OF EUROPE, UNBOXING ARTIFICIAL INTELLIGENCE: 10 STEPS TO PROTECT HUMAN RIGHTS 9-10 (May 2019), https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64 ("The use of an AI system *must not only be made public* in clear and accessible terms, individuals must also be able to understand how decisions are reached and how those decisions have been verified. Oversight over an entire AI system must also be enabled by transparency requirements. This can be either in the form of *public disclosure of information on the system in question*, its processes, direct and indirect effects on human rights, and measures taken to identify and mitigate

As stated above, many FOIA regimes (e.g. in Australia, the UK, the European Union, Spain, and Mexico) usually contemplate a twofold approach: (1) providing the public with an enforceable right to request access to public records, according to which public institutions shall disclose any information requested, unless it falls under any statutory exemptions (the right of access); and (2) the obligation of making available to the public relevant information in electronic format by means of public registers, disclosure schemes, or tailored official web portals of transparency (public disclosure).[159]

It should be noted regarding this second approach that some jurisdictions are moving toward enacting specific legislation or adopting proposals to amend existing FOI regimes to include mandatory disclosure of source code and algorithms.

A. From the Constitutional Value of Access to Public Disclosure: Lessons from Parcoursoup

As we said before, the *Loi Lemaire* in France came to codify the well-established doctrine of the CADA on the access to the source code and algorithms used by public administrations. But the amendment of the Code of Relations between the Public and Administration (CRPA) operated by the *Loi Lemaire*[160] went further and sought to give greater transparency to algorithmic processing.

The CRPA provides for two systems of access to administrative documents: on the one hand, the "communication," by exercising the right of access (*droit à communication*),[161] and on the other, the "public dissemination," which entails the mandatory or spontaneous releasing

---

against adverse human rights impacts of the system, or in the form of an independent, comprehensive, and effective audit.") (emphasis added).

159. *See generally*, TOBY MENDEL, FREEDOM OF INFORMATION: A COMPARATIVE LEGAL SURVEY (Unesco ed., 2008), https://law.yale.edu/sites/default/files/documents/pdf/Intellectual_Life/CL-OGI_Toby_Mendel_book_%28Eng%29.pdf (analyzing comparative FOIA regimes providing mandatory obligation for public bodies to publish key information, promoting open government and regulating procedures to facilitate access to public records and requests of information). *See also*, Manuel Palomares Herrera, *Estudio comparado sobre transparencia y derecho de acceso en el ámbito internacional y su influencia en España* [Comparative study on transparency and right of access in the international sphere and its influence in Spain], 6 IUS HUMANI. REVISTA DE DERECHO 123-153 (2017) (Spain) (analyzing comparative legislation enhancing proactive disclosure of public records or imposing obligations to make specific information held by governments available to the public on the one hand, and on the other, the right to access to public records).

160. *See supra* note 54.

161. *See supra* note 55 (including provisions governing the right of access to administrative documents in Articles L-311 to R-311-15 CRPA).

of documents by electronic means (*diffusion des documents administratifs*).[162]

From the perspective of the right to access, the Article L.311-3-1 CRPA says that the individual decisions based on an algorithmic processing shall include an explicit notice of such processing to the interested party. The rules defining the algorithmic processing and the main features of its implementation shall be solely communicated to the interested party upon request. This provision has been completed by a regulation of the French State Council.[163]

First, Article R.311-3-1-1 CRPA specifically stipulates that the individual administrative decision shall contain a notice of the purposes of the algorithmic processing, the right to obtain the communication of the rules defining the processing and the main characteristics of its application, as well as the modalities of exercising the right to communication and of review, if appropriate, before the CADA.[164]

Second, at the request of the addressee of an individual decision, and pursuant to Article R.311-3-1-2 CRPA, the notice shall include in an "intelligible form": (1) the extent to which and how the algorithmic processing has contributed to the decision; (2) the data processed and their sources; (3) the processing parameters, and, where appropriate, their weighting, applied to the individual situation of the interested party; and (4) the operations carried out by the processing.[165]

From the perspective of public dissemination, Article L.312-1-3 CRPA compels public administrations to "publish online the rules defining the main algorithmic processes used in the accomplishment of their missions when they are the basis of individual decisions."[166]

Nevertheless, the legal framework described was not applied in the context of education, where once again opaque algorithmic processing

---

162. *Id.* (including provisions governing public dissemination of administrative documents by electronic means in Articles L-312-1 to D-311-11 CRPA).

163. Décret 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique [Decree 2017-330, of Mar. 14, 2017, on the rights of persons subject to individual decisions made on the basis of algorithmic processing], JOURNAL OFFICIEL DE LA REPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE] Mar. 16, 2017, p. 1. (Notice that the term «personnes» used by the French Law refers to individual and legal persons who may be affected by an administrative decision assisted or made by means of algorithmic processing.)

164. *See generally Etalab, Les Algorithmes Publics: Enjeux et Obligations*, EXPLIQUER LES ALGORITHMES PUBLICS, available at https://guides.etalab.gouv.fr/algorithmes/guide/#_1-a-quoi-servent-les-algorithmes-publics (making clear that this information shall be provided not only in individual administrative decisions noticed to the addressee, but also online to inform the general public).

165. *See supra* note 55.

166. *Id.*

operated by public institutions was questioned by some supervisory authorities, affected parties, interested third parties, and public opinion. This is the case of the algorithms deployed within the national platform Parcoursup to automate the pre-registration process for undergraduate studies.[167]

In fact, the use of Parcoursup and the algorithmic processing operated by the platform were challenged before administrative courts. Out of the ninety appeals filed between 2018 and 2019, there were at least forty-eight appeals against decisions of the universities dismissing the disclosure of the algorithms implemented to assess students' applications.[168]

Despite the provisions set forth in Articles L.311-3-1 and L.312-1-3 of the CRPA and the affirmative decisions issued by the CADA compelling public disclosure,[169] the reality was that Ministère de l'Enseignement Supérieur et de la Recherche et de l'Innovation (MESRI) had neither published the entire source code of Parcoursup nor all the algorithmic procedures implemented by the universities. This point was confirmed by an external audit commissioned by the French Court of Auditors (*Cour des Comptes*), which assessed the efficiency and fairness of the algorithmic processing carried out to rank candidates within the

167. The Loi n. 2018-166 du 8 mars 2018 relative à l'orientation et à la réussite des étudiants, also known as Loi ORE [Law on students' orientation and achievement] replaced the former Admission Post-Bac (APB) platform with a new one (Parcoursup) for enrolling in higher education programs, particularly those whose capacity was lower than the number of applications received.

168. COUR DES COMPTES, UN PREMIER BILAN DE L'ACCES A L'ENSEIGNEMENT SUPERIEUR DANS LE CADRE DE LA LOI ORIENTATION ET REUSSITE DES ÉTUDIANTS. COMMUNICATION AU COMITE D'ÉVALUATION ET DE CONTROLE DES POLITIQUES PUBLIQUES DE L'ASSEMBLEE NATIONALE 169 (Feb. 2020) (Fr.), https://www.ccomptes.fr/system/files/2020-03/20200227-rapport-premier-bilan-loi-ORE-3.pdf.

