Indiana Journal of Global Legal Studies

Indiana Journal of Global Legal Studies

VOLUME 30 ISSUE 1 2023

CONTENTS

DIGITAL TRANSFORMATIONS OF GOVERNMENT: TOWARDS A DIGITAL LEVIATHAN?

ARTICLES

STUDENT NOTES

Digital Transformation of Government: Towards A Digital Leviathan?

June 23 – 24, 2022 | Campus Puerta de Toledo, Universidad Carlos III de Madrid

ALFRED C. AMAN, JR.

INTRODUCTION

A warm welcome to you all. It is a great pleasure to be able to participate in this exciting collaboration between Universidad Carlos III de Madrid (UC3M) and Indiana University—a conference that the *Indiana Journal of Global Legal Studies* is publishing in celebration of its thirtieth issue. This is a milestone for us, and we could not be happier to celebrate it in this way. Let me begin with a few words about the nature of this journal and its scholarly goals over the years.

The *Indiana Journal of Global Legal Studies* is a peer-reviewed interdisciplinary journal focusing on the intersections of global and domestic legal regimes, technologies, markets, politics, societies, and cultures. The journal seeks to facilitate dialogue among international communities of scholars in law and other disciplines, with intersecting concerns bearing on new forms of law related to globalization processes, transnationalism, and their social effects.[1] By its very nature, seeing law in such terms challenges the conventional boundaries among subject disciplines and professional research practices, as well as the boundaries around sovereign state regulatory regimes.

In 1993, when the Journal published its first issue, a bright line between domestic and international law was already largely illusory. As a result, we needed fresh assessments of issues, such as the role and theory of the nation-state in the twenty-first century, the need for and

[1] Editor's note: Our journal receives submissions globally, and our goal and mission is always to preserve authorial voice and style. This issue follows our standard editing conventions while also preserving the authors' voices. Due to the nature of the symposium, we have allowed for greater authorial discretion in terms of citations. We have ensured the accuracy of these citations with the goal of maintaining the authors' discretionary style choices.

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

development of new international and global institutions, and, in particular, the kinds of domestic legal reforms necessary to mesh with or respond to global economic and political effects. But at that time, our global institutions were few—even the WTO, for example, had not yet been established. There was also a lingering sense that globalization was a single, all encompassing process affecting everyone, everywhere, at the same time.

Today we know that globalization is not a single process pitched toward harmonization, but something far more complex. We know it is not a unidirectional process in time or locale, a process that occurs only once, as if globalization were a straightforward yes/no question, but—again—something far more complex. As we embark upon our thirtieth issue and the symposium topic, the "Digital Transformation of Government: Towards a Digital Leviathan?," we have a timely opportunity to focus on digital technology and to reflect on how law will or should respond to a technology that is at once local and global, personal and impersonal. We know that digital technologies have the capacity to greatly enhance our abilities to creatively and humanely interact in a global world, but we also know that they have the capacity to undermine, if not eliminate, a concept of the public interest.

The answers to the many questions implicitly and explicitly posed by technological innovations, such as the internet, will not come from technology alone. This conference, by its theme and individual papers, underscores the range of engagement with the challenges these issues and others related to them pose for all of us. The themes of these papers point to some exciting new conversations—new theoretical innovations within and across academic disciplines, new institutional partnerships, new legal regimes, and forms of legal analysis. For example, do we need new constitutional rights to be able to live with the impacts of these new technologies? Do we need new regulatory structures? What are they and, more importantly, who will build them?

Questions such as these epitomize the aspirations of the *Indiana Journal of Global Legal Studies*. Over the past thirty years, globalism has meant many things. At the outset, the Journal dealt mainly with globalism as a set of challenges to traditional concepts of national sovereignty, regionalism, and citizenship as these were transformed by the compression of public and private interests on a large scale. Those challenges remain, but our symposium topic points to a major shift in globalization itself—since digital technologies are now global and intimate, potentially affecting the ways people inhabit time, place, and their own identities.

For now, again, I want to thank José Vida and colleagues here at UC3M for this collaboration, and all participants for their contributions.

The Risk of Digitalization: Transforming Government into a Digital Leviathan

José Vida Fernández*

Abstract

This paper provides an overview of the threats posed by digitalization, particularly with regard to the public sector. It starts by describing digital risks as true global risks and argues that their scope and severity have not been recognized until now. The most well-known challenges come from the transformation of the private sector (economy, society, and individuals) and the emergence of large private powers that dominate the digital environment (digital feudal lord). However, there are even greater challenges coming from the digitization of government, creating almighty public bodies detached from laws that kept them locked until now.

I. Digitalization: A New Global Risk

The unstoppable digital transformation that most countries are undergoing is giving rise to growing concern about the negative effects this process brings. Our society is increasingly dependent on digital technologies (traditionally known as information and communication technologies) that are modifying our activities (economic, social, personal) but which we understand and control less and less. The progressive increase in the relevance of digital technologies in our existence forces us to reflect not only on the advantages, but also on the risks they introduce and how they can pose a threat to our current way of life.

In fact, we can identify a new global risk category: digital risks.

* Professor of Administrative Law, Public Law Department, Universidad Carlos III de Madrid (jose.vida@uc3m.es). This work is part of the research projects "Artificial intelligence in the national health care system: solutions to specific legal problems" (PID2021-128621NB-I00) and "The Impact of Artificial Intelligence in Public Services: A Legal Analysis of its Scope and Consequences in Healthcare" (PGC2018-098243-B-I00) 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)

These are becoming part of the so-called "risk society"[1] insofar as they are risks derived from technological innovation that threaten our existence at a global level. This is similar to what happens with technological developments that lead to other threats such as climate change, epidemics, or terrorism. Thus, those digital risks are not merely threats of information networks and systems addressed by cybersecurity, but have a broader and deeper meaning. Digital risks refer to all transformations resulting from digitalization that can threaten basic aspects of our current life in economic, political, or social terms.

A. Digital Risk: Too Fragile an Acknowledgment

Digital risks are very unique in nature as they do not physically compromise our survival—this can be seen with environmental, health, or security risks. On the contrary, digital risks affect people's rights, political freedom (including the very functioning of democracy), and, ultimately, human dignity, in addition to data privacy and information security.

Thus, digital risks are very distinct and different from traditional global risks because of the object that is threatened. In the case of digital risks (considered as "risks from digital environments") the object to be protected is not the "digital environment" (which would be the source of the risk) but fundamental rights, political freedom, and human dignity, which can be affected in many different ways in digital environments (from violation of privacy, racial or gender discrimination, to social exclusion). Thus, when we speak of "health" or "environmental" risks, the object to be protected can be perceived straightforwardly as it is tangible (population health, natural environment) and an end in itself. On the contrary, in the case of digital risks we find that the object to be protected (fundamental rights and human dignity) is abstract and artificial, and it is not the digital environment that needs to be protected, since it is precisely something that threatens the process.

This unique nature of digital risks makes them more difficult for citizens to identify. It is therefore harder to engage in a public debate on digital risks in order to address them through governance and

[1] Risk society is the way our society deals with hazards and insecurities induced and introduced by modernisation itself. *See* ULRICH BECK, RISK SOCIETY: TOWARDS A NEW MODERNITY 50 n.1 (1992) ("In social science's understanding of modernity, the plough, the steam locomotive and the microchip are visible indicators of a much deeper process, which comprises and reshapes the entire social structure."); *see also* ULRICH BECK, WORLD RISK SOCIETY (1999).

regulation. So, there is "too fragile an acknowledgment" of digital risks.[2] It is very difficult to identify digital risks and become aware of them, unlike traditional global risks that can produce physical damages. Conversely, as the complexity of the technological world increases, our understanding of it and of the reality around us decreases.[3] Digital risks are becoming more uncertain, more complex, and, therefore, more difficult to identify.

Even if these digital risks are identified, the fact is that no real harm is perceived to be caused by them. Digital disasters with serious damage on a global scale—such as the NSA's Prism Surveillance system revealed by Edward Snowden or the Cambridge Analytica affair of Facebook[4]—have not provoked a citizens' global mobilization similar to those in defense of the environment or health. These digital scandals have dissolved over time and therefore citizens are not on their guard. People are unaware and underestimate the damage caused, although it is clear these massive violations of privacy and manipulation have had fatal consequences. The problem is that the damages suffered in digital environments are not perceived as real damages since freedom and rights die without humans being physically hurt.[5]

When these digital risks are considered real risks with concrete harms, they are largely consented. Indeed, "dataisms" are widespread in society, so it is assumed as something natural to give up rights and freedoms in order to reach a higher stage in evolution (homo deus) through big data and artificial intelligence.[6] Without entering into this debate, it can be seen that most citizens are slipping into "dataism," as they assume the risks and even stoically bear the damages of digital

[2] *See generally* Ulrich Beck, *The Digital Freedom Risk: Too Fragile an Acknowledgment*, 22 QUADERNS DE LA MEDITERRÀNIA, 141, 141–44 (2015) (contrasting the difficulty in perceiving the damage suffered in the digital environment with events such as the Chernobyl disaster, global warming or the COVID-19 pandemic, in which a catastrophic situation occurs with concrete physical damage that generates awareness and the adoption of measures to address these risks).

[3] *See generally* JAMES BRIDLE, NEW DARK AGE: TECHNOLOGY AND THE END OF THE FUTURE (2018).

[4] *See generally* EDWARD SNOWDEN, PERMANENT RECORD (2019) (discussing PRISM, the program of the U.S. National Security Agency (NSA)); BRITTANY KAISER, TARGETED: THE CAMBRIDGE ANALYTICA WHISTLEBLOWER'S INSIDE STORY OF HOW BIG DATA, TRUMP, AND FACEBOOK BROKE DEMOCRACY AND HOW IT CAN HAPPEN AGAIN (2019) (discussing Cambridge Analytica and the U.S. 2018 elections).

[5] Beck, *supra* note 3, at 144.

[6] *See* Chris Anderson*, The End of Theory*, WIRED (June 23, 2008, 12:00 PM) (discussing the new era of dataism), https://www.wired.com/2008/06/pb-theory/; *see also* YUVAL NOAH HARARI, HOMO DEUS: A BRIEF HISTORY OF TOMORROW (2017).

services as a fair price to enjoy their functionalities.[7]

B. The Growing Threat of Digital Risks

The unique nature of digital risks, together with the late development of the technological revolution have meant that we have not been aware of their relevance until very recently, despite their growing and serious threat.

Digital risks are not limited to the violation of our privacy, but can reach much deeper and affect free will, limiting or making our own human condition disappear. Digital technologies fight to capture our attention[8] and can trap us in a certain ideological frame or "filter bubble."[9] It is a "friendly Big Brother" that knows us better than we know ourselves and can condition our thoughts and opinions.[10] Even more invasive technologies are being developed that can record mental data from brain impulses and manipulate them, leading to the recognition of neuro-rights to preserve the physical and psychological integrity of the individual.[11]

The evolution, scope, and consequences of digital risks are very different from traditional global risks not only because of their unique nature, but also because they occur in a hitherto unprecedented scenario. Digital innovation is not centered on the control and exploitation of nature, as is the case with other global risks. On the contrary, we are in the "age of surveillance capitalism"[12] in which

[7] This is the case of Google Maps, which can track its users to offer them the best routes, or in the case of intelligent assistants (such as Alexa, Siri or OK Google), which can be allowed to monitor conversations in exchange for the use of all their functionalities.

[8] *See generally* TIM WU, THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS (2016) (discussing the digital struggle to capture attention).

[9] *See generally* ELI PARISER, THE FILTER BUBBLE: HOW THE NEW PERSONALIZED WEB IS CHANGING WHAT WE READ AND HOW WE THINK (2012).

[10] *See* BYUNG-CHUL HAN, PSYCHOPOLITICS: NEOLIBERALISM AND NEW TECHNOLOGIES OF POWER 59–64 (2019); *see also* BYUNG-CHUL HAN, INFOCRACY: DIGITIZATION AND THE CRISIS OF DEMOCRACY (2022).

[11] Chile was the first country to recognize neuro-rights in its Constitution through the modification of Article 19, number 1. In the case of Spain, the Charter of Digital Rights, approved by Agreement of the Council of Ministers on July 13, 2021, dedicates its section XXVI to neuro-technologies, establishing that the limits and guarantees for the implementation and use of neuro-technologies on people can be regulated by law.

[12] SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 16 (2019) ("Just as industrial civilization flourished at the expense of nature and now threatens to cost us the Earth, an information civilization shaped by surveillance capitalism and its new instrumentarian power will thrive at the expense of human nature and will threaten to cost us our humanity.")

humans are controlled and exploited by large corporations which, for technological, economic, and jurisdictional reasons, are beyond the reach of the public authorities.

Finally, it should be noted that the unstoppable process of digitalization has intensified in recent years, exponentially increasing the level of risks involved. It is an irreversible process and all countries, companies, and individuals will depend more and more on digitalization. But the most paradoxical thing is that digitalization, which is the source of risk, has also become the panacea, even for overcoming other global risks.[13] A dangerous inverse correlation is generated whereby the reduction in the level of the traditional global risks (environment, health, or safety) goes through the increase of digital risks.[14]

II. RISKS OF DIGITALIZACION:

A. Between Techno-Feudal Lords and a Digital Leviathan

1. Leaving the Digital Laissez-Faire Era

Identifying digitalization as a global risk is a first step toward taking it seriously and initiating a public debate on it. There is no turning back from this process and the digital Luddites will not be able to stop it. The question is how to deal with the digitalization that is reshaping our society, both in the public and private spheres, and generating new forms of power.

Digital transformation has so far taken place under the principle of freedom (*laissez-faire*) letting innovation unfold without limits. Information and communication technologies have developed freely and produced enormous advances, such as personal computers and the internet, but without specific legislation. Only ad hoc measures have been taken to address specific issues raised by these developments (child protection, copyright, content liability, etc.)[15] from a negative, ex

[13] For fighting climate change (digital transition for decarbonization), curving pandemics (apps for COVID), for guarantying safety (surveillance devices).

[14] *See* Thomas A. Hemphill, *The Innovation Governance Dilemma: Alternatives to the Precautionary Principle*, 63 TECH. SOC'Y 7 (2020) (recommending as the main tool for innovation (and risk governance in general) the adoption of artificial intelligence and data analytics for risk management and regulatory adjustment, without realizing the risks that such a remedy entail).

[15] There is not an Internet Act as such, but legislation on child protection (Children's Internet Protection Act of 2000), copyright (Digital Millennium Copyright Act of 1998), liability (Communications Decency Act of 1996), and so on.

post, and strictly reactive-corrective approach.

However, the strategic nature of digital transformation and the emergence of relevant risks is leading many countries to abandon their passive and negative approach to digitalization. During the last decade countries are deploying alternative strategies for digital governance that involve positive, ex ante and proactive-preventive measures. We are at a turning point where an innovative legal framework for the new digital society and digital government is being forged.

In any case, countries are following quite different strategies to deal with this digital transformation, which can be classified into three main models. On the one hand, in the case of the Unites States, the aim is to maintain free competition and minimal intervention, although regulations are inevitably increasing. On the other hand, in totalitarian countries such as China or Russia, government takes over the digital sector, which becomes an instrument of power. In between, the European Union is developing an open, flexible, and adaptive model of governance that respects free market while ensuring security and trust, which is essential for digital innovation.

B. Techno-Feudal Lords

There are many issues arising from the digitization of companies and citizens that has led to a change in market, work, education, and personal life. A myriad of problems need new solutions as many are emerging from new types of markets (platform economy), labor relations (platform workers), consumers (prosumers), media (streaming platform), and political forum (social networks). However, one aspect all of these challenges have in common is the presence of a digital intermediary (platform, network, search engine) that serves as the basis for the development of the activity. Thus, the extraordinary private powers of the large technology companies (big tech) arise and they in turn dominate this digital environment.

This is an extraordinary gamble because, for the first time, a private empire governed by large companies is being generated with unprecedented levels of control, both by society and individuals. Big tech companies decide on public debate, cancel opinions, and help candidates win elections. They feed from citizens' data and can make people become transparent. Never before in history has such singular, global, and intense power been concentrated into so few companies without the presence or intermediation of government.[16] These large technology companies operate outside the law acting as lords of the digital environment they own and with an absolute dominion over citizens, in a kind of techno-feudalism system.[17]

Faced with this new scenario, the traditional rights of citizens are insufficient to cope with the power of the big tech companies. The right to privacy is too narrow to cover the numerous issues that arise in the new digital environment. The problem is that fundamental rights were conceived as limits to the power of government and not to control big business. This is the case, for example, of large platforms such as Facebook or Twitter that manage citizens' freedom of expression. It is also the case with the right of communication, which is now controlled by platforms such as HBO, Netflix, or YouTube.

This situation is leading to major changes in the legal approach to managing and controlling these new private powers and the risks they pose. In both the United States and the European Union, it has become clear that competition rules are not sufficient to contain big tech. As a result, new regulation of large intermediary providers is emerging that incorporates real ex ante measures to try to contain their extraordinary power.[18]

C. Digital Leviathan Rising

Digitalization is also transforming the public sector and, although it is not as attractive and glamorous, it is increasing to such an extent that it deserves specific attention. Most of the best-selling essays and academic literature focus on the problem of big tech's control over digital society,[19] but do not concern themselves with the power of a digital government. It should be noted that the government has more

[16] Global Risks Report of the World Economic Forum ranked the risk of "digital dependencies" and "digital power concentration." *See* WORLD ECONOMIC FORUM, THE GLOBAL RISKS REPORT 2022, at 95 (16th ed. 2022).

[17] Technology platforms have characteristics similar to feudal fiefdoms as companies are rent-seekers and control preferences, purchases and behavior without being accountable. *See* Yanis Varoufakis, *Techno-Feudalism is Taking Over*, DIEM25 (July 9, 2021), https://diem25.org/techno-feudalism-taking-over/.

[18] *See Proposal for a Regulation on a Single Market for Digital Services (Digital Services Act)*, at 1, COM (2020) 925 final (Dec. 15, 2020) (making the platforms liable for content (with the threat of large fines));)*; Proposal for a Proposal for a Regulation on Contestable and Fair Markets in the Digital Sector (Digital Markets Act),* at 1, COM (2020) 842 final (Dec. 15, 2020) (limiting the activities of some of Big Tech, particularly the gatekeepers, that control access to the market and dictate how markets operate). *See generally* STAFF OF H. COMM. ON THE JUDICIARY, 117th CONG., INVESTIGATION OF COMPETITION IN DIGITAL MARKETS (Comm. Print 2020).

[19] *See* ZUBOFF, *supra* note 12; HAN, INFOCRACY, *supra* note 10.

data, more resources, and more power than any company, so digitalization can lead to unknown consequences.

Information and Communications Technology has been transforming government for decades favoring an agile, simple, and transparent functioning, essentially thanks to online government based on internet. The e-government allows permanent and unlimited distance access and interaction with citizens. However, it is only an instrument as it has solely affected the means for governmental action but has not produced changes in the essence of government.

In the last decade, disruptive technologies—such as artificial intelligence, blockchain, and cloud and edge computing—are spreading throughout governments. These new technologies are no longer simply tools to facilitate governmental activity, but they are transmuting and reshaping government functions. Digital transformation into i-government is affecting the core functions of government, as they are even being applied to decision-making.

These new technologies are beginning to be widely used in the executive branch. More agencies and administrative bodies are using disruptive technologies both to provide their services and to make decisions. But in addition, these disruptive technologies are beginning to be used in other branches of government, which creates more problems in terms of implications for democratic processes (in the case of the legislative branch) or due process (in the case of the judicial branch).

Although there is e-government regulation, there is still no legislation for the use of these disruptive technologies in government. This makes sense since these are innovations—such as artificial intelligence—that do not yet have any general regulations, so there cannot be a special one regarding their use in government. Therefore, these innovations are being incorporated under the previous legal framework that is obsolete and does not provide answers to the innumerable challenges and problems that arise in all branches of government.

Digital transformation is overflowing the constitutional and administrative framework, one which has taken centuries to devise and has led to the model we have for current governments—the rule of law and democracy principle. Now a Digital Leviathan is emerging for which we still do not have rules. It may become a more powerful subject without the checks and balances approach models we have used up until now.

D. Some Thoughts on Digital Government

These were some concerns discussed at the Conference "Digital Transformation of Government: Towards a Digital Leviathan" (June 25–26, 2022) organized by the Universidad Carlos III de Madrid and Indiana University and which I co-directed with Alfred C. Aman, Professor of Law at the Maurer School of Law (Indiana University). The papers are published in the *Indiana Journal of Global Legal Studies* (Volume 30, Issue 1).

A good starting point for addressing the risks and challenges of digitalization is to examine the bills of digital rights. It is not so much a matter of recognizing new rights but rather adapting existing rights to the digital environment. This is the approach of the EU Declaration on Digital Rights and the national charters of digital rights.[20] The new generation of digital rights is being promoted by some scholars, both from scientific and legal fields. This is the case of Rafael Yuste, a neurobiologist in Columbia, who has been advocating for years for the recognition of neuro-rights to protect free will,[21] and of Tomás de la Quadra-Salcedo, an emeritus professor of law who led the group that drafted the Spanish Charter of Digital Rights.[22]

Among the disruptive technologies, artificial intelligence stands out as the technology that is posing the most challenges in its use by government. Several proposals of AI regulation are moving through the legislative process in the EU and US and in other countries. In the case of the EU, the proposed AI regulation (AI Act) is based on securing trust to promote AI use, as Antonio Estella points out.[23]

However, the use of AI in administration is spreading in the absence of regulation, so it is developing under previous general rules—on data privacy, e-government, etc.—that are clearly inadequate to address the challenges and problems it poses. Therefore, there is an urgent need to

[20] These are not merely Internet rights (like the 2014 Online Bill of Rights in Brazil or the 2015 Italian Declaration of Internet Rights) but go further as digital rights. *See generally Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Establishing a European Declaration on Digital Rights and Principles for the Digital Decade*, COM (2022) 27 final (Jan. 27, 2022); CARTA DERECHOS DIGITALES (CHARTER OF DIGITAL RIGHTS) (2021) (Sp.).

[21] *See generally* THE NEURORIGHTS FOUNDATION, https://neurorightsfoundation.org.

[22] *See* Rafael Yuste & Tomás De La Quadra-Salcedo, *Neurorights and New Charts of Digital Rights: A Dialogue beyond the Limits of the Law*, 30 IND. J. GLOBAL LEGAL STUD. (2023).

[23] *See* Antonio Estella, *Trust in Artificial Intelligence: Analysis of the European Commission Proposal for a Regulation of Artificial Intelligence*, 30 IND. J. GLOBAL LEGAL STUD. (2023).

enact specific legislation on the use of AI in government to bring it in line with the constitutional principles of the administrative state, as I claim.[24]

The difficulties of fitting AI into the traditional categories of administrative law are obvious. This mismatch between the new technology and the old rules is analyzed by Gilles Guglielmi,[25] who considers algorithm not as rules and technology but as soft law always under the rule of law.

One of the most problematic aspects of AI to us is transparency and access to the algorithm in AI used in government. Experiences to date have rejected or hindered access to the algorithm, preventing accountability and violating transparency and reason-giving, which are common principles of government action. This makes it necessary to rethink the conditions of access to algorithms to ensure transparency and the right to have an open and clear public decision, as Estrella Gutiérrez David claims.[26]

Another controversial issue regarding the use of AI is facial recognition. Although some forms of facial recognition are being banned, it is something that is becoming more and more widespread and allows control in a way that has never existed before. Governmental power is bearable because there are areas in which citizens can be left out of its reach. However, an omnipresent government is unbearable and even more so when it is used to control certain sectors of the population, giving rise to what Antonio Pelé calls necropolitics.[27]

The digitization of government is not limited to AI as there are other disruptive technologies that are also transforming its essence. Specifically, blockchain is one of the technologies with the most potential to improve administrative activity and give rise to hitherto unknown possibilities. Migle Laukyte analyzes the potential of blockchain use in government and how it can make a decisive

[24] *See* José Vida Fernández, *Artificial Intelligence in Government: Risks and Challenges of Algorithmic Governance in the Administrative State,* 30 IND. J. GLOBAL LEGAL STUD. (2023).

[25] *See* Gilles Guglielmi, *The Contentious Issues of Governance by Algorithms*, 30 IND. J. GLOBAL LEGAL STUD. (2023).

[26] *See* María Estrella Gutiérrez David, *Government by Algorithms at the* Light *of Freedom of Information Regimes: A Case-by-Case Approach on ADM Systems within Public Education Sector*, 30 IND. J. GLOBAL LEGAL STUD. (2023) (analyzing the MIUR, Ofqual, Parcoursup, and Houston cases).

[27] *See* Antonio Pelé, *On Facial Recognition, Regulation and 'Data Necropolitics'*, 30 IND. J. GLOBAL LEGAL STUD. (2023).

contribution to the right to good administration.[28]

Digitization has always had a special impulse in those areas where huge information processing is required and it contributes to increase public revenues. This is the case of the tax system, which has undergone a special development in almost all countries, and particularly in Denmark, as Peter Koever Schmidt and Louise Fjord explain.[29]

IV. CONCLUSION

Digitalization is transforming our society at an accelerating pace. Digital transformation has enormous advantages but it also brings great threats that must be considered. Digital risks must be considered to be true global risks in order to initiate a public debate on their consequences and how to regulate them.

The digital society has arrived but we do not have the institutions in place to govern this new digital world. The use of disruptive technologies in government is taking place without specific regulations to ensure their use. It is important that the public debate is not limited to the new forms of economy, work, and social relations, but also reaches the new digital government. Otherwise it may be too late and we may find ourselves in the hands of a Digital Leviathan.

[28] *See* Migle Laukyte, *Blockchain and the Right to Good Administration: Adding Blocks to or Block'ing of the Globalization of Good Administration?*, 30 IND. J. GLOBAL LEGAL STUD. (2023).

[29] *See* Louise Blichfeldt Fjord and Peter Koerver Schmidt, *The Digital Transformation of Tax Systems: Progress, Pitfalls and Protection in a Danish Context*, 30 IND. J. GLOBAL LEGAL STUD. (2023).

Neuro-Rights and New Charts of Digital Rights: A Dialogue Beyond the Limits of the Law

RAFAEL YUSTE AND TOMÁS DE LA QUADRA-SALCEDO[*]

ABSTRACT

In this article, the authors address some of the most pressing issues that stem from the relationship between the technological advancements of the twenty-first century and legal regulation. The development of neurotechnology and artificial intelligence (AI), while offering considerable opportunities for the betterment of social life, also poses unprecedented risks. These challenges manifest in a wide variety of topics. Areas such as human rights treaties, antitrust law, property law, and labor law are affected by these developments. The risks associated with the unregulated use of neurotechnology and AI do not cease at the sectorial stage. Some of the values upon which current democratic systems and governance models are built could be equally threatened. In anticipation of the harming potential of unmitigated technological advances, some governments and international institutions have enacted legal provisions to regulate the current digital landscape. These normative instruments, including the Chilean Constitutional Amendment and European Charts of Digital Rights, are also analyzed in the following pages. The purpose of this article is not purely descriptive,

[*] Rafael Yuste is Professor of Biological Sciences and Director of the NeuroTechnology Center (NTC) at Columbia University and Chairman of the NeuroRights Foundation.

Tomás de la Quadra-Salcedo is Emeritus Professor of Administrative Law, Universidad Carlos III de Madrid, Former Ministry of Justice and Former President of the Council of State.

Translated by Miguel García Fernández, Researcher at Instituto Pascual Madoz del Territorio, Urbanismo y Medio Ambiente, Universidad Carlos III de Madrid. . This work is part of the research projects "Artificial intelligence in the national health care system: solutions to specific legal problems" (PID2021-128621NB-I00) and "The Impact of Artificial Intelligence in Public Services: A Legal Analysis of its Scope and Consequences in Healthcare" (PGC2018-098243-B-I00) 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)

but rather to spark a debate among legal scholars and experts in their respective fields. The approach followed here, dialogical in its nature, may provide a model for further collaboration. It is the authors' understanding that the regulation of neurotechnology and AI requires an interdisciplinary approach that is transnational in its scope.

I. INTRODUCTION

It can be argued that one of the characteristic features of social life in the twenty-first century is the pervasiveness of technology. Words such as blockchain, AI, or data that were previously ostracized to the margins of specialised journals have now become mainstream. The technological developments that have been taking place in the last decades have changed the social, economic, and political landscape in an unapologetically and decisive way. It cannot be denied that some of these inventions have had many positive consequences. Technology has been shown to be extremely adept at fostering productivity and improving human connectivity. However, its destructive potential is equally impressive.

The consolidation of a relatively new branch of science—neurotechnology—could be added to the list of promising tools in the pursuit of human enhancement. The advancements in this area have allowed scientists to achieve an unprecedented knowledge of the way the brain functions and its structure. This information is susceptible to abuse by different subjects: unregulated corporations, autocratic governments, or other bad actors present in the global sphere. In the near future, essential values could be threatened by unregulated and inhuman technological development. The potential harm caused by the misuse of this type of technology is immeasurable: the generalized loss of privacy, the deterioration of democratic systems, and the erosion of societal bonds are part of a future in which the legal system does not adapt to accommodate the needs of citizens in the digital era. Hence, neurotechnology carries the same opportunities and risks accompanying the aforementioned advancements.

Another consequence of these technological advancements is the increasing instability of our legal systems. When faced with the prospect of an everchanging reality, such as technology, some of the cracks in traditional legal institutions are revealed. Law, both as a discipline and as a social construct, is particularly prone to outdatedness. The ossified nature of legal rules is hardly reconcilable with the imperatives derived from technological progress. The constant evolution of technology is a trend that does not show symptoms of exhaustion. This context raises many questions, some of which are presented here: What should be the role of law in this globalized and deeply unstable context? How to strike a balance between the different, and often contradictory, interests at stake? Among the different possible options (soft law, regulation, charts of rights) what should be the preferred normative instrument to tackle these challenges?

These are some of the questions that inspired the organization of the academic seminar *Digital Transformation of Government: Towards a Digital Leviathan?*, a joint initiative between the *Indiana Journal of Global Legal Studies* and the University Carlos III de Madrid (UC3M). Even though the proposed questions do not have an easy answer, there are some principles that should guide any normative reaction to this issue. First of all, it was evident from the beginning that the solution to the many challenges posed by the surge of new technological developments worldwide demands a transnational and interdisciplinary approach based on cooperation. The following article has been written in the spirit of these considerations.

The main goal of this article is to facilitate dialogue from various areas of expertise with intersecting concerns relating to digitalization processes. Specifically, the dialogue is set up between two disciplines: neurotechnology and law, the respective fields of expertise of its two authors, Rafael Yuste and Tomás de la Quadra Salcedo. The contents of the following pages consist of an adaptation of some of the ideas that were expressed during those seminars. Their nuanced thought gives testimony to the complexity of the subject at hand, that being the task of dealing with the regulatory risks resulting from the processes of digitalization and the development of new branches of science, such as neurotechnology. To focus the debate and introduce the reader to some of the ideas that will be analysed below, the article now describes the structure and some of the main ideas that make up the core of the position of both authors.

Rafael Yuste is a renowned scholar and scientist specialised in neurotechnology, who works at Columbia University. Since the launch of the BRAIN initiative,[1] a programme aimed at developing neurotechnology to map and alter brain activity, which he inspired, he is regarded as one of the most authoritative voices in his area. His leadership of the Morningside Group has placed him in a privileged position as an interlocutor in matters related to the creation of a corpus of new legal rights. One that includes a new category of rights of which he is an ardent proponent. The concept that articulates his contribution is that of "neuro-rights." In his opinion this new category is the key to

[1] *See* Alivisatos A.P. et al., *The Brain Activity Map and the Challenge of Functional Connectomics*, 74 NEURON 970, 970–74 (2012).

harmonise the contradicting interests derived from recent technological developments. Therefore, it should be the backbone of any future regulatory strategy. Such an approach, based on the principle of human dignity, could allow the scientific community to preserve the freedom required to continue with the greatly needed research that is being carried out daily (the treatment of neurological diseases that are, to this day, incurable; the enhancement of human capabilities; etc.) while erecting the necessary safeguards against the banalization of technology. As the reader will have the chance to discover, the different neuro-rights that are described by the author are an ingenious solution to the duality present in every scientific advancement of significance from its challenges to its opportunities. By the end of his presentation the author alludes to the positive experience with the Republic of Chile and the approved amendment to article 19 of the constitution of a provision aimed at protecting cerebral activity and the information drawn from it. This example is complemented by the attempt to update the Human Rights Charter by the United Nations. These examples shed an optimistic light on the position of those that advocate in favour of the consolidation of neuro-rights at the international stage.

If Rafael Yuste's exposition perfectly centres the debate by encapsulating the main challenges that society will face in the upcoming decades, the contribution of Tomás de la Quadra-Salcedo complements that of his coauthor by providing a complete analysis of the legal responses that are currently being enacted in anticipation of those same challenges. Thus, the debate that is currently taking place at the European level in relation to the guiding principles of the digital society is one of the topics that stands out from his exposition. De la Quadra-Salcedo also reflects on the evolution of the regulation on this area, and he analyses some of the previous attempts to base the legal system on individual rights adapted to the digital reality. However, his contribution to the current debate does not limit itself to a mere recollection of past regulatory proposals. After said compilation, the author introduces one of the issues that to this day perplex those that enter the debate on the regulatory needs of AI: the adequacy of previous legal categories in the digital landscape. Through a series of examples derived from diverse areas of law, he pushes forward the thesis that there is a need to recontextualize traditional legal forms and bring them to the present. In the author's opinion, it is imperative to overcome the notion that reduces the immense regulatory problems of the present to the mere concept of "data." This idea is then reinforced in the following epigraph. As it is stated by the author: "All the scenarios posed by the new digital reality require an exercise of reflection that will probably lead to the redefinition of many of the traditional rights." In the same vein as his coauthor, de la Quadra-Salcedo concludes his exposition by remarking on the importance of an approach based on human dignity. Such a principle constitutes the founding pillar upon which the European Bill of Digital Rights is built.

The work of both authors constitutes an exemplary invitation to collaboration between different legal traditions on both sides of the Atlantic. From the birth of a new generation of rights to its positivization through the legal instrument of the charts of rights, this article attempts to delineate some normative proposals to the challenges of the twenty-first century. The concept of "neuro-rights" and the broader category of "digital rights" provide insight into the nature of a legal system respectful to human dignity and technological progress. The manner in which that future will unravel will depend exclusively on the decisions taken by public powers in the following decades. If the authors of this article are right, perhaps the best way to approach this issue is through the optics of the revolutionary creation of human rights. The path to a future in which technology is implemented for the exclusive benefit of humankind is set. Thus, the thesis of this article is that it is the moral responsibility of academics and scientists to advocate in favour of a legal and scientific culture based on humanism and technological accountability.

II. THE NEED FOR NEURO-RIGHTS: RAFAEL YUSTE[2]

I would like to begin my presentation by remarking on the importance of considering the scientific side of neuroscience and AI in the pursuit of sound and effective legal regulation. The

[2] Sara Goering & Rafael Yuste, *On the Necessity of Ethical Guidelines for Novel Neurotechnologies*, 167 CELL 882, 882–85 (2016); Rafael Yuste et al., *It's Time for Neuro-Rights: New Human Rights for the Age of Neurotechnology*, 18 HORIZONS 154, 154 (2021); Cori Bargmann & Rafael Yuste, *Toward a Global BRAIN Initiative*, 168 CELL 956, 956–59 (2017); Clara Baselga-Garriga et al., *Neuro Rights: A Human Rights Solution to Ethical Issues of Neurotechnologies, in* 49 PROTECTING THE MIND: ETHICS OF SCIENCE AND TECHNOLOGY ASSESSMENT (López-Silva P & Valera L. eds., 2022); RAFAEL YUSTE, LAS NUEVAS NEUROTECNOLOGIAS Y SU IMPACTO EN LA CIENCIA, MEDICINA Y SOCIEDAD (Lecciones Cajal ed. 2019); Marcello Ienca et al., *Towards a Governance Framework for Brain Data*, 15 NEUROETHICS 20, 23–24 (2022); M.F Ramos et al., *A Technocratic Oath, in* 49 PROTECTING THE MIND: ETHICS OF SCIENCE AND TECHNOLOGY ASSESSMENT (López-Silva P & Valera L. eds., 2022); Rafael Yuste et al., *Four ethical priorities for neurotechnologies and AI*, 551 NATURE 159, 159–63 (2017); Sara Goering et al., *Recommendations for Responsible Development and Application of Neurotechnologies* 14 NEUROETHICS 365, 365–86 (2021); Alejandra Zúniga-Fajuri et al., *Neurorights in Chile: Between neuroscience and legal science, in* 4 DEVELOPMENTS IN NEUROETHICS AND BIOETHICS 165 (2021); Timo Istace, *Neurorights: The Debate About New Legal Safeguards to Protect the Mind*, 37 L. & MED. 95 (2022).

interconnectedness between law and science is becoming increasingly evident with the progressive development of new technologies and the consequent regulatory challenges that stem from this evolution. That is why I would like to thank the organisers of these seminars for taking into account the perspective of scientists. Hopefully, the interaction between scientists and leading academics specialised in human rights and legal issues more broadly is not a passing trend but rather a staple of future research projects on the impact of AI in the public sphere.

I consider that the importance of this collaboration can be better illustrated through a short story closely concerning one of the most intimidating inventions of modern history. A story whose origin can be traced back to the street where I have carried out my research as a neuroscientist during the past few years. I work at Columbia University, and my laboratory is located right in front of a building which has been included in the National Registry of Historic Places in the United States. The building I am referring to is the Pupin Hall Laboratory. The reason for the inclusion of this building in the registry is that the first atomic reactor was built in its basement. The work carried out by some of the physicists responsible for this achievement would go on to become the foundation of the Manhattan Project. The development of atomic energy changed the history of humankind in unprecedented ways.

Perhaps paradoxically, some of the physicists behind the Manhattan Project and the discovery of the processes necessary to create the atomic bomb were among the most fervent defenders of the need to regulate atomic energy. Thus, they carried out an impressive lobbying campaign aimed at the UN and the international community. Through said lobbying and the support of President Eisenhower, the UN created the Atomic Energy Commission in Vienna—an international organisation tasked with the regulation and control of atomic energy to this day. In my opinion, this story perfectly encapsulates the dual nature of science. Technological developments and science are morally neutral. They have the potential to be used for good or for bad.

With a mere change of application, technology that had the potential to bring humans to the brink of extinction allowed for the expansion of civilization. The same atomic energy that was used for devastating consequences only a decade prior could hold the promise to solve the perennial issue of energy shortages. It demonstrated the potential to provide the world with unlimited, free energy forever. If only we could figure out a way to control nuclear fission.

This is how I would like to frame the main topic of my presentation: How to tackle regulatory challenges in instances where science advances faster than expected? How should society adapt to the development of these new technologies?

A. Neurotechnology: A Path Toward Understanding

To answer these questions, it is important to define some of the technologies that are at the heart of the issue. Most of these inventions can be subsumed under the category of "neurotechnology." This is a term that alludes to the broad range of methods and devices that could be electronic, optical, magnetic, acoustical, or chemical in nature and that are aimed at two different objectives: (a) to merely record the activity of the brain or (b) to alter such brain activity. Neurotechnology is important for three main reasons.

First of all, its object of study is one of the most, if not the most, important organs in the human body. The brain is formed by eighty billion neurons inside the skull, whose activity is so complex that scientists have been unable to decipher some of the mysteries regarding the processes involved in its functioning. Nonetheless, these mysteries have not deterred experts from studying some of its characteristics. With current understanding, it is clear that what was historically understood as the "mind" is a product of brain activity. This activity includes all your thoughts, your memories, your imagination, your decisions, your behaviour, and your emotions. As such, the brain is inextricably linked with human identity. Some of the most promising advancements in the field relate to the invention of technologies that enable us to write and project information into the human mind. This technology is not science fiction. This sort of activity is currently being implemented in the lab and used with experimental animals. These practices allow us to further understand the way the brain works, but its utility cannot be reduced to scientific curiosity.

The second reason why the development of neurotechnology is important is the existence of numerous neurological diseases, such as Parkinson's, Alzheimer's, schizophrenia, epilepsy, depression, ALS, strokes, intellectual disability, etc., which reflect alterations of brain activity. To understand how to treat these disorders and cure patients with mental or neurological diseases, we need to further our understanding of this organ—something which poses important challenges. As it stands today, we lack the technology to delve into the brain, analyse what is happening, and change it. Bearing that in mind, medical clinical reasons can be considered another source of interest in this specific field. The development of new technologies is an urgent matter. Everyone knows at least one family member or friend that suffers from a mental or neurological disease. In fact, according to the World Health Organization (WHO), one in every eight people in the world suffers from mental or neurological diseases with our current methods providing limited assistance.[3]

The third reason why neurotechnology is important has to do with the economy and with harnessing the potential of algorithms that are already present in our brain. By deciphering how the brain works, we may be able to create new technologies that would supersede the information technology that we currently understand as AI. So why am I participating in a panel about the implementation of AI?

B. Harmonizing AI and Neurotechnology: The Human Rights Approach

AI has the potential to decode and ultimately change brain activity. As previously mentioned, this is not just a matter of speculation or science fiction. These experiments are currently being carried out with laboratory animals as well as human patients. However, the impact of algorithms is not limited to medical trials. In the current social media landscape, where algorithms are ubiquitous, most of us, not only medical patients, are affected by these technologies. It is the case that these new technologies are now being driven by large investments throughout the world through both public and private funding. The end goal of some of the projects is to create noninvasive interfaces that interact directly with the brain. It is a step further than the development of peripheral devices, such as glasses or earphones, that were so prevalent over the last decade. This development raises many ethical and societal issues.

In response to some of the challenges posed by the application of AI technologies that could have over time a detrimental effect on society, we created the Morningside Group and organised periodic meetings at the Morningside Campus at Columbia. One of the first conclusions that was reached during the meetings was the need to implement an approach to the topic based on human rights. So why do we say that this is a human rights issue?

Neuro-Rights

We are concerned about four different types of potential abuses derived from the use of neurotechnology. As a response to these challenges, we advocate in favour of the creation of a new category of rights aimed at the protection of the minds of citizens. We designate

[3] *See* WORLD HEALTH ORGANIZATION, https://www.who.int/news-room/fact-sheets/detail/mental-disorders (last visited Feb. 22, 2023).

them by the name: "neuro-rights."[4] In our opinion they could be classified in the following manner:

> (a) The right to mental privacy: the content of our mental activity should not be decoded without the consent of the person subject to these new technologies. This mental privacy includes both conscious thinking and the subconscious. Most brain activity is actually subconscious; we are not even aware of its existence, yet it determines our way of life and who we are. Despite its "hidden" nature, subconscious mental activity can be deciphered in the same way, given that it is generated by neurons.
>
> (b) The right to mental identity: consciousness and the concept of self do not come out of thin air—it is generated by the brain. It has already been proven, by recent experiments and evidence derived from clinical studies, that stimulation of the brain can cause identity changes. There are some anecdotal cases of Parkinson's patients that have deep brain stimulators that are switched on to alleviate their symptoms. These experiences prove that through stimulation personality changes may be induced. This anecdote means that, at least in principle, we should be able to change the identity of a person. This possibility clashes with one of the most fundamental principles of social life: the need to establish protections that guarantee the preservation of this inner sanctum of identity that determines who we are. The right to mental identity is intertwined with the next right on the list.
>
> (c) The right to agency or free will: this means that human decisions belong squarely in our brains, and they should not be interfered with from the

[4] *See* THE NEURORIGHTS FOUNDATION, https://neurorightsfoundation.org/ (last visited Oct. 24, 2022) (additional information available at this website).

outside through the use of new technology. Once more, the focus is placed on the idea of considering the brain as a sanctuary from external influences and intrusive external devices. As it has been stated before, none of these realities are science fiction. This intrusion is something that people already do with animals. In the group that I am part of, we can program and implant into the brains of mice images of things that they have not experienced. Nonetheless, the subjects of the experiment behave as if they had truly seen these images. We have reached these results by using optical neurotechnology.

(d) A general right to equality and justice in a context in which mental augmentation is part of our lives: this possibility is unavoidable. In about ten to twenty years, we will live through the creation of noninvasive devices that can connect us to the internet, something which could open the possibility of hybrid human beings. A significant part of the cognitive and mental processing of these individuals would be done from outside of their brains, using AI or external databases capable of enhancing mental processes. The application of technology aimed at the improvement of human life is not something new. Humans as a species have been improving and enhancing themselves from the beginning, from the discovery and application of fire or the invention of several instruments, such as the wheel, clothing, transportation units, and computers. Technology has the potential to improve human capabilities, but it poses a great challenge to the value of equality. The implementation of this sort of technology could have the unintended consequence of fracturing society by creating two types of human beings: humans that are augmented and humans that have not been enhanced. There is a need to establish regulations that prevent the most pernicious effects of a phenomenon that is likely

> to occur in the next couple of decades. Access to mental augmentation should be regulated under the universal principle of justice.

With all these challenges on the horizon, the need to advocate for initiatives, such as the Neuro Rights Foundation, becomes apparent. Our main goal for this project is to protect the brain and human life by updating the existing bodies of human rights currently inscribed in the international treaties that have been signed by most of the countries in the world. The addition of special provisions that will include these new neuro-rights is required so that we can enter the future with a solid protection of human nature, one that is based on a human rights approach.

The Universal Declaration and other additional human rights treaties define what it means to be human better than any other document in history.[5] They define the basic characteristics and rights of a human being. Against the backdrop of unbound technological advance, the inability of law to adapt to these changes presents itself as particularly pernicious. As society and technology changes, so should human rights. Consequently, these provisions should be updated to the standards necessary to overcome the challenges posed by technological developments that are going to change the concept of what it means to be human in a fundamental way and ready them for the twenty-first century. This is a conversation that we should start right now because these technologies have been in development for decades.

C. Conclusion: Inspiring Experiences in the Current Legal Landscape

I would like to finish my presentation with a general comment about the importance of the expansion of scientific advancement in this field. As I have tried to show, it can be argued that this scientific advancement is a human rights issue. In fact, some countries following this approach have jumped ahead. Such is the example of the Republic of Chile. Due to the efforts of its senate and its Committee of the Future, an amendment to article 19 of the constitution was approved unanimously by the senate and chamber and signed by the president of the republic. This amendment provides protection to cerebral activity

[5] G.A. Res. 217 (III) A, Universal Declaration of Human Rights (Dec. 10, 1948), https://www.ohchr.org/en/universal-declaration-of-humanrights#:~:text=The%20Universal%20Declaration%20o%20f%20Human%20Rights%20(UDHR)%20is%20a%20milestone%20,rights%20to%20be%20universally%20protected.

and the information that comes from it.[6] The amendment was approved with unanimous support from both the National Congress of Chile and the Senate. By being embedded in the constitution, it has become a human right for the Chilean people.

In the same vein, the United Nations has shown interest in the inclusion of a new provision in the existing national treaties.[7] The Neurorights Foundation is collaborating with the organisation and its current Secretary General, António Guterres, who, after his reelection last year, declared that one of his main objectives was to update the Human Rights Charter on this matter. It appears that the recognition of human rights in relation to technology is going to be part of the priorities of the UN for the next six years.

During my presentation, I have attempted to present an approach to technology through the lens of human rights. It may sound like something completely unexpected for some of you; however, this proposal is an interesting perspective on how things could be. We are in the midst of a technological revolution. The tipping point is near and the consequences of deregulation could be catastrophic. Perhaps a human rights approach, such as the one that has been introduced in these pages, might confront the future with more certainty and safeguards against all possible risks. If we are able to regulate these technologies (AI, robotics, surveillance technologies, the metaverse) within a larger framework, we will be able to capture all the possible unintended negative effects and ethical and societal consequences. A body of regulation aimed at the risk posed by neuroscience and the technologies currently applied in the field could be the spearhead for a larger Charter of Digital Human Rights that enables us to create the necessary guarantees to develop these technologies in a sensible and conscious way. Instead of waiting for the atomic bomb to be detonated, perhaps we should learn from the past and act to prevent potential problems by having our house in order. In this case, our human rights house in order.

[6] Law No. 21.383, art. 1, Octubre 25, 2021, Diario Oficial [D.O.] (Chile) (modifying the fundamental charter to establish scientific and technological development at the service of people), https://www.diariooficial.interior.gob.cl/edicionelectronica/index.php?date=25-10-2021&edition=43086-B&v=2; Allan McCay, *Neurorights: The Chilean Constitutional Change*, AI & SOC'Y, Mar. 2, 2022, at 1, https://doi.org/10.1007/s00146-022-01396-0.

[7] U.N. Secretary-General, *Roadmap for Digital Cooperation*, U.N. (June 2020).

III. NEW CHARTS OF DIGITAL RIGHTS IN THE EUROPEAN UNION AND SPAIN: TOMÁS DE LA QUADRA-SALCEDO[8]

The question with which I would like to introduce my presentation is: Why and for what purpose should there be charters of rights in Europe? Why talk about this topic? Probably because, as of January 26, 2022, the European Union released a proposal on a *Declaration on Digital Rights and Principles*[9] to reflect on the importance of digital rights at the highest European level. Among other things, the sheer scope of the challenge reveals the pressing nature of this debate. When we talk about digital rights, we are no longer talking about something that affects one country or one region but rather something that affects the entire world. The solution to the many challenges posed by the surge of new technological developments worldwide demands a transnational approach based on cooperation. My aim is to encourage such an attempt by providing collective solutions to these new problems that arise from what has been labeled as the "digital world" or "digital society."

The problem that Professor Yuste has raised in his presentation is intimately related to the great achievements and opportunities presented by intensive research aimed at mapping the brain and discovering how knowledge is created and stored. They are experiments with a vast potential to discover opportunities to cure diseases, and perhaps, even to improve human life in a more fundamental way. As with all important technological improvements, these discoveries entail many risks. This is the discussion in which we have been immersed since January 2022. We are currently at the centre of a European-level debate concerning the model of digital society that we aspire to build. What digital rights should be recognized to prevent a future in which

[8] Tomás de la Quadra-Salcedo Fernández del Castillo, *¿Por Qué Una Carta de Derechos Digitales?,* REVISTA REGISTRADORES DE ESPAÑA, https://revistaregistradores.es /por-que-una-carta-de-derechos-digitales/ (last visited Oct. 24, 2022); Tomás De La Quadra-Salcedo Fernández Del Castillo et al., *Sociedad Digital y Derecho*, BOLETÍN OFFICIAL DEL ESTADO (Ministerio de Industria, Comercio y Turismo), Nov. 2018; Tomás De La Quadra-Salcedo Fernández Del Castillo et al., *Sociedad Digital y Derecho*, BOLETÍN OFFICIAL DEL ESTADO (Ministerio de Industria, Comercio y Turismo), Nov. 2018, at 21–86; Tomás De La Quadra-Salcedo Fernández Del Castillo, *La Carta de Derechos Digitales*, VIMEO (Oct. 18, 2021), https://vimeo.com/635253955; Rafael de Asís, *Sobre la Propuesta de los Neuroderechos, in* 47 DERECHOS Y LIBERTADES 51 (Dykinson ed., 2022); Diego Alejandro Borbón et al., *Critical Analysis of Neurorights to Free Will and to Equal Access to Mental Augmentation*, 6 IUS ET SCIENTIA 3 (2020); Txetxu Ausín et al., *Neuroderechos: Derechos humanos para las neurotecnologías*, 43 DIARIO LA LEY 1 (2020); Elisa Moreu, *The Regulation of Neuro-Rights*, 2 EUR. REV. OF DIGIT. ADMIN. & L. 149 (2021).

[9] Commission Declaration 28, Jan. 26, 2022, European Declaration on Digital Rights and Principles for the Digital Decade.

humans become the servants of our own creations? The objective is to create a landscape in which science is used rationally for the betterment of society as a whole, a future guided by general interest.

A. Constitutionalizing Digital Rights: Past and Present

As far as the question of risks is concerned, this is not the first attempt there has been in the European Union to regulate digital rights. There was a solid project that took place earlier and deserves to be highlighted. The authors of this article are alluding to the proposal made by Professor Stefano Rodotà before the Italian Chamber of Deputies.[10] This proposal was aimed at the creation of a Constitution for the Internet[11] and inspired the Declaration of Rights and Duties on the Internet of the Commissione per i diritti e i doveri on the Internet.

The existence of alternative terminology—Constitution for the Internet/Bill of Rights for the Digital Era—bears witness to the different approaches that can be taken regarding this problem. While some scholars have argued in favour of the constitutionalization of digital society through the creation of an entirely new body of rights, others consider that the traditional legal principles are sufficient to tackle the challenges posed by this new environment. However, the question remains: Is there a need to constitutionalize this new field and establish legal guarantees? The main limitation of Professor Rodotà's proposal, if it is to be extrapolated to the present day, is that this project of constitutionalism was confined only to the margins of the internet. But, as Professor Yuste has stated in his magnificent presentation, we are no longer talking only about the internet. The challenge facing the law today is much more significant. It is essential to define the role of humans in the new digital society. A mere compilation of past regulatory proposals will not suffice.

This debate is not new. On the other side of the Atlantic, these

[10] Stefano Rodotà, *Towards a Declaration of Internet Rights,* AREA OF FREEDOM SECURITY & JUSTICE (Nov. 18, 2014) https://free-group.eu/2014/11/18/towards-a-declaration-of-internet-rights/.

[11] Mauro Santaniello et al., *Mapping the Debate on Internet Constitution in the Networked Public Sphere,* 3 COMUNICAZIONE POLITICA 327, 354 (2016); NEURON EDOUARDO CELESTE, DIGITAL CONSTITUTIONALISM: THE ROLE OF INTERNET BILLS OF RIGHTS, 1 (Routledge Publishing, 2022); Internet Rights and Principles Coalition, Matthias C. Kettermann, Forza Internet Rights: IRPC Charter as Source of Inspiration for Innovative Italian Declaration of Internet Rights (Sept. 9, 2022), https://internetrightsandprinciples.org/forza-internet-rights-iprc-charter-as-source-of-inspiration-for-innovative-italian-declaration-of-internet-rights/; *see generally* Politecnico di Torino, Nexa Center for Internet & Society, (Oct. 13, 2014), https://nexa.polito.it/declaration-internet-rights.

issues have been raised for decades. As early as 2001, legislation comparable to a Digital Bill of Rights was introduced in the US Congress.[12] It was a clear precedent for the regulatory instruments that were about to be developed in the decades since. I was able to witness the development of said bill firsthand in 2011 when I was a visiting professor at the Cardozo Law School (New York), and subsequently in 2015 at the Maurer School of Law in Bloomington (Indiana). Similar projects have been developed in Europe. In particular, a German foundation presented a proposal to the European Parliament for the elaboration of a Digital Constitution for Europe.[13] Since then, there have been several attempts to undertake such a project. One of the proposals that could be highlighted is the Declaration of Digital Rights, which was approved in Spain on July 14, 2021.[14] This text has had a notable impact in Europe, possibly serving as inspiration for the European Commission's declaration published in January 2022.

B. Revising Outdated Legal Categories

Everything seems to point to the existence of a series of challenges arising from the development of new technologies that are of concern to the main political institutions of the EU. The catalogue of rights we have had up until this point in time does not seem to suffice. These shortcomings should be alleviated by incorporating new concepts, such as the notion of "neuro-rights" proposed by Professor Yuste in his presentation. This notion encompasses an important part of reality that has been overlooked until now. The previous approach based on the notion of "data" is quite poor. The current problem extends far beyond that limited concept. Consequently, solutions must go beyond the notion of simple data protection strategies. The question of identity is at stake.

[12] *See* KeepTheWebOpen, *A Digital Bill of Rights at the Personal Democracy Forum*, YOUTUBE (June 14, 2012), https://www.youtube.com/watch?v=eNkb3w8Q8Is (showing Representative Darrell Issa and Senator Ron Wyden's presentation on the Digital Bill of Rights at the Personal Democracy Forum).

[13] *See Charter of Fundamental Digital Rights of the European Union*, WE DEMAND BASIC DIGITAL RIGHTS, http://www.digitalcharter.eu/ (Proposal of Digital Bill of Rights); *see also* Eur. Parl. Doc. (LIBE_PV(2016)1205_1) (2016) (Meeting minutes including information on Charter of Digital Fundamental Rights); *see also Committee on Civil Liberties, Justice and Home Affairs*, EUR. PARL. (May 12, 2016), https://multimedia.europarl.europa.eu/en/committee-on-civil-liberties-justice-and-home-affairs_20161205-1500-COMMITTEE-LIBE_vd (video of Parliament Session discussing Digital Fundamental Rights).

[14] *Carta Derechos Digitales* [*Digital Rights Charter*], GOBIERNO DE ESPAÑA, https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf (Spain).

The brain is the most sacred organ of the human person. If science discovers a way to connect neural networks to machines, we could find ourselves in a reality in which the subconscious itself becomes accessible to third parties, even against the individual's volition. Artificial intelligence has a dual nature—it can be used to cure diseases, transmit information, and even improve the cognitive capacities of individuals. But it can also be instrumentalized for the purpose of controlling those same subjects. All this poses much deeper challenges than the mere notion of data that has characterized the debate so far. To curb the most harmful consequences of the development of these new technologies, it is necessary to enshrine positive rights in legally binding texts. Precisely one of the first questions discussed in the preamble to the Digital Bill of Rights is whether this reality requires the recognition of new rights or whether the debate can be redirected to the classic question of human dignity—inherent to the idea of personhood—and its multiple manifestations.[15]

In this sense, traditional bills of rights would seem to have proved sufficient in the past to guarantee the protection of citizens' rights even in periods of profound social and technological change. The Spanish draft of the bill of digital rights raises some novelties, such as the adaptation of Spanish legislation on data protection to the standards required by European regulation.[16] This text also included a chapter on the issue of rights. Some of these rights are expressly mentioned, such as in the case of the right to digital disconnection. Each of these rights is accompanied by the subsequent questions. For example, the right to digital disconnection raises questions, such as the following: Is this right really a new right or is it merely an extension of the right to rest that has for decades been part of employment legislation since its initial inclusion in the Workers' Statute? Although such an interpretation is possible, it would be more accurate to state that a new law has been established as a result of the adaptation of the general principles of employment law to a new situation brought about by technological development. These are new situations that call for an innovative regulatory exercise.

The same applies to other rights, such as the right to a "digital will." The term alludes to the right of individuals to determine the way in which the digital heritage of a deceased person should be managed by their heirs. I would like to illustrate this concept with a real example. It is a case that recently arose in Germany and started an intense national debate about the limits of privacy and the ownership of accounts in

[15] *Id.*

[16] *Id.*

social networks. The case that sparked the debate was the possible suicide (rather than an accident) of a teenage girl in the Berlin subway.[17] After her death, her parents wanted to access the content of her social media accounts to find out what factors had led her to end her life.[18] They suspected that the mental state of their daughter could have been affected by someone from the school where she studied.[19] Facebook denied them access to her account on the platform.[20] The parents then decided to take legal action.[21] In the first instance, a Berlin court ruled in their favour.[22] The company appealed that decision of the first instance, which was overturned by the appellate court.[23] Finally, the Federal Court of Justice of Germany (Bundesgerichtshof) settled the case definitively by ruling in favour of the parents and allowing access to their daughter's Facebook account.[24]

Initially, we might think that this case could be resolved using the classic legal categories. However, the resolution reached by the court raises problems. Perhaps the teenager's account contained information and contacts that she would not have wanted her parents to know. Nowadays, the internet contains a vast amount of personal information that we might never want to see disclosed, such as our ideas, contacts, etc. Perhaps the will of the victim was that the contents of her social media accounts were never revealed to anyone, yet her right to privacy was quashed in this instance by the right of her parents to obtain material justice through an official investigation. The fact that this issue has arisen in multiple jurisdictions, with different legal systems and often contradictory guiding principles, seems to indicate that the debate is not settled yet.[25]

[17] *Facebook Ruling: German Court Grants Parents' Rights to Dead Daughter's Account*, BBC (July 12, 2018), https://www.bbc.com/news/world-europe-44804599.

[18] Bundesgerichtshof [BGH] [Federal Court of Justice] July 12, 2018, III ZR 183/17 1, 2–5 (Ger.)

[19] *Id.*

[20] *Id.*

[21] *Id.*

[22] *Id.*

[23] *Id.*

[24] *Id.*; *Germany: Federal Court of Justice Clarifies Scope of Postmortem Access to Social Media Accounts*, LIBRARY OF CONGRESS (2020), www.loc.gov/item/global-legal-monitor/2020-09-30/germany-federal-court-of-justice-clarifies-scope-of-postmortem-access-to-social-media-accounts/.

[25] Kristin Nemeth & Jorge Morais Carvalho, *Digital Inheritance in the European Union*, 6 J. EUR. CONSUMER & MKT. L. 253, 253 (2017); Giuseppe Marino, *La Successione Digitale*, 1 OSSERVATORIO DEL DIRITTO CIVILE E COMMERCIALE 165, 202 (2018); Alberto B. Lopez, *Posthumous Privacy, Decedent Intent and Post-Mortem Access to Digital Assets*, 24.1 GEO MASON L. REV., 183, 183–85 (2016).

C. Contextual Awareness and the Regulation of the Digital World: Risks and Opportunities

As a society, we must reflect on the content and implications of the notion of human dignity in this new reality. From the legal scholar's perspective, the idea of "context" is essential. Heidegger coined the term "Dasein" to refer to this notion.[26] Other philosophers have also alluded to this matter. Ortega and Gasset famously said: "*I am I and my circumstance*."[27] Our current circumstance today is determined by a world where the digital element is becoming increasingly important. The challenges posed by this new reality cannot always be circumscribed to the rigid margins of the classic idea of dignity. It is necessary to find solutions to the debate raised by the conflicting rights of the parents and their daughter.

In the same way, it is necessary to give context and set the boundaries of the concept of neuro-rights proposed by Professor Yuste, which possibly constitute the most novel and disruptive principle of those implicit in the Declaration of Digital Rights. Section 26 of the text itself hosts a series of reflections on the impact of neurotechnologies.[28] These scientific developments have the potential to cure diseases, such as Alzheimer's, depression, and Parkinson's. The possibility of curing diseases is presented as something very positive and uncontroversial. However, the development of technology may have other pernicious and unintended consequences on social life.

Equality Considerations: As the American philosopher Michael Sandel showed in his work *The Case Against Perfection*,[29] one of the main challenges posed by the possibility of enhancing humans through the application of these technologies is the deterioration of the principle of equality. We run the risk of creating a society divided between individuals who have been augmented and the rest of the people. Would a society based upon such stark inequalities be considered legitimate and fair?

Human Agency: Under a model of syllogistic thinking, such as the one that currently characterises human thought, we could move toward a scenario based on endless storage and reproduction of data. Imagine the case of a person who has been enhanced since childhood. The individual's mind is connected from their early years to an almost

[26] MARTIN HEIDEGGER, BEING AND TIME 28–31 (John Macquarrie & Edward Robinson trans., Blackwell Publishers Ltd. 1st ed. 1962).

[27] JOSE ORTEGA Y GASSET, MEDITACIONES DEL QUIJOTE (3d ed. 1914).

[28] *Carta Derechos Digitales*, *supra* note 14, § 26.

[29] MICHAEL SANDEL, THE CASE AGAINST PERFECTION: ETHICS IN THE AGE OF GENETIC ENGINEERING 10–24 (PHOTO. REPRT. 2009) (2007).

unlimited library of knowledge. Is that person really free or will they feel for their entire life dependent on the statistical and factual dictates of the machine? Such a way of thinking could be problematic from the perspective of innovation and human progress. Statistics and mere factual reproduction stagnate knowledge.

In short, all the scenarios posed by the new digital reality require an exercise of reflection that will probably lead to the redefinition of many of the traditional rights. Ideally, this process will be based on the principle of human dignity. In his essay, "The Outdatedness of Human Beings," Ghünter Anders, whose writing was motivated by the development of the atomic bomb, already anticipated this new reality and its effect on the human condition.[30] For Anders, the modern man that coexisted with the great scientific revolutions that took place in the twentieth century has become outdated; we are no longer the authors of our own destiny.

The Massachusetts Institute of Technology (MIT), in its initiative the Moral Machine, provides a very illustrative example in this regard.[31] In said initiative, people were asked for their opinion on self-driving cars. They were presented with a series of scenarios that posed serious ethical dilemmas comparable to the well-known "Trolley Problem." The questions followed this pattern: In the event of a failure of the vehicle's brakes, who should we try to avoid first, an elderly person or a mother and her children? Who should make this decision? Is it legitimate to delegate it to the machine itself or to the AI designer who governs the machine? Would this mean abdicating our responsibility as moral agents? Does this imply the deterioration or total loss of our moral status?

In addition to this issue, there are other very important risks related to the preservation of democratic institutions. As far as democracy is concerned, we have already witnessed some of the dangers derived from the implementation of these new technologies. We could observe it with the role played by social networks during the electoral campaign that led to Donald Trump occupying the White House.[32] This occurrence is in addition to the role played by Cambridge Analytica in the referendum on the United Kingdom's withdrawal from the European Union. In these electoral processes, the electorate was influenced through individually targeted propaganda aiming to exploit their political biases. This propaganda was accompanied by the

[30] *See* GÜNTER ANDERS, THE OUTDATEDNESS OF HUMAN BEINGS (1956).

[31] *See* MORAL MACHINE, https://www.moralmachine.net/ (last visited Oct. 29, 2022).

[32] *See* Michael Landon-Murray, et al., *Disinformation in Contemporary U.S. Foreign Policy: Impacts and Ethics in an Era of Fake News, Social Media, and Artificial Intelligence*, 21 PUB. INTEGRITY 512 (2018).

targeting of specific sectors of the population to promote or discourage participation depending on the socioeconomic and political profile of their members. Democracy itself could be in danger as it was acknowledged by the Spanish Constitutional Court in its ruling, STC 76/2019, de 22 de mayo de 2019.[33]

This battle also takes place in the economic sphere, specifically through antitrust law. In recent times, significant sanctions have been imposed on large intermediaries in the market.[34] These sanctions could indicate an increase in the attempts to manipulate free competition. The existence of markets governed by free competition is essential for a democratic society. This part of the legal system makes it possible to avoid economic concentrations whereby one or a group of operators could accumulate sufficient power and influence to condition society. This problem is not merely an economic issue, but also a political problem. Only if we ensure that economic power is not in the hands of a powerful minority will we be able to say that we live in a truly free society.

Therefore, when we talk about the Digital Bill of Rights, we are not only talking about the internet. The internet is a tool that completely alters the way in which social relations have been previously organised. This new scenario opens up possibilities that need to be regulated. The group that developed the Charter of Digital Rights, in which Ricard Martinez and I have had the opportunity to participate, tries to contribute to this conversation. One of the essential dimensions to face this challenge is the following issue: how to reconcile the ethical considerations that are involved with the development of these new rights? It is necessary to develop a convincing ethical discourse that, regardless of religious conceptions, serves as a meeting point for all those involved in the resolution of the problems arising from the development of all these new technologies.

This ethical reflection is the basis of the Charter of Digital Rights, which is intended to cover all the issues that have been mentioned above: digital will, neuro-rights, competition law, etc. All these considerations should inspire the regulatory framework for the digital economy. This regulation is a prescient topic when discussing the concept of "backward compatibility," that is, the ability of new electronic

[33] S.T.C. June 25, 2019 (T.C. No. 151, p. 67680–82) (Spain), https://www.boe.es/buscar/doc.php?id=BOE-A-2019-9548.

[34] *See* Nicolas Petit & David J. Teece, *Innovating Big Tech Firms and Competition Policy: Favouring Dynamic Over Static Competition*, 30 INDUSTRIAL & CORP. CHANGE 1168 (2021); *see also* Xavier Vives, *El Paradigma de la Competencia en el Sector Bancario Después de la Crisis* [*The Paradigm of Competition in the Banking Sector After the Crisis*], IESE PUBLIC-PRIVATE SECTOR RESEARCH CENTER (2011).

devices to adapt to previous versions. This concept goes hand in glove with the idea of sustainability. We need to ensure that things are used for as long as possible, rather than them being discarded after their first use. The same goes for energy efficiency and other principles and values that, if left to the sole discretion of the market, could be severely undermined. The ethics of sustainability must be transferred to industry and research.

These dilemmas also apply to the area of information. Today it is said that people have access to the largest number of information sources in history. However, there is a common belief that the information to which we have access does not have sufficient guarantees of veracity and quality.[35] We do not know who is informing us. No one takes responsibility for the information they transmit. This poses serious problems. The EU has made some attempts to alleviate this situation through various pieces of regulation: The Digital Services Act,[36] the Digital Market Act,[37] and the Digital Governance Act.[38] All of them are legal texts that analyse the same issue from different perspectives. Information, which is a precondition for the proper functioning of democracy, has undeniable economic ramifications. It is an essential resource for the defence of free competition and the market.

This is not a completely new issue. In Spain, there are historical antecedents in the matter of regulation that predate the twenty-first century. In the Spanish Constitution itself, the following prescription is contained (article 18.4): "The law shall limit the use of information technology to guarantee the honour and personal and family privacy of citizens and the full exercise of their rights" (my translation).[39] Even though at that time the Spanish Constituent Assembly could not anticipate the scope of the challenge posed by new technologies, the need to regulate this area of reality was already considered. The fear that computerization, if left uncontrolled, could cause significant damage was already foreseen. This consideration shows that the legislature was aware of the destructive potential of computers. On no

[35] *See* Jakob-Moritz Eberl, *Lying Press: Three Levels of Perceived Media Bias and their Relationship with Political Preferences*, 44 COMMC'NS 1 (2018).

[36] Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive, No. 2000/0361 (COD) of 15 Dec. 2020, at 1–2.

[37] Proposal for a Regulation of the European Parliament and of the Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act), No. 2020/0374 (COD) of 15 Dec. 2020, at 1–3.

[38] *Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act)*, No. 2020/0340 (COD) of 25 Nov. 2020, at 1.

[39] C.E., B.O.E. n. 311, art. 18.4, Dec. 29, 1978 (Spain).

other occasion has the legislature addressed any other human instrument in such a way—there is no mandate in the constitutional text for the legislature to limit the use of knives, for example. Normally the generic prohibition of harming others sufficed without the need to specifically relate the things that could be used to cause such harm. In regards to the internet, that was not the case. With information technology, we recognised the need to anticipate because we sensed some of the implications that the development of this type of technology could have on human life.

The courts have subsequently developed the content of some of the principles contained in this clause (personal and family honour, privacy, etc.). These are autonomous and complex rights that cannot be completely analysed from the perspective of the data. The data must be considered in conjunction with society. The demands of the Digital Constitution, or the Digital Bill of Rights, represent a global reflection on the ways in which the digital world constitutes that new scenario to which Heidegger referred when he spoke of the idea of "dasein." It is the circumstance that determines the risks to be faced. It contains this action of unveiling. The technique reveals an immanent reality in nature. Human beings were unaware of the nature and properties of water until technology revealed it. Soon reservoirs were built, and hydroelectric power was consequently discovered and exploited. What is paradoxical about the current situation and what distinguishes our era from past times is that technology is no longer a mere by-product of science. In the digital world, we no longer control technology, but it rather controls us. The machine has somehow become autonomous. Unlike in years gone by, now the technology itself creates and helps to discover with relative independence. In other words, what we must ask ourselves is what is the ethical framework that governs the driverless car? Who created this technique? Who created these ethics?

D. *Conclusion: Regulation for the General Interest or Public Efforts to Tame the Digital Leviathan*

These are some of the reflections that inspired the Charter of Digital Rights and that the European Union has assumed as its own in the declaration of principles and digital rights published in January 2022. Spain has drawn up its own charter since the European Declaration is more general and suffers from a lack of detail. The Spanish Charter goes into detail in specific areas, such as the right to education, social participation, neuro-rights, information, freedom of expression, and privacy. It even speaks to the question of identity. In this context, marked by the emergence of new technologies capable of permanently altering the brain, the question of identity acquires even greater depth. Identity is no longer exclusively that with which we identify ourselves and our memories, but rather a reality that can be altered from the outside. It becomes essential to answer the question of who constructs our identity. Identity is such an important and intimate dimension of the human experience that it should not affected by anyone but ourselves. The charter addresses this question as well.

Our "Dasein," the digital world, is what conditions the interpretation of the new rights that are to be enshrined. As has been stated previously, the global context that we inhabit requires solutions on a transnational scale. At the moment, there are various approaches to the debate on the digital landscape: Chinese regulation, the European model, and US regulation all provide possible approaches to the problem. All of them respond to very different models. Personally, I believe that Europe, through its Data Protection Regulation initiative, has shown that it can be an example to follow and inspire other nation states. The Charter of Digital Rights is a firm step in the right direction. However, the truth is that, ultimately, it would be beneficial to have a joint project at the global level. In the event it is not possible to incorporate all the major international players, cooperation between the United States and Europe would be welcome, as China may want to pursue its own approach. The lack of harmony between the different countries entails risks. In a jurisdiction that is less protective and respectful of human rights, it may be possible to achieve more rapid change. However, the risk is far too great.

In conclusion, the Digital Bill of Rights is not only a commitment to the European model but also an invitation to collaboration on both sides of the Atlantic. This is why the initiative of the Indiana Journal of Global Legal Studies to raise the subject of debate with universities and other participants from both continents seems to me particularly adequate. The disturbing notion of a digital leviathan, as a metaphor that captures the problems of the digital transformation of government, is very apt. It is certainly one of the major issues that will shape the course of the twenty-first century.

Trust in Artificial Intelligence Analysis of the European Commission proposal for a Regulation of Artificial Intelligence

ANTONIO ESTELLA*

I. INTRODUCTION

According to the European Commission, one of the main objectives of the regulatory framework that this EU institution is currently proposing in the field of Artificial Intelligence is to "increment trust in the use of artificial intelligence."[1] Therefore, this paper explores the issue of trust and AI. The questions that it attempts to answer are the following. Why is trust important? Why is trust important, in particular, in the domain of AI? How does the EU Commission intend to achieve the objective of incrementing trust in the use of AI? Will the proposed regulatory framework achieve its proclaimed end?

To answer these questions, this article proceeds as follows. I shall start by reflecting on the importance that trust has for society (section 2). From there, I will define what is to be understood in this paper by trust (section 3). I shall then review the basis of trust (section 4) and shall make a reference to the main sources of evidence on trust (like, surveys and laboratory experiments), and to some of the results that these sources reveal on interpersonal and institutional trust (section 5). In the next section (section 6), I shall go on to analyse specifically the issue of trust in AI, will refer to the existing evidence on the matter, and will review some of the most recent literature on this topic. In the remaining sections (sections 7 and 8), I will describe and analyse the European Commission's proposal for a regulation of AI, and in particular, the part of that proposal that deals with trust in AI. In the last section of this article, I will wrap up the whole argument of this paper and make some conclusions (section 9). The main argument that

[1] 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 {SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}. Brussels, 21.4.2021 COM(2021) 206 final 2021/0106 (COD).

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

will be developed in this paper is that it is inconsequential to speak of trust in AI systems.

II. THE IMPORTANCE OF TRUST

Trust has been defined by some authors as the "lubricant of society"[2] and by others as "a kind of glue that makes society function."[3] Political scientists, economists, and also lawyers have recently centred their intellectual efforts on trying to understand how trust impacts economic growth, development, democracy, justice, and even interpersonal relationships. One particularly clear expression of this renewed interest in trust is the setting up by the OECD of a High Level Group on the measurement of economic performance and social progress.[4] The Group started working in 2013. This group convened eight workshops during the years 2014 to 2016. The latest one took place in Paris in June 2016 and was titled: "Measuring Trust and Social Capital." The outcome of this workshop was published in 2018, together with the rest of the reports of the other workshops that have been mentioned, under the title "Trust and Social Capital."[5] In this paper, Algan gives ample evidence of how trust is positively correlated with economic growth in general and with economic development in particular.[6] The idea is that the more trustworthy a society is, the more it grows and develops in economic terms. The findings of this paper are important since this is the first time that an international institution like the OECD argues that trust should be a necessary component for the measuring of how the nations of the world grow in economic terms.[7]

[2] *See generally* JON ELSTER, EXPLAINING SOCIAL BEHAVIOR (2d ed., 2015).

[3] *See* SCIENCESPO, Joseph Stiglitz on the Importance of Trust in Economics, https://www.sciencespo.fr/en/news/joseph-stiglitz-about-importance-trust-economy (last visted Jan. 1, 2023).

[4] Organization for Economic Co-operation and Development [OECD], *High Level Expert Group on the Measurement of Economic Performance and Social Progress* https://www.oecd.org/statistics/measuring-economic-social-progress/aboutthehigh-levelexpertgroup.htm.

[5] *See generally* Yann Algan, *Trust and Social Capital*, *in* FOR GOOD MEASURE: ADVANCING RESEARCH ON WELL-BEING METRICS BEYOND GDP 283 (2018).

[6] Algan, *supra* note 4.

[7] *Id.*

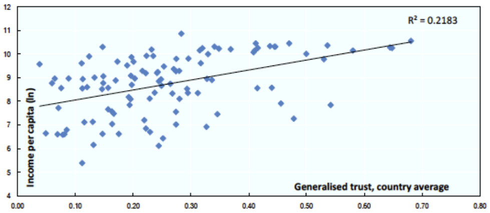


Figure 1: Inter-personal trust and income per capita[8]

On the basis of the previous Figure 1, Algan argues that "countries with higher levels of trust tend to have higher income." For example, Norway has very high levels of trust and has one of the highest incomes per capita of the countries that are included in the previous analysis. An opposite example would be Zimbabwe, with very low levels of inter-personal trust and comparatively low levels of income per capita. Algan acknowledges that there might be problems of reverse causality in analyses on the correlation between trust and economic growth: "one concern has been that this correlation . . . could go the other way around, i.e., from income to trust."[9] However, Algan and other authors have implemented statistical strategies to avoid this effect and try to figure out what direction causality takes in this area: "By focusing on the inherited component of trust, the authors avoid reverse causality. By providing a time-varying measure of trust over long periods, they can control for both the omitted time-invariant factors and other observed time-varying factors such as changes in the economic, political, cultural and social environments." The question is therefore a complex one that needs more refined analyses. However, it is probably safe to say that the positive impact of trust on economic growth and development is undisputed today. Still open to debate and analysis are the specific micro-mechanisms of such correlation.

Similar analyses are being made on trust and democracy, trust and justice, etc. In regards to democracy, the classical reference is Putnam.[10] According to this author, trust is a key component of social capital; therefore, when trust decreases social capital decreases as well, which has a negative impact on democracy. In turn, the impact of trust in the justice system is receiving a lot of attention from different academic

[8] *Id.*

[9] *Id.* at 302.

[10] *See* ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY 144–45 (2000); *see* also ROBERT D. PUTNAM ET AL., MAKING DEMOCRACY WORK: CIVIC TRADITIONS IN MODERN ITALY 169–70 (1993).

quarters today.[11] On the one hand, some of these analyses are worrisome of the decline of trust in the judicial system shown by surveys. On the other hand, other analyses have a more positive outlook since, compared to other branches of government, courts seem to be doing better in terms of trust. The debate on the impact of trust on justice and, in general terms, on the legal order is still open; more research still needs to be done in this important area.

III. WHAT IS TRUST?

As we have seen in the previous section, trust is important in different spheres of society, so therefore, we may turn now to the definition of trust.[12] Trust is a very intuitive concept: we all understand what we are talking about when we refer to trust. However, the definition of this concept at a theoretical level is much more elusive. In my opinion, the main reason for this rests in the confusion that exists between trust and cooperation. Trust and cooperation are treated, in many analyses, as co-terminus. However, it is important to differentiate them. Cooperation always stems from interest. I cooperate with you because I have a certain interest in doing it. You cooperate with me because you have a certain interest in doing it. Instead, trust does not necessarily stem from interest. I trust you irrespective of my interest in trusting you. In certain cases, like in blind-trust, trusting someone can even run against my interests. Therefore, the difference between trust and cooperation is that the latter one needs interest, whereas the former one does not.

Starting from this basic differentiation between trust and cooperation, we may use, for example, the definition that the Cambridge English Dictionary gives for trust. According to the CED, trust is "to believe that someone is good and honest and will not harm you, or that something is safe and reliable." I have chosen the CED definition of trust because it puts the focus on one important aspect: that trust is considered as a belief. For the time being, let us restrict the ensuing analysis to interpersonal relationships. When I say that "I trust you," what I am implying is that I believe that you will do what you say you would do. If you say, "I will be back at home at midnight," and I say, "I trust you," what I am saying in just a couple of words is that I believe that you will honour your promise and therefore, you will be back at home at midnight. Trust is therefore a belief, or an expectation, that my

[11] *See generally* ANTONIO ESTELLA, LEGAL FOUNDATIONS OF EU ECONOMIC GOVERNANCE (2018).