169. Comission d'accès aux documents administratifs [CADA] [commission for access to administrative documents], Sept. 6, 2018, 20182120 (Fr.) (granting access to functional specifications of Parcoursup platform sought by the applicant); Comission d'accès aux documents administratif [CADA] [commission for access to administrative documents] Sept. 6, 2018, 20182093, (compelling the applicant to submit his request of access to the algorithms used by universities to the universities rather than to the Ministry of Education); Comission d'accès aux documents administratifs [CADA] [commission for access to administrative documents] Sept. 8, 2018, 20182455 (ordering the public disclosure by electronic means of functional specifications of Parcoursup in order to make them accessible to anyone); Comission d'accès aux documents administratifs [CADA] [commission for access to administrative documents] Jan. 10, 2019, 20184400 (granting access to algorithmic procedures used by the decision tool implemented by the University of Aix-Marseille to process the applications of pre-enrollment in its bachelor degrees via Parcoursup platform as well as their source codes).

national platform.[170]

The audit found that two types of algorithmic processes had been put in place by the source code of the Parcoursup application. First, there were "local algorithms" embedded in a decision-making support tool that could be used—to their discretion—by local actors, such as the Academic Boards for the Assessment of Applications (CEVs)[171] or the Head Office at each university.[172] These local algorithms were run to automatically rank the students' applications for pre-registration in each degree program. Second, the source code of the platform implemented an algorithmic processing, the so-called "national algorithm," to calculate the final ranking of the students' applications based on the assessments previously made by the local actors, finally matching candidates' applications with the available spaces offered by universities within their degree programs. But only this national algorithm had been made public, despite the fact that—to the Court of Auditors—such information was of "limited interest to ensure the transparency of the entire system."[173]

Bearing this in mind, by running supervised machine learning techniques, and more specifically, random forests, the audit eventually identified and deciphered up to "15,000 local algorithms" implemented by the universities.[174]

The audit found that local algorithms applied "disparate" and "questionable" parameters for the assessment of applicants' academic records (e.g., reputation of secondary school, percentage of successful students at *baccalauréat*). In particular, it was revealed that the students' school (*lycée*) of origin as a criterion of eligibility was prioritized very often by the algorithm, and this resulted in a classification of the students' applications in such a manner that did not ensure the objectivity and fairness of the procedure.[175]

---

170. COUR DES COMPTES, *supra* note 168, at 53 (concluding that, in practice, the information published by the MESRI in the repository (available at https://framagit.org/Parcoursup/algorithmes-de-Parcoursup) only represented 1% of the lines of code and less than 2% of the JAVA and SQL files of the source code; and showing that the files and lines of code in SQL that had been made public were quantitatively less than lines published in Java). This point is relevant because SQL files allow structuring and analyzing data, while those written in JAVA allow developing applications and implementing algorithmic calculations in connection with SQL files.)

171. In French, *Commissions d'Examen des Vœux*.

172. The *Rectorat*, that is, the Chancellor or President's Office of each institution.

173. COUR DES COMPTES, *supra* note 168, at 53–54, 142.

174. *Id.* at 6.

175. *Id.* at 64–65 (quoting a decision of the French Ombudsman who had pointed out that the lycée of provenance criterion could amount to "a discriminatory practice if it results in candidates being treated differently and excluded for this reason, based on the geographic location of their *lycée*".) *See also* DÉFENSEUR DES DROITS, DÉCISION 2019-021

Among the conclusions drawn by the Court of Auditors' Report, there was a specific recommendation to the MESRI to engage in further public disclosure of the processing operated by Parcoursup, in order "to inform the public debate on 'local algorithms' and the . . . decision making through automated means."[176] In particular, the Court stressed that "with a view to greater transparency, there should be no objection to making public all the *parameters* of the decision tools used by the CEVs."[177] In fact, the Court of Auditors' recommendation endorsed the position held by some authorities, such as the CADA[178] and the CNIL,[179] that had respectively urged the universities to make public their local algorithms.

The issue was finally settled by the *Conseil Constitutionnel* in a judgment dealing with a preliminary ruling on an issue of unconstitutionality (*question prioritaire de constitutionnalité*) lodged by the Union Nationale des Étudiants de France (UNEF),[180] which had sought to challenge the statutory provisions that regulate the pre-registration procedure through Parcoursup on constitutional grounds.

The applicant-union considered that some provisions set forth in

---

DU 18 JANVIER 2019 RELATIVE AU FONCTIONNEMENT DE LA PLATEFORME NATIONALE DE PRÉINSCRIPTION EN PREMIÈRE ANNÉE DE L'ENSEIGNEMENT SUPÉRIEUR [Decision 2019-021 of January 18, 2019 concerning the Operation of the National Platform for Pre-Registration in the First Year of Higher Education] ¶ 89, https://juridique.defenseurdes droits.fr/doc_num.php?explnum_id=18303.

176. COUR DES COMPTES, *supra* note 168, at 54, 68 (arguing that the publication of the local algorithms used by the CEVs was desirable not only for pedagogical reasons, aimed at better informing the students how their previous education could be weighted in each university and university degree, but also to comply with a "mandate of transparency," insofar as the lack of public disclosure and the systematic reluctance of Universities to communicate the algorithms to third parties seeking access resulted in a clear "risk of mistrust.")

177. *Id.* at 66 (emphasis added).

178. Comission d'accès aux documents administratifs [CADA] [commission for access to administrative documents] Jan. 10, 2019, 20184400 (Fr.) (concluding that, though the French Code of Education does not impose an obligation on universities "to disseminate online the rules defining the main algorithmic procedures . . . when they are the basis of individual decisions . . . , *it does not restrain them from ensuring their spontaneous disclosure*.") (emphasis added).

179. COMPTES RENDUS DE LA COMMISSION DE LA CULTURE, DE L'EDUCATION ET DE LA COMMUNICATION: AUDITION DE MME MARIE-LAURE DENIS, PRÉSIDENTE DE LA COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS (Jul. 17, 2019) (Fr.), https://www.senat.fr/compte-rendu-commissions/20190715/cult.html#toc3 (recommending the public disclosure of the local algorithms as a policy of "good practice," notwithstanding the fact that the processing of the candidates' dossiers by the CEVs, in order to apply all the safeguards set in Art. 22 of the GDPR, was not fully automated).

180. Conseil Constitutionnel [CC] [Constitutional Council], decision No. 2020-834QPC, Apr. 3, 2020, JOURNAL OFFICIEL DE LA REPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE] April 4, 2020, 33 (Fr.).

Article L. 612-3 of the *Code de l'Éducation*[181] violated the rights and freedoms guaranteed by the French Constitution of 1958 and other fundamental texts to which the Preamble of the constitutional text refers, in particular the Declaration of the Rights of Man and of the Citizen of 1789. The contested provision of the Code of Education stated that

> "[i]n order to guarantee the appropriate protection of the secrecy of deliberations of pedagogical boards in charge of the assessment of the applications submitted as part of the national pre-registration procedure . . . the obligations arising from Articles L.311-3-1 and L.312-1-3 of the Code on Relations between the Public and the Administration shall be met insofar as the applicants are informed of the possibility of obtaining, upon request, the information relating to the criteria and methods used to assess their applications as well as the pedagogical reasons justifying the decision made."