[12] *See generally* RUSSELL HARDIN, TRUST 46 (2006).

interlocutor will respect her commitments.

The idea of commitment is therefore also important for the discussion on trust. As a matter of fact, the clearest way to introduce ourselves in the discussion of trust is to start from a commitment-structure. If you say that "I will be back at midnight," what I am actually doing is making the commitment that I will be back at midnight. Then it follows that if I say "I trust you," what is actually happening is that I am saying that "I trust that you will honour your commitment to arrive home at midnight." We see, therefore, that for trusting structures, we need at least two people. Therefore, we discard situations in which I would say, "I always trusted myself that I would be back at home at midnight." Trust structures only make sense in the framework of commitments, and commitments always involve at least two persons. Therefore, once we have a commitment in place, and trust is at stake, there are two relevant persons: the trustee and the trustor.

Another very common confusion in the domain of trust is to think that trust depends only or mostly on the trusted person. If the trusted person is trustworthy, then we will have all reasons to trust that she will honour her commitments. This is why many analyses on trust posit that in reality, we should speak about trustworthiness instead of speaking of plain trust. I think however that such analyses are misleading, to say the least. The reason is that the trustee and trustworthiness are of course important in a trust structure, but they are not the only relevant players in the game. The trustor is at least as important as the trustee. In particular, the capacity of the trustor to trust plays a fundamental role in this area. To drive the point home, think of the two extremes—and to a certain extent, pathological—cases of blind trust and no trust at all. It would be crazy, for example, to have blind trust in Hitler, as much as it would be odd not to have trust at all in Mahatma Gandhi. The first case would be a case of blind trust and the second case would be a case of pistanthrophobia—the fear to trust anyone. This means that both the trustee's trustworthiness and the trustor's capacity to trust are crucial in a trust structure.

IV. THE BASIS OF TRUST

So, why do we trust (or not)? The bases of trust are also important to consider. For some authors, the bases of trust are plainly rational. I trust you because I have analysed your behaviour and have concluded, on the basis of that evidence, that you are a trustworthy person. Again, it is important not to confuse between rational trust and cooperation and interest. In rational trust, the reason for trusting you is disconnected from my interest. I only trust you because you have said a thousand times that you would arrive home at midnight, and you have arrived home at midnight.

For other authors, trust has instead a moral component. I trust you because I think it is the right thing to do. In this case, trust is part and parcel of a wider set of principles and beliefs that are constitutive of what we could call the "moral personae." I tend to trust people because my ethics and my morals or my religious beliefs tell me to do so. I tend to trust people because I have a vision of the world in which this would be the right thing to do. Therefore, I do not trust you because I have observed your behaviour and have seen that you tend to honour your commitments, but because I have that moral predisposition to do so. Moral trust plays an important role above all in structures in which the information about the other person is lacking; or, to put it in a different way, in sequential games, in the first move. It is also important to note that rational trust and moral trust are not antagonist concepts: a person holding a moral vision of trust can distrust someone else if she sees that the other one is an untrustworthy person. It is therefore more realistic to think that both types of trust are supplementary. We could say that the person that has a moral vision of trust would need less rational trust and the person that does not hold a moral vision of trust may need more rational trust to trust. The point is here more analytical than normative: to understand why a person trusts, and the extent that she does it, it is important to try to understand what definition of trust she holds.[13]

V. SURVEYS AND LABORATORY EXPERIMENTS

We gather evidence about trust (whether people tend to trust or not, the extent that they do it, how they do it, etc.) through two basic methods: surveys on trust and social experiments. There are a number of surveys that ask about trust, for example, the World Values Survey (WVS).[14] This survey has been asking about trust for at least the last 25 years. In general terms, it can be said that these (and other) surveys differentiate between two basic kinds of trust: interpersonal trust and institutional trust. Interpersonal trust is trust in other people, whereas institutional trust is trust in particular institutions, like the parliament, the government, the political parties, or the judiciary. Additional surveys on trust in AI devices are starting to emerge.

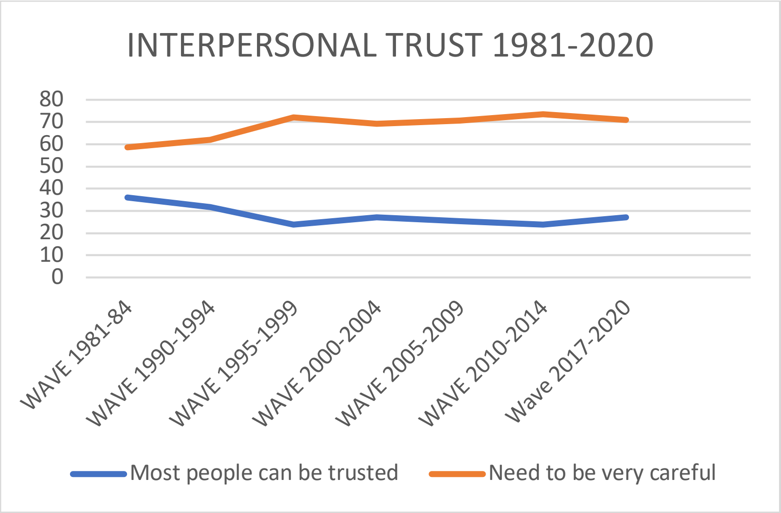
Laboratory experiments are another way to obtain evidence about

[13] HARDIN, *supra* note 11.

[14] Ronald Inglehart, et al., World Values Survey: All Rounds – Country Pooled Datafile, https://www.worldvaluessurvey.org/WVSContents.jsp (Last visited Jan. 19, 2023).

trust. These are game experiments that help to refine many of the hypotheses that we have about trust. They also give a more realistic, micro-funded, and dynamic picture of how trust structures work in practice. Ideally, surveys should be mixed with laboratory experiments to have a more fine-grained perception of how trust works. My problem with some laboratory experiments on trust is that, in many cases, it confuses trust and cooperation.[15]

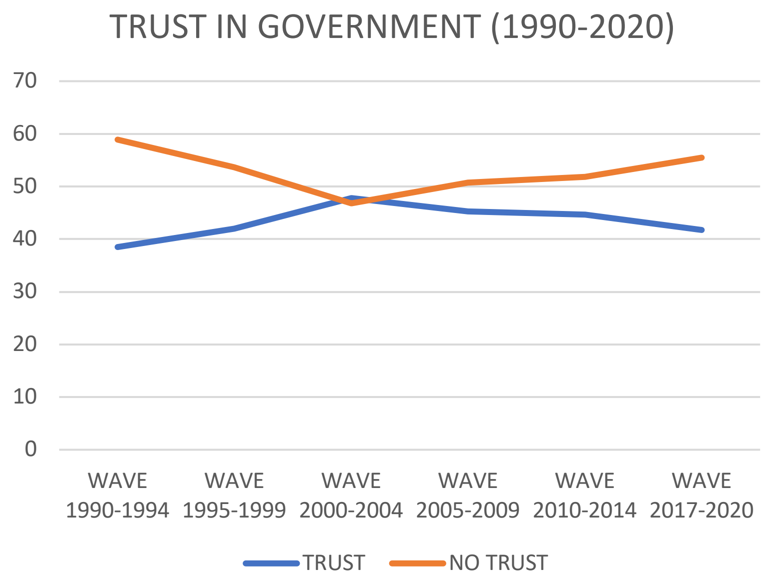
Unfortunately, departure surveys on trust do not provide good news. These surveys show that trust, both interpersonal and institutional, are being depleted all over the globe. Some authors even speak of a "cascade of trust destruction" that could haunt the world.[16] For example, Graph 1 shows the evolution of interpersonal trust from 1981 to 2020. We can easily see that levels of interpersonal trust have not been particularly high across the globe in this time series. We may, however, observe a certain amelioration of this trend since the wave of 2010-2014 to the wave of 2017-2020 (-4 percentage points). However, the difference between those who think that most people can be trusted and those who think that one needs to be very careful is more than 40 percentage points.



Graph 1: Interpersonal Trust (1981-2020)[17]

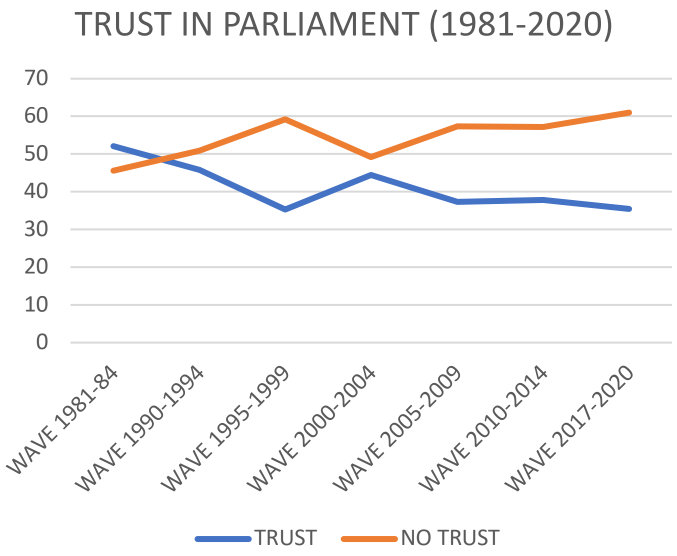
[15] HARDIN, *supra* note 11.
[16] SCIENCESPO, *supra* note 2.
[17] Inglehart, *supra* note 13.

If we turn now to institutional trust, things are not much better either. For example, in Graph 2, the WVS asks people whether they trust their governments or not. The people have been ambivalent across the years, but since the wave of 2000-2004 the trend is clear: mistrust in governments is skyrocketing around the world. The same is true for trust in parliaments. Graph 3 shows that except for the wave of 1981-1984, mistrust in parliaments has been the rule and, once again, since the wave of 2000-2004, it has been growing steadily. The apparent exception to this trend would be the courts. As shown in Graph 4, we observe a reversal of the previous trend of mistrust in this institution after the wave of 1995-1999. Since then, more people seem to trust the courts than not. This finding (why people trust courts and not the other two branches of government: executives and parliaments) is still open for explanation. However, in general terms, we may conclude that both interpersonal and institutional trust are probably at their lowest. This poses problems for our understanding of democracy and for the functioning of the economy, as many analysts have already remarked.

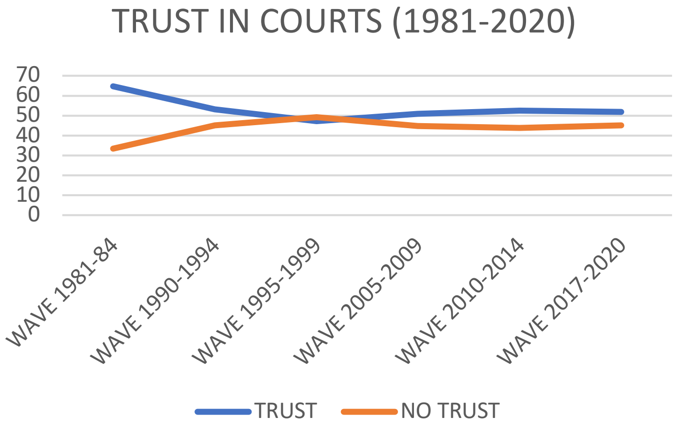


Graph 2: Trust in Government (1990-2020)[18]

[18] Inglehart, *supra* note 13.



Graph 3: Trust in Parliament (1981-2020)[19]



Graph 4: Trust in Courts (1981-2020).[20]

[19] *Id.*

VI. TRUST IN AI

This is the context in which we should analyse and understand the issue of trust in AI. The question is: if people mistrust other people and the most basic democratic institutions around the globe, why should they trust AI? After all, AI devices are made by humans, not by other machines. Perhaps we should therefore expect that this current wave of mistrust would be replicated in AI.

There are already some (partial) surveys on this matter. All of them point to the same result: in general terms, people tend not to trust AI. In the context of the general mistrust wave that has been previously analysed, this should come as no surprise. For example, Klynveld Peat Marwick Goerdeler (KPMG) conducted a survey on trust and AI in 2021.[21] The surveyed countries were the United States (US), Canada, Germany, United Kingdom (UK), and Australia— five of the most important economies of the world. The outcome of this survey is dismaying for the prospects of AI. In effect, as shown in Figure 2, only 28% of the sample would be willing to trust in AI, the highest being Australia (32%) and the lowest being the UK (26%). This survey also asks about trust in AI healthcare devices, which presents somewhat better results: 37% of the sample would be willing to trust healthcare AI. According to this report, trust, or rather lack thereof, in AI is influenced by four major causes: beliefs in the capacity of the regulatory system to make AI use safe; beliefs in the perceived impact of AI in jobs; familiarity and understanding of AI; and beliefs on the uncertain impact of AI on society. Of the four causes, regulation is clearly the strongest driver. This means that if people believed that the AI regulation in place was adequate, then they would have a better opinion on the other three items. In other words, at least according to this survey, the EU Commission has all reasons to focus on the regulation of AI as a way to enhance trust in AI.

[20] Inglehart, *supra* note 13.

[21] Nicole Gillespie et al., *Trust in Artificial Intelligence: A Five Country Study*, KPMG (March 2021) https://home.kpmg/de/en/home/insights/2021/06/artificial-intelligence-five-country-study.html.

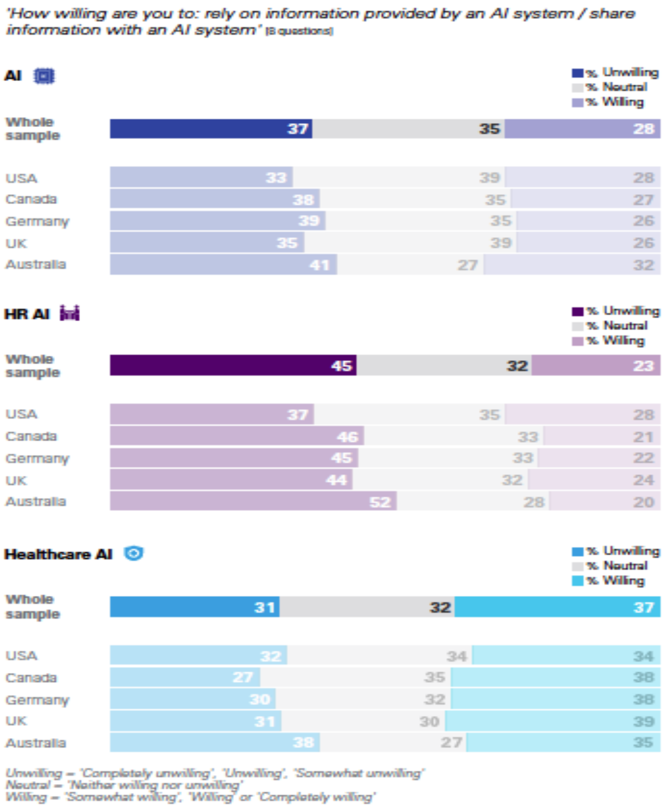


Figure 2: KPMG survey on Trust in AI[22]

In turn, a survey conducted by Ipsos in 2022 nicely complements the previous KPMG survey.[23] According to the Ipsos survey, only 50% of the sample would "trust companies that use artificial intelligence as much as they would trust other companies." Asked about the benefits and drawbacks of using AI, the percentage of people who think the benefits outweigh the drawbacks are the following: UK 38%; Australia 37%; Germany 37%; USA 35%; and Canada 32%. This means that for a vast majority of the people in these five important economies of the world, the drawbacks of using AI devices are much higher than the benefits. It is possible to think these findings correlate with trust in AI.

[22] *Id.*

[23] *See* Nicolas Boyon, *Global Opinions and Expectations about Artificial Intelligence*, IPSOS (January 5, 2022) https://www.ipsos.com/en/global-opinions-about-ai-january-2022.

Choung, David, and Ross have explored the issue of the impact of trust in AI voice assistants like Siri, Alexa, and Google Assistant. Voice assistants have been designed to be more human-like and therefore more trusted devices.[24] The previous authors develop two studies in their paper. In the first study, they find that trust is key to build positive attitudes towards AI voice assistants. Therefore, "people are inclined to regard a technology beneficial if they trust it." In contrast, "a lack of trust could raise concerns about the potential threats and risks of the technology instead of its benefits." The paper finds that in using AI voice assistants, ease of use and perceived usefulness are better predictors than trust. However, they also point out that trust can influence the other two factors.

In the second study, the impact of the "human-like dimension of trust" and the "functionality dimension of trust" on AI is tested. The human-like dimension of trust is to attribute human characteristics to AI (like the social and cultural values of the algorithms, for example, or human physical characteristics, as in robots). The functional dimension of trust is that AI works properly. According to this second study, the two factors are significant in predicting the perceived usefulness and positive attitude towards smart technology, which in turn predicts greater usage intention.

In turn, Lockey proposes with particular clarity the problems derived from some of the perversions of trust, like blind trust, or blind faith, as they call it.[25] "A foundational tenet of trust theory is that . . . it should be based on "good reasons"; trusting with no good reasons is no trust at all." These authors analyse five challenges that are directly related to trust in AI: 1) transparency and explain-ability; 2) accuracy and reliability; 3) automatism versus augmentation; 4) anthropomorphism and embodiment; and 5) mass data extraction. These authors find that the enhancement of all these factors increases trust, but with some qualifications. For example, over-explaining, in particular contexts like in AI assessment grading tools, may serve to decrease trust. Another finding is that accuracy is not enough. In some contexts, like in large, street-based games, the perception of accuracy may be as important as accuracy itself. Further, the issue of automatization versus augmentation is particularly relevant for AI in healthcare. In a series of experiments that are reported by the authors, it was found that people tend to trust less automated advices in healthcare than augmented ones (that is, advices that are made by AI

[24] *See generally* Hyesun Choung et al., *Trust in AI and its Role in the Acceptance of AI Technologies*, INT'L J. OF HUM.-COMPUT. INTERACTION (forthcoming March 2022).

[25] *See* Steven Lockey et al., *A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities, and Future Directions*, 54 L. HAW. INT'L CONF. SYS. SCI. 5464–67 (2022).

but supported by a human physician). The "human in the loop" approach is also preferred in the field of financial services. In turn, anthropomorphism is seen by the authors as a double-edge sword; in principle it increments trust, but it can also develop into over-trust. For example, the authors report that a study on an anthropomorphic health-care robot was perceived as less trustworthy than a machine-like robot. Finally, in regards to mass data extraction, the issue of privacy, and the use of data when using AI clearly impacts trust in AI; however, the authors report that more empirical work needs to be done in this area.

Winfield and Jirotka explore, at a more theoretical level, the connection between ethics and trust. According to the authors, a more inclusive, transparent, and agile form of governance would serve to build and maintain public trust in AI and ensure that AI is developed for the common interest. In this connection, these authors make a number of recommendations that range from publishing ethical codes of conduct and providing ethics training for all and being transparent about ethical governance of AI.[26]

In turn, Afroogh highlights, in his analysis on trust and AI, that mistrust in AI is a crucial barrier for its development.[27] According to this author, "any future development, implementation and usage of AI are tightly related to the public trust and supportive stance."[28] He therefore proposes a probabilistic theory of trust, the core of which is the distinction between four kinds of situations: an AI agent's trust in another AI agent; a human agent's trust in an AI agent; an AI agent‘s trust in a human agent; and an AI agent's trust in an object. His probabilistic theory would be formulated as follows: "A (including a human agent, AI, etc.) trusts B (including a human agent, AI Intelligence, etc.) or A believes that B is trustworthy only if there is a high degree of imprecise probability that B represents the proper functions or competence in nearby possible worlds." According to Afroogh, his formulation would integrate the four kinds of situations that I have mentioned before.

In sum, the previous review of surveys on trust and of the most recent academic literature that deals with this evidence points at two directions: the first one is that in general terms, people tend to distrust AI. The second is that it is possible to think that trust in humans or institutions is probably a different phenomenon from trust in intelligent

[26] Alan F.T. Winfield & Marina Jirotka, *Ethical Governance is Essential to Building Trust in Robotics and Artificial Intelligence*, PHIL. TRANS. R. SOC'Y A., Aug. 21, 2018, at 1, 1, 10 (2018).

[27] Saleh Afroogh, *A Probabilistic Theory of Trust Concerning Artificial Intelligence: Can Intelligent Robots Trust Humans?*, AI & ETHICS, June 2, 2022, at 1, 1, 13–14.

[28] *Id.*

machines. Maybe our theoretical understanding of trust should be further refined or even reformulated to integrate structures in which humans try to trust in AI. Or maybe it is not only a matter of qualification or even reformulation: perhaps when we speak about trust in machines, even if they are intelligent machines, we would be actually thinking in a different situation. For example, Afroogh differentiates between trust and reliance, and questions whether we should speak more of reliance in AI machines than in trust in AI machines. On the other hand, it is clear that we project the idea of trust to agents that are non-human (like institutions), and we think that this is not a contentious issue. Maybe when we say that we trust an institution, what we are implying is that we have trust in the persons that compose that institution. For the same token, maybe when we say that we trust an AI healthcare device, what we are implying is that we trust the humans that fabricated it and that are behind the machine. The whole thing would of course get much more complicated when, and if, AI machines become completely independent from humans.

VII. THE EUROPEAN COMMISSION'S PROPOSALS FOR ENHANCING TRUST IN AI

Concerned with the problems of trust in AI, the European Commission proposed a new regulatory framework that attempts to mitigate the detected problem of mistrust in AI devices. To this end, the European Commission issued a White Paper in February 2020, "On Artificial Intelligence-A European Approach to excellence and Trust."[29] This White Paper was followed by a Commission Communication of March 2021, "Fostering a European Approach to Artificial Intelligence,"[30] which was published together with a proposal for a regulation "laying down harmonised rules in Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts" of March 2021.[31] We shall review these three documents in the

[29] See generally Commission White Paper on Artificial Intelligence: A European Approach to Excellence and Trust, https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (Feb. 19, 2022). See also White Paper on Artificial Intelligence: A European Approach, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12270-White-Paper-on-Artificial-Intelligence-a-European-Approach/public-consultation_en (June 14, 2020).

[30] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Fostering a European approach to Artificial Intelligence. Brussels, 21.4.2021 COM(2021) 205 final.

[31] Commission Proposal, *supra* note 1.

following subsections.

A. The Commission's White Paper on Artificial Intelligence and Trust

The point of departure of the European Commission's White Paper is the assumption that trust is a prerequisite for the uptake of AI. Accordingly, the White Paper presents a number of policy options to enhance trust in AI. One idea is floating over the whole document: this is the notion of creating an "ecosystem of trust." For the creation of an ecosystem of trust in the field of AI, the White Paper tries to identify the main risks that may yield problems of trust in the domain of artificial intelligence. We shall see later on that the whole European Commission's regulatory framework in this field pivots around the idea of risks for trust. To deal with these risks, the Commission proposes to adapt the existing EU legislation on product safety and liability to the requirements of building trust in AI. A second proposal is to adopt a specific regulatory framework in the field of AI.

The White Paper was open for public consultation and comments since its publication in February 2020 until May 2020. The result of this public consultation was 1216 comments, which mainly came from citizens (30%), undertakings (18%), and academic institutions (12%). The rest of the comments had a diverse origin (entrepreneurial associations, NGOs, public administrations, and extra-European Union citizens).[32]

B. The European Commission's Communication on AI

In the European Commission's Communication on AI, the Commission announced that it is proposing a regulatory framework on trust in AI, and it also explained the main philosophy behind this new regulatory framework. As said before, the whole European Commission's edifice in this field revolves around the idea of risks. The Commission indicates that there are three kinds of risks: risks that are considered to be unacceptable, and therefore, are banned; high risks, that are to be highly regulated; and other (minor) risks that have a more lenient regulation. For example, the use of AI to contravene the European Union's values and violate its fundamental rights are to be banned. A particular case that the Commission mentions in its communication is that of remote biometric identification systems. An

[32] *Commission White Paper*, *supra* note 29.

example would be the real time use of AI for law-enforcement purposes, which would in principle be prohibited, unless when exceptionally authorised by law. This authorisation would be subject to specific safeguards.

In regards to high-risks, the European Commission specifically mentions the example of AI systems intended to be used to recruit people or to evaluate their creditworthiness and also the case of judicial decision making. These high-risks AI systems would not be prohibited but would be subject to the fulfilment of strict requirements and obligations. Finally, regarding the other (minor) risks, the uses of AI would be subject to the compliance of minimal transparency requirements. The European Commission cites, in particular, the examples of chatbots, emotion recognition systems, and deep fakes as examples belonging to the category of "minor risk." The European Commission summarizes its regulatory approach on trust and AI as "enabling trust without preventing innovation."[33]

C. The European Commission's proposal for a regulation on Artificial Intelligence

One of the declared fundamental aims (but of course, not the only one) of the European Commission's proposal for a regulation in the field of AI[34] is to mitigate the problems of mistrust in AI that the Commission, and other stakeholders in this area, have well identified. In this area, and as has been said before, this long proposal for a regulation (more than eighty articles) pivots around the notion of risks associated with the problem of trust in AI. The three categories are unacceptable risks, high risks, and other risks.

Article 5 of the proposal sets up a list of "prohibited AI practices." The idea is, therefore, not to ban specific types of AI but the use of specific types of AI. The prohibited practices are the following:

> - The use of AI systems that distorts a person's behaviour in a manner that causes, or is likely to cause, physical or psychological harm to humans.
>
> - The use of AI systems exploiting vulnerabilities of a specific group of persons linked to age and disabilities, so that AI materially distorts the behaviour

[33] Communication from the Commission, *supra* note 30.

[34] Commission Proposal, *supra* note 1.

> of one person belonging to that group in a way that produces or is likely to produce harm.
>
> - The use of AI systems for the evaluation or classification of the trustworthiness of natural persons, so that the outcome is detrimental for those persons.
>
> - The use of "real time" remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, unless and as far as such use is strictly necessary for a number of objectives which are related to individual and collective security threats.

Of the four cases, the hardest one is the last case, since it admits exceptions. According to the proposal, the exception is subject to a prior authorisation which shall be granted by either a judicial authority or an independent administrative authority of the Member State in which the use is to take place. It is for the Member States to develop the precise procedural requisites that are to be applied to authorisations, within the limits set by the proposal. These limits are, in essence, the following: the authorising agency has to take into account the seriousness, probability, and scale of the potential harm in the absence of the use of the AI system; and it has to take into account what consequences would be derived from the use of such system in terms of fundamental rights and freedoms.

In turn, Title III of the proposal (articles 6 to 51) is the lion's share of the regulation. It regulates the so-called "high-risks." The structure of this Title is the following. It is divided into five chapters, which deal with the following issues: classification of AI systems as high-risk (Chapter 1); requirements for high-risks AI systems (Chapter 2); obligations of providers and users of high-risk AI systems and other parties (Chapter 3); notifying authorities and notified bodies (Chapter 4); and standards, conformity assessment, certificates and registration (Chapter 5).

Chapter 1 defines what is to be considered as an AI system use that has a high-risk. To be considered as having a high-risk, the AI system use has to fulfil two conditions: first, that the AI system is used as a safety component of a product, or is itself a product, listed in Annex II of the proposal; and secondly, that the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a conformity assessment according to Annex II of the proposal. Further, all the AI systems listed in Annex III of the proposal are considered to be of high-risk. Therefore, the proposal remits us to Annexes II and III of the regulation. Annex II includes the list of "European Union harmonised legislation based on the new AI Legislative Framework" as well as the list of "other European Union harmonised legislation." This is a list of directives and regulations relating to the field of AI. One should therefore take a look at each and every one of these pieces of legislation to try to make sense of which AI uses are considered high-risk. It would have been more transparent to extract those uses from the previous legislation and incorporate them in an annex. Instead, Annex III includes a proper list of the AI uses that are considered to have a high risk. Some examples are: biometric identification and categorisation of persons, management and operation of critical infrastructure, education and vocational training, and administration of justice and democratic processes. For example, regarding the latter, all AI systems that are intended to assist a judicial authority in researching, interpreting facts, and applying the law to a concrete set of facts would be considered high-risk.[35]

Chapter 2 regulates the requirements for high-risk AI systems. These requirements have to be complied with by users and providers of AI systems (depending on the requirement). Some of these requirements are the following: to establish, implement, document and maintain risk management systems for AI; to set up training, validation and testing of data sets for AI systems; to draw up technical documentation for high-risk AI systems; to design AI systems with the capability of automatic recording of events; to design and develop high-risk AI systems in such a way as to ensure that their operation is sufficiently transparent, to enable users to use the AI systems in a proper way; and to design and develop high risk AI systems in such a way that they are accurate.

Let me draw the readers' attention to the requirement that is established in Article 14 of the proposal. This requirement is human oversight. According to this article, high-risk AI systems have to be designed and developed in such a way that they allow for effective human oversight. The key obligation established in this article is found in letter "e", of paragraph 4 of the commented Article 14. It reads as follows: "[H]umans overseeing a high-risk AI system must] be able to

[35] *See* Invertia: El Español, Justice Awards Telefónica a Project that will allow 3,000 Judges to Issue Sentences with Voice and Artificial Intelligence, https://www.elespanol.com/invertia/empresas/tecnologia/20220606/justicia-adjudica-telefonica-permitira-sentencias-inteligencia-artificial/677432697_0.html (Last visited 7 June 2022) (the Spanish online daily "El Español" reports that the Spanish Department of Justice has granted a contract to Telefonica, one of the most important Spanish telecommunications companies, to assist judges to write judicial decisions with the help of Artificial Intelligence. This would be a case that would probably fall in Annex III.)

intervene on the operation of the high-risk AI system or ***interrupt the system through a "stop" button or a similar procedure.***" (emphasis mine). This means that behind any high-risk AI system there must be, at the end of the day, a person. This crucial point has more implications than it initially seems regarding our understanding of the relationships between trust and AI, especially when the AI systems are considered to potentially yield high risks. It means that when we are speaking of trust in AI in reality we are speaking of trust in the person that is behind the AI system. As the proposal for regulation clearly contemplates, the problem is when intelligent machines acquire complete independence from humans. This fundamental point of the whole European Commission's regulatory edifice on AI is discussed in section 8.

In turn, chapter 3 of the proposal establishes a number of obligations upon providers and users of high-risk AI systems. It is difficult to draw a line between requirements and obligations. The logic of the proposal seems to be that a requirement is a characteristic that the high-risk AI system must have, whereas obligations are imposed on persons, natural or legal. However, it is obvious that many requirements imply, be it indirectly, correlative obligations. In any case, the obligations set up by chapter 3 are the following: to ensure that the requirements that have been seen before are complied with; to have a quality management system in place; to draw-up technical documentation of the high-risk AI system; to ensure that the high-risk AI system undergoes the relevant conformity assessment procedure, prior to its marketing; to comply with a number of registration obligations; to take the necessary corrective measures; to inform the national competent authorities of the Member States of the non-compliance and the corrective measures that have been adopted; to affix the CE marking to the high-risk AI systems; and to show the high-risk AI system's conformity with the requirements previously seen.

Once again, the human factor is the key in the domain of these obligations. In effect, article 29(2) says that "the obligations . . . are without prejudice of the user obligations under European Union or national law . . . for the purpose of implementing the human oversight measures indicated by the provider.". Therefore, it is for the provider to set up the "stop button" and for the user to press it if there is a need. On the one hand, this is a very good illustration of what has been indicated before: requirements include implicit obligations. Here, the obligation is for the provider to establish a disconnection system in the high-risk AI system. In turn, the chapter on obligations, chapter 3, specifies that it is for the user to disconnect the intelligent machine.

Chapter 4 makes a difference between notifying authorities and notified bodies. The idea is that competent national authorities, which the Member States must designate, have to extend accreditations for bodies that evaluate the conformity with the proposal's requirements of high-risk AI systems. The proposal for regulation establishes a clear dividing line between national authorities and conformity bodies: the former must avoid conflict of interests with conformity assessment bodies, and they must ensure the impartiality and objectivity of the operations undertaken by the latter. Here it would have been important to explicitly prohibit any revolving door system between the two.

Finally, chapter 5 regulates the substance of the so-called "conformity assessment" for high-risk AI. As points 63 and 64 of the preamble of the proposal indicate, the idea here is two-fold: first, conformity assessments should be carried out according to the sectoral legislation relating to AI (for example, the Machinery Regulation). The second idea is that to minimize the economic impact of conformity assessments upon AI providers, AI Providers should carry out their own conformity assessments under their own responsibility as a general rule. The main exception is to be found in AI systems intended to be used for remote biometric identification of persons, for which a third-party conformity assessment is made compulsory by the proposal. For these third-party conformity assessments, notified bodies should be designated, as has been seen before. In turn, Annex VI of the proposal specifies what the conformity assessment procedure based on internal control is. In this procedure, the drawing up of a specific technical documentation is the key. The technical documentation to which this annex, and also article 11 of the proposal, refer is established in Annex IV of the proposal. Further, article 48 of the proposal establishes the obligation for the provider to draw up an "EU declaration of conformity." The content of such declaration is established in Annex V of the proposal. Additionally, article 49 establishes that the Conformité Européenne (CE) shall be affixed visibly, legibly, and indelibly for high-risk AI systems.[36] Finally, article 51 specifies that providers of high-risk AI systems of article 6(2) (those listed in Annex III of the proposal) shall register the information established in Annex VIII of the proposal in the EU database that is established in article 60 of the proposal. In sum, the conformity system that is established by the proposal does not differ much from other products' conformity assessment systems, although it has some exceptions.[37] The system is therefore very much based on the individual responsibility of providers. Taking into account the specifics

[36] Commission Regulation, Product Requirements: CE Marking https://europa.eu/youreurope/business/productrequirements/labelsmarkings/cemarking/index_en.htm (Nov. 21, 2022) (providing the EU general conformity system described in general terms).

[37] *Id.*

of the AI field, this part of the proposal can be open to criticism.

Title IV of the proposal regulates "the other" uses of AI systems; that is, the AI systems' uses that do not belong to the two categories that have been reviewed so far (prohibited uses of AI systems and high-risk AI systems). This Title is composed of only one article, article 52, which simply establishes transparency obligations for providers of AI systems. The first obligation is that AI systems shall be designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obviously the case. The second obligation is that users of an emotion recognition system or a biometric categorisation system must inform of the system's operation to the natural persons that are exposed to them. A third obligation imposed on users as well is a disclosure obligation: users of AI systems that may yield so-called "deep fakes" must disclose that contents have been artificially generated or manipulated. Article 52 establishes a number of exceptions for each of these cases; it also establishes that it cannot affect the requirements of Title III, previously analysed.

VIII. ANALYSIS

The proposal for a regulation of AI tries to address the issue of trust in AI, among other objectives. We have seen in the previous section that the approach taken by the Commission is decremental: the Commission identifies uses of AI systems that generate so many risks that are prohibited; further, it identifies uses of AI systems that generate high risks, so as to need a conformity assessment plus other compliance requirements; and finally, it identifies uses of AI systems that are only made subject to transparency obligations.

As has been suggested in the previous section, the main requirement or obligation that the proposal establishes is that of human oversight. A human has to be behind the intelligent machine. She has to ensure that the intelligent machine is under control. She also has to insert mechanisms in the intelligent machine that allow the human to interrupt its operation, and she has the obligation to report malfunctioning. This is the so-called "human in the loop" approach.

From a theoretical perspective, the fact that AI systems have to rely on the "human in the loop" approach poses some fundamental questions relating to the relationship of trust and AI. In turn, this theoretical perspective has clear practical and legal implications. The theoretical problem that is posed by this approach is that, in reality, it is inconsequential to speak of trust in AI. The practical effect is that if it is not possible to speak of trust in AI systems, then the Commission's attempts to enhance trust through risk configuration will hardly be successful.

Let me start with the first point— the theoretical point. Trust structures are always, and I would add, only, conceived among human actors. When we speak of trust, we are speaking of an attitude, an emotion, or a rational expectation that can be only proclaimed for humans towards humans. We can speak of trust in humans because we know more or less how human rationality and behaviour work. Therefore, we can elaborate on expectations about humans' behaviour. Precisely the outliers of common rational behaviour are understood by social sciences as deviations from the rule— as pathological states of mind— that need a different treatment. For example, Elster's analysis on addictions[38] is important because addictions escape from our traditional understanding of common rationality. The problem is that we cannot apply this scheme to intelligent machines. We simply do not know how machines are going to behave when they acquire autonomy and independence from humans. We simply do not know how the intelligent machines' rationality (if we can speak of a machine's rationality) is going to evolve in the future. The issue of independence from humans is what is at stake here. Therefore, trust is projected from humans to humans since we know how humans act. However, it is impossible to project trust from humans to intelligent machines since we do not know how intelligent machines act. The matter is, as said before, not only practical, but above all theoretical. Assume that we would understand in the far future AI's rationality, but even in this case, it would not be a human rationality. Therefore, we could not speak of trust in AI systems even if this assumption was ever held.

Maybe a different perspective can help to clarify this theoretical point. This perspective is that of animals. As has been convincingly argued by some authors, animals are "sentient" beings.[39] This means that they are able to feel. This fact has been tested in laboratory experiments with animals and today is beyond any reasonable doubt. From this evidence, many authors have promoted an "animal rights' agenda," that has made headway in some states of the world. Now animals have more protection from humans than was the case a century ago. The point is that asserting that animals have sentiments is to indirectly say that animals have a kind of rationality. In effect, the most recent approaches in rationality argue that the divide between feelings, or emotions, and rationality, is plainly absurd: emotions are part and parcel of our rationality. Being the case, then animals have a certain kind of rationality. This rationality is a sort of human-downgraded

[38] JOHN ELSTER, STRONG FEELINGS: EMOTIONS, ADDICTION, AND HUMAN BEHAVIOR 190-91 (Francois Recanati ed. 1999).

[39] PETER SINGER, ANIMAL LIBERATION 47, 49 (2d ed. 1995).

rationality. This means that it is close to human rationality but it is, in general terms at least, much less developed than the human rationality. Can we therefore say that I can trust animals? Can we therefore say that I trust my dog, or my horse? The answer to this question would be positive. Yes, we can say that we can trust an animal, because animals have a kind of rationality that is very close to human rationality. It is a kind of rationality that we more or less understand and more or less can control. Of course, this type of trust will be much more nuanced and qualified than trust in humans. But the fact that animal rationality is close to human rationality makes it possible to say that we can trust an animal.

This is not the case for machines. Machine learning is the problem in this domain. According to many analysts, once certain algorithms are in place, some intelligent machines learn in a way that we can simply not understand.[40] And this is not going to stop here: on the contrary, in the future, we are going to have many more instances in which this is going to be the case. If we do not know how machines learn, then we do not know about their rationality since learning is the main gateway to rationality. This means that we should discard any discussion of trust from humans towards AI systems. As Afroogh suggests, we might rather speak of reliance in AI systems instead of trust, for example.[41] In other words, the discussion on trust in AI seems misplaced from a theoretical perspective.

This has, as previously announced, clear practical implications. The main implication is that, as some authors argue, a good dose of mistrust in machines whose rationality we cannot comprehend would be a case in point. Said in other terms, for human rationality, mistrust in AI systems (I insist once again: in systems which rationality we cannot understand) is the appropriate outcome. Therefore, my proposal is to move to a different practical dimension and speak of other concepts that are closer to the world of machines. It is obvious that an AI system is a product different than other products. But from there it is difficult to give human qualities and characteristics, such as trust, to AI systems. In other words, we cannot simply speak of such a humane activity regarding machines as trust is. The idea, from a practical perspective, would therefore be to place this debate in the realm of reliability, accuracy, efficiency, correctness, technical competence, control, etc. In

[40] Bradley (2017), writing for Forbes, reports that "Facebook shut down an artificial intelligence engine after developers discovered that the AI had created its own unique language that humans can't understand". See the article here: https://www.forbes.com/sites/tonybradley/2017/07/31/facebook-ai-creates-its-own-language-in-creepy-preview-of-our-potential-future/?sh=139144ec292c

[41] Afroogh, *supra* note 26.

other terms, what I want from a machine, even from an intelligent machine, is that it works properly and accurately under my control. Therefore, it would be important to apply a healthy measure of mistrust towards it, above all if it has intellectual properties that may easily escape my control.

IX. CONCLUSIONS

I have analysed in this paper the European Commission's proposal for a regulation of AI systems, in particular the part of this proposal that is aiming at mitigating the problems of human mistrust in AI systems. This analysis cannot be done if it is made outside the context of a wider discussion on trust. This is why I started this paper by making a number or reflections on trust: its meaning, its importance, the attention that it is receiving today from the academic world. I defined trust as an expectation about whether or not my interlocutor will honour her commitments. Trust analyses only make sense within commitment structures, in which A commits to doing X, and then B trusts (or not) that A will honour her compromise. Commitment structures are two-person games, as are trust structures. I also underlined the importance of bearing in mind that both the trustor and the trustee (and not just the trustee) are important to understand how trust works. The trustee has to have trustworthiness, but trust is impossible when the trustor is uncapable of trusting, even in contexts in which the trustee's trustworthiness is Mahatma Gandhi-like.

All this projects a picture in which trust is understood, as a matter of both theory and practice, as a very human behaviour. This is why this paper's main argument is that it is odd to speak of trust in the context of machines, even if they are intelligent. Expectations of trust are based on a given common knowledge about how rationality works and what rationality is. This is something that we cannot and will not be able to predicate about intelligent machines.

The Commission's approach to solve the problems derived from mistrust in AI systems is based on the conception that the regulation of risks impacting trust will be enough to at least mitigate the current wave of mistrust that society has towards AI systems. Once these risks are properly regulated, then the outcome should be an enhancement of trust in intelligent machines. Therefore, the European Commission proposes to ban certain uses of AI systems, it configures a number of risks that the European Commission (and other stakeholders) think are too high and subjects them to strong regulation, and it conceives other risks that are only subject to a minor regulatory stretch. The European Commission's reasoning is rather linear and can be summarised in the following formula:

> **Mistrust because of AI risks → Identify, systematize and regulate AI risks → Trust as a result of enforcement of AI risks regulation**

In particular, the current proposal can be commented on from different perspectives. To start with, an explanation is lacking about the criterion or criteria laying behind the categorisation of risks. It is, for example, surprising that the risks of AI devices used for medical purposes are not contemplated in a specific way in the proposal. A second point has to do with the conformity assessment procedure. This procedure relies almost exclusively on the individual responsibility of the provider of the AI system. It is therefore a "private" conformity system assessment. To be sure, there are exceptions to the rule, but these exceptions are not as important as to trump the previous general rule. It is however unclear why high-risk AI systems are not subject to a third-party conformity assessment which would be in turn supervised by public administrations. Taking into account the specifics of AI devices, the treatment for high-risk AI systems should be completely different from the regulatory treatment that is currently given to regular products. A final, but not less important, point is that the difference between requisites and obligations is unclear. Some requisites at least imply obligations for the user and the provider of AI systems. This is not to say that a difference between requisites and obligations makes no sense; it only means that the demarking line between the two should be made clearer in the proposal.

The main argument of this paper can be summarised by saying that the current European Commission's proposal for the regulation of AI probably rests upon a misconception. Trust in AI systems will be impossible to achieve for the reasons that have been pointed out before—we do not understand AI systems' rationality and, what is possibly more important, the kind of rationality that AI systems will develop will not be similar to human-like rationality. In this context, it makes little sense to speak of trust in AI devices. Also, it makes no sense to try to regulate risks for trust as a way to solve this problem. In fact, a good deal of structural mistrust in AI systems might be beneficial for all.

CITED BIBLIOGRAPHY

Saleh Afroogh, *A Probabilistic Theory of Trust Concerning Artificial Intelligence: Can Intelligent Robots Trust Humans?*, AI & ETHICS, June 2, 2022, at 1, 1, 13–14.

Yann Algan, *Trust and Social Capital*, *in* FOR GOOD MEASURE: ADVANCING RESEARCH ON WELL-BEING METRICS BEYOND GDP 283, 286, 302 (Joseph E. Stiglitz, Jean-Paul Fitoussi, & Martine Durand eds. 2018).

Yann Algan & Pierre Cahuc, *Trust, Growth and Well-being: New Evidence and Policy Implications*, 2 HANDBOOK OF ECON. GROWTH (forthcoming June 2013).

Tony Bradley, *Facebook AI Creates its Own Language in Creepy Preview of Our Potential Future.*, FORBES (July 31, 2017, 11:20 AM), *https://www.forbes.com/sites/tonybradley/2017/07/31/facebook-ai-creates-its-own-language-in-creepy-preview-of-our-potential-future/?sh=79504d9f292c.*

Hyesun Choung et al., *Trust in AI and its Role in the Acceptance of AI Technologies*, INT'L J. OF HUM.-COMPUT. INTERACTION (forthcoming March 2022).

JOHN ELSTER, STRONG FEELINGS: EMOTION, ADDICTION, AND HUMAN BEHAVIOR 190–91 (Francois Recanati ed. 1999).

See generally ANTONIO ESTELLA, LEGAL FOUNDATIONS OF EU ECONOMIC GOVERNANCE (2018).

See generally RUSSELL HARDIN, TRUST 46 (2006).

Steven Lockey et al., *A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities, and Future Directions*, 54 L. HAW. INT'L CONF. SYS. SCI. 5464–67 (2022).

ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY 144–45 (2000).

ROBERT D. PUTNAM ET AL., MAKING DEMOCRACY WORK: CIVIC TRADITIONS IN MODERN ITALY 169–70 (1993).

PETER SINGER, ANIMAL LIBERATION 47, 49 (2d ed. 1995).

Alan F.T. Winfield & Marina Jirotka, *Ethical Governance is Essential to Building Trust in Robotics and Artificial Intelligence*, PHIL. TRANS. R. SOC'Y A., Aug. 21, 2018, at 1, 1, 10 (2018).

Artificial Intelligence in Government: Risks and Challenges of Algorithmic Governance in the Administrative State

José Vida Fernández*

Abstract

This article analyzes the legal implications of using artificial intelligence in government and how it is challenging the foundations of the administrative state. It begins by demonstrating that a new model of government is emerging, based on information and intelligence (i-Gov). To understand the nature and scope of this new i-Gov model, this article will explain what artificial intelligence really is and analyze the applications that are currently being carried out in the US and the EU. Next, it will review the regulatory framework that is emerging that regulates government use of artificial intelligence in both the US and the EU. Finally, the article concludes by identifying and analyzing the main legal and policy problems involved in the use of artificial intelligence in government. It challenges values, principles, and institutions of the traditional administrative state and also requires us to think of new frameworks for constitutional and administrative law to guarantee citizens' rights and public interest.

I. Introduction

Artificial intelligence (AI) is set to transform human life in all its dimensions. Although this may sound somewhat exaggerated and disturbing, it is a process that has already happened with other disruptive technologies. That is the case with the development and

*Professor of Administrative Law, Public Law Department, Universidad Carlos III de Madrid. This work is part of the research projects "Artificial intelligence in the national health care system: solutions to specific legal problems" (PID2021-128621NB-I00) and "The Impact of Artificial Intelligence in Public Services: A Legal Analysis of its Scope and Consequences in Healthcare" (PGC2018-098243-B-I00) 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)

expansion of the internet since the beginning of the century, which has brought about major changes in our economies, societies, politics, and personal lives.

Nevertheless, there is something different in this new disruptive innovation that has led to an obsession with AI. Although the impact of AI is not yet widespread and overt, there is a worldwide debate on how it will change our work, health, education, entertainment, personal relationships, and many other aspects of our lives. Yet, the key point of this debate is not about the timing or intensity of this transformation, but on how AI may transform human nature and its role in our lives. AI does not involve a transformation in the sense of how we carry out our activities (economic, social, personal) with the removal of physical constraints (distances, storage, etc.) as the internet does, but rather AI affects how the end product is achieved since it performs the activities by replacing the human factor.

Concern about this technology has led to an increasing number of studies on the legal implications of AI systems. However, most of this analysis focuses on the legal consequences of AI in the private sector and how it impacts individuals (companies, families, or citizens) and their rights (privacy, competition, intellectual property, work conditions, liability). In contrast, few studies focus on the application of AI in the public sector, particularly in government functions, and how it impacts the exercise of public power and citizens' rights.

This paper analyzes the legal implications of using AI in government from a general perspective, including all three branches of government, but it focuses on administrative activities, from government decisions to service provisions. Although each country faces digital transformation in line with their own constitutional and administrative tradition, there are some common challenges related to the use of AI in the public sector which can be considered as global issues. Therefore, the purpose of this piece of work is to identify these global challenges by carrying out a comparative analysis of the United States (US) and the European Union (EU).

The premise is the change that is taking place in the digitalization of government, which has gone from an online government (e-Gov) to an information-intelligent government (i-Gov), changing its nature and characteristics (Section II). In order to understand the meaning and scope of this change that leads to an i-Gov, AI is analyzed from a legal perspective as this disruptive technology must be understood in order to capture the legal implications of the change that it produces (Section III). It is also necessary to explore how AI is being used in government action to understand the real implications that it may have, without relying on science fiction scenarios (Section IV). Once we determine what AI is and how it is used in government, it is necessary to identify the legal framework that applies to it, which is emerging in both the US and the EU in new ways (Section V). To conclude, the analysis of the legal implications of AI use in government highlights the inadequacies of the current administrative state as it cannot adequately handle the many challenges that arise. These inadequacies require new tools and strategies to guarantee constitutional rights and values (Section VI).

In any case, it should be noted that this work does not attempt to find solutions to the challenges for AI use in government at this time, as they are not well-defined enough. The aim is to offer a good diagnosis by identifying and understanding these challenges and their context, as a preliminary step in the search for solutions that can save the administrative state as we know it today.

II. A NEW ERA IN GOVERNMENT DIGITALIZATION: FROM E-GOV TO I-GOV

The Digital Revolution is reshaping our world, and it affects both the private and public sector. The impact of new information technology is well known in its private dimension, and although technological changes are quickly digested, we can still marvel at the transformation in our economy (with a digital global market without limits or distances, both for companies and individuals), our learning and entertainment (accessing an endless amount of content on different platforms), our personal relationships (reaching whoever we want and interacting with millions of people through social networks), and so on.

Digital advances are also transforming government and, in particular, public service performance, although the achievements are much less spectacular and glamorous than in the private sector. The problem is that government has to follow many regulations that include restrictions and requirements, so it is not free to incorporate the technological innovations that corporations or individuals can. First, although many government activities are similar in substance to those of companies and other private entities (information management, decision-making, service provision, etc.), it implies the use of government power, therefore the incorporation of technological innovations must be previously validated and approved. In addition to this, there are a significant limitations when new digital solutions are acquired under public procurement rules, and there is also the challenge of training public civil servants and personnel.

A. Online Government: e-Gov

Despite all these difficulties, governments have not remained oblivious to technological change and have incorporated information technologies that are also transforming the venerable administrative state that is leading to a change of the model. However, digital transformation for government has so far been limited to so-called online government (e-Gov), which essentially consists of putting government online, as it is based in one specific technology, such as the internet, and its sole purpose is to enhance interaction by eliminating the spatial and temporal barriers that separate government from its citizens.[1] E-Gov is purely instrumental, but not substantial, as it is limited to considering interactions between the government and its citizens by streamlining information distribution and service provision, but without the ability to change the model or essence of government.[2]

This is the experience in the US, where the digital transformation of the federal government began at the turn of the century and has been limited to generalizing e-Gov services and processes promoting the use of the internet and other information technologies to provide increased opportunities for citizen participation. In particular, the E-Government Act of 2002 was passed to enhance citizen access to government information and services and improve government transparency and decision-making through the use of the internet.[3] The Office of E-Government and Information Technology was created to promote the use of internet-based technologies to make it easier for citizens and businesses to interact with the federal government, save taxpayer dollars, and streamline citizen participation.[4]

[1] This approach to e-Gov still prevails internationally as can be seen in the United Nations' e-Gov development index that is based primarily on the Online Services Index. *See* UNITED NATIONS, E-GOVERNMENT SURVEY 2022: THE FUTURE OF DIGITAL GOVERNMENT (2022). On the concept of e-Gov, *see* J. E. J. Prins, *Electronic Government. Variations on a Concept, in* DESIGNING E-GOVERNMENT. ON THE CROSSROADS OF TECHNOLOGICAL INNOVATION AND INSTITUTIONAL CHANGE, 1–5 (2001). *See also* Robert M. Davison et al., *From Government to E-government: A Transition Model*, 18 INFO. TECH. & PEOPLE 280–99 (2005).

[2] E-Gov has also been defined as an interaction of a "managerial" nature, which dominates over the "consultative" and "participatory" interaction models. *See* Andrew Chadwick & Christopher May, *Interaction Between States and Citizens in the Age of the Internet: "E-Government" in the United States, Britain, and the European Union*, 16 GOVERNANCE: INT'L J. POL'Y, ADMIN. & INST. 271 (2003).

[3] E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2902 (defining electronic government as the "use by the Government of web-based Internet applications and other information technologies, combined with processes that implement these technologies, to . . . enhance the access to and delivery of Government information and services to the public, other agencies, and other Government entities; or . . . bring about improvements in Government operations that may include effectiveness, efficiency, service quality, or transformation").

[4] For further information on Federal e-Gov strategy, *see* Office of the Federal Chief Information Officer, THE WHITE HOUSE (last visited Feb. 6, 2023), https://www.white

At the same time, the EU has been promoting e-government policies for its development in the member states since 2000.[5] It should be noted that the EU does not have the capacity to implement e-Gov, so its development across Europe has been fragmented at the national level. The EU can only support the actions of the member states; it cannot enforce how national agencies are organized or function. Therefore, the EU has promoted the expansion of e-Gov in member states through coordination and benchmarking actions,[6] under a model based on online access through the internet to eliminate distances and reduce time in government access.[7]

Therefore, as in the American and European models, IT has so far been used worldwide as a passive instrument in government, either to improve internal activities (computers, databases) or to facilitate interaction with citizens and to provide permanent access (online services). Thus, up to now, digital technologies have been just a means for governance, and have not been an instrument for administrative reform as the government's activity has remained essentially unchanged, even though it has developed through IT.[8]

B. Disruptive Technologies

However, major changes are underway as IT innovations are accelerating and leading to developments that are increasingly far-reaching and transformative in nature. Big data, cloud computing,

house.gov/omb/management/egov/; *see also* Rachel Silcock, *What is E-Government*, 54 PARLIAMENTARY AFFS. 88 (2001); John C. Reitz, *E-Government*, 54 AM. J. COMP. L. 733, 733 (2006); Shannon Howle Schelin, *E-Government: An Overview*, *in* G. DAVID GARSON, MODERN PUBLIC INFORMATION TECHNOLOGY SYSTEMS: ISSUES AND CHALLENGES 110, 113 (2007).

[5] *Communication from the Commission to the Council, the European Parliament, The European Economic and Social Committee and the Committee of the Regions – the Role of e-Government for Europe's Future,* at 7, COM (2003) 567 final (Sept. 26, 2003) (defining e-Government as "the use of information and communication technologies in public administrations combined with organisational change and new skills in order to improve public services and democratic processes and strengthen support to public policies").

[6] The successive European Union digital strategies have included EGOVERNMENT ACTION PLANS (2005–2011, 2011–2015, and 2016–2020) as a specific instrument to coordinate and pool the efforts of member states' e-Government strategies and activities.

[7] On the E.U. perspective of e-Gov, *see* Clara Centeno et al., *A Prospective View of e-Government in the European Union*, 3 ELEC. J. E-GOV'T 59, 62 (2005); PAUL G. NIXON & VASSILIKI N. KOUTRAKOU, E-GOVERNMENT IN EUROPE, RE-BOOTING THE STATE (2007).

[8] *See* Kenneth Kraemer & John Leslie King, *Information Technology and Administrative Reform: Will E-Government Be Different?* 2 INT'L J. ELEC. GOV. RSCH. 1 (2006) (arguing that IT has never been an instrument of administrative reform, rather, it has been used to reinforce existing administrative and political arrangements).

blockchain, and artificial intelligence have developed strongly in the last decade. These are disruptive technologies that will lead to great economic, social, and political transformation in the coming years.

Within all of these new IT advances, AI stands out to the point that there is real AI fever and excitement. The present relevance of AI is explained, on the one hand, by its huge technological development in recent years, and, on the other hand, by the extraordinary capabilities that it has acquired, making it the technology with the greatest capacity for transformation.[9]

Regarding the technological development of AI, it should be noted that although it is a technology that has been around since the middle of the last century, only in the last decade has there been a real push for three overlapping factors: firstly, advances in deep learning that make it possible to solve new problems; secondly, the explosion of big data, which, thanks to cloud computing, makes it possible to capture, store, share, and manage large amounts of higher quality data; and finally, the constant growth of computing power that allows AI to solve problems in less time.

With regards to transformative capacity, there is a general consensus that AI is permanently at the forefront of disruptive technologies because of its enormous disruptive capacity in all industries, including agriculture (productivity forecasting or autonomous tractors), to health care (refining diagnosis or discovering new drugs), and education (personalized learning, etc.).[10]

C. AI National Strategies

All countries have taken the thrust of AI seriously, and this can be seen in the numerous strategy memorandums and legislation on AI that have been adopted since the end of the last decade with different approaches.

In the case of the United States, the priority for plans and legislation adopted since 2016 has been to ensure continued US

[9] *See* DARRELL M. WEST & JOHN R. ALLEN, TURNING POINT: POLICYMAKING IN THE ERA OF ARTIFICIAL INTELLIGENCE (2020); *see also* KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021); ERIK J. LARSON, THE MYTH OF ARTIFICIAL INTELLIGENCE: WHY COMPUTERS CAN'T THINK THE WAY WE DO (2021).

[10] *See Disruptive Technologies: Advances That Will Transform Life, Business, and the Global Economy*, MCKINSEY DIGITAL 18, 35, 47, 58, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/disruptive-technologies (last visited Jan. 13, 2023). After ten years, AI still remains at the head of disruption in *McKinsey Technology Trends Outlook 2022*, at 22, https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-top-trends-in-tech (last visited Jan. 13, 2023).

leadership in the development and use of AI systems both for public and private sectors.[11] In 2020, Congress passed the National AI Initiative Act to coordinate a program within the federal government to accelerate AI research and its application for national economic prosperity and security. It also included the creation of the National AI Initiative Office to coordinate and support the National AI Initiative.[12] Thus, the US strategy has been devoted to having more and stronger AI systems driving innovation, so as not to lose momentum to other innovative countries like China. Special consideration has been given to AI in the public sector. The 2020 AI in Government Act was passed to facilitate, improve, and coordinate the adoption and use of AI within the federal government under the AI Center of Excellence program.[13]

The EU has also adopted a Europe-wide strategy that seeks a different kind of leadership for AI in Europe. This is demonstrated in the 2018 AI Commission Communication for Europe, the 2020 White Paper on AI, and the 2021 Communication Fostering a European Approach to AI.[14] The objective is to create an AI "made in Europe" that is distinguished by being trustworthy, secure, and ethical. For this purpose, a proposal for a regulation laying down harmonized rules on AI (the so-called AI Act) is in the pipeline.[15] Regarding the use of AI in the public sector, the EU has not prepared specific legislation or plans since

[11] The National Science and Technology Council prepared a strategic plan, NETWORKING & INFO. TECH. RSCH. & DEV. SUBCOMM., NAT'L SCI. & TECH. COUNCIL, NATIONAL AI RESEARCH AND DEVELOPMENT STRATEGIC PLAN (2016), that defined strategic priorities for AI R&D. Later, in 2019, it was signed into executive order, Exec. Order No. 13859, 84 Fed. Reg. 3967 (Feb. 14, 2019), and in 2020, Congress passed the National Artificial Intelligence Initiative Act of 2020, H.R. 6216, 116th Cong. (2020).

[12] *National Artificial Intelligence Initiative*, ai.gov (last visited Jan. 13, 2023).

[13] AI IN GOVERNMENT ACT OF 2020 was preceded by Exec. Order No. 13960, 85 Fed. Reg. 78939 (Dec. 8, 2020), which established principles for a common and expert use of AI within the federal government. The AI Center of Excellence was created as a program within the General Services Administration that operates within GSA Centers of Excellence (CoE). *See The Centers of Excellence*, IT MODERNIZATION CENTERS OF EXCELLENCE, https://coe.gsa.gov (last visited Jan. 13, 2023).

[14] The E.U. strategy on AI is included in the *Communication from the Commission on Artificial Intelligence for Europe*, COM (2018) 237 final (Apr. 25, 2018); *Artificial Intelligence – A European Approach to Excellence and Trust*, COM (2020) 65 final (Feb. 19, 2020); *Communication for the Commission on Fostering a European Approach to Artificial Intelligence*, COM (2021) 205 final (Apr. 21, 2021). Member states have also adopted their own national AI strategy, such as Germany, STRATEGIE KÜNSTLICHE INTELLIGENZ DER BUNDESREGIERUNG (November 2018), France, DONNER DU SENS À L'INTELLIGENCE ARTIFICIELLE: POUR UNE STRATÉGIE NATIONALE ET EUROPÉENNE (Mar. 2018) or Spain, ESTRATEGIA NACIONAL DE INTELIGENCIA ARTIFICIAL (Nov. 2020).

[15] *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. 21, 2021).

it has no competence in this area, so member states have their own strategies for implementing AI in government.[16]

D. Toward Information and Intelligence: i-Gov

As can be seen, the concern about the impact of AI worldwide is growing and affects the public sector. In fact, there is a belief that AI can be as transformative in the private sector as in the public. The use of AI-based tools in decision-making, adjudication, enforcement, and public services can take government digitalization to a new level beyond human decision-making limitations.[17]

AI can lead to many positive developments. It can help improve government processes and procedures; design and meet strategic goals; reduce costs and environmental impacts; combat fraud, waste, and abuse by enhancing oversight of public funds; increase efficiency and mission effectiveness; improve quality of services; improve safety; and support decision-making.

According to the characteristics of AI, digitalization will no longer be passive but active, as it affects administrative decision-making and the service delivery process. In many uses of AI, the technology ceases to be a mere instrument as it can assume the essence of government decisions; however, there is a risk that AI stops being a means, and that in reality, it becomes an end.

According to the transformative capacity, it is far from discussion that the use of AI initiates a new period in the digital transformation of government that may modify its nature. Therefore, we are leaving the e-Gov behind and entering the i-Gov era, a government based on flows of information and intelligence.[18] The range and relevance of this change is yet to be defined, as it will depend on how AI is incorporated into government action. In any case, it is necessary to reflect now on the possible challenges and the perils of this process because it is likely to transform the foundations and principles of the administrative state, as

[16] The national AI in government strategies in EU member states are usually included in broader strategies on government digitalization: in France, PUBLIC ACTION 2022; in Italy, THREE-YEAR PLAN FOR IT IN THE PUBLIC ADMINISTRATION; in Spain, PUBLIC ADMINISTRATION DIGITALIZATION PLAN 2021-2025.

[17] Physical limitations include memory capacity, fatigue, aging, impulse control, perceptual inaccuracies; biases include endowment effect, loss aversion, system neglect, hindsight bias, availability bias, confirmation bias, framing, anchoring, susceptibility to over persuasion and implicit racial and gender biases. CARY COGLIANESE, A FRAMEWORK FOR GOVERNMENTAL USE OF MACHINE LEARNING, 8–20 (2020).

[18] GOVERNANCE AND INFORMATION TECHNOLOGY: FROM ELECTRONIC GOVERNMENT TO INFORMATION GOVERNMENT (Viktor Mayer-Schonberger & David Lazer eds., 2007); *see also* CORIEN PRINS ET AL., IGOVERNMENT (2011).

discussed below.

III. WHAT REALLY IS ARTIFICIAL INTELLIGENCE? AN APPROACH FOR LAWYERS

To better understand the actual scope of this transition from e-Gov to i-Gov, it is necessary to have a good understanding of the technology that is causing it. There is a great deal of confusion surrounding AI, and it is common to believe that this technology can solve problems and do things that humans are not capable of.[19] Science fiction literature has anticipated problems regarding AI, and even has proposed the first solutions—as the Asimov's robotic laws[20]—but all this refers to an imaginary world more than an actual technology. Therefore, the scope of this section is devoted to explaining the technological grounds of AI, but in a way that is accessible to non-specialists.

A. Demystifying AI

It is important to start demystifying AI, affirming that AI is not intelligence, or more precisely, actual human intelligence. As it emulates human cognitive functions, it causes great confusion, even some people consider AI to have not only human skills but also qualities.[21] Without going that far, it is true that many people attribute human-like reasoning capabilities to AI systems, considering that they can carry out activities like any person, as AI is defined as programs with abilities that normally require human intelligence.[22]

In order to close the debate about AI and its human skills, it must be clear that the so-called general or strong AI that resembles human intelligence (developing general and abstract thinking to perform different tasks) has not yet been created and it will probably not be

[19] *See* ERIK J. LARSON, THE MYTH OF ARTIFICIAL INTELLIGENCE-WHY COMPUTERS CAN'T THINK THE WAY WE DO (2021).

[20] Isaac Asimov, *Runaround, in* I, ROBOT 27 (Gnome Press, 1950).

[21] *See* Nico Grant & Cade Metz, *Google Sidelines Engineer Who Claims Its A.I. Is Sentient*, N.Y. TIMES (June 12, 2022), https://www.nytimes.com/2022/06/12/technology /google-chatbot-ai-blake-lemoine.html (providing an example of an engineer who believed an AI showed lifelike qualities).

[22] *Artificial Intelligence*, OXFORD REFERENCE, https://www.oxfordreference.com/ view/10.1093/oi/authority.20110803095426960; Francesca Bigami, *Artificial Intelligence Accountability of Public Administration*, 70 AM. J. COMP. L. 312, 313–15 (2022) (discussing the definition of AI from a legal perspective).

created in the near future.[23] The AI that exists today has very specific functions, and it can solve only specific problems, so it can be merely applied to very specific tasks.[24] AI systems can perform these specific tasks with better results than humans—as it is the case of playing chess, photo recognition, or, in a near future, driving cars—and they are even able to solve different problems with the same program. But these systems are not able, up to now, to interrelate knowledge or produce abstract thinking. Therefore, it is not possible for an AI to write a real novel— that includes original elements like humor or irony—or produce theories that interpret or explain reality.

Even considering only narrow AI, it is possible to find many definitions of AI from different perspectives—from the philosophical to the economic or technical point of view. To avoid never-ending debates about the nature and essence of AI, it is better to focus on the legal definition settled both in the United States (National AI Initiative Act of 2020)[25] and in the European Union (Proposal of AI Act of 2021).[26]

According to these legal definitions of AI, we can conclude that these systems consist of software that is run on computers. This seems to be a simple conclusion, but it is an important starting point as it allows us to identify AI systems as a chain of commands that are run by machines and not necessarily requiring or needing physical assistance. So, we can

[23] *See* AMNON H. EDEN, ET AL., SINGULARITY HYPOTHESES: A SCIENTIFIC AND PHILOSOPHICAL ASSESSMENT 1 (2013) (discussing the ethical, social, and legal challenges, or "technological singularity," of general AI).

[24] STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 27–31 (2016).

[25] The National Artificial Intelligence Initiative Act of 2020 defined AI as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments" that "use[s] 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." 15 U.S.C. § 9401(3) (2022).

[26] Article 3(1) of the Proposal of AI Act defines AI as "software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with." 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, at 34 SEC (2021) 167 final (Apr. 4, 2021). Annex I lists the following AI techniques and approaches: "(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." *Id.*

distinguish between AI and robotics, as the majority of robots operate without AI, such as an industrial robotic arm in a factory, and many AI systems only operate virtually, like the Google photo recognition or Netflix prediction.

B. AI Algorithm

Despite this distinction, it must be said that AI systems are ultimately nothing more than software. It is true that they are a special kind of program as they are based on a new generation of algorithms. In fact, what really defines AI is the type of algorithms that are used in its programming, considering an algorithm as a process or a set of rules to solve a problem or perform a calculation.[27]

In conventional algorithms, programs are created manually by providing input data and the rules to follow, so the algorithm produces the output by automatically performing a task as instructed by the programmer. These conventional algorithms are used in entirely deterministic systems that are self-executing. Therefore, they are fully predictable as they basically consist of simple or complex decision trees; this is the so-called code-drive regulation.[28]

In AI or predictive algorithms, the input and output data are fed to the algorithm, so it creates the rules to solve the problem as it is coded to learn to perform a task autonomously. The singular design of AI algorithms gives them some exclusive functions, such as providing predictions, recommendations, or decisions to achieve specific objectives. They do so by continuously learning about data from the environment or from the results of its actions. These algorithms are informed by the data on which they have been trained instead of being informed by a programmer that has translated their insights into code. There is no deterministic code, and it introduces a new type of discretion, situated in the design choices made when training the algorithms; this is the so-called data-driven regulation.[29]

C. Further Remarks

At this point, some observations are needed on the nature of the predictive algorithms on which AI systems are built. Although AI

[27] *See* Justice Against Malicious Algorithms Act of 2021, H.R. 5596, 117th Cong. § 2(a)(1)(7) (2021) (defining algorithms, which is valuable because no legal definition of algorithms exists yet).

[28] *See* Mireille Hildebrandt, *Algorithmic Regulation and the Rule of Law*, 3 PHIL. TRANSACTIONS ROYAL SOC'Y 376 (2018).

[29] *Id.*

systems are probabilistic (nondeterministic) and they create the rules to solve problems, they are not autonomous and even less creative. These systems can only solve specific problems within a given set of human-defined objectives. So, AI systems do not find and solve problems by themselves, and they do not look for solutions beyond the objectives and ranges previously defined by humans, as they are part of a narrow AI.

It should also be noted that AI systems produce outputs, such as content, predictions, recommendations, or decisions influencing the environments they interact with. In this regard, AI systems do not actually interact with their environment to influence or modify it deliberately. AI systems act as imitators of the human mind, and therefore, must be able to "notice" what is going on around them, process that information, and be able to draw conclusions from it, while inferring new conclusions that have not been previously preprogrammed by a human being. As they have no freewill, they interact with the environment by merely receiving new input or output data within the terms that have been determined by humans.

The main technique behind the AI systems is machine learning, so-called as these systems are capable of changing their behavior to enhance their performance on some tasks through experience.[30] But these AI machine-learning based systems do not "learn" in the way that humans do but instead undergo mathematical "training" and "improve" their results in statistical terms. They are used to detect patterns in data in order to automate complex tasks or make predictions. They can produce automated results similar to those that would have been made by a human, so it would appear that they are learning and that they are "intelligent."[31] Machine learning divides into two models, supervised and unsupervised learning, differentiated by the degree of human intervention in the algorithm learning process.[32]

[30] STUART RUSSELL & PETER NORVIG, ARTIFICIAL INTELLIGENCE: A MODERN APPROACH 693 (3d ed. 2010); *see* Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1156–60 (2017); David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 669–702 (2017) (discussing machine learning and how it works).

[31] Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 90 (2014).

[32] In supervised learning, algorithms work with labeled data, trying to find a function that, given the input data, assigns the appropriate output label. The algorithm is trained with a "history" of data and thus "learns" to assign the appropriate output label to a new value, i.e., it predicts the output value (this model is used for email spam filters). Unsupervised learning systems are trained with raw, unlabeled data, so we only know the input data, but there is no output data corresponding to a given input. Therefore, we can only describe the structure of the data, to try to find some kind of organization that simplifies the analysis in an exploratory way (this model is used in recommendation systems). *See* Osvaldo Simeone, *A Very Brief Introduction to Machine Learning with*

AI systems produce several problems and pose several questions when used for decision-making.[33] First, the problem of transparency, since many of these algorithms are black boxes, which means that it is not possible to know how the given problem is solved. The second problem is that of bias, as AI systems are probabilistic and tend to perpetuate trends without taking into account principles such as equality or equity. All of these technical problems become relevant legal problems when AI systems are used for government, as we shall see.

IV. FROM SCIENCE FICTION TO REALITY: ACTUAL AI USES IN GOVERNMENT

Once AI has been defined as a software with singular characteristics that offers new functionalities that can help or even substitute human actions, we can explore what the actual AI uses in government are. The purpose of this section is to show the differences in the use of AI within government. AI systems are used with different purposes, so the legal implications of AI depend on how they are embedded in government actions.

In general, AI can be used by all three branches of government, but it should be noted that its relevance is very different within each branch of government. The challenges are also of a different nature depending on whether AI is used by the legislative, executive or judicial branch.[34] We will now take a preliminary approach to the use of AI in the different branches of government, and then focus on the analysis of its application in the executive branch, since this branch offers the greatest potential for its use and, consequently, the one that poses the most problems.

A. AI in Congress

In the case of the legislative branch, AI will have a very narrow range in which it can directly or indirectly affect the deliberative process that is inherent to democracy. However, the growing complexity of reality and the need for more precise and technical legislation create

Applications to Communication Systems, 4 IEEE TRANSACTIONS COGNITIVE COMMC'NS & NETWORKING 648 (2018).

[33] Tal Zarsky, *The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making*, 41 SCI., TECH., & HUM. VALUES 118 (2016).

[34] *See* Ephraim Nissan, *Digital Technologies and Artificial Intelligence's Present and Foreseeable Impact on Lawyering, Judging, Policing and Law Enforcement*, 32 AI & SOC'Y 441 (2015).

new room for AI solutions. AI can be a tool to help Congress make laws more effective, as long as it always remains an ancillary tool since it can distort the legislative process and threaten democracy.

Although there are no specific initiatives in the United States or the European Union, AI systems could be used both *ex ante* to simulate the impact of the proposed legislation and *ex post* to monitor the actual impact of enacted legislation.[35] It is likely that AI systems will soon be one of the standard tools used for law making, as impact assessment has become part of the legislative process. In the United States, AI systems could be used by the Congressional Budget Office (CBO) in carrying out cost analysis about the likely effects of proposed legislation on the federal budget. In the EU, the commission could also use AI systems for impact assessments to examine whether there is a need for EU action and analyze the possible impacts of available solutions.

B. AI in the Courts

The judicial branch has been more open to the use of AI, so there are precedents of use for AI in different jurisdictions from a long time ago.[36] There are many ways in which AI can be used by courts[37]: AI can be internally used to assist with information management (digitizing court records and organizing legal information); to assess external circumstances that can be used in judging; and to provide full advice to courts or even be an alternative through online dispute resolution systems.

In particular, predictive AI systems are very relevant in criminal justice as it allows for the possibility to assess recidivism. There are AI systems that aid human decision-making in criminal cases with respect

[35] *See* Joe Mariani, *AI for Smarter Legislation*, DELOITTE INSIGHTS (Sept. 22, 2022), https://www2.deloitte.com/us/en/insights/industry/public-sector/artificial-intelligence-can-benefit-the-legislative-process.html.

[36] *See* JUDICIAL APPLICATIONS OF ARTIFICIAL INTELLIGENCE (Giovanni Sartor & L. Karl Branting, eds., 1998) (explaining there are many examples from the Dutch Rechtwijzer (Roadmap to Justice) designed for couples who are separating or divorcing, to the British Columbia Civil Resolution Tribunal that provides a full suite of dispute resolution services); *see* John Zeleznikow & Fernando Esteban de la Rosa, *Artificial Intelligence as a New Component of the Justice System: How it Creates New Possibilities, but Has Limitations Especially with Regards to Governance*, *in* JUSTICE, TRADE, SECURITY, AND INDIVIDUAL FREEDOMS IN THE DIGITAL SOCIETY 59 (Fernando Esteban de la Rosa et al., eds., 2021).

[37] Cary Coglianese & Lavi M. Ben-Dor, *AI in Adjudication and Administration*, 86 BROOK. L. REV. 798 (2021); *see* James E. Baker et al., AI FOR JUDGES, CENTER FOR SECURITY AND EMERGING TECHNOLOGY (2021) (exploring other ways AI can affect judges); *see also* A. D. (Dory) Reiling, *Courts and Artificial Intelligence*, 11 INT'L J. CT. ADMIN. 4 (2020).

to questions of bail, sentencing, and parole, like PATTERN, LSI-R, or COMPAS.[38] COMPAS is the most relevant of these systems as it has been already used by courts in forty-six states to assess a defendant's likelihood to reoffend, and it adopts pretrial release decisions challenging due process.[39]

Therefore, predictive AI systems can be useful for the judiciary as they can help judicial decision-making in many ways. However, judging cannot be based only on predictions, as it is a very complex function that includes balance and fairness. The use of AI in courts raises many questions that need to be analyzed in detail,[40] as it affects the basic guarantees on the right of access to a court, the adversarial principle, the equality of arms, the impartiality and independence of judges, the right to counsel, and so on.[41]

For these reasons neither the United States nor the European Union have yet adopted AI systems in courts to make the ultimate, fully automated determination on a legal or factual question substituting human decisions.[42] Indeed, some judicial claims challenging the court's trial use of AI systems have been dismissed, as it has been considered that the risk assessment algorithms are merely a tool that courts can

[38] PATTERN (Prisoner Assessment Tool Targeting Estimated Risk and Needs) is used for risk assessment in federal parole decisions; LSI-R (Level of Services Inventory-Revised) aims to predict a defendant's risk of recidivism; COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), is an AI system for pretrial decisions.

[39] From 1998 COMPAS has been used as a criminal risk assessment tool to assess more than one million offenders in US courts. COMPAS has been accused of racial biases and inaccuracy, *see* Julia Angwin et al., *Machine Bias*, PROPUBLICA, (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; *see also* Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, SCI. ADVANCES (2018).

[40] *See* ETHAN KATSH & ORNA RABINOVICH-EINY, DIGITAL JUSTICE: TECHNOLOGY AND THE INTERNET OF DISPUTES (2017); John Zeleznikow, *Can Artificial Intelligence and Online Dispute Resolution Enhance Efficiency and Effectiveness in Courts*, 8. INT'L J CT. ADMIN. 30, 36–37 (2017); Andrew Lee Park, *Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing*, UCLA L. REV. (2019), https://www.uclalawreview.org/injustice-ex-machina-predictive-algorithms-in-criminal-sentencing/; RICHARD SUSSKIND, ONLINE COURTS AND THE FUTURE OF JUSTICE (2019); Ray Worthy Campbell, *Artificial Intelligence in the Courtroom: The Delivery of Justice in the Age of Machine Learning*, 18 COLO. TECH L. J. 323 (2020); TANIA SOURDIN, JUDGES, TECHNOLOGY AND ARTIFICIAL INTELLIGENCE: THE ARTIFICIAL JUDGE (2021).

[41] *See* EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE, EUROPEAN ETHICAL CHARTER ON THE USE OF ARTIFICIAL INTELLIGENCE IN JUDICIAL SYSTEMS AND THEIR ENVIRONMENT (Dec. 4, 2018) (identifying five principles regarding the use of AI in judicial systems: 1. Principle of respect for fundamental rights; 2. Principle of non-discrimination; 3. Principle of quality and security; 4. Principle of transparency, impartiality and fairness; 5. Principle "under user control").

[42] *See* Cary Coglianese & Lavi M. Ben-Dor, *AI in Adjudication and Administration*, 86 BROOK. L. REV. 791, 795, 798 (2021).

use to enhance their evaluation before sentencing.[43]

C. *AI in the Executive*

Finally, the use of AI in the executive branch will be enormous and critical considering the number and variety of missions and responsibilities that departments' (ministries and agencies) administrations have. Administrative agencies can use AI systems to develop new rules on guidance and adjudicate, enforce, or otherwise implement statutory policies. The possibilities for governmental use of AI are vast, including the use for military purposes, which must be considered separately because of the implications it presents.[44]

Looking at the United States, it can be found that federal agencies are today using AI systems gradually. Some academic studies reviewed the different uses of AI by agencies, highlighting and analyzing the most relevant examples to show the implications of its uses and provide some recommendations.[45] These studies were the only way to know how US agencies used AI as there were no official records or reports on AI uses in federal agencies until Executive Order 13960. Since 2021, agencies have been required to create an inventory of AI usage.[46] The problem of fragmentation in the use of AI is exacerbated by the US federalist structure. This structure means national and local agencies can incorporate AI independently.[47] It also shows that the application of AI in the US government is taking place without any determined plan at the global level and, above all, without a common legal framework or control over its deployment in the public sector.