Hitherto, the French State Council (Conseil d'État) has construed Article L.612-3 in the sense that the Code of Education precluded the application of the general regime provided in the CRPA, laying down a qualified access to information instead. Such qualified access would be an enforceable right only by those candidates who submitted a request for such information following the final decision granting or dismissing the pre-registration application, and only in relation to the criteria applied to them individually.[182]

The applicant-union argued that such interpretation was contrary to the right to access administrative documents recognized in Article 15 of the Declaration of the Rights of Man and the Citizen of 1789, because it excluded the access of third parties or any candidate willing to know at any time the algorithmic processing put in place by the universities. In the union's view, neither the deliberative secrecy of CVEs nor any other reason could justify such exclusion. Moreover, that provision would violate the right of the candidates to a judicial remedy by precluding them from challenging not only the lack of communication of

181. Ordonnance n. 2000–549 du 15 juin 2000 relative à la partie Législative du code de l'éducation [Ordinance n. 2000-549 of June 15, 2000 regarding the Legislative part of the code of education], JOURNAL OFFICIEL DE LA REPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE] n. 0143 du 22 juin 2000, https://www.legifrance.gouv.fr/loda/id /JORFTEXT000000583540.

182. *See* supra note 180, at ¶11–12.

the algorithms implemented, but also the pre-registration denials.[183]

First, the Constitutional Council conceded that, pursuant to Article 15 of the Declaration of 1789, the "[s]ociety has the right to hold any public official accountable for his office." This provision guarantees the right of access to administrative documents. It is for the lawmaker to establish statutory limits to this right according to constitutional requirements or justified by the general interest, provided that such limitations are not disproportionate to the aim pursued. Second, the Council argued that the national pre-enrollment procedure was not fully automated because the use of algorithms was a discretionary decision of the universities and, in such a case, the decision taken on each application could not be based exclusively on the algorithmic processing but required an individual assessment of the merits of the candidates by CEVs and, then, by the head of the university to ensure human oversight. Third, the candidates affected by dismissals could obtain from the university, upon request, the criteria applied by the algorithmic processing implemented by CEVs.[184]

However, in the view of the Constitutional Council, this qualified access would only benefit the candidates. In consequence, once the national procedure of pre-registration is finished, precluding third parties from seeking information on the criteria and procedures applied by universities constituted a disproportionate infringement of the right guaranteed by Article 15 of the Declaration of 1789 in relation to the general interest arising from the protection of the secrecy of the deliberations of the CEVs. Consequently, the contested provision cannot be construed as exempting universities from the obligation to publish the criteria upon which the pre-enrollment applications were assessed. In addition, universities shall also specify, if applicable, to what extent algorithmic processing was used to carry out the assessment of candidates' applications, published in the form of a report.[185]

B. Towards Public Disclosure in Comparative Law

The "constitutional value" of the judgment—to the CADA[186]—cannot be ignored. In establishing an interpretation of the contested provision consistent with the Constitution, the French Constitutional Council connects the right of access to administrative documents, as read in the Universal Declaration of 1789, with an active obligation of the

183. *Id.* at ¶ 2.
184. *Id.* at ¶ 8, 13–16.
185. *Id.* at ¶17.
186. *See generally* Comission d'accès aux documents administratif [CADA] [commission for access to administrative documents] Jan. 13, 2022, 20213847 (Fr.).

administration to publish relevant information on the algorithmic processing applied in individual decisions affecting citizens. Though the Council did not go further and only required *ex post* transparency of the algorithmic processing put in place by CEVs—rather than *ex ante* duty of information—the judgment underscored the importance of transparency measures addressed to the public at large.

Accordingly, some jurisdictions are moving towards transparency measures seeking to make available to the public at large (and not only to the affected persons) relevant information of the algorithmic systems. These measures differ across jurisdictions in relation to the relevant information to be published and the instruments used for such publicity.

Following the taxonomy proposed by some studies,[187] transparency mechanisms across jurisdictions aimed at civil society and citizens may fall into any of the following categories or a combination of them: (1) statutory requirements of public disclosure for source code, algorithms or relevant information of ADM systems (Canada, France, Germany, Valencia) relying on FOI regimes; (2) public registries or inventories of algorithmic systems (Canada, United States, European Union); and (3) specific provisions in sectoral legislation requiring explanations of algorithmic logics, seeking to allow the public and policymakers to understand how an algorithmic decision was reached (European General Data Protection Regulation, France, Canada).

1. Requirements for Public Disclosure Relying on FOI Regimes or Sectoral Legislation

For the time being, the Canadian ADM Directive[188] is arguably one of the pieces of legislations in comparative law that imposes the greatest transparency measures on ADM systems and algorithmic processes. Section 6.2 of the Directive is entirely devoted to regulating the transparency measures for ADM systems of Level I, II, III, and IV impact.[189]

---

187. ADA LOVELACE INSTITUTE, AI NOW INSTITUTE & OPEN GOVERNMENT PARTNERSHIP, *supra* note 1, at 18–19.

188. This Directive applies to any system, tool, or statistical model used by federal government to recommend or make an administrative decision about a client, with the exception of National Security Systems. The Directive imposes a set of requirements on the federal government's use of ADM systems which implement AI to make, or assist in making, administrative decisions on a risk approach basis and in a manner that is compatible with core administrative law principles such as transparency, accountability, legality.

189. Appendix B of the ADM Directive ranges automated systems from Level I to IV in relation to the impacts (little to no impact, moderate, high-, or high-risk impact) and the reversibility and duration thereof on the rights of individuals or communities, the health

Transparency measures imposed by the Canadian Directive shall include: (1) providing a prior notice in plain language through all service delivery channels in use (internet, in-person, mail, or telephone) that the decision rendered will be undertaken in whole or in part by an ADM System of Level II, III, or IV; (2) providing notices prominently and in plain language, pursuant to the Canada.ca Content Style Guide;[190] and (3) releasing custom source code owned by the government of Canada, as per the requirements specified in the "Enterprise Architecture Framework" (EA framework)—unless it processes data classified as secret, top secret, or protected. The disclosure would otherwise be exempted or excluded under the Access to Information Act, or an exemption is provided by the Chief Information Officer of Canada.

According to the EA framework, when implementing application architecture practices and transitioning from legacy systems, the Government of Canada shall evolve significantly to the use of reusable and open-source solutions hosted in a public cloud. This includes selecting existing solutions that can be reused over custom-built and registering open-source software to the Open Resource Exchange.[191]

In addition, Sec. 6.1.4 of the ADM Directive also imposes the obligation to release the final results of Algorithmic Impact Assessments (AIA) in an accessible format via Government of Canada websites and any other services designated by the Treasury Board of Canada Secretariat pursuant to the Directive on Open Government.[192]

Bearing in mind that much of the algorithmic systems used by public institutions have been developed by third-party contractors, the Council of Europe argues that the provision of entire algorithms or the source code to the public is an unlikely solution due to the existence of

---

or well-being of individuals or communities, the economic interests of individuals, entities, or communities, or the ongoing sustainability of an ecosystem.

190. In addition, for Level II and III ADM systems, it shall be mandatory for authorities under the application of the Directive to publish documentation on relevant websites about the automated decision system, in plain language, and describing: (1) how the components work; (2) how it supports the administrative decision; (iii) results of any reviews or audits; and (3) a description of the training data, or a link to the anonymized training data if this data is publicly available.