[43] See the cases at the state court level (Wisconsin, Indiana, Kansas) that support the use of AI system in courts but recognize the right to access to the report and to the algorithm; s*ee id.* at 807–13.

[44] See the references on AI military use in Coglianese & Ben-Dor, *supra* note 42, at 792 n.4 (2021).

[45] A very useful tool in a first approach is the 2020 Stanford University Report prepared for the ACUS on the "Use of AI in Federal Administrative Agencies" that offers a broad picture of government use of AI. *See* DAVID FREEMAN ENGSTROM ET AL., GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES (2020); *see also* Coglianese & Ben-Dor, *supra* note 42, at 791–96.

[46] *See* Exec. Order No. 13960, *supra* note 13, § 5; *see also* DEPARTMENT OF HEALTH AND HUMAN SERVICES, *Artificial Intelligence Use Cases Inventory,* https://www.hhs.gov/about/agencies/asa/ocio/ai/use-cases/index.html (last visited Jan. 14, 2023); THE DEPARTMENT OF ENERGY, *Agency Inventory of AI Use Cases*, https://www.energy.gov/sites/default/files/2022-07/DOE_Agency_Inventory_of_AI_Use_Cases.pdf; THE DEPARTMENT OF AGRICULTURE, *Inventory of USDA Artificial Intelligence Use Cases,* https://www.usda.gov/data/AI_ Inventory.

[47] Coglianese & Ben-Dor, *supra* note 42, at 793.

While the use of AI in US federal agencies is very limited at present,[48] there are examples of its usage in the full range of governance tasks to aid human decision-making.[49] In particular, there are cases of AI usage in agency policymaking as a tool for regulatory research, analysis, monitoring, and collecting or processing information.[50] Additionally, AI systems for enforcing regulatory mandates are used to identify or prioritize targets of agency enforcement action.[51] AI is also used in adjudicating benefits and rights performing tasks that support formal or informal agency adjudication.[52] AI systems usage is expanding in public service provision—it identifies needs and facilitates communication with citizens.[53] Finally, AI is largely used in internal management to support agency management of resources. This management includes human resource management, public procurement, and the maintenance of technology systems.[54]

In the case of the EU, it should be noted that the deployment of AI in public administration is an internal matter for each member state as the union has no direct competence in this specific area. Although the EU can regulate AI in general and promote its use at a national level, it

[48] ENGSTROM ET AL., *supra* note 45, at 88 (finding only 157 cases in 64 US Federal agencies, and only 20 cases could be considered of higher level of sophistication).

[49] According to some studies, the US has not yet instituted an AI system providing for total decision-making by algorithm, leaving the human "out of the loop" in the decision. However, it is not clear the role of AI in the final decision, activity, or service provided by the agencies. Coglianese & Ben-Dor, *supra* note 42, at 795.

[50] This is the case of the Consumer Financial Protection Bureau AI system for analysis of consumer complaints; the Bureau of Labor Statistics coding of worker injury narratives; and the Food and Drug Administration analysis of adverse drug events. ENGSTROM ET AL, *supra* note 45, at 53–58, 59–64.

[51] Some examples are the Securities and Exchange Commission, Centers for Medicare and Medicaid Services, and Internal Revenue Service predictive enforcement tools; also, the Customs and Border Protection and Transportation Security Administration facial recognition systems; and finally, the Food Safety and Inspection Service prediction to inform food safety site testing. *See* ENGSTROM ET AL., *supra* note 45, at 30–37.

[52] Such as the Social Security Administration system for correcting adjudicatory errors or the US Patent and Trademark Office tools for adjudicating patent and trademark application. ENGSTROM ET AL, *supra* note 45, at 37–45, 46–52.

[53] This is the area in which AI expansion is most likely to take place performing tasks that support the direct provision of public services to the citizens or facilitate communication with the public for regulatory or other purposes. There are several examples as the US Postal Service autonomous vehicles project and handwriting recognition tool, the Department of Housing and Urban Development and US Citizenship and Immigration Services chatbots or the Agencies analysis of submitted rulemaking comments. ENGSTROM ET AL, *supra* note 45, at 59–64, 65–69.

[54] Among the examples are the Department of Health and Human Services tool to assist procurement decision-making; the General Services Administration tool to ensure legal compliance of federal solicitations; and the Department of Homeland Security tool to counter cyberattacks on agency systems. ENGSTROM ET AL, *supra* note 45, at 30–36.

cannot impose a generic AI model for the governments of all member states.

EU countries are gradually including AI systems in government, so there is an increasing number of cases of AI use by national public administrations. The European Commission released in 2022 a report on "Artificial Intelligence in Public Services"[55] that offers a complete overview of its use and impact in member states. The report found 686 user cases of AI in twenty-seven member states, with the cases increasing each year in a very fragmented and unevenly distributed way reaching all government functions.[56]

Member states are using AI systems to provide public services and engagement (service personalization, engagement management, service integration, and data sharing management); enforcement (smart recognition processes, predictive enforcement processes, supporting enforcement processes, management of auditing and lodging); analysis, monitoring, and regulatory research (prediction and planning, information analysis processes, and monitoring policy implementation); internal management (internal support and primary processes); and also for adjudicating (deciding on benefits).[57]

Although the use of AI is still very limited in the EU considering the size and variety of government actions of member states, it is increasingly expanding to new areas, and it has already taken part in critical activities. In fact, the use of AI in government led to the resignation of the Dutch Prime Minister in 2021 after thousands of families were wrongly accused of fraud due to a biased algorithm.[58] In Europe, AI systems take part in a wide range of public services (in

[55] In addition to the overview, the report analyzes the challenges, barriers, and risks of the use of AI in the public sector and provides policy recommendations in its adoption and implementation. *See Joint Research Centre Science for Policy Report*, *AI Watch: European Landscape on the Use of Artificial Intelligence by the Public Sector,* EUR 31088 EN (2022), https://joinup.ec.europa.eu/sites/default/files/inline-files/JRC129301_01-1.pdf.

[56] The number of cases is increasing each year (from 5 in 2015 to 167 in 2021) and unevenly distributed— Netherlands (123), Italy (75) and Portugal (60). Most of them are case of use of AI at national level (54%), based in machine learning (58%) and for provision of public services and engagement (36%). For an overview of cases, s*ee id.* at 35–45.

[57] For providing public services and engagement (36%), enforcement (26%), analysis, monitoring and regulatory research (22%); internal management (16%); and adjudicating (2%). *Id.* at 41.

[58] Thomas Erdbrink, *Government in Netherlands Resigns After Benefit Scandal,* N.Y. TIMES (Jan. 1, 2021), https://www.nytimes.com/2021/01/15/world/europe/dutch-government-resignation-rutte-netherlands.html; *see also* Gabriel Geiger, *How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud,* VICE (Mar. 1, 2021), https://www.vice.com/en/article/jgq35d/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud.

particular healthcare).[59] They will likely penetrate the public sector through the delivery of services and then will likely spread to legal decision-making (rulemaking, enforcement, and adjudication).

V. GOVERNMENT USE OF AI REGULATORY FRAMEWORK

AI is spreading in the public sector in a very fragmented and unsystematic manner so far as agencies are embedding AI systems in specific functions without a common plan or a complete regulation that guarantees their use in government. However, the use of AI is not beyond the law as it is classified under existing regulations (general and specific). These regulations are already applied to AI, so it is now necessary to determine what the government use of AI regulatory framework is both in the United States and the European countries that are engaged in this AI government usage revolution.

A. Government Use of AI Regulation in the United States

In the United States, there is no comprehensive federal legislation to date on AI as a whole. Although the United States has passed legislation both on AI and AI use in government (see Section III), these are very limited pieces.

During the Trump Administration, the approach to AI had been to focus on promoting and funding research development to ensure US leadership in this area. The National AI R&D Strategic Plan, released in 2016 and updated in 2019, establishes a set of strategic priorities for funded AI research,[60] including ensuring the safety and security of AI systems. Furthermore, the National AI R&D Strategic Plan does ask and propose AI regulation to be made a priority.

The National AI Initiative Act passed in 2020 continues on the same

[59] AI systems are already used in healthcare to design vaccination policies and to support emergency management. AI is also used in prevention (e.g., to predict future risk to suicidal ideation from social media data), diagnosis (e.g., voice-based diagnosis of covid) and treatment (e.g., personalized cancer care). For more information, *see Panel for the Future of Science and Technology, Artificial Intelligence in Healthcare: Applications, Risks, and Ethical and Societal Impacts,* EUR. PARL. RSCH. SERV. 3 (PE 729.512) (June 2022).

[60] Strategy 1: Make long-term investments in AI research; Strategy 2: Develop effective methods for human-AI collaboration; Strategy 3: Understand and address the ethical, legal, and societal implications of AI; Strategy 4: Ensure the safety and security of AI systems; Strategy 5: Develop shared public datasets and environments for AI training and testing; Strategy 6: Measure and evaluate AI technologies through standards and benchmarks; Strategy 7: Better understand the national AI R&D workforce needs; Strategy 8: Expand public-private partnerships to accelerate advances in AI.

path, as it is limited to ensuring US leadership in AI research and development by providing a set of initiative activities to be carried out by the president acting through the National Artificial Intelligence Initiative Office and the Interagency Committee. Far from proposing a specific regulation for AI, the AI Initiative Act supports the development of private instruments for the development and use of AI, such as voluntary standards, best practices, and benchmarks, including the development of a voluntary risk management framework for the trustworthiness of AI systems.[61]

Executive Order 13859, Maintaining American Leadership in AI, approved in 2019, expresses the negative approach to the regulation of AI. It also promotes and protects US advancements in AI. One of the strategic objectives is to reduce barriers to the use of AI technologies, which promotes their innovative application while protecting American technology, economic security, national security, civil liberties, privacy, and values.

In the same vein, the Office of Management and Budget (OMB) Memorandum on Guidance for Regulation of AI Applications released in 2020 sets out policy considerations that should guide the approaches to AI applications developed and deployed outside of the federal government.[62] Following the negative approach, the OMB Memorandum considers that AI applications "do not necessarily raise novel issues" and they can be promoted "through forbearing from new regulation" that should be only considered after deciding "that it is necessary."[63] The OMB Memorandum settles ten principles for the stewardship of AI applications that are a reproduction of the common principles of rulemaking[64]—without including any specific principles to face AI challenges—and prefers the use of nonregulatory approaches to AI[65] while promoting the reduction barriers to the deployment and use of

[61] These AI activities are under the responsibility of the National Institute of Standards and Technology according to Title III of the National Artificial Intelligence Initiative, H.R. 6216, 116th Cong. §§ 101, 301(a)–(e) (2020).

[62] OFF. OF MGMT. & BUDGET, EXEC OFF. OF THE PRESIDENT, MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES: GUIDANCE FOR REGULATION OF ARTIFICIAL INTELLIGENCE APPLICATIONS (2020), https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf.

[63] *See id.* at 3.

[64] The principles are public trust in AI (responding and mitigating risks), public participation (informing the public and promoting voluntary frameworks and standards), scientific integrity and information quality, risk assessment and management, benefit and cost, flexibility, fairness and non-discrimination, disclosure and transparency, safety and security, and interagency coordination. *See id.,* at 3–7.

[65] The non-regulatory approaches include sector-specific policy guidance or frameworks, pilot programs and experiments, voluntary consensus standards, and voluntary frameworks. *Id.* at 7–8.

AI.[66]

The Biden Administration's approach to AI is more protective and citizen based. The White House Office of Science and Technology Policy (OSTP) released a blueprint in 2022 for an AI Bill of Rights that included a set of five principles and associated practices that will help guide the design, use, and deployment of AI systems to protect the rights of the citizens: (a) safe and effective systems; (b) algorithmic discrimination protections; (c) data privacy; (d) notice and explanation; (e) human alternatives, consideration, and fallback.[67]

Government use of AI has been specifically considered in the United States providing some measures to promote it within federal agencies. The 2020 Government Act authorized the AI Center of Excellence within the General Services Administration to facilitate and improve the use of AI in federal government. Even more relevant is Executive Order 13960, Promoting the Use of Trustworthy AI in the Federal Government (December 3, 2020), which sets the following principles for the use of AI in federal government: (a) lawful and respectful of our nation's values; (b) purposeful and performance-driven; (c) accurate, reliable, and effective; (d) safe, secure, and resilient; (e) understandable; (f) responsible and traceable; (g) regularly monitored; (h) transparent; and (i) accountable.

These principles are referred to as the main problems that arise with the use of AI in government and show that there is increasing concern about the implication of AI use in government. This concern is the basis of agency guidelines for deploying AI tools adopted by the US Administrative Conference. The guidelines ask agencies to consider issues such as transparency, technical capacity bias, procurement, privacy, security, decisional authority, and oversight.[68] The Government Accountability Office has also issued an accountability framework identifying key practices to ensure accountability and responsible AI use by federal agencies.[69]

[66] Through access to Federal data and models for AI R&D; Communication to the public; Agency Participation in the Development and Use of Voluntary Consensus Standards and Conformity Assessment Activities; and International Regulatory Cooperation. *Id.* at 8.

[67] WHITE HOUSE OFF. OF SCI. AND TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 13 (Oct. 2022), https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

[68] ADMINISTRATIVE CONFERENCE OF THE UNITED STATES, ADMINISTRATIVE CONFERENCE STATEMENT #20 AGENCY USE OF ARTIFICIAL INTELLIGENCE 2–10 (Dec. 2020), https://www.acus.gov/sites/default/files/documents/Statement%2020%20Agency%20Use%20of%20Artificial%20Intelligence.pdf.

[69] U.S. GOV'T ACCOUNTABILITY OFF., GAO-21-519SP, ARTIFICIAL INTELLIGENCE: AN ACCOUNTABILITY FRAMEWORK FOR FEDERAL AGENCIES AND OTHER ENTITIES (2021).

B. Government Use of AI Regulation in the EU

In the EU, there is also no specific legislation on AI to date, although there are very relevant proposals going on. According to the strategy included in the White Paper on AI (2020) and the Communication Fostering a European Approach to AI (2021), the commission launched a proposal for a regulation laying down harmonized rules on AI in the EU (AI Act) in April 2021.[70] This regulation was completed in September 2022 with the proposal for a directive on adapting noncontractual civil liability rules to artificial intelligence (AI Liability Directive).[71]

The AI Act will presumably be adopted by the end of 2023 and will introduce a harmonized regulation that will be applicable to AI systems used both in the private and the public sector. However, it is not an extensive and detailed regulation on AI, but rather a minimal regulation that includes the prohibitions of certain AI practices, specific requirements and obligations for high-risk AI systems, and transparency rules for AI systems intended to interact with people.

For high-risk AI systems—that include most of the AI systems used in government—they will have to observe legal requirements in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy, and security. The high-risk AI systems will have to pass a conformity assessment procedure controlled by independent third-parties or notified bodies.

This specific legislation on AI is completed by several soft law instruments that set out substantive expectations but are not directly enforceable by government. The main instrument is the Ethics Guidelines and Assessment List for Trustworthy AI, developed by a High-Level Expert Group on AI in 2018, that is the only legal specific framework to deal with AI systems in the EU today.[72] The international standards, such as the ISO or the IEEE standards,[73] also have a

[70] *Proposal for a Regulation of the European Parliament and of the Council Concerning Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts,* COM (2021) 206 final (Apr. 21, 2021).

[71] *Proposal for a Directive of the European Parliament and of the Council on Adapting Non-Contractual Civil Liability Rules to Artificial Intelligence (AI Liability Directive),* COM (2022) 496 final (Sept. 28, 2022).

[72] High-Level Expert Group on AI, *Ethics Guidelines for Trustworthy AI* (June 2018). On the role of ethic in AI, *see* Jessica Morley et al., *Ethics as a Service: A Pragmatic Operationalisation of AI Ethics*, 31 MINDS MACH 239 (2021).

[73] For example, the ISO/IEC TS 4213: ASSESSMENT OF MACHINE LEARNING CLASSIFICATION PERFORMANCE; ISO/IEC CD 5259: DATA QUALITY FOR ANALYTICS AND

relevant advisory role as they provide a technical reference to the design of the AI system. There are also digital rights charts (such as the European and Spanish charts)[74] that include specific rights in the interactions with algorithms.

Beyond all this specific regulation, AI is subject to general regulation that is currently in effect and this includes the regulation on data (data protection, open data, data governance),[75] digital services,[76] cybersecurity,[77] product safety or consumer protection,[78] and, of course, fundamental rights.[79] Therefore, there is currently an extensive legal framework for AI, albeit a specific one.

Despite all the regulation that applies to AI in Europe, there is no specific regulation on the use of AI in government as the EU has no competence in this specific field. Member states are regulating government use of AI on a national basis, which means the regulation is highly fragmented and underdeveloped.

MACHINE LEARNING (ML); ISO/IEC DIS 5338: AI SYSTEM LIFE CYCLE PROCESSES; ISO/IEC CD 5339 GUIDELINES FOR AI APPLICATIONS; ISO/IEC CD 5392: REFERENCE ARCHITECTURE OF KNOWLEDGE ENGINEERING.

[74] *Communication from the Commission Establishing a European Declaration on Digital Rights and Principles for the Digital Decade*, at 1, COM (2022) 28 final (Jan. 26, 2022); Spanish Charter of Digital Rights art. 24, July 2021.

[75] Regulation (EU) 2016/679, 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) 2016 O.J. (L 119) 17; Council Directive 2019/1024, on Open Data and the Re-Use of Public Sector Information, 2019 O.J. (L 172) 74 (EU); Regulation (EU) 2022/868 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act) 2022 O.J. (L 152) 1; *IoT Data: Proposal for a Regulation on Harmonised Rules on Fair Access to and Use of Data (Data Act),* at 18, COM (2022) 68 final (Feb. 23, 2022).

[76] Council Directive 2000/31, on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market ('Directive on Electronic Commerce'), 2000 O.J. (L 178) 17 (EC); *Proposal for a Regulation on a Single Market for Digital Services (Digital Services Act)*, at 1, COM (2020) 925 final (Dec. 15, 2020)*; Proposal for a Proposal for a Regulation on Contestable and Fair Markets in the Digital Sector (Digital Markets Act),* at 1, COM (2020) 842 final (Dec. 15, 2020).

[77] Council Directive 2000/31, on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market ('Directive on Electronic Commerce'), 2000 O.J. (L 178) 17 (EC); *Proposal for a Regulation on a Single Market for Digital Services (Digital Services Act)*, at 1, COM (2020) 925 final (Dec. 15, 2020)*; Proposal for a Proposal for a Regulation on Contestable and Fair Markets in the Digital Sector (Digital Markets Act),* at 1, COM (2020) 842 final (Dec. 15, 2020).

[78] Council Directive 85/374 on the Approximation of the Laws, Regulations and Administrative Provisions of the Member States Concerning Liability for Defective Products, 1985 O.J. (210) 32 (EC); Council Directive 2006/42/EC on Machinery, and Amending Directive 95/16/EC (recast), 2006 O.J. (157) 14.

[79] Fundamental rights beyond those included in Member States' constitutions can be found in Charter of Fundamental Rights of the EU, Oct. 26, 2012, 2012 O.J. (C 326) 395; European Convention on Human Rights § 1 art. 2.

In the case of Spain, for example, there used to be a single article, article 41 in Act 40/2015, that regulated the use of AI in government and only required the identification of the competent body or bodies for the definition of specifications, programming, maintenance, supervision, and quality control.[80] Recently, a new article, article 23 in Act 15/2022, introduces new specific requirements for the use of AI in government decision-making (minimization of bias, transparency and accountability, applying impact assessment, and a quality seal for algorithms) but these are not compulsory for public administration.[81]

VI. RISK AND PERILS OF AN ARTIFICIALLY INTELLIGENT GOVERNMENT

There is serious concern about the expanding use of AI, which has led to the development of a growing body of regulation and soft law instruments, both in the US and the EU. However, no particular attention has been paid to the implications of the use of AI by government. In spite of this lack of attention, singular problems arise that are challenging constitutional and administrative principles, and they require specific principles and regulations.

These singular problems arise with particular complexity when AI is used in government decisions involving the use of power and affecting the public's rights, such as rulemaking (e.g., regulatory analysis), adjudication (e.g., grants, aids) or enforcement (inspection). On the other hand, internal management (e.g., application form managers), public engagement (e.g., chatbots), monitoring (e.g., analysis of adverse drug events), and public service provision (e.g., personalized diabetes care) are government tasks that are similar in nature to the use of AI in the public sector. Therefore, they can be sufficiently covered by general AI regulations and soft law instruments requiring human agency and oversight, transparency, safety, privacy, nondiscrimination, or accountability. Nevertheless, all these principles, although relevant, may not be enough for government decision-making. Use of AI in this area needs administrative and constitutional principles.

[80] On Legal Regime of the Public Sector Act 40/2015 art. 41, B.O.E. 2015, 236 (Spain); *see* Julián Valero Torrijos, *The Legal Guarantees of Artificial Intelligence in Administrative Activity: Reflections and Contributions from the Viewpoint of Spanish Administrative Law and Good Administration Requirements*, 1 (1-2) EUR. REV. OF DIGIT. ADMIN. & L. – ERDAL, 56–57 (2020); *see also* Itziar Sobrino-García, *Artificial Intelligence Risks and Challenges in the Spanish Public Administration: An Exploratory Analysis through Expert Judgements*, 11 ADMIN. SCI. 102 (2021).

[81] Comprehensive Law 15/2022 art. 23 (B.O.E. 2022, 167) (Spain).

A. *Human Factor Inside and Outside AI Decision-Making*

Regarding the use of AI in government decisions that involve the use of power, the first question that arises is whether AI systems can replace public officials or authorities in the decision-making processes, or will it be used as a tool that aids such processes. One of the requirements of a trustworthy AI is the idea of human agency and human oversight,[82] but governments can adopt automated decisions and use AI systems without a human-in-command as long as they are covered by national legislations. Neither the US nor the EU have explicitly excluded the use of AI systems in administrative agency decisions, but some European countries have banned or make the use of AI systems more difficult when exercising authority discretion.[83]

Although it has not yet been instituted in government, AI systems that provide for total decision-making by algorithm (that is, human "out of the loop" decisions),[84] it is not clear when computers are making a fully independent determination or when they can be merely supportive formally but determinative in fact.[85] It is important to clarify the role that AI systems can assume, and establish a rule indicating in which

[82] Regulation (EU) 2016/679, art. 26, General Data Protection Regulation, 2019 O.J. (L 151) 2 (explaining that human agency includes the right not to be subject to a decision based solely on automated processing when this produces legal effects on users or similarly significantly affects them—included in article 26 of GDPR—and human oversight based on the human-in-command (HIC) approach that allow to decide when and how to use the system in any particular situation including levels of human discretion during the use of the system and the ability to override a decision made by a system); *see Ethics Guidelines for Trustworthy AI,* at 16 (2018).

[83] In the US and the EU, AI in government regulation implicitly admit the use of human independent AI system. In the US there is no right to a human in the administrative decision-making process. *See* Francesca Bigami, *Artificial Intelligence Accountability of Public Administration*, 12 AM. J. COMP. L. (2022). In Germany, fully automated administrative acts are covered by VwVfG § 35a, but it excludes the use of automated systems for administrative acts wherever these require the use of discretion. *See* Elena Buoso, *Fully Automated Administrative Acts in the German Legal System*, 1 EUR. REV. OF DIGIT. ADMIN. & L. 113, 114 (2020). In Spain, article 43 of Act 40/2015 refers to automated decision, but the Section XVI of the Charter of Digital Rights requires that discretionary decision-making is reserved to persons, unless a specific law allows for the adoption of automated decisions in that particular area. *See* On Legal Regime of the Public Sector Act, *supra* note 80.

[84] *See* Coglianese & Ben-Dor, *supra* note 37 (noting that in the EU, it is not possible to confirm that any administrative body has instituted an independent AI system).

[85] *See* Cary Coglianese & David Lehr, *Transparency and Algorithmic Governance*, 71 ADMIN. L. REV. 1, 31 (2019); Cary Coglianese & David Lehr, *supra* note 30, at 1167–70 (exploring the difference between supportive and determinative algorithms); *see also* Lilian Mitrou et al., *Human Control and Discretion in AI-Driven Decision-Making in Government,* ICEGOV '21: PROCEEDINGS OF THE 14TH INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF ELECTRONIC GOVERNANCE 10 (2021).

cases they can act independently of humans, since these systems may outsource government decisions without a constitutional or even a legal reform.[86]

B. Problems with Transparency

A second challenge is reconciling public law's commitment to public participation and reason giving as there is a lack of transparency as well as a need to explain black box AI systems in advance. Citizens should participate in the design of AI algorithms as long as they affect their rights. They must receive an explanation about AI decisions and have access to the merits of decisions, especially if they appeal, and it should be in natural language (and not in machine code). This AI decision-making would not be a problem if AI systems were always supportive and if the final decisions were adopted by humans, but as noted before, the reality is that government decisions are increasingly relying on the result of an algorithm.[87]

Transparency is a basic principle in the use of AI in government. It has a broader scope than AI transparency in the private sector because it goes from the design of the public algorithm to the adoption and supervision of algorithmic decision-making.[88]

On one hand, government use of AI should include public participation in algorithm design in the same way that the notice and comments process allow citizens to participate in rulemaking.[89]

[86] Despite the use of human-independent AI is not banned in the US nor the EU it is clear that "right to a human decision," is one of the fundamental assumptions in legal systems. Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 615–20 (2020); *see also* CARY COGLIANESE, A FRAMEWORK FOR GOVERNMENTAL USE OF MACHINE LEARNING 51–52 (2020) (report to the Admin. Conf. of the US).

[87] *See* ENGSTROM ET AL, *supra* note 45, at 15–20.

[88] In the US, Executive Order 13960, *supra* note 13, § 3, includes among the principles of AI use in government that: AI should be sufficiently understandable by experts, users, and others; human users have a role and responsibility in documenting all the process of use of AI; AI performance should be regularly monitored and supervised; and there should be transparency and frequent disclosure of relevant information regarding the use of AI. The OMB MEMORANDUM M-21-06 (2020), *supra* note 62, ¶ 8, provides further guidance on transparency. In the EU, there are no specific principles for the use of AI in government but the Ethics Guidelines for Trustworthy AI prepared by the High-Level Expert Group refers to transparency in general which includes: traceability (data sets and the processes should be documented); explainability (technical processes of an AI systems should be understood and traced by human beings); and communication (AI systems should be perceived as such by humans to users). High-Level Expert Group on AI, *supra* note 72.

[89] OFF. MGMT. & BUDGET, *supra* note 62, at 3. On participation in algorithm design, *see* Francesca Bigami, *supra* note 83.

According to some opinions, AI algorithms should be considered rules and should be submitted to a political notice and comments rulemaking process.[90]

On the other hand, transparency includes the disclosure of information on algorithm design and performance as well as on the datasets and the training process. This disclosure requires that these processes be traceable and should include access to all information, making them sufficiently understandable for citizens. The problem is that it can be impossible to fulfill these requirements (traceability, access, intelligibility) for black box AI systems. Therefore, it has to be considered if using this kind of algorithm is compatible with constitutional processes and administrative procedure rights.

Beyond these problems with black box AI and in the case of common algorithms, further barriers for the transparency of AI government decisions remain such as copyright, privacy, national security, and other protected information can deny access to algorithm information.[91] There is no binary solution to these transparency problems, but to begin with,

[90] *See* ENGSTROM ET AL, *supra* note 45, at 77, 84 (discussing asking for notice-and-comment rulemaking for AI algorithms); *see also* David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. REGUL., 800, 836–39 (2020); Administrative Conference Statement #20: Agency Use of Artificial Intelligence, 86 Fed. Reg. 6616, 6618 (2020); Andrés Boix Palop, *Algorithms as Regulations: Considering Algorithms, When Used by the Public Administration for Decision-Making, as Legal Norms in Order to Guarantee the Proper Adoption of Administrative Decisions*, 1 EUR. REV. DIGIT. ADMIN. & L., 75 (2020) (providing a similar idea from a European perspective); Francesca Bigami, *Artificial Intelligence Accountability of Public Administration*, 12 AM. J. COMP. L., 1, 21, 23 (2022) (discussing the blurring difference between rules and adjudication in using AI).

[91] In the US, the main concern of court challenges of government AI decision has been transparency, not equal protection nor privacy. Aziz Z. Huq, *Constitutional Rights in the Machine-Learning State*, 105 CORNELL L. REV. 1875, 1879, 1903 (2020). Within the EU, France's Constitutional Court, in its decision of April 3, 2020, denied access to the code of the Parcoursup algorithm that assesses the applications higher education, alleging that limitation was justified by general interest and was not disproportionate. *See* Lucie Cluzel-Métayer, *The Judicial Review of the Automated Administrative Act*, 1 EUR. REV. DIGIT. ADMIN. & L., 101, 101–03 (2020). In the Netherlands, the Hague District Court decision of February 5, 2020, about the SyRi algorithm (program to fight tax fraud) also prevents the judge from controlling the algorithm. In Spain, the first case regarding government use of AI of 2022 also denies the access to the code. *See* Juli Ponce Solé, *The Energy Social Bonus and the Bosco Program: About Algorithms, Bugs and Source Code. Regarding The First Court Decision Handed Down in 2021: A Bad Judgment That We Hope Will be Corrected Soon,* LUMSA UNIVERSITÁ (Sept. 29, 2022, 3:34 PM), https://betteregulation.lumsa.it/repost-rednmr-energy-social-bonus-and-bosco-program-about-algorithms-bugs-and-source-code-regarding. For a similar discussion related to Italy, *see* Flavio Bravo, *Access to Source Code of Proprietary Software Used by Public Administrations for Automated Decision-Making: What Proportional Balancing of Interests?*, 1 EUR. REV. DIGIT. ADMIN. & L., 157 (2020).

it will have to be taken into account the context in which AI is used in government and the characteristics of the AI system.[92]

C. Gaming and Controlling AI Algorithms

Another challenge of AI use in government related to transparency is the risk of gaming with algorithms without manipulating them.[93] A full transparency of AI government systems can lead to a total disclosure of the public algorithms letting stakeholders look inside the government's brain. So, transparency allows large stakeholders to invest in the right technology, which will anticipate and control administrative algorithm-based decisions, and in doing so, dominate government's decision criteria.

This side effect of transparency can be admitted in some cases (for example adjudication of benefits), so the algorithm should be made public with no particular concern. In other cases (such as tax inspection),[94] the algorithm must be kept hidden to avoid giving advantages for government actions. But even in these cases there is an actual risk of "adversarial learning" as the government criteria can be identified and handled through reverse engineering that shows the decision model.

D. AI Providers Dependence

Another challenge for government AI is the dependence on technology providers. Creating AI systems within government might yield better tailored tools and generate internal capabilities to better handle the system, while obtaining AI systems from external sources might allow access sooner to more sophisticated tools and save some associated costs.[95]

In fact, it is likely that most of the AI systems used in government

[92] Administrative Conference Statement #20, *supra* note 90, at 6616; *see also* Coglianese & Lehr, *supra* note 85, at 2 n.2; Agustí Cerrillo I. Martínez, *How Can We Open the Black Box of Public Administration? Transparency and Accountability in the Use of Algorithms*, 58 REVISTA CATALANA DE DRET PUBLIC, 13 (2021) (discussing AI government transparency).

[93] This has been a collateral problem, noted in Administrative Conference Statement #20, *supra* note 90, at 6616; *see also* ENGSTROM ET AL., *supra* note 45, at 86–87.

[94] *See* Elise Degrave, *The Use of Secret Algorithms to Combat Social Fraud in Belgium*, 1 EUR. REV. DIGIT. ADMIN. & L., 167 (2020).

[95] *See* Administrative Conference Statement #20, *supra* note 90, at 6617.

will come from private companies.[96] They produce systems that are purchased and are also hired to design these systems for administrative bodies. The peculiar nature of AI systems and the role that they play in administrative decision-making leads to a kind of outsourcing for government functions.[97] Private companies create programs that can even replace government decision-making, and public bodies totally depend on these companies to make their decisions, so administrative decisions are actually taken outside government.

In this situation, public procurement will turn into a key tool not only to get better AI systems, but also to regulate the use and operation of AI systems in government. The terms and conditions in the procurement processes will have to have imbedded general AI regulations and this process will make up for the lack of AI use in government regulation.

E. AI Accountability and Oversight

The last and more relevant challenge is AI government accountability and oversight.[98] Internal oversight within agencies and public bodies allows proper use and functioning of the AI system, and external oversight ensures that government AI systems are lawful and respectful with constitutional values.

The most relevant external oversight instrument that guarantees that government use of AI is under the rule of law is judicial review. Courts may be overwhelmed by government AI for several reasons: lack of expertise on technical implications of AI systems, inaccessibility of algorithm and dataset information, unintelligibility of this information, or inability to explain the result. These are old problems in a new context that make them more serious as the AI system challenges the basic principles of administrative and constitutional law.[99]

Furthermore, if AI systems without humans in command are widespread in government, there is no human in control for providing reasons, and these AI systems are increasingly complex or, directly, cannot give explanations for their decisions, it will be impossible for

[96] In the US, it seems that most of the Federal AI systems (53%) are the product of in-house efforts by agencies. ENGSTROM ET AL*, supra* note 45 at 7. Nevertheless, AI systems will be outsourced as long as the turn more complex.

[97] *Id.* at 88–90.

[98] *See* Administrative Conference Statement #20, *supra* note 90. *See also* David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State*, 37 YALE J. REGUL., 800 (2020).

[99] ENGSTROM, ET AL., *supra* note 45, at 75–78. *See also* Rebecca Williams, *Rethinking Administrative Law for Algorithmic Decision Making*, 42 OXFORD J. LEGAL STUD., 468 (2022).

courts to provide judicial review in line with traditional models.

The limitations of courts to provide a full judicial review leads to calls to explore new alternatives to a government AI systems oversight beyond the courts' scope, including *ex ante* mechanisms (such as precertification); AI oversight boards (both within and outside public bodies); soft law rules (such as technical standardization or ethics guidelines); risk and impact assessment; and so on.

There are many other challenges regarding the use of AI in government, such as harmful biases, data privacy, and security.[100] These are challenges that are common to AI use both in the private and public sectors, so they do not defy existing administrative and constitutional principles. Although these challenges to AI use are relevant when projected in the public sector, they can be magnified and exacerbated as well as generate problems of discrimination and violation of privacy or insecurity as never known before. Therefore, they will require special attention.

VII. CONCLUSION

The use of AI in government is posing new challenges that go beyond the usual problems of digitalization (such as technical infrastructure, human capital, and regulatory barriers). The unique characteristics of this new technology are transforming the nature of government, as it is not used as a tool to facilitate its activity, but rather it affects administrative decision-making.

Indeed, AI is a technology that is not magic, but it has functionalities that were unknown until now, and they include generating outputs, such as content, predictions, recommendations, or decisions influencing the environments it interacts with. The use of predictive algorithms takes government action to a new level as it

[100] In the US, EO 13690 ask for an AI in Government be accurate, reliable, effective, safe, secure, and resilient. OFF. OF MGMT. & BUDGET, *supra* note 62; *see also* Administrative Conference Statement #20, *supra* note 90, at 6617–18. In Europe, the ETHICS GUIDELINES FOR TRUSTWORTHY AI, *supra* note 72, include the following principles for AI in general: technical robustness and safety (resilience to attack and security; fallback plan and general safety; accuracy; reliability and reproducibility); privacy and data governance (privacy and data protection; quality and integrity of data; access to data); and diversity, non-discrimination and fairness (avoidance of unfair bias; accessibility and universal design; stakeholder participation). On legal issues with governmental use of AI, *see* Coglianese & Lehr, *supra* note 86. On biased government AI algorithms, *see* ENGSTROM ET AL*, supra*, note 45, at 79–81; *see also* Kristen M. Altenburger & Daniel E. Ho, *When Algorithms Import Private Bias Into Public Enforcement: The Promise and Limitations of Statistical Debiasing Solutions*, 1 J. INSTITUTIONAL & THEORETICAL ECON. (2019).

ceases to be merely "automatic" and instead is "autonomous," becoming progressively detached from human decision.

In any case, the use of AI in government is still very limited, but it is spreading to new activities and services. It is foreseeable that it can be applied to all government functions. A number of examples of this exist both in the United States and the European Union, which demonstrates the enormous potential of using AI systems in government.

The development and expansion of AI systems in government is not accompanied by the introduction of specific regulation on the use of AI in the public sector. In the United States and the European Union, regulation of AI is emerging with a global perspective without addressing the specific problems that it raises when used in government. In fact, the incorporation of AI in government is occurring without a specific legal framework so the use of algorithms is only subject to traditional administrative law.

It is urgently needed to promote a new governance for the use of algorithmic AI by administrative bodies to meet the challenges it will pose for government. A first step is to identify what AI consists of and how it is being applied (and can be applied) in government functions. From there, it is necessary to analyze the incipient problems that are arising with respect to the role of humans in administrative AI decisions; the transparency and the possibility of accessing the reasons for AI administrative decisions; the dependence on third parties that provide external AI; and, above all, the difficulties of accountability and oversight of government AI action to ensure that it is lawful and respectful of constitutional values.

To conclude, it is clear that the use of AI in government is changing the rules of the game. A key area for public debate and academic inquiry is how to adapt existing principles of administrative and constitutional law to the new playing field. We have to be vigilant to this transformation and adopt the necessary measures in time. Otherwise, we may soon find ourselves trapped in the rationality of algorithms and missing some human arbitrariness.

The Contentious Issues of Governance by Algorithms

GILLES J. GUGLIELMI*

The development of computerized tools that lead to decision-making processes which apply locally defined parameters poses many questions about democracy. These questions stem from our very conception of the state and its role, going beyond the boundaries of typical administrative law. According to a popular notion that permeates the practices of most executive branches in liberal political regimes, democratic concerns are now competing with managerial concerns.

In order to analyze this idea, we must study the implementation of algorithms in administrative decision-making, underscoring both the changes to the characterization of administrative decisions and the questions raised about an administrative judicial review of litigation.

To summarize a French administrative law judge's review so far, the judge began by assessing the legality of using algorithms in administrative procedures. Secondly, the judge reviewed the legality of making administrative decisions on the basis of an algorithm.

Three issues now appear to be guiding the future of algorithm-based administrative decisions: (1) the security of legal transactions; (2) the compensation for harm or damage caused by the algorithms, and (3) the degree of in-depth review by the administrative judge.

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

INTRODUCTION

To confine ourselves to the most general administrative definition, an algorithm constitutes "the study of problem-solving by implementing series of elementary operations according to a predefined process culminating in a solution."[1]

The dematerialization of administrative procedures must precede the implementation of algorithms, both to obtain quantities of data, personal data in particular, held by users and render that data meaningful. This will also allow for the products and services supplied to users to be developed in a personalized manner. Nonetheless, the necessary aim regarding the strategic enhancement of such data was the development of algorithms in administrative circles.

Algorithms have gradually been introduced into administrative processes to get rid of repetitive tasks, detect correlations, identify risks, systematize internal control, help decision-making, and produce decisions that create rights. Thus, cases have emerged of algorithms used openly prior to a tax or social security assessment, offering an amount of compensation, assigning school-leavers to universities, or applicants to social housing.[2]

The development of computerized tools that lead to decision-making processes with locally defined parameters poses many questions about democracy that stem from our very conception of the state and its role,[3] thereby going beyond the boundaries of administrative law. According to a logic that henceforth permeates the practices of most executive branches in liberal political regimes, democratic concerns are now competing with managerial concerns.

* Gilles J. Guglielmi is a full-time professor at the University of Paris-Panthéon-Assas

[1] Arrêté 0216 du 27 juillet 1989 relatif à l'enrichissement du vocabulaire de l'informatique [Order of July 27, 1989 Relating to the Enrichment of Computer Vocabulary], Journal offciel de la République Française [J.O] [Official Gazette of France], June 27, 1989, p 11725. (Fr.)

[2] *See generally* Danièle Bourcier & Primavera de Filippi, *Les algorithmes sont ils devenus le langage ordinaire de l'administration?*, *in* LECTURES CRITIQUES DU CODE DES RELATIONS ENTRE LE PUBLIC ET L'ADMINISTRATION 193, 200-01 (Geneviève Koubi et al. eds., 1st ed. 2018) (discussing the use of algorithms in university assignments) (Fr.); Ivar Timmer & Rachel Rietveld, *Rule-Based Systems for Decision Support and Decision-Making in Dutch Legal Practice: A Brief Overview of Applications and Implications*, 103 DROIT ET SOCIÉTÉ 517 (2019), https://www.cairn.info/revue-droit-et-societe-2019-3-page-517.htm (discussing the use of algorithms in tax and social security assessments) (Fr.).