191. Treasury Board of Canada Secretariat, Government of Canada Enterprise Architecture Framework, https://www.canada.ca/en/government/system/digital-governm ent/policies-standards/government-canada-enterprise-architecture-framework.html#toc04 (last visited on Aug. 31, 2022).

192. Sec. 6 of the ADM Directive imposes the completion of an AIA prior to the production of any ADM System; and updating of the AIA when the system functionality or the scope of the ADM changes. An AIA is a "framework to help institutions better understand and reduce the risks associated with Automated Decision Systems and to provide the appropriate governance, oversight and reporting/audit requirements that best match the type of application being designed."

enforceable proprietary rights over such information. Instead, the Council is more inclined to make available to the public "key subsets of information about the algorithms . . . for example which variables are in use, which goals the algorithms are being optimi[z]ed for, the training data and average values and standard deviations of the results produced, or the amount and type of data being processed by the algorithm."[193] Yet the Council does not clarify what would be the appropriate instrument to render this information public.

In the view of the German Commissioners for Freedom of Information, the legislature should engage in taking the appropriate measures to ensure transparent and responsible use of algorithms and procedures. Most importantly, such measures should be embedded in the respective legislation on transparency and freedom of information. In detail, the Commissioners urged the federal and state legislators to implement specific requirements in their respective FOI regimes to ensure sufficient transparency about the algorithms used. This should include: (1) the data categories of the input and output data of the processing; (2) the logic contained therein, in particular the calculation formulas used, including the weighting of the input data, information about the underlying expert knowledge, and the individual user settings; and (3) the scope of the decisions based thereon and the potential impact of the processing. This information shall be published in a meaningful, comprehensive, and understandable manner.[194]

Similarly, in Spain, the legislature of the Autonomous Community of Valencia has recently enacted a new FOI statute, the Law 1/2022, of April 13, on Transparency and Good Governance of the Autonomous Community of Valencia, which makes an obligation for public administrations and public bodies subject to this statute to publish in their respective official web portals of transparency "the list of the algorithmic or artificial intelligence systems that have an impact on administrative procedures or the provision of public services with a comprehensible description of their design and operation, the level of risk involved, and a contact point to address in each case, in accordance with the principles of transparency and explainability."[195]

Taking into account that such a provision is neither provided in the state legislation nor the rest of the regional legislation of freedom of information, the regional initiative is a good starting point, but it fails to

193. COUNCIL OF EUROPE, *supra* note 11, at 38.

194. BUNDESBEAUFTRAGTE FÜR DEN DATENSCHUTZ UND DIE INFORMATIONSFREIHEIT ET AL., *supra* note 4, at 3–4 (clarifying that "[t]o the extent legally possible, this information should be published." which obviously refers to the possible application of statutory exemptions to protect other legitimate public or private interests).

195. *See* B.O.E. 2022, 119, https://www.boe.es/buscar/doc.php?id=BOE-A-2022-8187.

include relevant information such as the data categories used by the algorithmic systems as input and output data are generated. This information is also crucial to understanding the logic underlying the algorithmic system and the potential impacts of the processing on the governed.

2. Open-Source Code Repositories and Public Inventories or Registries of Algorithmic Systems

In some jurisdictions, governments have deployed online catalogues or repositories for making available open-source code developed and used by public administrations.

The Open Resource Exchange is a catalogue developed by the government of Canada which includes five main services that focus on sharing solutions in an open-source format which are freely available for use.[196] Most of the open-source code is published in Github, a code-hosting platform for software development, version control, and collaboration.

In France, pursuant to Articles L.300-2, L.311-1, and L.321-1 CRPA, source code produced by public bodies are communicable and reusable administrative documents unless statutory exemptions are applicable. And Article L.312-1-3 CRPA requires organizations to "publish online the rules defining the main algorithmic processes used in the accomplishment of their missions when they are the basis of individual decisions." According to this legal framework, in May 2018, the French Government published the *Politique de contribution aux Logiciels Libres de l'État*, aimed at setting the rules and principles to be respected for the opening of source codes and establishing best practices.[197] Since then, the French Chief Data Officer, Etalab, dependent upon the Interministerial Digital Directorate (DINUM), has been publishing the list of algorithms implemented by different departments,[198] the list of source code repositories open by public bodies in order to facilitate their reuse by third parties (business, developers, researchers) or anyone willing to do it, and the list of public organizations that are publishing

196. GOVERNMENT OF CANADA, Open Resource Exchange, https://code.open.canada.ca/en/index.html (last visited on Aug. 31, 2022).

197. *See* ETALAB, OUVRIR LES CODES SOURCES (Aug. 11, 2022), https://guides.etalab.gouv.fr/pdf/guide-logiciels.pdf. *See also* Etalab, Guide des algorithmes publics, https://etalab.github.io/algorithmes-publics/guide.html (last visited on Aug. 31, 2022).

198. ETALAB, ETALAB/LOGICIELS-LIBRE, https://git.sr.ht/~etalab/logiciels-libres/tree/master/codes-sources-algorithmes-publics.md (last visited Aug. 31, 2022).

the source code of the applications used by them.[199]

In the United States, the Executive Order of 2020, Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, requires federal agencies to conduct an annual inventory of their AI use cases, and to publish them to the extent possible, excluding AI use cases that are classified, sensitive, or used in defense or national security systems by the Department of Defense or Intelligence Community.[200]

Agencies started publishing their first annual inventories in June 2022. The inventories published to date include the Departments of State, Agriculture, Commerce, Energy, Health and Human Services, Homeland Security, Justice, Labor, Veterans Affairs, the NIST, the US Agency for International Development, and the US Environmental Protection Agency. The Federal Chief Information Officers Council has provided guidance to agencies on how to conduct their inventories.

The information to be published mandatorily includes: (1) AI use case name and agency/subagency or office; (2) contact information; (3) a short summary of what the AI does, including a high-level description of system inputs and outputs; and (4) lifecycle stage (planned, in production). On an optional basis, the agency may also publish the AI techniques used; the data approach (information about the origin of the training, validation, or test data, and if data is publicly available); or the name of the information system associated with the AI use case.[201]

Algorithmic registries are other instruments which are gaining prominence. These registries are directories providing relevant information on algorithmic systems used by organizations, including public authorities, agencies, or bodies.

If enacted, the EU Proposal of AI Regulation will establish a system for registering stand-alone high-risk AI applications in a public EU-wide database. This registration will also enable competent authorities, users, and other interested people to verify if a high-risk AI system[202]

---

199. ETALAB, CODEGOUV. BROWSE FRENCH PUBLIC SECTOR SOURCE CODE, available at https://code.gouv.fr/#/ (last visited Aug. 31, 2022).

200. Exec. Order No. 13960, 85 Fed. Reg. 78939–78943 (2020).

201. FEDERAL CHIEF INFORMATION OFFICERS, 2021 GUIDANCE FOR CREATING AGENCY INVENTORIES OF ARTIFICIAL INTELLIGENCE USE CASES, https://www.cio.gov/assets/resources/2021%20Guidance%20for%20Creating%20Agency%20Inventories%20of%20AI%20Use%20Cases%2010.06.2021.docx (last visited on Aug. 31, 2022).