[3] *See generally* Arnaud Sée, *La régulation des algorithmes: un nouveau modèle de globalisation?*, 5 REVUE FRANÇAISE DE DROIT ADMINISTRATIF [R.F.D.A.] 830 (2019), https://halshs.archives-ouvertes.fr/halshs-02450617/document.

I. FROM DEMOCRATIC CONCERNS TO MANAGERIAL CONCERNS

The computational power of algorithms can be a formidable aid to the decision-making power vested in executive and administrative authorities.[4] On the one hand, the power of algorithms increases the asymmetrical (or inegalitarian) nature of administrative law and its litigation processes, as individuals are truly on their own in the face of a machine whose means far outweigh the individual's means. On the other hand, it impinges on an ethical conception of administrative decisions, which results from a humanistic free will applied to a personalized—and thus unique—situation of its recipient.

The history of the review of administrative action shows that the main concern of administrative judges and legal theory was to make administrative decisions subject to the law for roughly a century, from the Council of State's 1860 procedural regulation to just after the Second World War.

Then, at a subsequent stage corresponding to the rapidly expanding adversarial principle (due hearing of both parties), then the right to a fair trial[5] from the Trompier Gravier ruling (1944) to the Didier ruling (1999),[6] administrative law and its litigation processes favoured a primarily procedural conception. This trend mirrored the global trend of defining globalized administrative law through transparency, participation, motivation of decisions, accountability, right to appeal, and some review standards, such as proportionality, matching the means to the end, the nonuse of needlessly restrictive means, and legitimate expectations.

Finally, in a third phase, which is particularly perceptible in France since the implementation of various public policies for reforming the state and public services from 1995 onwards, administrative decisions have been gradually guided by the notion of quality. This notion stems from company organization sciences and is based on the match between outcomes and objectives, the cut in operating costs, or the satisfaction of users. Administrative decisions are thus taking a primarily managerial turn, one in which due observance of substantive law and its

[4] *See generally* Sonia Desmoulin-Canselier & Daniel Le Métayer, DÉCIDER AVEC LES ALGORITHMES: QUELLE PLACE POUR L'HOMME, QUELLE PLACE POUR LE DROIT?, Dalloz, coll. "*Les sens du droit*" (2020) (Fr.).

[5] *See generally* Scarlett-May Ferrié, *Les algorithmes à l'épreuve du droit au procès équitable*, LA SEMAINE JURIDIQUE – EDITION GÉNÉRALE 1 (2018) (questioning compatibility of algorithms and right to a fair trial) (Fr.).

[6] *See* CE Sect., May 5, 1944, Rec. Lebon 133, 256; CE Ass., Dec 3, 1999, 207434, Rec. Lebon 399; REVUE FRANÇAISE DE DROIT ADMINISTRATIVE {FRENCH ADMINISTRATIVE LAW JOURNAL] [RFDA] 2000, 584, concl. Seban; AJDA, 2000, 126, chron. M. Guyomar & P. Collin.

fundamental justification takes second place to due observance of the procedure, which is to act as a safety umbrella, mollify the satisfaction of users, and meet the quantitative criteria of accounting efficiency. The managerial concern of administrative decisions is, for that matter, akin to that of the administrative jurisdiction that is supposed to review it. The latter has indeed already initiated the trial movement toward algorithms,[7] but this very specific question is not dealt with here.

II. THE KEY AREAS UNDER REVIEW

We must study the implementation of administrative decision algorithms, underscoring both the modifications they induce in the representation of an administrative decision,[8] and the questions they raise in an administrative judge's review of litigation.

For that purpose, there is no need to think in-depth about the notion of artificial intelligence (AI), the scope of which is both too broad and inconsequential to reasonably compare specific and specialized analytical tools. According to the experts, current AI applications are the product of weak AI.[9] AI processes are based on algorithms, lists of instructions, and rules that bring out decisions, either directly or with the aid of probabilities. Thus, it now suffices to focus materially on algorithms to set out the terms of a judge's review problem, including the focus on protecting human rights.[10] The task is facilitated, as it were, because, unlike certain fantasies, there is no paradigm shift.

The current, widespread trend consists in promoting professional ethics. Ethics and prevention tend to divert our attention from the real difficulties the theory of law encounters in comprehending the

[7] *See generally* Marc Clément, *Algorithmes au service du juge administratif: peut-on en rester maître?*, A.J.D.A. 2453 (2017), https://www.dalloz.fr/lien?famille=revues&dochype=AJDA%2FCHRON%2F2017%2F3339 (discussing current algorithm use in French databases and American sentencing and planned algorithm use in British online courts) (Fr.).

[8] *See generally* Ackiel Boudinar-Zabaleta, La décision administrative algorithmique, 5 La revue droit pub. approfondi 7, 8 (2017), https://en.calameo.com/read/0045851905e6d ec0abb2b (Fr.).

[9] *See generally* Axel Cypel, AU CŒUR DE L'INTELLIGENCE ARTIFICIELLE: DES ALGORITHMES À L'IA FORTE (1st ed. 2020) (Fr.) (We speak of strong AI when this discipline gives the machine a mind of its own, and beyond self-learning, instils in it a form of consciousness.).

[10] *See generally* COMMISSION NATIONALE CONSULTATIVE DES DROITS DE L'HOMME, OPINION ON THE IMPACT OF ARTIFICIAL INTELLIGENCE ON FUNDAMENTAL RIGHTS (A-2022-6), at 12–13 (2022), https://www.cncdh.fr/sites/default/files/2022-05/A%20%202022%20%206%20%20%20EN%20%20Artificial%20intelligence%20and%20fundamental%20rights%2C%20april%202022.pdf (the C.N.C.D.H. is the French National Consultative Commission on Human Rights).

modifications that algorithms make to administrative decisions. This idea is borne out by the ethical principles of fairness, confidence, and vigilance, which should encompass concepts whose exact legal nature is already uncertain, like human dignity or privacy, as underlined by both the French data protection authority (CNIL)[11] and legal theory.[12] The resulting self-regulation undeniably strengthens a beneficial preventive effect already found in law (Article 121 of the French data protection law [*loi informatique et libertés*]).[13] This self-regulation also follows the recent recommendation of the French National Consultative Commission on Human Rights (CNCDH) to carry out an impact assessment to introduce algorithms in the administrative decision-making process. However, like any type of "compliance" devised for private sector players, it appears to disregard the fact that, for administrative authorities, these obligations are included in their observance of the rule of law. As compliance is not primarily based on law, it cannot be the sole nor the best review mode.[14]

Accordingly, in terms of the law, the new problems that arise are of the same nature as these problems in the past.[15] This is because the question of responsibility and its apportionment is still posed: whether the decision is made by a machine or whether the machine decision is supported by the named person having authority. Whether the decision is made by a machine or a delegation, the question of the legality of the decision arises.[16] In either case, it is the outcome of the process, a legal instrument or a fact having a legal effect, that the legal system applies.

[11] *See generally* COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTES, COMMENT PERMETTRE A L'HOMME DE GARDER LA MAIN? LES ENJEUX ETHIQUES DES ALGORITHMES ET DE L'INTELLIGENCE ARTIFICIELLE [HOW TO ENABLE HUMANS TO STAY IN CONTROL? THE ETHICAL ISSUES OF ALGORITHM AND ARTIFICIAL INTELLIGENCE] (2017), https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf (reporting the public debate as part of its ethical reflection remit granted by the Law for a Digital Republic).

[12] *See generally* Fanny Grabias, *La transparence administrative, un nouveau principe?* [*Administrative Transparency, a New Principle?*], 50 JCP A 2340 (2018) (Fr.).

[13] *See* Loi 78–17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés [Relating to Data, Processing, Files and Freedon], JOURNAL OFFICIEL DE LF RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Jan. 7, 1978, p. 227.

[14] *See generally* David Forest, *La régulation des algorithmes, entre éthique et droit* [*The Regulation of Algorithms, Between Ethics and Law],* 137 LAMY DROIT DE L'IMMATÉRIEL 38 (2017) (Fr.).

[15] *See generally* Primavera de Filippi, *Repenser le droit à l'ère numérique : entre la régulation technique et la gouvernance algorithmique* [*Rethinking the Law in the Digital Age: Between Technical Regulation and Algorithmic Governance*], 3 DROIT ET MACHINE 33 (2017) (Fr.).

[16] *See generally* Jean-Baptiste Duclercq, *L'automatisation algorithmique des décisions administratives individuelles* [*The Algorithmic Automation of Individual Administrative Decisions*], 2019 RDP 295 (Fr.).

In either case, there is tension between the law and technology, between legal IT and IT law.

III. THE CONTENT OF THE REVIEW

To sum up the review conducted by a French administrative judge to date, the judge first assessed the legality of using algorithms in the administrative procedure. Second, the judge reviewed the legality of administrative decisions made on the basis of an algorithm.

For the procedure, the key points of the review are the transparency and intelligibility of the use of algorithms. This is because the procedure is indeed the point at which individuals become aware of how their applications or claims are dealt with, and it was also through the procedure that the French lawmaker began to regulate the use of algorithms.[17]

An additional point is the review of the algorithm itself, which proves to be a more delicate task. The transparency of the use of algorithms and an understanding of their scope in no way implies the transparency of the algorithm itself. Private sector operators have always refused to disclose their source code and other items protected in their view by patents or trade secrets. Furthermore, the algorithm is merely the form of data processing. So, the challenged act, the basis for the processing, must also include an authorization to implement the algorithm after it has been developed. Lastly, the algorithm, as an automated process that directs behaviour and leads to internal optimization standards for case-review criteria, could be governed by soft law as "guidelines."

For the review of a final administrative decision made on the basis of an algorithm, it is based first and foremost on the legal fiction that regards the competent administrative authority as the author of the administrative act enacted on the basis of an algorithm. The fact is that the administrative authority is not technically the author of the algorithm itself. In most cases, the administrative authority engages

[17] *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 RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Oct. 7, 2016 (Fr.) (creating Code des relations entre le public et l'administration [Code of Relations between the Public and the Administration] Article L311-3-1); *see* 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 March 14, 2017 relating to the rights of persons subject to individual decisions made on the basis of algorithmic processing], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], March 14, 2017 (Fr.) (creating Code des relations entre le public et l'administration [Code of Relations between the Public and the Administration] Article R311-3-1-2).

private sector operators to develop and supply such software under contractual terms that may allow the company supplying the algorithm to retain its intellectual property rights. In this sequence of operations, the question then arises of whether the final administrative decision really results from the will of its author.

The legal basis for a decision was also examined and revealed a hidden standard-setting level that results from a multitude of implicit microstandards. These microstandards simply supplement the legal requirements, because constituting a category of IT experts for the decision produces its own interpretation of concepts that thereby changes from legal language to natural language to computing language. The existence and legality of administrative decisions produced by an algorithm are thus not under threat, but it goes without saying that the administrative authority must then assume responsibility for its decisions if the use of the algorithms leads to inegalitarian, inappropriate, detrimental, or unlawful decisions.

CONCLUSION

Algorithms can thus be presented as part of a set of standards or as part of another approach that belongs to the category of soft law. In either case, their submission to the rule of law is not in doubt, but the practical arrangements for circumscribing them and analysing the judicial route to reach them are subtle and have yet to be mapped out.

And yet, this point is crucial because the use of algorithms by administrative authorities, in both service and review activities, can only intensify in the "public transformation," in two main ways. First, algorithms seem to be the most appealing digital tool for public administrators in crisis situations.[18] Second, algorithms integrate themselves into daily and repeated contact between individuals and administrative authorities.[19] Irrespective of the increasing complexity and density of standards that result from tools restricting access, freedom, and potential infringements of fundamental rights produced by total control of bodies and behaviours (as in social scoring or widespread

[18] *See generally* Véronique Guillotin et al., RAPPORT D'INFORMATION FAIT AU NOM DE LA DÉLÉGATION SÉNATORIALE À LA PROSPECTIVE (1) SUR LES CRISES SANITAIRES ET OUTILS NUMÉRIQUES : RÉPONDRE AVEC EFFICACITÉ POUR RETROUVER NOS LIBERTÉS [INFORMATION REPORT MADE ON BEHALF OF THE SENATORIAL DELEGATION FOR FORESIGHT (1) ON HEALTH CRISES AND DIGITAL TOOLS: RESPONDING EFFECTIVELY TO REGAIN OUR FREEDOMS], S. REP. NO. 673 (2021), https://www.senat.fr/rap/r20-673/r20-6731.pdf (discussing algorithm use during times of crisis in Asian and European countries) (Fr.).

[19] *See generally* Boris Barraud, *L'algorithmisation de l'administration* [Algorithmization of Administration], 150 REVUE *LAMY DROIT DE L'IMMATÉRIEL*, 42 (2018)(Fr.).

biometric recognition in public spaces), the software or communication tools used led public corporations to use almost exclusively private sector operators for most of such activities. Relations with individuals are thus heavily weighed in favour of the administrative authorities, which control and impose increasingly restrictive and intrusive procedures, going as far as favouring arbitrariness through perfectly prepared decisions that preclude any human adaptation.

Three issues now appear to be guiding the future of administrative decisions when they rely on algorithms. The first issue pertains to the security of legal transactions, which requires digital tools to be reliable enough to be the foundation for foreseeable decision without eroding the confidence citizens have in public authorities. The second issue centers around creating a compensation system, shaped by judicial review and appropriate principles, for any harm or damage caused by the algorithms. The final issue, which presents a problematic question at this stage, concerns the degree of in-depth review completed by an administrative judge who traditionally resists examining expert consideration. At the same time, through preventative ethics, a regulating program is developing that would give administrative law the ability to regulate all of the powerful executive branch's administrative activity—a guarantee of the effectiveness of democracy.

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.defenseurdesdroits.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-government/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.

On Facial Recognition, Regulation, and "Data Necropolitics"

ANTONIO PELE & CAITLIN MULHOLLAND*

ABSTRACT

This paper argues for actual and legal regulation of artificial intelligence (AI) and facial recognition. These new technologies represent great opportunities to improve the welfare of societies. However, some of their uses can also enhance discrimination and, eventually, lead to violence. From a comparative approach (examining the European Union and Brazil), we address the current and future aspects of facial regulation, AI, and personal data. This paper shows that regulation is relevant to protect the rule of law, free markets, and individual freedoms. It also examines the looming risks unfolding from the unregulated uses of new technologies. Our concept of "Data Necropolitics" defines a predatory form of digital governance that exploits and discriminates against vulnerable populations.

INTRODUCTION

Increasing literature has been highlighting how our societies and subjectivities are being modified and threatened by new technologies,[1]

*Antonio Pele is Associate Professor at the Law School at the Pontifical Catholic University at Rio de Janeiro, Brazil (PUC-Rio), and Marie Curie Fellow EHESS/IRIS, Paris (2021–23) with the E.U.-funded project HuDig19. DOI: 10.3030/101027394. Email: apele@puc-rio.br

Caitlin Mulholland is Associate Professor and Head of the Law School at the Pontifical Catholic University at Rio de Janeiro, Brazil (PUC-Rio).

Email : caitlinsm@puc-rio.br

[1] *See generally* Antoinette Rouvroy & Bernard Stiegler, *The Digital Regime of Truth: From Algorithmic Governmentality to a New Rule of Law*, LA DELEUZIANA, no. 3, 2016, at 6; BERNARD E. HARCOURT, EXPOSED: DESIRE AND DISOBEDIENCE IN THE DIGITAL AGE (2015); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019); FRANCK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); CÉDRIC DURAND, TECHNO-FÉODALISME - CRITIQUE DE L'ÉCONOMIE NUMÉRIQUE (2020);

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

including "Algorithmic Governmentality" (A. Rouvroy), "Expository Society" (B. Harcourt), "Black Box Society" (F. Pasquale), "Surveillance Capitalism" (S. Zuboff), "Techno-Feudalism"(C. Durand). The present article is inserted in these debates and examines more particularly the role of legal regulation regarding AI and facial recognition. From a comparative approach, it explores the regulation of such fields in Brazil and in Europe. This paper argues that regulation is essential since it is the only way to protect the fundamental basic rights of individuals (e.g., privacy) while avoiding potential discrimination unfolding from socioeconomic and racial biases. Those questions will be addressed in the first part (Part I) of the paper. The second part (Part II) argues that the lack of regulation can lead to violence and, eventually, death. Exploring specific cases where new technologies are related to digital surveillance and military activities, we highlight the dangers of what is called "Data Necropolitics," namely, a predatory and digital form of governance.

PART I

I. STARTING THE DEBATE: PERSONAL DATA PROTECTION AND ALGORITHMIC NONDISCRIMINATION

In mid-2010, a Taiwanese family purchased a camera from Nikon and found a malfunction.[2] The product had a feature to prevent selfies with eyes closed, which were confused with the eyes of Asians. As a result, whenever family members tried to photograph each other, a message flashed on the screen asking: "Did someone blink?" This led them to think, at first, that the camera was broken. However, the messages stopped when one of the brothers posed with his eyes wide open. There, it was possible to verify that the intelligent face detection technology, initially designed to make photography more efficient, had a design error that exhibited an occasional bias towards the faces of Caucasians.

Such face detection technology was a feature that quickly gained traction in various smart technological devices. In 2015, Google launched Google Photos, a sharing and storage service designed to provide users free, unlimited photo and video storage. This service applies the technology of markings on images through its AI software

ÉRIC SADIN, L'INTELLIGENCE ARTIFICIELLE OU L'ENJEU DU SIÈCLE: ANATOMIE D'UN ANTIHUMANISME RADICAL (2021).

[2] Adam Rose, *Are Face-Detection Cameras Racist?*, TIME (Jan. 22, 2010), http://content.time.com/time/business/article/0,8599,1954643,00.html.

with the computer vision technique.[3] For example, in one of the automatic tagging processes, the application labelled two black men as "gorillas."[4] At the time, the company justified the problems in recognizing images due to "obscured faces" and the need for "different contrast processes for different skin tones and lighting," and presented promises of long-term fixes.[5]

Still, computer vision and facial recognition have been applied in policing several cities around the world. In Brazil, the practice started in December 2018 by the secretary of public security of Bahia in the cities of Feira de Santana and Salvador.[6] Since its implementation, facial recognition technology has led to approximately 200 arrests in the region.[7] There were also false positives among the more than 4.3 million recorded images.[8] For example, a seventeen-year-old teenager was approached inside a subway station to comply with an arrest warrant for drug trafficking. Upon arriving at the police station, police discovered that the boy's identity was incompatible with the subject identified by the recognition system and that they had apprehended the boy in error.[9] In another situation, a twenty-five-year-old man with

[3] Computer vision is the field of AI that trains computers to interpret and understand the visual world. Depending on programming, machines can identify and classify elements such as objects, animals and people, through images and videos and, together with deep learning models, even react to what they see. In other words, they are systems designed for rapid detection and reaction to visual stimuli. *Computer Vision: What it is and Why it Matters*, SAS, https://www.sas.com/pt_br/insights/analytics/computer-vision.html (last visited Jan. 20, 2023).

[4] Jana Kasperkevic, *Google Says Sorry for Racist Auto-tag in Photo App*, THE GUARDIAN (July 1, 2015, 1:52 PM), https://www.theguardian.com/technology/2015/jul/01/google-sorry-racist-auto-tag-photo-app.

[5] *Id.*

[6] It works through a comparison system: if the images captured in real-time are more than 90% compatible with those available in the wanted database, alerts are generated to professionals who call teams on the streets to confirm the identity of the suspects and follow up to the execution of the arrest warrant. Marcia Santana, *Facial Recognition Completes One Year and is a National Highlight*, SECRETARIA DE SEGURANÇA PÚBLICA DE ESTADO DA BAHIA (Dec. 18, 2019), http://www.ssp.ba.gov.br/2019/12/6981/Facial-Recognition-completes-one-year-and-and-national-highlight.html (Braz.).

[7] *Homem é preso em Salvador após ser identificado pelo sistema de reconhecimento facial*, G1 (Mar. 14, 2021, 8:24 AM), https://g1.globo.com/ba/bahia/noticia/2021/03/14/homem-e-preso-em-salvador-apos-ser-identificado-pelo-sistema-de-reconhecimento-facial.ghtml.

[8] Samuel Celestino, *Facial Recognition System Has Already Recorded More than 4.3 Million Images*, BAHIA NOTÍCIAS (Feb. 24, 2020, 8:00AM), https://www.bahianoticias.com.br/noticia/244624-sistema-de-reconhecimento-facial-ja-registrou-mais-de-43-milhoes-de-imagens.html (Braz.).

[9] Tarcízio Silva, *Reconhecimento Facial na Bahia: mais erros policiais contra negros e pobres* [*Facial Recognition in Bahia: More Police Errors Against Blacks and the Poor*],

special needs was approached by police forces because the facial recognition system pointed him out as someone with an outstanding arrest warrant.[10]

Although such facial recognition technology (FRT) is not a novelty, having already been used in security systems of banking applications and cell phones, for example, the potential of its use for specific purposes—such as investigation and criminal prosecution—has brought about debates over control and surveillance, which takes us back to Bentham and the Panopticon theory,[11] and Foucault and his theory on social control and the history of the penitentiary systems.[12]

Machine-learning programs allow the development of facial recognition technology that promotes autonomous decision-making ability free from human interference. It becomes possible through the treatment of bulk data (pictures of people, for example) and self-learning development of machines (i.e., programs and systems) that allow the achievement of specific results (outputs) independently of any mediation by a human being. Such a decision could concretely deny or impede rights or generate abusive or illegitimate discrimination. However, machine-learning applications "are adopting machine-learning systems at unprecedented rates due to the technology's ability to radically improve data-driven decision-making at a cost and scale incomparable to that of humans."[13] As a consequence, their comprehensiveness makes them play an essential role in regulating our lives. For example, the judicial system can use them to assess the probability that a subject will relapse into a particular crime. Banks can decide whether or not an individual should be granted a mortgage. Governments can rely on machine learning to determine market reallocation strategies. It is this scope of situations, and the possible effects their results have generated, that have intensified questions about transparency and accountability.

These questions are natural because those technologies are not easily understandable to humans, especially in the ways they function

TARCÍZIO SILVA (Nov. 21, 2019), https://tarciziosilva.com.br/blog/reconhecimento-facial-na-bahia-mais-erros-policiais-contra-negros-e-pobres (Braz.).

[10] Amanda Palma & Clarissa Pacheco, *'O policial já foi com a arma na cabeça dele', diz mãe de rapaz confundido por reconhecimento facial* [*'The Policeman Already Came with a Gun Pointed to his Head', Says The Mother of a Boy Identified by Facial Recognition*] CORREIO (Jan. 5, 2020, 9:00am), http://glo.bo/3TFduBt (Braz.).

[11] *See* JEREMY BENTHAM, THE PANOPTICON WRITINGS (Miran Božovič ed., 2011).

[12] *See* MICHEL FOUCAULT, DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON 82 (Alan Sheridan trans. 1977).

[13] Bryan Casey et al., *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, BERKELEY TECH. L.J. 145, 150 (2019).

and how their results are justified. Another concern revealed in the study of algorithms, AI, and facial recognition is the belief that "predictive algorithms rationalize the decision-making process by summarizing all relevant information in a more efficient way than the human brain."[14] The myth about the objectivity, neutrality, rationality, and impartiality of the application of such technology has been gradually deconstructed. Research has shown that existing biases in human culture are inevitably replicated in technology, as they produce, on a large scale, prejudices and stereotypes that negatively affect the mediation between the human and the machine.[15] Just as we humans are subject to heuristics and biases in our decision-making, the algorithms are too.[16]

Allied to this false idea of technology neutrality is the exponential growth in the ability to process personal data of the most diverse orders, precisely because of the advent of advanced artificial intelligence technologies, with the use of sophisticated algorithms and the possibility of machine learning. The treatment of "big data"—literally, large databases—through increasingly developed computational techniques can lead to probabilistic results that, while reaching the interests of a specific part of the population, take away the individual's capacity for autonomy and their right of access to goods, services, public policies, for example.

In this sense, the principle of nondiscrimination (provided, for example, in Article 6, IX of the Brazilian General Data Protection Law) must be reflected in all circumstances in which the use of data, whether sensitive or not, generates some misjudgment or inducement to results that would be unfair. Accordingly, this principle should serve as a basis for sustaining the protection of sensitive data, especially when we are faced with exercising democracy and access to social rights, such as the right to work, health, and housing.

One of the practices with a high potential to cause discrimination is

[14] ANGELE CHRISTIN ET AL., COURTS AND PREDICTIVE ALGORITHMS 1 (2015), https://www.law.nyu.edu/sites/default/files/upload_documents/Angele%20Christin.pdf.

[15] *See generally* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION (2016); ELI PARISER, THE FILTER BUBBLE (2011); Camila Souza Aranjo, Wagner Meira Jr. & Virgilio Almeida, *Identifying Stereotypes in the Online Perception of Physical Attractiveness, in* 1 SOCIAL INFORMATICS 419 (Emma Spiro & Yong-Yeol Ahn eds., 2016); Aylin Slam-Caliskan, Joanna J. Bryson & Arvind Narayanan, *Semantics Derived Automatically from Corporate Language Necessarily Contain Human Biases*. 356 SCIENCE, 183-86 (2017); Joy Buolamwini, *How I'm Fighting Bias in Algorithms*, TED (last visited August 8, 2022), https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms.

[16] *See generally* PLOUS SCOTT, THE PSYCHOLOGY OF JUDGEMENT AND DECISION MAKING (1993) (discussing the influence of heuristics on human decision-making).

profiling, where the controller creates the data subject's profile, which is intended to serve as an evaluation parameter on some aspects of the subject's personality. In this scenario, it is evident there is a need for forms of controlling these practices to avoid and even mitigate risks of potential discrimination, illegality, or abuse in processing personal data.

From this perspective, when faced with the processing of data that makes use of algorithmic probability and machine-learning models for decision-making, indeed what is in dispute, depending on the legal interest involved in the decision, is whether the data controller will or will not be denying or even promoting the fruition of a fundamental right to data protection. Therefore, it will be essential to know if the process of decision-making was discriminatory concerning the data subject or a social group that the subject represents (people with disabilities, the elderly, and BIPOC, among others). This evaluation is necessary to verify that the result of applying the controller's algorithm not only refrains from committing this discrimination but also whether it fails to adequately promote the right to data protection.

Considering that these applications are increasingly having a substantial impact on sensitive social areas, such as the use of data providing for the development of humanitarian aid, accurate medical diagnosis, or rationality to decisions,[17] these automated decisions may affect individual and collective rights (Article 5 of the Brazilian Federal Constitution) of data subjects, but also their social rights (Article 6 of the Brazilian Federal Constitution).

Furthermore, the principle of equality is identified as one of the axiological substrates of the general clause for the protection of the human person, foreseen as one of the foundations of the Democratic State of Law in Article 1, III of the Brazilian Federal Constitution. More than the right to equal treatment, respect for differences and unequal treatment are forms of materialization of the dignity of the human person. Observing the constitutional context, the legal protection of personal data in the Brazilian legal system is due to the need to preserve the principle of equality—and the consequent principle of noncompliance discrimination—to support eventual, existential vulnerabilities.

Considering that the collection of personal data and the creation of social profiles may lead to discrimination, data protection should be seen as "the protection of life choices against any form of public control and social stigma" (L. M. Friedman) and as "vindication of the

[17] Danilo Doneda & Virgilio Almeida, *O que é a governança de algoritmos*, *in* TECNOPOLÍTICAS DA VIGILÂNCIA: PERSPECTIVAS DA MARGEM [SURVEILLANCE TECHNOPOLITICS: PERSPECTIVES FROM THE MARGIN] 141, 143 (Fernanda Bruno et al. eds., 2019).

boundaries protecting each person's right not to be simplified, objectified, and evaluated out of context" (J. Rosen).[18] Therefore, it is concluded that personal data protection—as a result of the general clause of protection of the human person, the right to privacy, and the principle of equality—is an essential requirement for democratic exercise.

A. *FACIAL RECOGNITION AND REGULATION IN BRAZIL*

In 2018, the Brazilian General Data Protection Law (LGPD) was passed, aiming to protect the rights of holders of personal data and impose a series of obligations to be complied with by those who process data in the country. Despite not mentioning at any time the facial recognition technology or even AI systems as an object of regulation, the LGPD is the law applicable to situations where such technologies are used. The LGPD applies because these technologies use personal data to achieve the desired results. Whereas people's images (specifically the face) are understood as biometric data, facial recognition systems meet the regulatory framework already established in the LGPD. In this sense, we can indicate some aspects of the LGPD that are guidelines for regulating facial recognition technology. The first concerns the principles applied to personal data processing activities (Article 6, LGPD). Here, we can consider three principles as being of direct relevance: the principle of prevention, which matters in the adoption of measures to prevent the occurrence of damages due to the treatment of personal data; nondiscrimination, which prohibits processing for unlawful discrimination; and responsibility and accountability, which requires the data processing agent to demonstrate the adoption of effective measures to comply with personal data protection rules.

Another relevant aspect of the LGPD is the recognition of the data subject's right to request a review of decisions made solely based on automated processing of personal data that affects the person's interests or aspects of the person's personality (Article 20, LGPD). In addition, the data processing agent must provide clear information regarding the criteria and procedures used for the automated decision. Furthermore, LGPD recognizes the right of the Data Protection National Authority for carrying out an audit to verify discriminatory aspects in the automated processing of personal data.

On the other hand, a series of bills intended to regulate AI. In

[18] Stefano Rodotà, *Data Protection as a Fundamental Right*, *in* REINVENTING DATA PROTECTION? 77, 78 (Serge Gutwirth et al. eds., 2009). *See generally* Stefano Rodotà, *Some Remarks on Surveillance Today*, 4 EUR. J.L. & TECH. (2013), https://www.ejlt.org/index.php/ejlt/article/download/277/388?inline=1.

Brazil, following what is happening in Europe, there is a specific bill, the PL 21/20, which is currently being debated in the federal senate. The bill establishes foundations, principles, and guidelines for developing and applying AI in Brazil. The project has received much attention, especially for its characteristic of being a principled and conceptual law, contributing little to the concrete regulation of situations in which AI is used.

However, two references must be made to the bill: (i) the inclusion of the security and prevention principle, which requires the person who provides the AI system to use technical, organizational, and administrative measures that allow the mitigation of risks from the operation of artificial intelligence systems, as well as (ii) the obligation imposed on public administration to implement concrete risk management, taking into account the definitions of the need for regulation of artificial intelligence systems and the appropriate level of intervention. The references to the management and mitigation of risks, considered beacons for the use of AI systems and the protection of fundamental rights, generate the obligation of a continuous assessment of AI uses and applications that require thoughtful analysis of the proportionality and adequacy in the use of such systems when opposed to the fundamental interests of the human person. It is precisely for this reason that we seek to assess whether the use of facial recognition systems—notably in applications used to provide public security and allow an "efficient" criminal prosecution—is proportionate and adequate to constitutionally guaranteed fundamental rights.

B. FACIAL RECOGNITION AND REGULATION IN EUROPE

Regulatory debates on AI and facial recognition technologies are already quite mature in Europe. In 2021, a bill was proposed, called the AI Act, which aims to ensure that Europeans can benefit from new technologies developed and functioning according to European Union values, fundamental rights, and principles.

The regulation follows a risk-based approach and differentiates between uses of AI that create: (a) an unacceptable risk, (b) a high risk, and (c) a low or minimal risk. In addition, the AI Act, in Title II, establishes a list of prohibited AI practices. The list includes all AI systems whose use is considered unacceptable for violating the values of the EU— for example, violating fundamental rights. The bans cover practices with significant potential to manipulate people through subliminal techniques that go unnoticed or explore the vulnerabilities of specific groups, such as children or people with disabilities, to materially distort their behaviour in a way that is likely to cause psychological or physical harm to them or another person. Other manipulative practices or exploratory approaches that are made possible by AI systems and that affect adults can be covered by legislation on data protection, consumer protection, and digital services, which ensures that individuals are adequately informed and are free to decide not to be subject to profiling or other practices that may affect their behaviour. The proposal also prohibits social classification based on AI for general use by public authorities. Finally, the use of "real time" remote biometric identification systems (FRTs) is not permitted in spaces accessible to the public when the objective is to maintain public order. This practice is considered particularly intrusive on the rights and freedoms of the data subjects, as they can affect the private life of a large part of the population, give rise to a sense of constant mass surveillance, and indirectly deter the exercise of freedom of assembly and other fundamental rights.

Considering the high risk that the use of FRTs brings to the exercise of democratic rights, the European Data Protection Board (EDPB) has called for FRTs to be banned from use under the proposed EU AI Act. The EDPB considers AI-supported facial recognition systems categorizing individuals based on their biometrics into clusters according to ethnicity, gender, and political or sexual orientation as incompatible with the European Charter of Fundamental Rights. In addition, the EDPB considers that "processing of personal data in a law enforcement context would rely on a database populated by a collection of personal data on a mass scale and in an indiscriminate way, e.g., by 'scraping' photographs and facial pictures accessible online,"[19] in particular those made available via social networks, would, as such, not meet the strict necessity requirement provided for by Union law.

On the other hand, there is another proposal for a moratorium that intends to be sent to the European Parliament to regulate the uses of AI in criminal law and its use by police and judicial authorities in criminal matters (2020/2016(INI)).[20] The parliament aims to regulate the uses of AI technologies, specifically, the FRT, which is already being used to search databases of crime suspects, in addition to carrying out forecasting (predictive policing and analysis of crime points) with behaviour detection tools. According to parliament, applications of AI

[19] EUROPEAN DATA PROTECTION BOARD, GUIDELINES 05/2022 ON THE USE OF FACIAL RECOGNITION TECHNOLOGY IN THE AREA OF LAW ENFORCEMENT (2022), https://edpb.europa.eu/system/files/202205/edpbguidelines_202205_frtlawenforcement_en_1.pdf.

[20] Resolution of 6 October 2021 on Artificial Intelligence in Criminal Law and its Use by the Police and Judicial Authorities in Criminal Matters, EUR. PARL. DOC. A9-0232/2021 (2021).

technology to law enforcement can have varying degrees of reliability and accuracy that can impact fundamental rights and the dynamics of criminal justice systems.

According to that document, the European Data Protection Board and the European Data Protection Supervisor request a moratorium on the "the deployment of facial recognition systems for law enforcement purposes that have the function of identification, unless strictly used for the purposes of identification of victims of crime."[21] Such a motion aims to set a deadline within which the technical standards for the use of this technology must be examined with full respect for fundamental rights and must not lead to prejudice and discrimination that could hinder the exercise of democracy.

C. INITIATIVES FOR A REGULATION OF THE USE OF FACIAL RECOGNITION TECHNOLOGIES

Among the initiatives to regulate facial recognition technologies, some can already be put into practice. First, laws that protect personal data or declarations of fundamental rights bring moral postulates that are recognized as principles (i.e., transparency, accountability, equality, etc.). On the one hand, the recognition of these principles is paramount for the protection of fundamental rights; on the other hand, their low enforceability leaves something to be desired when delimiting the uses of technologies. Nevertheless, in the absence of a "hard law," those ethical or moral postulates are welcome as a first effort to regulate the use of FRTs.

The first proposal concerns the so-called principle of necessity and data minimization that intends to limit the collection and storage of personal data to the essential minimum to achieve the purposes indicated in the processing of personal data. Moreover, as a result of the recognition of the principle of transparency and accountability, it is required of organizations that use personal data processing technology to establish clear rules on the purpose and legal bases for the processing of those data, that is, the purpose of their use. Consequently, the holder can reject its use if it is employed in an abusive or illegal manner. One way to implement the principle of accountability is precisely the definition of transparent rules for data sharing, informing the holder of personal data in advance of such procedures. It is also of paramount importance to limit the processing of biometrics data in a single database and ensure that security systems information is robust and follows standards established by the community, in addition to allowing

[21] *Id.* at art. 27.

that external bodies audit databases and personal data processing operations.

However, when regulatory standards are not implemented, our society lives in a legislative vacuum that has allowed the increasingly invasive use of technologies of facial recognition.

II. PART II

In this part, we argue that some use of technologies and digital surveillance—especially facial recognition—can lead to violence and, eventually, death. To explain our approach, we rely on Achille Mbembe's notion of necropolitics. Through our updated and novel interpretation of Mbembe's insights, we hold that new necropolitical interventions are relying on the use of data to subjugate, discriminate, and, eventually, eliminate given individuals. We call this phenomenon data necropolitics. To unpack our argument, we will first briefly explain Achille Mbembe's idea of necropolitics.

A. "THE POWER TO TAKE LIFE"

The insight that death and violence still play a relevant political role in governing given populations has fostered numerous academic debates. Since Foucault's perspective on biopower, authors such as Giorgio Agamben, Roberto Esposito, Mike Hill, and Warren Montag have insisted, respectively, on how the "power to take life" is still pervasive in the exercise of sovereignty, modern science, and liberal economics.[22] Achille Mbembe has radicalized these perspectives, since it would be possible to understand genocides, famines, refugee crises, civil war, and so on, under a common paradigm, namely, necropolitics. This idea refers to the "subjugation of life to the power of death."[23] Indeed, following Mbembe, a different sort of "weapons are [now] deployed in the interest of maximum destruction of persons and the creation of *death-worlds*, new and unique forms of social existence in which vast populations are subjugated to conditions of life conferring upon them the status of *living-dead*."[24]

It is possible to detect in Mbembe's scholarship that the making of these *death-worlds* is produced through the interplay of at least four

[22] MIKE HILL & WARREN MONTAG, THE OTHER ADAM SMITH 235-342 (2014). *See generally* GIORGIO AGAMBEN, HOMO SACER: SOVEREIGN POWER AND BARE LIFE (Daniel Heller-Roazen trans., 1998). *See generally* ROBERTO ESPOSITO. BÍOS. BIOPOLITICS AND PHILOSOPHY (Timothy C. Campbell trans., 2008).

[23] *Id.*

[24] ACHILLE MBEMBE, NECROPOLITICS 92 (Steven Corcoran trans., 2019).

factors.[25]

First, necropolitics relies on *necroeconomies*. Late capitalism and neoliberalism would have produced an excess of populations that could not be exploited anymore, and, as a consequence, require management through their constant exposure to dangers. The climate crisis, erosion of socioeconomic rights, and unstable working conditions would be the most illustrative examples of this necroeconomy**.**

Second, necropolitics implies the confinement of given populations in specific territories, namely, campsites. Mbembe holds that the camp-form (refugees, prisons, banlieues, suburbs, favelas) are now the dominant technique to govern undesirable populations.

Third, necropolitics keeps on expanding in societies thanks to racism. This can have different forms (institutional, systemic, subjective), and it enables discrimination and humiliation of "anyone considered not to be one of us."[26]

Fourth, necropolitics aims at producing "death on a large scale."[27] State terror, wars, and predation of natural resources "manufacture an entire crowd of people who specifically live at the edge of life, or even on its outer edge—people for whom living means continually standing up to death . . . ."[28]

Mbembe's necropolitics have been applied and discussed in numerous fields of academic research, such as the latest pandemic, the conditions of inmates, the conditions of asylum seekers, the marginalization of indigenous people, and the climate crisis.[29] It is only very recently that scholars have intended to examine the pervasiveness of death under our current "digital revolution." Evelyn Wan refers to necropolitics to define the mining of minerals necessary to our digital

[25] *See* Antonio Pele, *Achille Mbembe: Necropolitics,* CRITICAL LEGAL THINKING (March 2, 2020), https://criticallegalthinking.com/2020/03/02/achille-mbembe-necropolitics/.

[26] MBEMBE, *supra* note 24, at 54.