202. See Recital (32) of the Proposal, where the concept of high-risk, refers to AI systems that pose a "high risk of harm to the health and safety or the fundamental rights of persons taking into account both the severity of the possible harm and its probability of occurrence." Article 6 of the Proposal identifies two main categories of high-risk AI systems: (1) AI systems intended to be used as safety component of products that are subject to third party ex ante conformity assessment; and (2) other stand-alone AI systems with mainly fundamental rights implications that are explicitly listed in Annex III of the Proposal (i.e. biometric identification and categorization of natural persons; education and

complies with the requirements laid down in the proposal to enhance oversight over these systems.

To feed this database, AI providers, regardless of if they are public or private organizations,[203] will be obliged to provide meaningful information about their systems, before placing them on the market or otherwise putting them into service.

Among the information to be included in the registry shall be: (1) name, address, and contact details of the provider; (2) AI system trade name and any additional unambiguous reference allowing identification and traceability of the AI system; (3) description of the intended purpose of the AI system; (4) status of the AI system (on the market, or in service; no longer placed on the market/in service, recalled); (5) member states in which the AI system is or has been placed on the market, put into service, or made available in the Union; (6) a copy of the EU declaration of conformity referred to in Article 48[204]; (7) electronic instructions for use, with the exception of high-risk AI systems in the areas of law enforcement and migration, asylum, and border control management; and (8) a URL for additional information (optional).

3. Explanations of AI-driven Decisions

It is important to make it clear that administrative transparency pursued by FOIA regimes (a legal principle) cannot be confused with algorithmic transparency (technical concept). At the same time,

---

vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private services and public services and benefits; law enforcement, migration, asylum and border control management; and Justice and democratic processes).

203. It should be noted that, pursuant to Article 3(1) of the EU Proposal, an AI provider means any "natural or legal person, *public authority, agency or other body* that develops an AI system or that has an AI system developed with a view to placing it on the market or *putting it into service* under its own name or trademark, whether for payment or free of charge." (emphasis added). Under this definition an AI provider could be any public authority, agency or body that develops in-house AI systems; or any contractor of the public authority, agency of body.

204. A "conformity assessment" is an *ex-ante* process of verifying whether the requirements set out for high-risk systems in Title III, Chapter 2 of the EU Proposal have been fulfilled (Article 3 (20) of the EU Proposal). Among the requirements imposed on high-risk systems, the Proposal include: (a) the quality of data sets used to train, validate and test the AI systems in order to ensure that they are relevant, representative, free of errors, complete and with appropriate statistical properties (Recitals 44 and 45, Article 10); (b) Technical documentation (Recital 46, Article 11, and Annex IV); (c) automatic record-keeping of events (logs) (Article 12); d) transparency and provision of information to users (Recital 47 and Article 13); (e) human oversight (Recital 48 and Article 14); and (f) robustness, accuracy and cybersecurity (Recitals 49 to 51, and Article 15).

algorithmic transparency is related to explainability.

The frequent overlapping between "administrative transparency" and "algorithmic transparency"—at least in some legal literature—is what probably makes an important sector of scholars consider that FOI regimes do present many constraints to guarantee the transparency of ADM systems and to open the *black boxes* of public administrations. Yet, this approach simply departs from a wrongful premise. The goal of administrative transparency is not to guarantee the understandability and reasonableness of administrative decisions (algorithmically driven or not) in themselves. On the contrary, FOIA regimes are intended to ensure the public scrutiny of administrative decisions and how the power has been exercised. And such public scrutiny is what makes it possible to ascertain whether administrative decisions were made according to legality (i.e., consistent with the factual premises and consequences prescribed in the applicable law) and on reasonable grounds. No more and no less.

Broadly speaking, in AI systems there is an inverse relationship between interpretability and performance, whereby simple models are more interpretable, but have a lower predictive capacity and vice versa.[205] The branch of AI science, called "Explainable Artificial Intelligence" (XAI), is devoted to developing techniques aimed at generating more explainable models and differentiates between the following concepts: interpretability, explainability, and transparency.

On the one hand, "interpretability" is a passive attribute of a model which means how understandable or intelligible an algorithmic model is to a human observer. The interpretability of a model is higher if it is easy for a person to reason and trace in a coherent way why the model made a particular prediction.[206]

On the other hand, "explainability" is an active attribute of the model that refers to the ability to generate an explanation of the model's behavior based on the data used, the results obtained, and the entire decision-making process according to the audience for which the explanation is intended (e.g., authorities, experts, third-party auditors, certification bodies, public at large, and addressees of an individual decision). Explanations are instruments by which the decisions of an algorithmic model can be explained in a more clear, understandable, transparent, and interpretable manner. Therefore, if interpretability is the ultimate goal, explanations are tools to achieve the interpretability

---

205. Alejandro Barredo et al., *Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI*, 58 INFORMATION FUSION 82, 100 (2020).

206. *Id.* at 84; Diogo V. Carvalho, et al., *Machine Learning Interpretability: A Survey on Methods and Metrics*, 8 ELECTRONICS 8, 10 (2019).

of the model.[207]

In turn, a distinction must be made between models that are "interpretable by design" (i.e., "transparent models") and models that, not being interpretable *prima facie*, can nevertheless be explained by means of different techniques which extract relevant information from the model to generate explanations.[208]

Consistently, the "transparency" of AI models is determined by the degree of intrinsic interpretability of a specific model. Therefore, from a technical point of view, transparency is an attribute of the model that defines the degree of comprehensibility that a model itself has for a human observer. Transparency can be measured at three levels.[209] First, in relation to the model as a whole ("simulability"), transparency means that the model can be reproduced or replicated by a human in a reasonable time from the data and parameters of the model.[210] Second, in relation to its individual components ("decomposability"), transparency means that the components of the model, inputs, parameters, and calculations admit an intuitive explanation. Third, in relation to the training algorithm implemented by the model ("algorithmic transparency"), this means the ability to understand the process operated by the model to produce a specific outcome from the data.

Consequently, an AI model is considered transparent if it is interpretable by itself (i.e., if the overall performance of the model, its individual components, and its learning algorithm are intelligible or understandable to a human). The overall transparency of a model will depend, in any case, on an appropriate balance between simulatability, decomposability, and algorithmic transparency.[211]

This technical approach has been embraced by the High-Level Expert Group on Artificial Intelligence of the European Commission (HLEGAI) by requiring AI systems for being trustworthy to comply with

---

207. Carvalho, *supra* note 206, at 15.

208. Brent Mittelstadt, Chris Russell & Sandra Wachter, *Explaining Explanations in AI*, 19: PROCEEDINGS OF THE CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 2 (2019); Barredo, *supra* note 205, at 83.

209. Mittelstadt, Russell & Wachter, *supra* note 208, at 2; Zachary C. Lipton, *The Mythos of Model Interpretability*, 16 ACM QUEUE, 3, 12 (2018); Bruno Lepri et al., *Fair, transparent and accountable algorithmic decision-making processes. The premise, the proposed solutions, and the open challenges*, 31 PHILOS. TECHNOL 619 (2018); Barredo, *supra* note 205, at 88–100; ICO & ALAN TURING INSTITUTE, *supra* note 142, at 61–63, 115–18.

210. Hous. Fed'n of Tchrs., 251 F. Supp. 3d. at 1174. (This is precisely why the Houston Court considered the EVAAS system non-transparent as a blackbox, because the EVAAS scores could not be replicated at all.)

211. Barredo, *supra* note 205 at 90; ICO & ALAN TURING INSTITUTE, *supra* note 142, 67–68.

the principle of explicability. This principle entails that the models "need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions—to the extent possible—explainable to those directly and indirectly affected." The degree to which explicability is needed is highly dependent on the context and the severity of the impacts if outputs are erroneous or inaccurate.[212]

When dealing with the problem of interpretability, explainability, and transparency of decisions made by AI models, some jurisdictions have started to introduce specific provisions in sectoral legislation or in FOI regimes requiring explanations of algorithmic logics to allow the public, policymakers, and other relevant stakeholders to understand how an algorithmic decision was reached.

In the European Union, the legislation on personal data protection seems to address this issue. Taking into account that most of ADM systems are applied to individual persons, it is said that the guarantees provided by the data protection regulation, at least in the European context, are enough to ensure the transparency and explainability of algorithmic processing carried out by processors, including public authorities.[213] But this is true only to a certain and limited extent.

In fact, neither the European General Data Protection Regulation ("GDPR")[214] nor the Directive 680/2016 ("Enforcement Directive")[215] seem to satisfactorily address algorithmic processing based on personal data.

Articles 13.2 (f), 14.2 (g), 15.1 (h) and 22 GDPR contemplate specific safeguards applied to automated decision-making, including profiling. These safeguards include: (1) providing a meaningful amount of

---

212. HLEGAI, ETHICS GUIDELINES FOR TRUSTWORTHY AI (European Commission, 2019) 13, 19, https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html.

213. *See* Bryce Goodman & Seth Flaxman, EUROPEAN UNION REGULATIONS ON ALGORITHMIC DECISION-MAKING AND A "RIGHT TO EXPLANATION" 38 AI MAGAZINE 3, Oct. 10. 2017, 1–5 (arguing that the GDPR creates a "right to explanation," whereby individuals can ask for an explanation of an algorithmic decision that was made about them); Andrew D. Selbst & Julia Powles, *Meaningful Information and the Right to Explanation*, 7 INT'L DATA PRIV. L. 4, 233, 235-237 (2017) (purporting that a plain and contextual reading of recital (71) and Articles 13(2)(f), 14(2)(g), 15(1)(h), and 22 supports a right to explanation).

214. *See* Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data and Repealing Directive 95/46/EC (General Data Protection Regulation), O.J. (L. 119) 1–88.

215. Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data by Competent Authorities for the Purposes of the Prevention, Investigation, Detection of Criminal Prosecution of Criminal Offences or the Execution of Criminal Penalties, and on the Free Movement of Such Data, O.J. (L. 119) 89–131.

information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject, either complying with mandatory and *ex ante* transparency in prominent, meaningful and timely privacy notices available for anyone, or ensuring the right of access of data subject to such information; and (2) ensuring the right of the data subject to obtain human intervention on the part of the controller, to express their point of view, and to challenge the automated decision.

But safeguards provided by GDPR are exclusively applied for the type of automated decision-making referred to in Article 22.1, namely, decisions based *solely* on automated processing (including profiling), which produce legal effects on the data subject or similarly significantly affects them. This means that individual decisions made on partial automated processing (because there exists some degree of human intervention) would be out of the scope of such safeguards. The Article 29 Data Protection Working Party of the European Commission (A29WP) has produced some guidance on the interpretation and scope of the said GDPR provisions and the meaning of automated individual decision-making pursuant to Article 22.1.[216]

Nevertheless, scholars have long criticized the great ambiguity of the A29WP Guidelines and described relevant pitfalls with regards to algorithmic processing of personal data within GDPR's Articles 13.2 (f), 14.2 (g), 15.1 (h), and 22. According to them, the GDPR includes restrictions such as: (1) carve-outs for intellectual property and trade secrets; the limited scope of the GDPR safeguards, exclusively applied to individual decisions made by fully automated systems, including profiling, which produce "legal" or similarly "significant" effects; (2) the timing of such safeguards in relation to the decision being made; the non-binding provision in Recital (71), which further includes the right to obtain "an explanation of the decision reached after [the) assessment" made by the solely automated processing (including profiling); (3) lack of clear-cut requirements for such explanations, leading to substantial legal uncertainty; (4) the extent of the human oversight, how to ensure the human-in-the-loop principle and the practical difficulties in knowing when or how automated decisions are being made; (5) the relative ease with which "meaningful" human intervention can be diluted within the automation-induced complacency; or (6) the real impacts on individuals and sensitive collectives, particularly in relation to "smart" environments, such as IoT applications or online platforms, in relation to the full compliance with transparency obligations set forth in Articles

216. *See* European Commission, Guidelines on Automated Individual Decision-making and Profiling for the Purposes of Regulation 2016/679 29 (2018), https://ec.europa.eu/newsroom/article29/items/612053.

12-14 GDPR in privacy notices. Such pitfalls can be even more problematic in AI environments as Recital (71) GDPR does not establish mandatory requirements to open the *black box* nor an enforceable right to obtain an explanation.[217]

Pitfalls described by scholars in GDPR can be even more challenging when automated decision systems are implemented by governments as long as some specific safeguards of Article 22 GDPR (the right of the data subject to obtain human intervention on the part of the controller, to express their point of view, and to challenge the decision) are only applicable when lawful basis for processing relies on a contract between the data subject and a data controller or the explicit consent of the data subject. But such safeguards are not established for personal data processing where the legal basis applied is "the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller" (Article 6.1 (e) GDPR), given that this legal basis is the most frequently applied in processing carried out by public administrations.[218] Even when automated decisions of Article 22.1 GDPR are authorized by national legislation of the EU Member States, the GDPR only requires such legislation to lay down "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests," without describing what is meant by "suitable measures." This inevitably leads to a great legal uncertainty.

For example, Article 41 of the Spanish Law 40/2015 allows public administrations to engage in "automated administrative action," i.e., "act or action carried out entirely by electronic means by a public administration within an administrative procedure and in which a public employee has not intervened directly." Section 2 of the same Article dictates that, before engaging in any automated administrative action, the competent body or bodies for the definition of the specifications, programming, maintenance, supervision, quality control,

---

217. Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, 16 DUKE L. & TECH. REV. 18, 21 (2017). *See* also Sandra Wachter, Brent Mittelstadt, Luciano Floridi, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, 7 INTERNATIONAL DATA PRIVACY LAW 2, 76, 79–82 (2017); Sandra Wachter, Brent Daniel Mittelstadt & Chris Russell, *Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR*, 31 HARV. J.L. & TECH. 2, 841, 862 (2018); María Estrella Gutiérrez, *Personal data processing ex machina in sharing tourism platforms. Awareness and Foreseeability by Means of Privacy Policies*, REVISTA DE PRIVACIDAD Y DERECHO DIGITAL, 22, 57, 90–91, 94–100 (2021) (Spain).