[27] *Id.* at 37

[28] *Id.*

[29] *See* Bárbara L. C. V. Dias & Jean-François Y. Deluchey, *The "Total Continuous War" and the COVID-19 Pandemic: Neoliberal Governmentality, Disposable Bodies and Protected Lives*, *Law, Culture and the Humanities*, L., CULTURE & HUMANITIES 4 –5 (2020); Frédéric Le Marcis, *Life in a Space of Necropolitics*, 84 ETHNOS 74, 74–77, (2019); Ariadna Estévez, *The Politics of Death and Asylum Discourse: Constituting Migration Biopolitics from the Periphery*, 39 ALTS. 75, 77 (2014); Carl Death, *Africanfuturist Socio-Climatic Imaginaries and Nnedi Okorafor's Wild Necropolitics*, 54 ANTIPODE 240, 240–42, 245–46, 250–52, 254–55 (2022). *See generally* Sophia Martensen, *Necropolitics, Colonialism, and Indigenous Peoples in Canada,* 3 YORK UNIVERSITY CRIMINOLOGICAL REV. (2021).

infrastructure.[30] Vural Ozdemir et al. have explored how "digital death and grieving" are becoming commodities of digital culture.[31] Francesca Maria Romeo refers to "digital necropolitics" to examine "how images of the dead and the dying circulate within various digital contexts . . . ."[32]

Our discussion on data necropolitics intersects these debates and is also more ambitious since we argue that the current production and exploitation of digital data can produce a novel production of death targeting growing, vulnerable populations. Mbembe holds that necropolitics can be twofold. It is "*the generalized instrumentalization of human existence and the material destruction of human bodies and populations*."[33] Under this perspective, necropolitics implies, on the one hand, exploiting and consuming human lives through socioeconomic exploitation, and, on the other, destroying human existences through the lack of access to basic rights, or even physical elimination.

In this part, we hold that data necropolitics oscillates between these two dimensions. First, data can produce and normalize the vulnerabilities that given populations have been facing (i.e., racial bias). Second, it can legitimize and turn invisible the violence and death those same populations have been suffering. Violence should not be understood as "mere" physical aggression or violation of private property rights. It is also socioeconomic and symbolic. When we refer to data necropolitics, we have in mind not only the physical elimination of certain individuals but also a predatory/digital form of governance that exposes and produces social violence, vulnerability, and, eventually, (social) death. It circulates below and sets the foundations of our technological welfare. We will examine different fields where data necropolitics can be deployed. First, we will examine how facial recognition in Latin America and Brazil, in particular, can be understood within a data-necropolitical framework since it relies on legal vacuums and targets vulnerable populations. Second, we will interpret specific military and intelligence activities (i.e., drones) as other forms of data necropolitics. Finally, regarding health inequalities, we will understand how data necropolitics can work not only through an excess of data but also a (voluntary) lack of data concerning a given

[30] Evelyn Wan, *Labour, Mining, Dispossession: On the Performance of Earth and the Necropolitics of Digital Culture*, 15 INT'L J. PERFORMANCE ARTS & DIGIT. MEDIA 249, 251–52 (2019).

[31] Vural Özdemir et al., *Thanatechnology and the Living Dead: New Concepts in Digital Transformation and Human-Computer Interaction*, 25 OMICS 401, 402, 404 (2021).

[32] Francesca Maria Romeo, Towards a Theory of Digital Necropolitics 7 (June 2021) (Ph.D. dissertation, University of California, Santa Cruz), https://escholarship.org/uc/item/1059d63h.

[33] MBEMBE, *supra* note 24, at 68.

population.

B. *FACIAL DATA NECROPOLITICS*

According to Mbembe, necropolitics relies on “[i]nsidious techniques of mass surveillance” that create “a segmented planet of multiple speeds” where the basic (digital) rights of vulnerable populations are bluntly ignored.[34]

Facial recognition has slowly but surely been deployed in Latin America, and this example shows the prescient insights of Mbembe. The use of facial recognition in Latin America has been mostly implemented “without any kind of public consultation” and thanks to “deficient regulatory context[s],” according to the latest report of AlSur, a consortium of eleven civil society and academic organizations from Latin America.[35] Regarding the areas of application of facial recognition, public security and surveillance of public spaces are the most relevant.[36] It is also worth mentioning other areas, such as transportation, social care, and health.

In Brazil, three examples of facial recognition deployment can illustrate these trends: transportation, public security, and health care. Since 2018, the metro of São Paulo has been gathering data—through facial recognition— without the consent of its users. It was only in 2021 when the systems were deactivated, thanks to court orders (ViaQuatro and Edital de Licitação do Metrô de São Paulo).

As a second example, twenty Brazilian cities have been experimenting with facial recognition for law enforcement purposes. Brazil’s federal public authorities have designed a pilot project (Em Frente Brasil) providing, since 2019, specific public funding to cities interested in this initiative. This project relies on partnerships with foreign tech companies (mostly from China, Europe, and Israel) that have offered their surveillance equipment to this public program.[37]

Finally, the discreet but sustained deployment of facial recognition in Brazil appears in the intriguing case of the Brazilian NGO, the Central Única das Favelas (CUFA). For more than twenty years, this NGO has promoted art, education, sport, music, and leisure among Brazil’s vulnerable youth communities. Like many other NGOs, CUFA launched an initiative to distribute free food baskets in the favelas

[34] *Id.* at 50, 101.

[35] ALSUR, FACIAL RECOGNITION IN LATIN AMERICA: TRENDS IN THE IMPLEMENTATION OF A PERVERSE TECHNOLOGY 7, 8 (2021).

[36] *Id.* at 7.

[37] Jonas Valente, *Face Recognition Tech Gains Ground in Brazil*, AGÊNCIA BRASIL (Sept. 20, 2019, 2:14 PM), https://bit.ly/3KKXrOf.

during the COVID-19 pandemic. However, in contrast to other similar initiatives, CUFA also planned to use facial recognition to register the potential two million beneficiaries. A partner tech company offered its expertise to collect all the biometric data. Amid critiques raised by activists and scholars regarding the final use of the collected data, CUFA decided to give up the use of facial recognition.[38]

Those cases reveal how AI and facial recognition still rely on and produce racial bias and criminalize Afro-Brazilian and other Brazilian vulnerable populations. The cases also show the lack of transparency in the collection and storage of data.

Despite the relevance of these questions, another issue should be addressed. The lack of efficient national regulation and legal vacuums regarding the precise use of facial recognition is designed to foster the deployment of these technologies. In other words, data necropolitics, namely, the circulation of predatory and digital forms of power, depends on a deficient regulatory framework to gather data from vulnerable populations.

While the Global North, as we have seen above, has adopted relatively strong regulations regarding facial recognition and AI, like the upcoming EU regulation on AI, these technologies are being tested in Latin America and in the Global South in areas that are forbidden in the Global North. It is also with the help of companies situated in Europe, China, Israel, and the United States that data necropolitics can be performed. So far, as we have seen above with Brazil, these technologies are deployed in areas such as transportation, public security areas, and public health. Data necropolitics penetrates precisely into the breach of the social and institutional weakness of the Global South, namely, criminality/violence and socioeconomic inequalities. It is at this intersection where data necropolitics is the most predatory since it targets the most vulnerable populations of the world. Here, data necropolitics is disguised by what we call "techno philanthropic capitalism." Technological donations and trial run technological experiments aim at filling the social and economic vacuum of many Latin American and Global South societies. Some tech companies intend to consolidate their foothold, building a strong relationship with officials while massively collecting data from citizens to improve their technologies.[39] It is not only the violent data extraction

[38] Alessandro Feitosa Jr., *Por que a Cufa interrompeu o uso de reconhecimento facial após polêmica* [*Why Cufa Stopped Using Facial Recognition after Controversy*], *G1* (Apr. 27, 2021, 8:17 PM), http://glo.bo/3KIcYOW.

[39] Leo Schwartz, *Major Surveillance Firms are 'Gifting' Tools to Find a Foothold in Latin America*, REST OF THE WORLD (Aug. 12, 2021), https://bit.ly/3q7COlQ.

of "data colonialism,"[40] but also, foremost, a seeming techno-philanthropic ethos that pretends to fix state failures and help vulnerable communities.

These ongoing strategies turn the Global South and Latin America into giant and open laboratories for the experimentations of AI, facial recognition, and mass data surveillance. Because of legal weakness and political complacency, these populations are becoming the digital guinea pigs of data necropolitics. Facial recognition (and other technologies) are indeed insidious techniques that segment the planet into different populations that can be, more or less, observed and manipulated.

The effectiveness and the lack of a legal regulatory framework play a relevant role in the deployment of this predatory form of data necropolitics. Brinks, Levitsky, and Murillo have presented a comprehensive approach to *The Politics of Institutional Weakness in Latin America,* bringing to light "limited enforcement, insufficient state capacity, or societal cooperation."[41] Among the roots of "institutional weakness" in this region, the authors have underlined socioeconomic inequality, low state capacity, and economic/political volatility.

"Thus, much of Latin America may be suffering from a self-reinforcing cycle in which social inequality and economic and political instability generate institutional weakness, which, in turn, reinforces inequality and instability."[42] It is possible to add that data necropolitics relies on Latin America's institutional weakness, a process that would ultimately bring about more inequalities and suffering among the vast majority of the Latin American population.

After having examined facial recognition in Latin America through data necropolitical lenses, we will explore, in the following part, the functions of the drone and mass surveillance.

C. ON DRONES AND DIGITAL SURVEILLANCE

"By creating new military markets, war and terror have transformed into modes of production, period."[43] Necropolitics is, therefore, entrenched in late capitalism and neoliberalism. From Mbembe's interpretation, it is possible to unfold how data economy is also related to necropolitics and wars.

[40] *See generally* Nick Couldry & Ulises A. Mejias, *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*, (2019) (defining "data colonialism" and detailing how it is used in the current era of pervasive datafication).

[41] DANIEL M. BRINKS ET AL., THE POLITICS OF INSTITUTIONAL WEAKNESS IN LATIN AMERICA (2020).

[42] *Id.* at 291.

[43] MBEMBE, *supra* note 24, at 36.

The more prominent roots of data necropolitics are related to military activities and intelligence activities. The relationships between the military industries, intelligence services, and big tech are even more critical. The current global race on AI supremacy and the economic stakes underpinning data surveillance show those core political issues.

The Pentagon's Project Maven is currently involving Silicon Valley companies, such as Google, to boost and apply AI technologies in the defence project.[44] The UK intelligence services have recently signed a contract with Amazon to store sensitive data in the cloud of the US-based firm.[45] A similar agreement was signed in 2015 between the French intelligence services and the US-based firm Palantir.[46] Also, two French tech companies have been charged with complicity in torture for selling surveillance equipment to Libya and Egypt.[47]

These examples certainly reveal the competition (and collaboration) between tech companies to access profitable public contracts. Regarding the issue of this paper, these examples show how a myriad of public and private actors are collaborating (and competing) "to produce total information, the first and most important prong of counterinsurgency paradigm."[48]

Following the prescient analysis of Bernard E. Harcourt, the counterinsurgency strategies (once used in the battlefields in colonial settings and after 9/11) are now a model of national governance in most countries. Counterinsurgency tactics with the deployment of massive surveillance programs and hyper-militarized policing are now deployed against groups that are not active insurgent minorities, namely, asylum seekers, refugees, Muslims, Afro-American protesters, eco-activists, etc. Harcourt mentions three main counterinsurgency strategies: first, to collect all data and achieve total awareness; second, to eradicate the active minority; and, finally, to gain the consent of the majority of the population.[49]

It is possible to understand the increasing collaboration between

[44] Tom Simonite, *Pentagon Will Expand AI Project Prompting Protests at Google*, WIRED (May 29, 2018, 7:00 AM), https://www.wired.com/story/googles-contentious-pentagon-project-is-likely-to-expand.

[45] Helen Warrell and Nic Fildes, *Amazon Strikes Deal with UK Spy Agencies to Host Top-Secret Materials*, FINANCIAL TIMES (October 25, 2021), https://on.ft.com/3Q6oyEH.

[46] Mathieu Rosemain, *A French Alternative to Palantir Would Take Two Years to Make, Thales CEO Says*, REUTERS (October 23, 2020, 1:34 PM), https://reut.rs/3ReumNM.

[47] Sarah Elzas, *French Executives Face Torture Charges for Selling Spy Gear to Libya, Egypt*, RFI (June 22, 2021, 1:13 PM), https://www.rfi.fr/en/france/20210622-french-executives-face-torture-charges-for-selling-spy-gear-to-libya-egypt-amesys-nexa-human-rights.

[48] BERNARD E. HARCOURT, THE COUNTERREVOLUTION (2018)

[49] HARCOURT, *supra* note 48, at 13–14.

tech companies, intelligence services, and the military, under Harcourt's counterrevolution paradigm. Indeed, "the boundaries between counterinsurgency as foreign policy and counterinsurgency as domestic governance begin to crumble as more and more data is necessary for more effective data mining. As the battle against terror goes global, so do the populations to target—including our own."[50]

As a consequence, counterrevolution produces an increasingly social, political, and digital vulnerability that targets the behaviour of given populations. Timnit Gebru and the DAIR Institute have revealed how AI can foster racism and may harm vulnerable groups.[51] Shaka McGlotten advances the idea of "Black data" to grasp how Black people are marginalized by big data through race.[52] There is a growing scholarship examining racial, ethnic, and socioeconomic bias in the digital world.[53] In any case, our notion of data necropolitics intersects Harcourt's concept of counterrevolution and both shape forms of governance that enhance the discriminations that vulnerable populations have been suffering.

One of the radical forms of data necropolitics, namely, the ability to kill remotely and automatically, is epitomized by drone strikes. "Death by data" shows the role of algorithms in targeted killings.[54] A 2021 report of the UN Panel of Experts on Libya suggests that in March 2020 the first attack launched automatically by an AI-based drone was registered.[55] While Western countries and China are massively investing in Lethal Autonomous Weapons Systems (LAWS), the UN has declared that their use should be prohibited by international law. However, powerful states refuse any sort of regulation since these new forms of weapons are becoming crucial to their respective and alleged

[50] *Id.* at 66.

[51] *See Research Philosophy*, DAIR INST., https://www.dair-institute.org/research (last visited Oct. 29, 2022).

[52] Shaka McGlotten, *Black Data*, THE SCHOLAR AND FEMINIST ONLINE (Feb. 13, 2014), https://sfonline.barnard.edu/shaka-mcglotten-black-data.

[53] *See generally* LIZZIE O'SHEA, FUTURE HISTORIES: WHAT ADA LOVELACE, TOM PAINE, AND THE PARIS COMMUNE CAN TEACH US ABOUT DIGITAL TECHNOLOGY (2019); Greta Byrum & Ruha Benjamin, *Disrupting the Gospel of Tech Solutionism to Build Tech Justice*, STAN. SOC. INNOVATION REV. (June 6, 2022), https://ssir.org/articles/entry/disrupting_the_gospel_of_tech_solutionism_to_build_tech_justice; Stephen Kearse, *The Ghost in the Machine: How new technologies reproduce racial inequalitie*s, THE NATION (June 15, 2020), https://www.thenation.com/article/culture/ruha-benjamin-race-after-technology-book-review.

[54] Jennifer Gibson, *Death by Data: Drones, Kill Lists and Algorithms*, E-INTERNATIONAL RELATIONS (Feb. 18, 2021), https://www.e-ir.info/2021/02/18/death-by-data-drones-kill-lists-and-algorithms.

[55] Joe Hernandez, *A Military Drone with a Mind of its Own was Used in Combat, U.N. Says*, NPR (June 01, 2021, 3:09PM), https://n.pr/3Q3BHym.

national security.

Experts and activists have, therefore, warned against the non-prohibition of LAWS, since they would potentially trigger more violence. Indeed, in words that are tragically similar to Mbembe's necropolitics, LAWS could "facilitate violence on a large scale."[56] Additionally, "with facial recognition and other technologies, they can target individuals or groups . . . which could appeal to violent groups and state militaries committing political assassinations and ethnic cleansing."[57] Finally, "LAWS may make it easier for those who control them to hide their identities."[58]

As Gregoire Chamayou presciently suggests in *A Theory of the Drone*, while ethics, in general, refers to the set of doctrines of living well and dying well, a "necroethics," namely, the ability of "killing well," is shaping our understanding of current and future wars.[59] The "necroethics of the drone [and LAWS] abandon[] any discussion of fundamental issue" since "the targets are presumed guilty until they are proved innocent—which, however, can only be done posthumously."[60] Consequently, following Chamayou, "by ruling out the possibility of combat, the drone destroys the very possibility of any clear differentiation between combatants and noncombatants."[61]

Simultaneously and more profoundly, the development of autonomous weapons has broader consequences for our societies. Indeed, it is worth reminding that "[t]he State's dependence on the bodies of the lower classes to wage war was also one of the factors that made it possible for those classes to establish a durable bargaining of power."[62] In other words, the history of the welfare state is a result of warfare as Thomas Piketty and Michel Foucault have notoriously shown it under biopolitical lenses.[63] It is relevant to understand that under the deployment and logic of LAWS "the promise to preserve national lives goes hand in hand with the *increased social vulnerability and precariousness of many of those lives*."[64] Therefore, with our notion of

[56] Robert F. Trager, *Killer Robots Are Here—and We Need to Regulate Them*, FOREIGN POLICY (May 11, 2022, 1:46 PM), https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation.

[57] *Id.*

[58] *Id.*

[59] GRÉGOIRE CHAMAYOU, A THEORY OF THE DRONE 146 (Janet Lloyd, trans., 2015).

[60] *Id.*

[61] *Id.* at 147.

[62] *Id.* at 193.

[63] *See* THOMAS PIKETTY, CAPITAL IN THE TWENTY-FIRST CENTURY (Arthur Goldhammer, trans., 2017); MICHEL FOUCAULT, THE BIRTH OF BIOPOLITICS (Michael Senellart, ed., Graham Burchell, trans., 2010).

[64] CHAMAYOU, *supra* note 59, at 194 (emphasis added).

data necropolitics, we can understand how the development of autonomous and AI-based weapons is entrenched on the socioeconomic pauperization of vulnerable communities.

It is not only the massive collection of data that might trigger discrimination and injustice but also the insufficient existence of data regarding given populations. Data necropolitics is entrenched not only to pervasive surveillance but also to a lack of data, namely, what we call a "digital or data gap."

D. MISSING DATA AS NECROPOLITICS

Data necropolitics can operate not so much from an excess of data and surveillance on a given vulnerable population, but, on the contrary, through the absence or deficient use of data.

The COVID-19 pandemic has brought to light how the lack of data can enhance social and racial injustice. Regarding health inequalities during the pandemic in the United States, Rashida Richardson holds that "government data practices in the public health sector represents one extreme where insufficient collection, use, and reporting of ethnoracial health data can disguise underlying problems and tacit discrimination that aggravate and hasten racial inequities and harms including excess death."[65] Similarly, in Brazil, the federal government tried to withdraw data concerning the pandemic's daily infections and deaths.[66] Death by reporting date and epidemiological week were not published, just like the curve of new cases by reporting date and epidemiological week.[67]

Also, the first epidemiological reports regarding COVID-19 did not take into account the racial impact of the virus, an approach that is legally compulsory in any official public health information in Brazil. Consequently, and just like in the United States, the mortality impact of the virus on black, brown, and indigenous populations was underreported.[68] In a prescient work regarding France's management of the latest pandemic, Mathieu Arminjon and Régis Marion-Veyron have

[65] Rashida Richardson, *Government Data Practices as Necropolitics and Racial Arithmetic*, GLOBAL DATA JUSTICE (Oct. 8, 2020), https://globaldatajustice.org/gdj/1977/ (emphasis added).

[66] Dom Phillips*, Brazil Stops Releasing Covid-19 Death Toll and Wipes Data from Official Site*, THE GUARDIAN (June 7, 2020, 1:40PM), https://www.theguardian.com/world/2020/jun/07/brazil-stops-releasing-covid-19-death-toll-and-wipes-data-from-official-site.

[67] *News Organizations Team Up to Provide Transparency to Covid-19 Data*, O GLOBO (June 6, 2020) http://glo.bo/3pZk2wZ.

[68] Márcia Pereira Alves dos Santos et al., *População negra e Covid-19: reflexões sobre racismo e saúde* [The Black Population and COVID-19: Reflections on Racism and Health], 34 ESTUDOS AVANÇADOS 225, 225–43 (2020).

highlighted the lack of data regarding social vulnerability to COVID-19 and more generally France's myopia regarding biostatistics, becoming factors that have also normalized health injustice.[69]

Data necropolitics evolved through data gaps, where data are insufficiently collected. This situation normalizes health injustice and, eventually, death. Didier Fassin's scholarship has been exploring how health inequalities do not succeed by accident, but are the results of political and social choice. "Bio inequalities" shape different hierarchies of human lives.[70] The missing data and/or the deficient use of data regarding the morbidity of given populations in times of the pandemic have revealed and enhanced the moral and political hierarchies of individual lives regarding their racial and socioeconomic profiles.

CONCLUSION

Our paper has examined some potential risks unfolding from the nonregulation of specific uses of AI, facial recognition, and, more generally, digital data. Our approach has compared specific cases in the Global North and in the Global South. It has demonstrated the implementation of new technologies and their respect of basic rights, depending on legal and regulation frameworks. We have also shown how the lack of regulation can unfortunately lead to discrimination, injustice, and violence. Data necropolitics is a reality for many individuals belonging to vulnerable populations. It is therefore important to keep addressing these issues and bring forward public and private initiatives that keep on building the rule of law, the common good, and the respect of human rights.

[69] Mathieu Arminjon & Régis Marion-Veyron, *Coronavirus biopolitics: the paradox of France's Foucauldian heritage*, 43 HIST. AND PHIL. LIFE SCIS 1, 3 (2021), https://link.springer.com/article/10.1007/s40656-020-00359-2. *Cf.* Daniele Lorenzini, *Biopolitics in the Time of Coronavirus*, 47 CRITICAL INQUIRY 40, 40–45 (2021),https://www.journals.uchicago.edu/doi/10.1086/711432 (explaining how the COVID-19 pandemic has revealed various ways society relies on systemic economic and racial inequalities); Antonio Pele & Stephen Riley, *For a Right to Health Beyond Biopolitics: The Politics of Pandemic and the 'Politics of Life,'* L., CULTURE AND THE HUMANITIES 1 (2021), https://doi.org/10.1177/1743872120978201 (discussing how prioritizing a human right to health can function as a shield against discrimination).

[70] *See* Didier Fassin, *Another Politics of Life is Possible*, 26 THEORY, CULTURE & SOC'Y 44, 60 56); DIDIER FASSIN, LIFE: A CRITICAL USER'S MANUAL 66 (2018).

ACKNOWLEDGEMENTS

This work has been partly conducted under the European Union Marie Skłodowska-Curie Action "HuDig19," Grant agreement ID: 101027394 led by Professor Antonio Pele (EHESS/IRIS, Paris & The Columbia Center for Contemporary Critical Thought, New-York)

The authors have no conflict of interest to disclose.

Blockchain and the Right to Good Administration: Adding Blocks to or Blocking of the Globalization of Good Administration?

MIGLE LAUKYTE[*]

ABSTRACT

In this article, the author addresses the complex and multifaceted relationship between the right to good administration enshrined in the Charter of Fundamental Rights of the European Union and the uses of blockchain technology by the public administration, which is in charge of making the right to good administration real. The opportunities and threats come hand in hand, and there is an urgent need to push forward a public debate on the uses and misuses of blockchain to guarantee public services, so much so that many aspects of blockchain are not compatible with citizens' expectations in relation to the public sector. Although the focus is on Europe, and the right to good administration is not technically recognized on the international level, the globalization produced by technological advancements on the one hand, and the emergence of global administrative law on the other hand, makes this debate relevant to the rest of the democratic states that want to foster human-centric technologies for the well-being of their citizens.

I. INTRODUCTION

Blockchain has not yet become a mainstream technology, and many people in Europe, the United States, and other countries still do not understand what it means and what it does.[1] Surely many have heard the term, particularly in relation to one of the most popular uses that blockchain was put to— cryptocurrencies. As a matter of fact, statistics show that more than three hundred million people in the world owned

[*] Tenure Track Professor of Cyberlaw and Cyber Rights, Law Department, Pompeu Fabra University (Barcelona, Spain). The project leading to these results has received funding from "La Caixa" Foundation, under agreement CF/PR/PR16/51110009.

[1] Although the most correct and representative term would perhaps be Distributed Ledger Technology, in this article, I will use the term *Blockchain*.

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

cryptocurrencies in 2021.[2]

However, blockchain is not just cryptocurrencies. The public sector is exploring a variety of possibilities that blockchain offers, and this paper focuses on these possibilities: it addresses them through the lens of the right to good administration, a principle and a right established in the European Union (EU) within the framework of the Charter of the Fundamental Rights of the European Union (Charter).[3]

In particular, this article looks at the right to good administration as both a self-standing right and a guiding principle[4] as it applies not only to blockchain in particular, but to any technology that is currently emerging and could be considered useful within the public sector, such as Artificial Intelligence (AI), robotics, interfaces between human brain and digital devices (brain-computer interfaces), the metaverse, and many others. The promises that these technologies bring are not always possible to fulfill, not only because of the objective reasons, such as insufficient digitalization of public services[5] or lack of digital literacy of the population, but also because the price to fulfill these promises in terms of fundamental rights is (or, for those that may occur in the future, might be) too high. Indeed, no digital technology is possible to implement in the EU public sector if it does not comply with EU values—accountability, transparency, privacy, and personal data protection, just to name a few—and fundamental rights, established as core elements and nonnegotiable assets of the community's coexistence.

This rule, *sine qua non*, is reflected in many EU acts, among many, the most recent European Declaration on Digital Rights and Principles for the Digital Decade,[6] which, in terms of digital public services online,

[2] Jordan Tuwiner, *63+ Crytocurrency Statistics, Facts & Trends*, BUY BITCOIN WORLDWIDE (July 15, 2022), https://buybitcoinworldwide.com/cryptocurrency-statistics/.

[3] Charter of Fundamental Rights of the European Union, 2000 O.J. (C364), art. 41, Dec. 18, 2000, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2F TXT [hereinafter Charter].

[4] In fact, to consider the right to good administration as just a principle would be an error. Diana-Urania Galetta, *Digitalizazzione e Diritto ad una Buona Amministrazione*, 3 REV. INTERDISCIPLINARE SUL DIRITTO DELLE AMMINISTRAZIONI PUBLICHE 197, 198 (2021); *see generally* Jaime Rodríguez-Arana, *La Buena Administración Como Principio y Como Derecho Fundamental en Europa*, 6 MISIÓN JURIDICA 23 (2014); *see also* Marc Clement, *Breach of the Right to Good Administration: So What?*, 1 ELTE L.J. 19 (2018) (finding the correct qualification is to see the Right to Good Administration both as a right and also as a principle of EU law).

[5] *See* Oliver Large & Hilda Barasa, *Digital Government in Europe: In Pursuit of Cross Border Functionality*, TONY BLAIR INST. GLOB. CHANGE, (Apr. 11, 2022), https://institute.global/policy/digital-government-europe-pursuit-cross-border-functionality (finding only 35% of public sector in EU had an organization-wide digital skills program).

[6] *European Declaration on Digital Rights and Principles for the Digital Decade*, COM (2022) 28 final (Jan. 1, 2022) [hereinafter Declaration].

clearly establishes that:

> Everyone should have access to all key public services online across the Union. Nobody is to be asked to provide data more often than necessary when accessing and using digital public services.[7]

The various uses of technologies in the general digitalization of public administration and the application of specific technologies, such as AI, has raised many questions, hopes, doubts, litigation, uncertainties, and a loss of trust in the state and its institutions across the world. The questions range from the more theoretical ones, related to automated administrative state as such and its legitimacy, to the more specific ones related to certain applications that promised more than they delivered and, in addition, harmed the weakest social groups.[8] However, there are also voices that see AI as a tool to make a change for the better and, in relation to the topic of this article, could help make the right to good administration effective and more efficient, on condition that the human stays in the loop and does not leave the AI-based application to function without supervision.[9] But what about the blockchain?

To understand the use of blockchain in the public sector (also called

[7] *Id.* at 4 (demonstrating the EU's commitment to "ensuring that all Europeans are offered an accessible, secure and trusted digital identity that gives access to a broad range of online services, ensuring wide accessibility and re-use of government information, facilitating and supporting seamless, secure and interoperable access across the Union to digital health and care services, including health records, designed to meet people's needs.").

[8] *See* Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797 (2021); Stephan Grimmelikhuijsen & Albert Meijer, *Legitimacy of Algorithmic Decision-Making: Six Threats and the Need for a Calibrated Institutional Response*, XX PERSPECT. PUB. MANAG. GOV. 1 (2022); Nicolas Kayser-Bril, *Spain: Legal Fight over an Algorithm's Code*, ALGO. WATCH (Aug. 12, 2019), https://algorithmwatch.org/en/spain-legal-fight-over-an-algorithms-code/ (describing the problem of algorithm that allocates electricity bonuses for the socially vulnerable families); *see generally* Sascha van Schendel, *The Challenges of Risk Profiling Used by Law Enforcement: Examining the Cases of COMPAS and SyRI*, REGULATING NEW TECHNOLOGIES IN UNCERTAIN TIMES 225 (2019), https://research.tilburguniversity.edu/en/publications/the-challenges-of-risk-profiling-used-by-law-enforcement-examinin (addressing the well-known cases of COMPAS in the US and SyRI in the Netherlands, which in both cases were systems that were proved to be discriminatory, unfair and unreliable).

[9] Izabela Wrobel, *Artificial Intelligence Systems and the Right to Good Administration*, 49 REV. EUR. & COMP. L. 203, 218 (2022).

"messy world of public sector IT,")[10] and to address the different aspects of how blockchain and the right to good administration could mutually reinforce each other, the paper is organized as follows. In part 2, I focus on the right to good administration, which is established in the EU as a fundamental right, but also recognized directly or indirectly in other parts of the world, making it possible to talk about its global recognition.[11] In part 3, I succinctly explain what blockchain is and how it works, detailing how the increased levels of technological complexity challenge citizens' ability to understand and question blockchain and similar technologies. In part 4, I turn to the ways in which blockchain technology could strengthen the right to good administration, whereas in part 5, I focus on weaknesses that blockchain introduces for the achievement and realization of the right to good administration. The article finishes with concluding remarks.

Before we start, and for the purposes of contextualization, the following remark is due: we should bear in mind that although it was (also, but not only) thanks to new technologies that we started to really understand what globalization is,[12] certain areas of human knowledge—such as public law in general and administrative law in particular—have resisted globalization processes, leaving them for international law to address. International law, however, applies to specific themes, such as trade, armed conflicts, environment, or intellectual property, but does not deal with issues so dear to a sovereign state, like its internal mechanisms and procedures, that are the essence and heart of public administration. This resistance reflects the intention of states to keep certain aspects of its internal mechanisms exclusively national, but the question is for how long. The advancement of what is known as Global Administrative Law—that is, the kind of administrative law that

[10] Michael Veale, Max Van Kleek, & Reuben Binns, *Fairness and Accountability Design*
Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making, CHI 2018 CONFERENCE PAPER 440, 2 (2018), https://arxiv.org/abs/1802.01029.

[11] In fact, the general legal principles of the EU are recognized as a global rule of law; among these principles, we can find the obligation to provide reasons for decisions and the right to be heard, which are constituent rights of the Right to Good Administration as described in the following sections of this article. *See generally* Marco Macchia, *The Rule of Law and Transparency in the Global Space*, *in* RESEARCH HANDBOOK ON GLOBAL ADMINISTRATIVE LAW 261, 269 (2016) (exploring the dynamics between the rule of law, global institutions and the state).

[12] For example, one of the possible visions is that globalization depends on information technologies because the technologies enabled international trade and foreign direct investment. *See generally* JEFFREY JAMES, GLOBALIZATION, INFORMATION TECHNOLOGY AND DEVELOPMENT (1999) (arguing that globalization is mainly a technological phenomenon, driven by influences exerted on international trade and foreign investment by various forms of information technology).

through a body of basic rules mediates between states and supranational rules and rulers[13]—promises the end of an entirely national understanding of what administration of the state is all about. This article argues that the right to good administration could be one of these basic rules that should permit nations to build an international community where human rights are at the center and where technologies—Internet, AI, blockchain, metaverse, or any other—serve to achieve this goal and not to make it even more difficult to bring into being. In particular, blockchain, this "democratizing escape from the failings of territorial legal systems,"[14] has a particularly promising role in this regard.

II. RIGHT TO GOOD ADMINISTRATION

The right to good administration is established in article 41 of the Charter in the following terms:

> 1. Every person has the right to have his or her affairs handled impartially, fairly and within a reasonable time by the institutions, bodies, offices and agencies of the Union.
> 2. This right includes:
>   a. the right of every person to be heard, before any individual measure which would affect him or her adversely is taken;
>   b. the right of every person to have access to his or her file, while respecting the legitimate interests of confidentiality and of professional and business secrecy;
>   c. the obligation of the administration to give reasons for its decisions.
> 3. Every person has the right to have the Community make good any damage caused by its institutions or by its servants in the performance of

[13] These global rules and rulers are represented by the 2,000 global regulatory regimes, 60,000 international non-governmental organizations, and over 100 international courts. *See* Sabino Cassese & Elisa D'Alterio, *Introduction: The Development of Global Administrative Law*, *in* RESEARCH HANDBOOK ON GLOBAL ADMINISTRATIVE LAW 1, 1 (Sabino Cassese ed., 2016); *see also* Benedict Kingsbury, et al., *The Emergence of Global Administrative Law*, 68 LAW & CONTEMP. PROBS. 15 (2005) (describing the field of global administrative law as a field of study).

[14] Kevin Werbach, *Trust, but Verify: Why the Blockchain Needs the Law*, 33 BERKELEY TECHNOL. L.J. 487, 489 (2018).

> their duties, in accordance with the general principles common to the laws of the Member States.
>
> 4. Every person may write to the institutions of the Union in one of the languages of the Treaties and must have an answer in the same language.[15]

The first part of the article establishes a few principles—impartiality, fairness, and efficiency in terms of time necessary to address a particular matter—whereas the following parts articulate rights that compose the right to good administration, namely the: (a) right to be heard, (b) right to information broadly construed, (c) right to remedy, and (d) freedom to choose communication language, as long as this language belongs to those languages in which the EU treaties have been written in. The right to information is broadly construed because, for purposes of this article, it means not only the right of a citizen to access the information that the public administration has on them, but also the right to demand the public administration to explain its decisions.

The right to good administration is also reflected in many constitutions of EU member states, such as the Spanish Constitution (1978),[16] the Italian Constitution (1948),[17] the Lithuanian Constitution (1992),[18] and many others. That is to say, this right does not refer only and exclusively to the EU institutions, but also reverberates through the legislations of member states, where its foundations were already established constitutionally before the Charter came into force. In addition, it also reflects the general principle of good administration that belongs to EU law.[19] This right on the EU level is guaranteed by

[15] Charter, *supra* note 3, at art. 41.

[16] *See* Jaime Rodríguez-Arana, *El Derecho Fundamental a la Buena Administración en la Constitución Española y en la Unión Europea*, 40 REV. GALLEGA DE ADMINISTRACIÓN PUBLICA 233 (2010) (addressing the link between the right to good administration and representative democracy); *See* CONSTITUCIÓN ESPAÑOLA, art. 103, Dec. 29, 1978 (Spain) (linking the right to good administration and Spanish Constitution).

[17] COSTITUZIONE, art. 97 [COST.] (It.).

[18] LIETUVOS RESPUBLIKOS KONSTITUCIJA, art. 5, Oct. 25, 1992 (Lith.).

[19] Clement, *supra* note 4, at 19; *see* Consolidated Version Treaty on European Union, art. 10.3, June 7, 2016, 2016 O.J. (C202) 10.3 ("Every citizen shall have the right to participate in the democratic life of the Union. Decisions shall be taken as openly and as closely as possible to the citizen"); *see also* Consolidated Version Treaty on the Functioning of the European Union, art. 20, 24, Oct. 26, 2012, 2012 O.J. (C326) (providing the find the right to petition, applying to and addressing EU institutions such as Parliament, in "any of the Treaty languages and to obtain a reply in the same language"); *see also The Code of Good Administrative Behaviour*, THE EUROPEAN PARLIAMENT (Mar. 1, 2002); *see also* Treaty of Lisbon Amending the Treaty on European Union and the Treaty Establishing the European Community, Dec. 13, 2007, 2007 O.J. (C 306) 1; *see also Recommendation*

the European Ombudsman and by similar institutions in the member states respectively.[20]

But what is the essence of this right? Besides establishing a series of rights related to public administration and citizens' interaction with it, most importantly it places the citizen—in Rodríguez-Arana's words, "a real individual, a person, with the heap of circumstances that walk with him or her in his social environment"[21]—at the center of this interaction,[22] and requires that the discretionary powers of the public administration be used properly.[23]

Indeed, Juli Ponce describes the general idea of good administration adopted by the European Court of Justice as a procedure to follow before making a decision that has to include:

> [H]earing the people concerned; taking into account all the relevant factors and rejecting the irrelevant; weighing the interests involved; and explaining why [institutions] chose one alternative over another.[24]

No. R (80) 2 of the Committee of Ministers Concerning the Exercise of Discretionary Powers by Administrative Authorities, COUNCIL OF EUROPE (Mar. 11, 1980), https://rm.coe.int/cmrec-80-2-concerning-the-exercise-ofdiscretionary-powers-by-administ/1680a43b39.

[20] EU OMBUDSMEN, https://www.ombudsman.europa.eu/en/home (last visited Dec. 12, 2022) ("The European Ombudsman works to promote good administration at EU level. The Ombudsman investigates complaints about maladministration by EU institutions and bodies, and also proactively looks into broader systemic issues,"); DEFENSOR DEL PUEBLO, https: //www.defensordelpueblo.es/el-defensor/que-es-el-defensor/ (last visited Dec. 12, 2022) (defining the same office in Spain—called The Defensor del Pueblo—as "responsible for defending the fundamental rights and civil liberties of citizens by monitoring the activity of the Administration and public authorities."); LITHUANIA OMBUDSMEN, https://www.lrski.lt/en/ (last visited Dec. 12, 2022) (explaining that, in Lithuania, the Ombudsman's functions are carried out by Seimas Ombudsman Office, whose "primary constitutional duty […] is to protect a person's right to good public administration securing human rights and freedoms, to supervise fulfilment by state authorities of their duty to serve the people properly.").

[21] Rodríguez-Arana, *supra* note 4 at 256, *translated by* MIGLE LAUKYTE.

[22] *See* Rodríguez-Arana, *supra* note 16, at 235–36 (arguing that the citizen has stopped being inert and defenseless individual in front of the state powers that aim to control him or her and this change of vision pushed forward the idea of the modern administrative law); *see also* Rodriguez-Arana, *supra* note 4 (developing further the idea that centricity of citizen is linked to the new idea of the administrative law as a branch of legal system).

[23] *See* Juli Ponce, *Good Administration and Administrative Procedures*, 12 IND. J. GLOB. LEG. STUD. 551, 554 (2005).

[24] *Id.* at 558–59; *see also* E.U. AGENCY OF FUNDAMENTAL RIGHTS, E.U. CHARTER OF FUNDAMENTAL RIGHTS, at art. 41 (last visited Dec. 12, 2022), https://fra.europa.eu/en/eu-charter/article/41-right-good-administration (providing additional context on the

And also explains that this right is related to:

> The existence of a legal duty for public authorities to be in the best position to be able to make appropriate decisions, thereby resulting in a common European inheritance.[25]

Such a vision of the right to good administration leads us to see it as a part of new administrative law, as already described by Rodríguez-Arana, and links it to the public interest as an overall objective and *raison d'être* of public administration.[26]

However, the right to good administration, as such, is known only within the European Union's frontiers. In the United States, this right has developed in a different form and can be traced back to the V and XIV Amendments of the US Constitution, which both refer to limitations of the state's powers to deprive a person "of life, liberty, or property, without due process of law."[27] It is true though that this link is weak and, for some authors, even inexistent, as due process is:

> Simply a defensive tool, intended to protect citizens. For it to work, due process needs an entitlement, that is, a right given by a legal system to an individual . . . if there is a discretionary power, there is not an entitlement: there is unfettered discretion, and consequently due process fails to work.[28]

Indeed, to have a right recognized by the US legal system would mean that this right triggers the Due Process Clause. The US Supreme Court has explained how this right—an entitlement—does so by arguing that:

enormously rich case law of the European Court of Justice and European Court of Human Rights on the Right to Good Administration).

[25] *See* Ponce, *supra* note 23, at 561–62; Rodríguez-Arana, *supra* note 4, at 239; *see also* U.S. v. South-Eastern Underwriters Ass'n, 322 U.S. 533, 591 (1944) (explaining what the Supreme Court has called "the body of institutional experience and wisdom so indispensable to good administration.").