218. Manuel Guerrero Medina, *El Derecho a conocer los Algoritmos utilizados en la Toma de Decisiones. Aproximación desde la Perspectiva del Derecho Fundamental a la Protección de Datos Personales* [The Right to know the Algorithms used in Decision Making: An Approach from the perspective of the Fundamental Right to Personal Datat Protection], 49 TEORÍA Y REALIDAD CONSTITUCIONAL 141, 152–153 (Spain).

and, if applicable, auditing of the information system and its source code shall be established. Likewise, the body to be held competent for the purposes of challenging automated decisions shall be indicated beforehand as well. Thus, some pertinent questions arise. Are these provisions laying down "suitable measures" to ensure the transparency and the understandability of the automated decision-making affecting the addressee? What if the addressee of the administrative decision is a legal person, provided that the GDPR is only applicable to individual persons?

Moreover, most of the specific safeguards provided by GDPR are excluded or widely restricted in the context of data processing operated by law-enforcement authorities for the purposes of the prevention, investigation, detection, or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security within the European Union and the transfer of such personal data to third countries and international organizations.[219]

Still in Europe, French legislation resided in the CRPA constitutes one of the prominent explicit efforts to give effect to transparency and

---

219. *See* Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data by Competent Authorities for the Purposes of the Prevention, Investigation, Detection or Prosecution of Criminal Offences or the Execution of Criminal Penalties, and on the Free Movement of Such Data, O.J. (L 119) 89-131. For instance, the provision of the information to the data subject pursuant to Articles 13.2 (f), 14.2 (g), 15.1 (h) GDPR is not present in the Directive. And Article 11.1 of the Directive establishes that the European Union or Member State law may authorize individual decisions based solely on automated processing, including profiling, producing an adverse legal effect or significantly affecting the data subject, if the legislation in question provides appropriate safeguards for the rights and freedoms of the data subject, including "*at least the right to obtain human intervention on the part of the controller"* (emphasis added). Although Article 11.2 imposes a general prohibition on automated individual decision-making based on special categories of personal data (namely, data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, genetic data, biometric data for the purpose of uniquely identifying a natural person, or data concerning health or natural person's sex life or sexual orientation), the Union or Member State law may lift such a prohibition by including "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests." Once again, the content and the extent of the "suitable measures" are not described in the operative provisions of Directive, to the exception of the expanded wording set forth in the (non-binding) Recital 38: "The data subject should have the right not to be subject to a decision evaluating personal aspects relating to him or her which is based solely on automated processing and which produces adverse legal effects concerning, or significantly affects, him or her. In any case, such processing should be subject to *suitable safeguards, including the provision of specific information to the data subject and the right to obtain human intervention, in particular to express his or her point of view, to obtain an explanation of the decision reached after such assessment or to challenge the decisión."* (emphasis added).

some degree of explainability of algorithmic decisions.[220]

The provisions described in CRPA contribute to the explainability of automated individual decisions, ensuring *ex ante* (L311-3-1, 312-1-3) and *ex post* (L311-3-1, R311-3-1-1, R311-3-1-2) information. In addition, the CRPA also requires algorithm accountability based on the *understandability*, thus ensuring that the individual decision laying in algorithmic processing is explained in an intelligible form to the person affected.[221]

Two important aspects should be noted from the French legal framework.

First, in recognizing this right of the interested parties, the provisions of the CRPA are applied to any algorithmic processing, be it deterministic or predictive, AI-driven or not, because legal and reglementary provisions do not make any difference between the types of algorithmic techniques. This is of particular importance as algorithms not based on AI techniques may also have social and individual adverse impacts (e.g., the MIUR's and Ofqual's algorithms that haven discussed *supra*).

Second, the right of access to information related to the algorithmic processing can be exercised by both natural and legal persons affected by such processing,[222] thus superseding the scope of the GDPR, which is only applicable to individual persons. Interested parties and addressees in administrative procedures and decisions can be either individual, legal persons, or even entities without legal personality.

By the same token, Canadian ADM System Directive must be welcome in the sense that it provides transparency measures aimed at ensuring some degree of explainability and accountability of algorithmic decisions rendered by public authorities, by imposing *ex ante* information by providing notice before decisions (Sec. 6.2.1 and 6.2.2) and *ex post* "meaningful explanations" after the decisions are made (Sec. 6.2.3).[223]

---

220. *See* Section V.A *supra*.

221. Gianclaudio Malgieri, *Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations*, 35 COMPUT. L. & SEC. REV. 5, 22–23 (2019) (referring to the "notion of legibility").

222. *See* ASSEMBLEE NATIONALE, PROJET DE LOI POUR UNE REPUBLIQUE NUMERIQUE. ÉTUDE D'IMPACT, 10-12 (2015), https://www.assemblee-nationale.fr/14/pdf/projets/pl3318-ei.pdf (noting that the statutory provision seeks "to strengthen the transparency of public activity, by giving citizens and legal persons a new opportunity to understand the algorithmic basis of decisions that affect them.")

223. In addition to any applicable legal requirement, for Level I ADM Systems, a meaningful explanation via a Frequently Asked Questions section on a website; for Levels II, III and IV a meaningful explanation shall be provided with any decision that resulted in the denial of a benefit, a service, or other regulatory action.

VI. CONCLUSIONS

ADM systems, AI-driven or not, are being increasingly used by governments in critical sectors, such as law enforcement, health, or education.

Assumptions, data, learning models, statistical inferences, and/or purposes underlying such systems may not only be inappropriate for the intended use cases, but also have adverse effects on individuals or collectives.

Indeed, there is growing evidence which shows how automatization and algorithmization do have adverse impacts on human rights: equal treatment under the law and non-discrimination, fair trial and due process, privacy and data protection, freedom of expression, freedom of assembly and association, effective remedy, social rights, access to public services, and so on.

In this sense, special attention should be drawn to Recital (35) of the European Proposal of AI Regulation: "AI systems used in education or vocational training, notably for determining access or assigning persons to educational and vocational training institutions or to evaluate persons on tests as part of or as a precondition for their education should be considered high-risk, since they may determine the educational and professional course of a person's life and therefore affect their ability to secure their livelihood. *When improperly designed and used, such systems may violate the right to education and training as well as the right not to be discriminated against and perpetuate historical patterns of discrimination*."[224]

At the same time, ADM systems are severely affected by frequent algorithmic opacity in two senses: the widespread lack of public awareness of the uses that the administration makes of ADM systems and the impossibility of understanding the "why" and "how" of the automated decision-making process, thus resulting in a serious problem for judicial control and a risk of abandonment of the core principles governing public administration. Precisely, lack of transparency becomes a common grievance in the MIUR, Ofqual, Parcoursup, or Houston cases.

The futility of FOI regimes to address the two-tier risks of algorithmization—adverse impacts on rights and algorithmic opacity—has largely been argued. Nevertheless, cases herein analyzed evidence how freedom of information may contribute to rendering government's ADM systems (AI-driven or not) accountable in two ways: by disclosing, by request of any citizen seeking access, the source code,

224. See *supra note* at 30 (emphasis added).

the algorithms, and/or relevant documents explaining them (the right of access); or by making available to the public relevant information thereof, either proactively or under statutory obligations provided in FOIA or sectoral legislation (public disclosure schemes).

It is undisputed that source code and algorithms implemented by governments are public records for the purposes of FOI regimes, though its legal status is being discussed (rulemaking, adjudication, internal instructions)—at least in some civil law systems and FOIA cases (the MIUR and Parcoursup's algorithms). Asserting their legal status is crucial to determine their legal regime and even their degree of submission to FOI regimes. No aprioristic answers should be given because the legal status of the source code or algorithm will depend on the functionalities that have been attributed to them for each case.

FOIA cases analyzed (the MIUR and Ofqual's algorithms, Parcourpsup) and Houston decision on grounds of due process clause also evidence that the public education sector is being exposed to the two-tier risks described above: (1) the existence of discriminatory bias and individual or collective adverse impact on rights and freedoms (the MIUR and Ofqual's algorithms, Houston); and (2) the algorithmic opacity (MIUR's algorithm, Houston, local algorithms of Parcoursup).

More specifically, the analysis of the FOIA cases (the MIUR and Ofqual's algorithms, Parcoursup) shows how access to public records—not only the source code or the algorithm but also the functional and technical specifications, or third-party audits—allow public scrutiny of ADM systems, detection of their pathologies (errors in programming, lack or defective validation of models) and better understanding of their adverse impacts, individual or collective.

On the contrary, third-party proprietary rights and trade secrets on algorithms used by the government (Houston) pose relevant problems of opacity constraining not only the possibility of challenging individual decisions affecting the rights and legitimate interests of those affected but also hindering administrative transparency and accountability.

Though the remedy applied by the Houston Court was to overturn the district's policy, while leaving the trade secrets intact, this might not be the case in other courts or jurisdictions, where judges are more prone to give deference to the government's automated decisions and algorithms.

In Spain, the BOSCO issue is a clear example of this: though there was enough evidence of malfunctioning of the computer application, by having unfairly excluded applicants who met the requirements to be qualified as vulnerable consumers, the Administrative Court upheld the Ministry decision of withholding the source code on grounds of intellectual property and public security of the information systems. In

contrast, the T.A.R. judgment in the MIUR's algorithm or decisions made by the FOI Authority in Catalonia (Spain), also concerning the education sector, show that there are procedural alternatives (qualified access or conditioned access) to overcome the collision of interests between intellectual property and transparency of automated decision-making.

It may be argued that even if intellectual property exemption would not be applicable, public security exemption still would be under the "Security Through Obscurity (STO)" principle, as applied in BOSCO,[225] in order to prevent government information systems from being attacked. But in this respect, the findings of the CADA in the Parcoursup saga are illuminating again: "The commission also points out that the communication of the source code is a factor in making information systems more reliable and secure, as it allows the code to be compared with users' feedback . . . . Indeed, the security of information systems is supposed to be protected by perimeter security devices, which are not within the scope of the software or application concerned, and therefore not intended to be written back into the source code."[226]

Unlike BOSCO, the decision of the CADA was precisely to uphold the public release of the source code of Parcoursup on grounds of the opposite principle. Instead of the STO principle, the French Commission endorsed that of "transparency by default," while reconciling it with the requirements of security of information systems to the extent strictly necessary. The communication by means of online publication of the full source code was granted, "but redacting or segregating the fragments of the code which technically described those elements deployed for the security and functional management of the infrastructure, insofar as they are vectors of risk for the security of information systems."[227]

The undeniable value of the judgment made by the French Constitutional Council in Parcoursup resides in the strong liaison between the right of access to administrative documents with an active obligation of the administration to publish relevant information on the algorithmic processing applied in individual decisions affecting citizens. Though the Constitutional Council did not go further and only required ex post transparency about algorithmic processing put in place, rather

225. Juz. Cont. Adm. 143/2021, n. 8 §3 (Dec. 30, 2021) (Spain).

226. Commission d'Accès aux Documents Administratifs [CADA] [Commission for Access to Adminstrative Documents], Jan. 1, 2022, 20213847 §1 Fr.) (arguing that, according to the expert view of the CNIL and the Agence Nationale de la Sécurité des Systèmes d'Information (ANSSI), "when administrations use appropriate techniques to secure their software and respect certain coding rules, the communication of source codes does not present any risk in terms of security.")

227. *Id.* at § 1.

than ex ante information, the judgment underscored the importance of transparency measures addressed to the public at large, not only the addressees affected by a particular automated decision-making.

In this sense, diverse transparency mechanisms aimed at civil society and citizens are starting to be implemented or proposed across jurisdictions: (1) statutory requirements of public disclosure for source code, algorithms (Canada, France), or relevant information of ADM systems (Autonomous Community of Valencia, German Commissioners, Council of Europe) relying on FOI regimes or sectoral legislation; (2) public repositories of open-source codes and algorithms (Canada, France), public inventories (United States), or registries of AI systems (European Union) deployed and used by public administrations; and (3) specific provisions in sectoral legislation requiring ex ante or ex post explanations, seeking to allow the public and policymakers to understand how an algorithmic decision was reached (GDPR, France, Canada).

At least from the European perspective, it will be necessary to wait for the Court of Justice of the European Union to establish a consolidated case law to know how the relevant provisions of the GDPR applied to automated decisions are to be interpreted, especially in relation to the so-called right to obtain "an explanation of the decision taken" of Recital (71) and the rest of the guarantees set out in the normative provisions. Whether the interpretation will be far-reaching or will adhere to the wording of Article 22 remains unclear.

FOIA cases raised herein do illustrate why the opacity of ADM systems (AI-driven or not) should be addressed urgently. Though falling within the realms of criminal justice and sentencing, it is worth recalling the frank yet worrying acknowledgment of Judge Abrahamson while joining the majority of the Loomis Court: "[T]his court's lack of understanding of COMPAS was a significant problem in the instant case. At oral argument, the court repeatedly questioned both the State's and defendant's counsel about how COMPAS works. Few answers were available . . . . Such an explanation is needed . . . ."[228]

Bearing in mind that many ADM systems used by public administrations have been developed by third-party contractors, any "consideration could be given to the possibility of having the code, the generated data—as far as they are non-personal—and the trained model made public by default upon agreement with the developer, in order to guarantee transparency, enhance cybersecurity and enable the reuse thereof so as to foster innovation."[229]

---

228. State v. Loomis, 881 N.W.2d 749, 774 (Wis. 2016).

229. European Parliament Resolution of 20 October 2020 with Recommendations to the Commission on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and

As the Spanish State Council of Transparency has observed, there is an undeniable public demand for transparency of ADM systems and algorithms used by public administrations as an inexcusable condition to preserve accountability and oversight of the decisions of public authorities and, ultimately, as an effective guarantee against arbitrariness or discriminatory biases in fully or partially automated decision-making. Significantly, the Spanish Council has noted that, until other mechanisms are put in place to achieve the goals of transparency, accountability, and oversight with equivalent guarantees—such as independent audits or supervisory bodies—"the only effective remedy for such purpose is access to the algorithm itself, to its code, so that it can be audited both by those who may feel harmed by its results and by the general public in the interest of the adherence to ethical principles and justice."[230]

Related Technologies ¶86 (2020/2012(INL)), O.J. C. 404 (Oct. 6, 2021) 63, 76 (emphasis added), https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html.

230. Consejo de Transparencia y Buen Gobierno [CTBG], *supra* note 277.