[26] Rodríguez-Arana, *supra* note 4, at 236–38.

[27] *See* U.S. CONST. amends. V, XIV.

[28] *See* Ponce, *supra* note 23, at 576–77; *see also* Javier Barnes, *Buena Administración, Principio Democrático y Procedimiento Administrativo* [*Good administration, democratic principle and administrative procedure*], 21 REV. DIGITAL DE DERECHO ADMINISTRATIVO 77, 79 (2019) (defining the rights that define the Right to Good Administration as "defensive rights").

> Food-stamp benefits . . . "are a matter of statutory entitlement for persons qualified to receive them" . . . Such entitlements are appropriately treated as a form of "property" protected by the Due Process Clause . . . .[29]

However, there is no right to good administration recognized in the United States and, therefore, it cannot be linked to Due Process. As a matter of fact, the Supreme Court provides a few insights on the matter that could help us to understand how the right to good administration could be understood in the United States. For example, in *United States v. L.A. Tucker Truck Lines*,[30] the Supreme Court has argued that on a variety of previous occasions, it has established that:

> [O]rderly procedure and good administration require that objections to the proceedings of an administrative agency be made while it has opportunity for correction in order to raise issues reviewable by the courts.[31]

Therefore, the *idea* (not the right!) of good administration is not unknown and has been adopted in different cases, not only by the Supreme Court but also by Congress. Therefore, according to this case, we could establish indirect references to the right to good administration, or rather, a duty of good administration inherent in the judicial and legislative understandings of the state's functioning.

Furthermore, the right to good administration—and, in particular, the right to be heard—could also be traced to the following statement by the Supreme Court in an earlier case, *N.L.R.B. v. Electric Vacuum Cleaner Co.*, where the Court confirmed that "[h]andling of complaints as quickly as is consistent with good administration is of course essential."[32]

Having seen these different interpretations of the right to good administration, could we claim that this right is a global right? If we look at the international law and focus on the most important international organizations, we will find references to the constituent rights of the right to good administration. That is the case, for instance, with the Agreement on Safeguards as part of the Annexes to the Uruguay Round Agreements of the World Trade Organization (WTO):

[29] Atkins v. Parker, 472 U.S. 115, 128 (1985).

[30] 344 U.S. 33 (1952).

[31] *Id.* at 37.

[32] N.L.R.B. v. Electric Vacuum Cleaner Co., 315 U.S. 685, 699 (1942) (questioning the interpretation of what is understood as essential and whether any obligations are attached to it).

according to this Agreement, during the investigation, the parties have a right to be heard and the authorities have a duty to publish reports with their findings and motivated conclusions.[33]

Perhaps though, the point is to make the question on globality of the right to good administration more abstract and, therefore, reformulate the question and ask whether we can talk in general about civic values and democracy without talking about good administration?[34] Could a state be considered objectively democratic without guaranteeing its citizens this right? Of course, the guarantees have to be real and effective: that is to say, declarations of this right are not sufficient if the state does not guarantee mechanisms to bring it into being. From this perspective then, we invert the deduction of Ponce, that the right to good administration leads to a legal duty of public authorities,[35] and reach the conclusion that the duty to implement, preserve, and guarantee good administration is where the right to good administration emerges from, and its origins are as old as democratic institutions themselves.

But as old as these origins could be, the contemporary technological advancements and speed of innovation is another matter: public administrations have been dealing with digitalization issues for quite a lot of time already, and the right to good administration was not excluded from these debates.[36]

[33] *See* Marrakesh Agreement Establishing the World Trade Organization, Apr. 15, 1994, 1867 U.N.T.S. 274 ("A Member may apply a safeguard measure only following an investigation by the competent authorities of that Member pursuant to procedures previously established and made public in consonance with Article X of GATT 1994. This investigation shall include reasonable public notice to all interested parties and public hearings or other appropriate means in which importers, exporters and other interested parties could present evidence and their views, including the opportunity to respond to the presentations of other parties and to submit their views, *inter alia*, as to whether or not the application of a safeguard measure would be in the public interest. The competent authorities shall publish a report setting forth their findings and reasoned conclusions reached on all pertinent issues of fact and law.").

[34] Rodríguez-Arana, *supra* note 4, at 38 (arguing that democracies do not belong to politicians nor public officers but to the public domain and citizens whose common needs (public interest) are the priority of the democratic state); Barnes, *supra* note 28, at 79.

[35] Ponce*, supra* note 23, at 561–62.

[36] *See* Galetta, *supra* note 4, at 198 (suggesting that the public administrations should be free to use any technologies that could be functional to improve impartiality and transparency of administrative procedures and highlights the importance of responsible officer in linking digitalization of public sector with good administration); *see also* Tuomas Pöysti, *Trust in Digital Administration and Platforms*, SCANDINAVIAN STUD. L. 321, 322 (2018), https://scandinavianlaw.se/pdf/65-19.pdf (describing the situation in Finland where good administration is a foundation for trust in digitally enhanced public administration). *See generally* Claudia Elena Marinică, *Digitalization – The Key for Adapting Good Administration to a Better Governance*, 8.2 ACAD. J.L. & GOVERNANCE 111

In what follows of this article, I address one of the most novel technologies, blockchain, that has already been tested in a variety of public sector applications. However, blockchain, unlike many other technologies, such as Internet, mobile apps, and e-payment systems, has been neither widely adopted, nor fully explored. I address this technology through the lens of the right to good administration and question its impact on this right for better—as a tool to strengthen it—and for worse—as a tool to weaken it. But let us first understand what we talk about when we talk about blockchain.

III. BLOCKCHAIN[37]

Blockchain is a relatively young, very complex, and continuously evolving technology that emerged in the financial sector.[38] Its creator, Satoshi Nakamoto, described blockchain as a technology that enables the functioning of cryptocurrencies called bitcoins.[39] Soon after Nakamoto's paper was published, the first bitcoins were released in 2009. What happened next is probably known to everyone, and the crises, crashes, booms, and collapses of cryptocurrencies are part of news programs weekly, if not daily, all over the world.

However, nowadays, to think that blockchain is just for cryptocurrencies would be a mistake: blockchain is much more than that, although cryptocurrencies remain its most famous use, at least for the time being.

According to Khandelwal, blockchain is:

> An immutable, distributed, decentralized; peer-to-peer ledger replicated across multiple nodes connected in a network, making it possible to record data about any event or transaction as it happens. It consists of blocks

(2020) (explaining the dynamics of public sector digitalization with good administration as a guiding principle carried out in Romania).

[37] Because of limitations of space, I will not address Blockchain exhaustively and therefore many functionalities (mining), features, stakeholders (miners), their economic incentives and other dynamics are not described here. However, the limited explanations should help to understand the essence for those who are unfamiliar with the technology and the references of this section provide with sufficient bibliographic material for those interested to understand the "back office" of Blockchain more in detail.

[38] *See generally* BLOCKCHAIN.COM, https://www.Blockchain.com/explorer (last visited Dec. 12, 2022).

[39] Satoshi Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System* (2008), https://bitcoin.org/bitcoin.pdf.; *see* Jonathan B. Turpin, *Bitcoin: The Economic Case for a Global, Virtual Currency Operating in an Unexplored Legal Framework*, 21 IND. J. GLOB. LEG. STUD. 335, 337–39 (2014) (viewing bitcoins from a legal perspective).

> in a chain used to record as digital assets using a secure algorithm.[40]

Figure 1 explains this definition graphically.

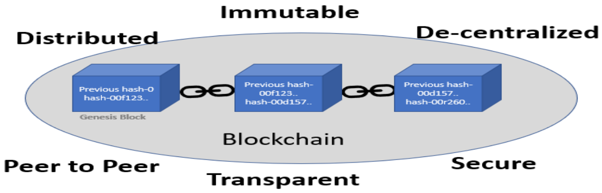


Figure 1. The essence of blockchain technology[41]

In figure 1, the blue cubicles with the writing "previous hash . . ." are data on transactions (purchases, bills, etc.), parties to the transaction (companies, public administrations, individuals under pseudonyms), and the unique code called hash (described in more detail in the following section). The chain of blocks is then a public database of transactions that keeps record of each and every transaction that has been carried out.

Put differently, blockchain is a ledger—a place where we keep trace

[40] Renu Khandelwal, *A Simple Guide to Understand Blockchain*, Medium (Feb. 22, 2021), https://medium.com/swlh/a-simple-guide-to-understanding-Blockchain-8dd09356b153. *See generally* Nakamoto, *supra* note 39 (explaining blockchain from the technological perspective); Akira Sumers, Understanding Blockchain and Cryptocurrencies (2022) (explaining blockchain from a more recent perspective); Primavera de Filippi & Aaron Wright, Blockchain and Law at 33 (2018) (connecting blockchain and the law); Primavera de Filippi et al., *The Alegality of Blockchain Technology*, *Policy and Society* 1 (2022) https://academic.oup.com/policyandsociety/advancearticle/doi/10.1093/polsoc/puac006/6529327 (providing more information on blockchain); Rajesh Dhuddu & Srinivas Mahankali, Blockchain in e-Governance (2021), https://www.perlego.com/book/2661005/blockchain-in-egovernance-driving-the-next-frontier-in-g2c-services-pdf (addressing the possibilities offered by the Blockchain for public services, such as voting, healthcare, cybersecurity, smart cities, and others); Svein Ølnes et al., *Blockchain in Government: Benefits and Implications of Distributed Ledger Technology for Information Sharing*, 34 Got Info. Q. 355 (2017), https://www.sciencedirect.com/science/article/abs/pii/S0740624X17303155 (elaborating two perspectives—governance by Blockchain and governance of Blockchain—on how governments interact with Blockchain, bearing in mind all the benefits and promises of this technology for the public sector).

[41] Khandelwal, *supra* note 40 (providing the illustration of the essence of Blockchain technology). *But see* Ølnes, et al., *supra* note 40, at 360 (noting that some authors also argue that there is no such thing as Blockchain as it is a technology that comes in a variety of shapes, forms and properties).

of records—that permits to register every transaction on a block and add that block to a chain: once added, the block cannot be altered and cannot disappear as it is shared by all nodes, and all nodes have a copy of the latest version of blockchain. Should the block disappear for any reason (cyberattack is a typical example), it can be easily retrieved as the copies are distributed among the nodes of the blockchain network and are updated with every new block added to the chain. It is visible to the rest of the people who have access to this blockchain, and what is recorded on it cannot be changed. Therefore, the transactions are visible, but cannot be altered, without the consent of all the nodes.

Practically, the functioning of blockchain is as follows: I buy a book on Amazon, and this transaction is verified by a network of computers or "nodes" that constitute the particular blockchain. That is to say, these nodes verify, for example, that I have funds to buy a book and that the marketplace where I buy it is really Amazon. Afterwards, once the verification is over, this transaction is added to the block, and the block is "locked" with the help of hash. Once it is done, the block is added to the chain, and everyone has a copy of this new version of the chain of blocks (new because it was updated with a new Amazon-book transaction block).

What is so special about this technology then? First of all, it is completely transparent for those who belong to the blockchain network, in the sense that the transaction chain (the chain of blocks) is visible to everyone who participates in it. The fact that it is also decentralized stands against the traditional vision of transactions that have to pass through the central body, for example, a bank which confirms that a buyer has funds to carry out a particular transaction. Furthermore, blockchain is also very secure: blockchain is tamper-proof because of asymmetric cryptography, digital signatures, and, in particular, hash function. It is also a multistakeholder technology in the sense that, as a decentralized ledger, its network is made of peers and not based on a hierarchical structure.[42] These peers work together to verify, register, and share the data on this ledger, while earning at the same time.

Therefore, it is no surprise that for some, "[b]lockchain re-writes trust as we know it, replacing it with a platform of shared, verifiable integrity."[43] Indeed, we no longer must trust individuals or institutions

[42] *See* Marcella Atzori, *Blockchain Technology and Decentralised Governance: Is the State Still Necessary?*, 6 J. REGUL. GOV. 45, 51 (2017), https://virtusinterpress.org/IMG/pdf/10.22495_jgr_v6_i1_p5.pdf (This position could be challenged in certain particular cases, for instance, when 51% of nodes take over the blockchain network).

[43] NASCIO, *Blockchains: Moving Digital Government Forward in the States* 2 (May 16, 2017), https://www.nascio.org/wpcontent/uploads/2019/11/NASCIO20Blockchains

to transfer assets: the architecture of blockchain technology guarantees the successful outcome of a transaction. This trust could be further augmented if we are dealing with the permissioned—and not permissionless—blockchains.[44] In the case of the former, we have a network based on a group of "nodes" who can trust each other more than if they were in a permissionless blockchain because their access to form the blockchain network was monitored by a centralized authority or other entity. That is not the case with a permissionless blockchain, where anyone can enter and become a part of the network freely without identifying themselves. However, those who understand trust differently—(i.e., where loyalty and coherence play a role)—and who do not consider that decentralization, cryptography, and algorithms are enough to build it (Werbach calls it the "cryptoeconomic trust model"),[45] see the blockchain as a trustless, rather than a trustworthy, technology.[46]

In what follows in this article, I first focus on technological aspects of blockchain that explain its characteristics, and then I address some of the most promising and debated applications of blockchain in the public sector services.

A. Technology

The main technologies that blockchain are based on are: the unique code of hash and a consensus mechanism, that is, a way for all the nodes to agree on what is a valid transaction on a particular blockchain. Another key aspect of blockchain is a smart contract. Let us briefly address each of these technological aspects of blockchain that help us to understand blockchain's strengths and weaknesses.

1. Hash

Hash is an essential element of blockchain, without it, the whole blockchain technology could not exist. It is a unique code given to every block to "lock" it for good and make it very complicated to modify, change, delete, or in any other way alter the information it contains.

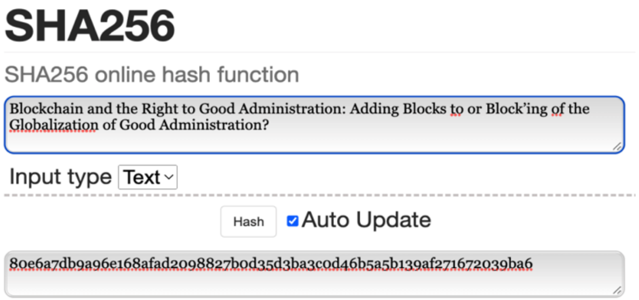
20in20State20Government.pdf (exploring trust in Blockchain and the role that law plays in this relationship); *see* Werbach, *supra* note 14, at 494.

[44] De Filippi & Wright, *supra* note 40, at 31.

[45] Werbach, *supra* note 14, at 495.

[46] Primavera de Filippi & Benjamin Loveluck, *The Invisible Politics of Bitcoin: Governance Crisis of a Decentralised Infrastructure*, 5 INTERNET POL'Y REV. (2016), https://policyreview.info/articles/analysis/invisible-politics-bitcoin-governance-crisis-decentralised-infrastructure.

There are different tools to generate the hash, one of the most known is Secure Hash Algorithm (SHA) 256. Figure 2 shows how the hash of the title of this paper would look like:[47]



Blockchain and the Right to Good Administration: Adding Blocks to or Block'ing of the Globalization of Good Administration?

Input type: Text

☑ Auto Update

80e6a7db9a96e168afad2098827b0d35d3ba3c0d46b5a5b139af271672039ba6

Figure 2: Example of how hash looks like if applied to the title of this article.

Once generated, hash is subsequently checked by the other nodes of the blockchain network, and once this relatively simple operation is over, the block is added to the chain of blocks.

Should there be any problems—for example, there are insufficient cryptocurrencies in a buyer's account—the blockchain network (nodes) reject the operation and the block is not created.

How does this long string of numbers and letters guarantee the safety and trustworthiness of the transactions on blockchain? The hashes on the blocks are connected, therefore, to manipulate the content of one block also means to manipulate the hash. Each block has two hashes: the hash of that particular block and the hash of the previous block. Let us say that we want to manipulate block 3: we need to change both hashes, that is, the hash of block 3 and the hash of the previous block (block 2). Even if we manipulate these two hashes, we will have to go to block 2 and manipulate its hash there as well (we cannot have the hash of block 2 in block 2 different from the hash of block 2 in block 3). This previous block 2 also has its own hash and the hash of other block (block 1) that comes before it and, therefore, manipulation of blocks means manipulation of hashes until we reach the very first block on the blockchain. This kind of backward hash-manipulation operation would require an enormous number of resources in terms of time and computational power, and it would be impossible to keep undetected.

[47] ONLINE TOOLS, https://emn178.github.io/online-tools/sha256.html (last visited Dec. 12, 2022) (tool used to generate this hash); *see also* Nakamoto, *supra* note 39.

This is why it is quite an impossible (although not unfeasible) endeavor. Indeed, the majority of nodes could agree to alter the blocks, but that would be a majority's decision and not the decision of one node. The same could happen in a permissioned blockchain where the governing authority might decide to perform such an operation. However, these scenarios of block alteration are more exceptions rather than rules of how blockchain operates. Usually we take for granted that once blocks are added they are not subject to alterations, modifications, or updates.

2. Consensus Mechanism

As mentioned before, consensus mechanism is an agreement on what constitutes a valid transaction on blockchain: it is a very important feature as it defines the security and validity of data stored. It goes without saying that agreement on how the nodes establish what is the state of affairs at every moment is of fundamental importance to the functioning of the whole blockchain network.

We cannot address all the different consensus mechanisms available, but suffice to say that, as of today, the most popular consensus mechanisms are Proof of Work (PoW) and Proof of Stake (PoS).[48]

PoW is the oldest consensus mechanism and was described by Nakamoto themself.[49] It is based on a competition between the nodes in solving cryptographical riddles, and the node who solves it first is the one who has the right to process the block and add it to the blockchain, thus earning some cryptocurrencies too. The problem is that these riddles are becoming more sophisticated, and their solutions require more computational power, substantial hardware, and software resources; therefore, the alternatives for PoW started to emerge.

PoS is one of such alternatives: it means that the nodes "stake" their own cryptocurrencies in exchange for a chance to validate the new transaction, add the block to the blockchain, and consequently, earn cryptocurrency. The PoS chooses the node at random, but the amount of stake matters: the interesting point here is that should the block be invalid, the node loses the stake, and therefore, the PoS mechanism involves risk for the node to not only not be selected and lose reward but also, even if selected, lose the stake. Furthermore, the ethical question emerges if the node with the highest stake is more eligible than one

[48] *See* Anastasiya Haritonova, *What Is the Difference Between Blockchain Consensus Algorithms?* PIXELPLEX (Mar. 31, 2022), https:// pixelplex.io/blog/best-Blockchain-consensus-algorithms/ (discussing the benefits and drawbacks of Proof-of-Authority and Proof-of-History, etc.).

[49] *See* Nakamoto, *supra* note 39.

with less: would it mean that "rich" nodes have a higher probability to validate blocks and become richer than "less rich" nodes? But a node that stakes a higher sum also means that node has earned more than others and could be more committed to the cause of blockchain.

Leaving ethical considerations aside, one of the main critiques of blockchain technology is the environmental impact it causes because of the computational power and electric energy it needs. Energy consumption is particularly high if the PoW is adopted.[50] Blockchain communities have come up with alternative solutions where different and more environmentally friendly consensus mechanisms are being deployed, for instance, the aforementioned PoS uses less energy than PoW.[51] Other options are also available, such as Proof of Ethic (PoE) consensus mechanism, that require even less energy than PoS.[52]

3. Smart Contract

Smart contract is best understood as "an agreement in digital form that is self-executing and self-enforcing."[53] When we use the term agreement, we do not refer to an agreement of a contract in the classical sense of this term, but to a software code written in programming language and inserted in the blockchain to negotiate an agreement between the parties according to certain preestablished terms and conditions.

The code becomes active once certain conditions are met: for instance, if I am renting my house using a smart contract, I send the code of entrance to the person only once she pays the first month of rent, and the payment enters my bank account. Therefore, payment triggers sending of the code, or rather, payment triggers the execution of the smart contract of rent. Indeed, different from the legal contract as we

[50] Haritonova, *supra* note 48; *see also* Marco Schletz, *Blockchain Energy Consumption: Debunking the Misperceptions of Bitcoin's and Blockchain's Climate Impact*, DATA DRIVEN ENVIROLAB (Aug. 25, 2021), https://datadrivenlab.org/climate/blockchain-energy-consumption-debunking-the-misperceptions-of-bitcoins-and-blockchains-climate-impact/.

[51] *See e.g.,* Celo Foundation, *A Carbon Negative Blockchain? It's Here and it's Celo*, THE CELO BLOG (May 26, 2021), https://blog.celo.org/a-carbon-negative-blockchain-its-here-and-it-s-celo-60228de36490 (discussing Celo, a carbon negative Blockchain that besides being based on PoS is also contributing to decrease its environmental impact by daily offsets through the Celo's protocol).

[52] Crypto Research, *How Helo™ is Solving Blockchain's Core Problems*, (June 15, 2022), https://cryptoresearch.report/crypto-research/how-helo-is-solving-Blockchains-core-problems/. *See generally* NUPAY, https://nupaytechnologies.com/ (providing more information on PoE and HeloTM).

[53] Kevin Werbach & Nicolas Cornell, *Contracts Ex Machina*, 67 Duke L.J. 313, 314 (2017) (analyzing the smart contracts *vis à vis* contract law).

know it, the smart contract does not require human presence, even in the stage of execution. It does it all by itself, that is where its smartness comes from, besides that it also makes it possible for people who do not know and consequently cannot trust each other, to enter into agreements. The smart contract is safely and permanently stored on the blockchain, thus ensuring the contracting parties the possibility to retrieve it, launch it again, or use it to claim any kind of damages or losses.

The aforementioned example of renting a house is an example of a nondeterministic smart contract, that is, a contract that needs information from the outside to be executed. In this case, the outside information is represented by the bank, which informs the smart contract about the payment made to my account. It is a different case with deterministic smart contracts that do not need external information. This is the case of a lottery: people buy lottery tickets by sending money to a smart contract account, and the smart contract has preestablished rules on how the lottery winner is established. Once the deadline to buy tickets comes, the smart contract executes the rule of establishing a winner and sends the money to him or her.[54]

The execution of a smart contract is not possible to interfere with and, thanks to its decentralization, blockchain does not have authority that could stop the smart contract and, as we will see later on, it might be a problem for the right to good administration.

In what follows, I further explain blockchain through the most promising applications that this technology can offer in the public domain.

B. Applications in the Public Sector

The potential blockchain in businesses and governments is widely known, although public administration concerns in the public sector are still problematic. Blockchain represents a promising tool to store and keep track of legally relevant information, such as different kinds of certificates (birth, death, ownership, university degree, vote, entitlement to social benefits, marriage, etc.), licenses (for instance, to open a bar, a shop, or a gym, to convert a flat into an office or vice versa, to occupy a public parking space with a truck, to move from one neighborhood to another, and so on and so forth), decisions and regulations of governments, ministries, regional and local authorities,

[54] Mary Lacity, *Crypto and Blockchain Fundamentals*, 73 ARK. L. REV. 363, 383 (2020).

and of course legislative acts of parliaments.[55]

In fact, all these applications describe different forms that the governance *by* blockchain can take, whereas a further challenge is to address the challenges of governance *of* blockchain, which is a completely different matter.[56] Governance of blockchain addresses how blockchain should work in terms of both architecture (what information is stored, how the accesses are managed, consensus reached, etc.) and interaction with citizens. For instance, if a citizen wants to register her newborn baby, depending on the choices that the public administration has taken regarding blockchain architecture, she might be able to either only see the registered data (in this particular case, the data submitted by the hospital where she gave birth) or also be able to insert the data, which means that it is the citizen and not the hospital who takes care of registering the baby. Then her data is confirmed by the blockchain nodes (hospital and registry of births).[57]

Therefore, governance by blockchain represents all that blockchain can do for public administration, whereas governance of blockchain means how blockchain should be built so that what it can do (governance by blockchain) can be carried out properly and with public interest and individual rights in mind (including the right to good administration).

The following examples in this section refer to the domain of governance by blockchain.

1. *Land and Real Estate Registries*

Blockchain has been used to build land registries in Sweden and some US states.[58] It is particularly useful in those countries where land ownership is difficult to detect, although it should be borne in mind that what blockchain guarantees is authenticity of the land title, not its

[55] Ølnes et al., *supra* note 40.

[56] *Id.* at 359.

[57] *See e.g.,* Illinois Department of Commerce & Economic Opportunity, *State of Illinois Partners with Evernym to Launch Birth Registration Pilot* (August 31, 2017), https://www2.illinois. gov/IISNews/14759-DCEO_Birth_Registration_Pilot_Release.pdf (explaining how the state of Illinois has launched an initiative on birth registries on blockchain).

[58] *See generally* Anetta Proskurovska & Sabine Dörry, *Is a Blockchain-Based Conveyance System the Next Step in the Financialisation of Housing? The Case of Sweden*, 17 LISER WORKING PAPERS (2018) (describing how Sweden is using Blockchain for its Land Administration System (LAS)); NASCIO, *supra* note 43, at 6 (describing the State of Illinois Blockchain Initiative).

accuracy.[59]

Coming back to the Swedish example, which relies on ChromaWay technology,[60] the changes in the procedure to purchase a small house by a private person via a real estate agent are evident: without blockchain, the land registry, although an institution with a very high credibility, gets actively involved in the process of purchase at a very late stage. In addition, the process is lengthy; the documents are not digital; checking of buyer's and seller's identities is manual; and documents have to be stored for ten years. As these documents are paper, their storage requires space and resources—not to say what it would take to search these documents for information. Applying the blockchain technology, the situation changes: the procedures that took four months are reduced to several days; manual checks are no longer needed; property registration is automatic; digital signatures resolve the identity issue; and all the documentation is digital, searchable, and easy to store and secure.[61]

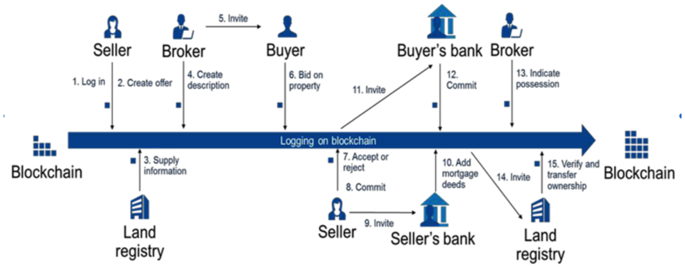


Figure 3. ChromaWay real estate transfer workflow[62]

The use of blockchain for land and real estate registries and their management opens up a wider discussion not only about how to implement blockchain into dynamics of land ownership, but more so

[59] Ølnes et al., *supra* note 40, at 357. S*ee generally* Mohammed Shuaib,Shadab Alam, Salwani Mohd Daud, et al., *Improving the Authenticity of Real Estate Land Transaction Data Using Blockchain-Based Security Scheme*, in ADVANCES IN CYBER SEC., 3 (2021) (discussing authenticity issues related to real estate management).

[60] CHROMAWAY, https://chromaway.com/ (last visited Sept. 29, 2022).

[61] LANTMÄTERIET ET AL., THE LAND REGISTRY IN THE BLOCKCHAIN-TESTBED, 40–55 (2017), https://static1.squarespace.com/static/5e26f18cd5824c7138a9118b/t/5e3c35451c2cbb6170caa19e/1581004119677/Blockchain_Landregistry_Report_2017.pdf.

[62] DAVID ALLESSIE ET AL., JRC SCIENCE FOR POLICY REPORT: BLOCKCHAIN FOR DIGITAL GOVERNMENT 27 (2019), https://joinup.ec.europa.eu/sites/default/files/document/201904/JRC115049%20blockchain%20for%20digital%20government.pdf.

about how the existing administrations and public organizations should re-arrange their tasks so as to accommodate blockchain technology within the flow of administrative procedures. This re-arrangement requires long-term strategical planning of the future public services because public services represent a network of interrelated data and information flows that continuously reverberate through different administrations and affect provision of these services. In the particular case of land and real estate registries, the importance of collaborations with third parties, such as banks or other financial institutions, becomes particularly relevant.

2. Voting

Democratic processes, in particular voting, could benefit enormously from the use of blockchain. Indeed, voting processes are particularly subject to fraud and manipulation, and the data integrity and no-repudiation that blockchain guarantees represent the strengths of this technology. Therefore, the use of blockchain for the purposes of electing representatives or making decisions in referendums seems to be a promising way to use blockchain in the public sector.

Practically, the blockchain-enabled voting would involve storage of votes on a blockchain network (distributed among the nodes) and an encrypted vote validated via a chosen consensus mechanism. Everyone could see the votes, different from classical voting, but without knowing who voted for whom or what (in the case of a referendum). The blockchain would ensure cryptographically the security and integrity of data and, therefore, reduce the possibilities to manipulate the votes. In addition, the costs of blockchain and organizing live elections are high; however, what changes in the case of blockchain-enabled voting is that the human involvement in processing votes is reduced to a minimum. This means less possibilities for human errors and discretion when interpreting unclear voting ballots, and also, in getting speedy and reliable results.[63]

To be sure, voting as a process is very complex, and in terms of blockchain architecture (the issue pertaining to the field of governance of blockchain), certain decisions, such as identity management or secrecy of one's vote, would be particularly stringent and differ substantially from the general idea of openness and transparency that is usually associated with blockchain applications. In addition to that,

[63] Uzma Jafar & Mohd Juzaiddin Ab Aziz, *A State of Art Survey and Research Directions on Blockchain Based Electronic Voting System*, *in* ADVANCES IN CYBER SEC. 248, 248 (2021).

as blockchain is still not widely known and even less understood, many voters might mistrust the technology. And that is not just because blockchain is a complex technology to grasp, but rather because blockchain-enabled voting overturns the classical dynamics of the voting process, which is black-boxed, centralized, and top-down, into a transparent, decentralized, and bottom-up process.[64]

As much as blockchain's benefits for the voting systems are widely discussed in the literature, there are still few cases that could help us to assess and fully understand whether blockchain-enabled voting in state, national, regional, or autonomic elections really works. Indeed, blockchain can be used for voting in other settings, such as voting in the meetings of organizations, as it happened in Abu Dhabi's Securities Exchange.[65] But voting at meetings is not the same as voting in public elections. The state of West Virginia was the first state in 2018 to offer the possibility to use Voatz, a blockchain-based voting application, yet in 2020, decided against its use in its primary elections,[66] because of security concerns that were pointed out by MIT researchers.[67]

There is still much work that needs to be done, and not only in terms of the technical viability of blockchain projects in the public sector. There is a lack of common understanding and agreement on basic concepts of blockchain, and it reverberates on the expectations related to its uses[68] by all the stakeholders—citizens, businesses, and public administration—involved. The expectations are particularly high in voting: the margins of error are very low, technological failures are inadmissible, and stakes are very high. Therefore, the reluctance to rely on blockchain in election processes is understandably cautious.

Having seen what the right to good administration is and also what kind of technology blockchain is, the challenge now lies in combining the two and addressing this combination by looking at the positive and negative sides of this interaction respectively.

[64] *See* Philip Boucher, *What if Blockchain Technology Revolutionised Voting?*, EUROPEAN PARLIAMENT (Sept. 29, 2016), https://www.europarl.europa.eu/thinktank/en/document/ EPRS_ATA(2016)581918.

[65] Karl Flinders, *Abu Dhabi Securities Exchange Uses Blockchain for E-Voting*, COMPUTER WKLY (Oct. 18, 2016, 1:00 PM), https://www.computerweekly.com/news/450401258/Abu-Dhabi-Securities-Exchange-uses-blockchain-for-evoting.

[66] Jed Pressgrove, *West Virginia Pauses Use of Voatz Voting App, Cites Security*, GOV'T TECH. (Mar. 3, 2020), https://www.govtech.com/products/west-virginia-pauses-use-of-voatz-voting-app-cites-security.html.

[67] Michael A. Specter et al., *The Ballot is Busted Before the Blockchain: A Security Analysis of Voatz, The First Internet Voting Application Used in U.S. Federal Elections*, 29TH USENIX SEC. SYMP. (2020).

[68] Silvia Semenzin et al., *Blockchain-Based Application at a Governmental Level: Disruption or Illusion? The Case of Estonia*, 41 POL'Y & SOC'Y 386, 394–95 (2022).

IV. BLOCKCHAIN AS A TECHNOLOGY THAT STRENGTHENS THE RIGHT TO GOOD ADMINISTRATION

In the beginning of this paper, the right to good administration was deconstructed into a few principles and specific rights, such as the right to be heard, the right to information, and the right to remedy. Therefore, the question now is how blockchain technology could foster, strengthen, or carry out these rights in the data-driven public sector.

The potential of blockchain to boost human rights in general has been already observed:

> [Blockchains] can enable new strategies for establishing and enforcing rights that, unlike the current regimes, do not rely on the assent of military-backed nation-states. Blockchains have the potential to create a new layer of global social contracts, in which human peers, more than territorial governments, are the protagonists.[69]

In addition to the above, blockchain seems to be on a different—more advanced?—wave than governments actually are in terms of human rights:

> With distinct and diverse governance designs, blockchains can help protect the kinds of rights that states are badly suited to defending. Human rights on blockchains can and should look different from those of nations. Blockchains worth having should expand our sense of what kinds of rights are reasonable to imagine and to expect for ourselves.[70]

Therefore, the question is whether blockchain can not only make existing human rights stronger, but also create new human rights.

As interesting and appealing as the idea of new human rights might

[69] Nathan Schneider, *How We Can Encode Human Rights in the Blockchain*, NOEMA MAG. (June 7, 2022), https://www.noemamag.com/how-we-can-encode-human-rights-in-the-Blockchain/. There are many blockchains that take human rights into account, *see e.g.*, DiginexLUMEN, DIGINEX (https://www.diginex.com/lumen), that helps to trace working conditions in supply chains. What we should consider is that tracking working conditions does not mean preventing illegal labour, poor working conditions, and other problems. On human rights and blockchain, *see also* William Crumpler, *The Human Rights Risks and Opportunities in Blockchain*, A Joint Strategic Report of the CSIS Strategic Technologies Program and Human Rights Initiative (Dec. 2021), https://www.csis.org/analysis/human-rights-risks-and-opportunities-Blockchain.

[70] Schneider, *supra* note 69.

be, these rights are not the object of this article: we focus on the rights that already (at least in theory and at least in some places of the world) exist and states' duties with respect to these specific rights, namely the right to good administration, and more specifically, the right to be heard, right to information broadly construed, and the right to remedy.[71]

The right to be heard—that is the right to be heard "before any individual measure which would affect him or her adversely is taken"[72]— could benefit from blockchain technology as all the actions related to a particular citizen's case could be inserted in the blockchain of public administration, and the citizen could access that blockchain. For example, if a citizen submits an application to receive an electric energy bonus, he or she not only should be able to trace where his or her application is within the administrative process of granting these bonuses, but also be able to see on this blockchain—besides being informed personally and directly—that should his application be impossible to satisfy, there is a procedure with a clearly established timeframe on when and how he or she can exercise the right to be heard. That is to say, the citizen would be enabled to explain the reasons why they qualify for this bonus although the public administration thinks that it is not the case. Therefore, and differently from the current practices in such cases, the citizen would know where his or her application is, as the blockchain would ensure the transparency of the procedure and if, for instance, the application does not advance in the administrative process, there is proof of that on the blockchain. In addition, the Ombudsman, who is in charge of making the right to good administration a reality and not a miracle, could also be aware of the processing of the application and see that the application was processed without any citizen involvement (without hearing him or her). This would ensure a double kind of auditability from both the citizen and the Ombudsman.

This way of processing of (in this particular case) applications for electricity bonuses would also ensure a higher level of control over how public administrations deal with social entitlements: to reject a citizen's application, a particular administration (its section, committee, or department) would need to add a transaction to blockchain about it.

[71] We could also speculate about the possible benefits for the right to communicate in one of the languages of the EU because blockchain should also be available in different languages and, in particular, in those languages that are of risk of extinction within the EU, such as Lithuanian, Estonian, or Hungarian. In this sense, blockchain could be an indirect way to contribute to a multi-lingual society and preserve linguistic heritage of the planet.

[72] Charter, *supra* note 3, at 41.

But, in order to do that, the administration would need to have a session to allow the citizen to present his or her claims because otherwise the blockchain would both be a proof that no session took place and that the public administration ignored the citizen's right to be heard. Either way, the blockchain would register illegal activity and should set off the alarms of auditors both internal to the public administration and external (the Ombudsman).

The right to information perhaps is the easiest to satisfy. If its essence is that the citizen has a right to access the information the public administration has on him or her, and at least some of this information is available on a blockchain, it should not be difficult for the administration to retrieve it or offer the citizen a way, for instance through a digital gateway to public services, to access it anytime and from anywhere. At the same time, access to the information on blockchain would be carried out with due guarantees of privacy and personal data protection, ensuring higher data quality as any data on the citizen that is inserted in the blockchain would need to undergo consensus of different public administrations' nodes.

In addition, linking different blockchains could also ensure accessibility of information through different points of entrance to the network of public administration. The access to one's information through the tax authorities should also lead to access to one's information on social welfare and permit the citizen to update his or her data (for example, the change of residence or family status). If the tax authority blockchain could "talk" to the social welfare authority blockchain, the functionality and efficiency of the public sector blockchain would increase significantly.[73]

As to the right of public administration to give reasons that would explain why a certain decision that concerns a citizen was taken, blockchain of course would not be able to give reasons instead of the public administration, but could register and keep a trace of these reasons and keep a record that this duty was carried out and respected time limitations (right now it is difficult to understand what the time limit to react to citizens' demands for information is).

The right to remedy, that is the right to have any damage repaired should this damage emerge from the actions or inactions of the public administration or its employees, could follow a similar path as described above. Blockchain could be used to register the claim for remedy and trace its processing through the system and thus provide the citizen

[73] *See also* Rafael Belchior et al., *A Survey on Blockchain Interoperability: Past, Present, and Future Trends*, ARXIV (MAR. 22, 2020), https://arxiv.org/abs/2005.14282 (providing a very detailed literature survey and analysis of the possibilities to seamlessly interconnect different blockchains).

with real time information where his or her claim is and what institution (department, section, etc.) is dealing with it, what the deadline to issue the remedy is, and other information.

We have seen that blockchain offers various ways to facilitate, expedite, and access the right to good administration. However, these opportunities do not come without a price, and dangers in using them without critically addressing their side effects would lead to citizens' subjugation rather than empowerment, which is enshrined in the very essence of the right to good administration as a fundamental right.[74]

In addition, and quite surprisingly so, there is no—to the knowledge of the author—literature on how blockchain could be used by citizens to make public administration more transparent, accountable, and better (in the sense of good administration and good governance). That is to say, the majority of debates focus on how government could use blockchain to assist citizens, yet what is lacking is how the citizens could use blockchain to understand their rights and keep public administrations accountable. Put differently, in the citizen-public administration relationship, it is always the public administration that shapes the ways of interacting with citizens, but a real citizen's empowerment and a trust-based, mature, and democratic relationship between citizens and public administration cannot evolve in only one direction (from public administration to citizen), but has to be bidirectional (from citizen to public administration and from public administration to citizen).

V. BLOCKCHAIN AS A TECHNOLOGY THAT WEAKENS THE RIGHT TO GOOD ADMINISTRATION

In what follows, I look at those applications and uses of blockchain technology that could be detrimental to the principles and rights that are covered under the umbrella of the right to good administration.

The very nature of blockchain seems to be more related to anti-government and anti-state stances which we normally link to people and social movements disengaged from democratic societies, and usually associated with extremisms.[75] The decentralization that blockchain is built upon is but one example. Its technologically-driven nature is another: as Primavera de Filippi and Benjamin Loveluck argue in their

[74] Galetta, *supra* note 5; Rodríguez-Arana, *supra* note 16; Rodríguez-Arana, *supra* note 4.

[75] *See generally* DAVID GOLUMBIA, THE POLITICS OF BITCOIN: SOFTWARE AS RIGHT-WING EXTREMISM (2016) (exploring how supporters of Bitcoin and its blockchain technology subscribe to a form of cyberlibertarianism that depends to a surprising extent on far-right political thought).

essay, the Bitcoin project in particular (they are not talking about blockchain as such, but are focusing on Bitcoin specifically) is an example of governance by infrastructure, which theoretically should, but practically cannot, substitute a platform that functions with and integrates institutional framework.[76]

Indeed, crypto anarchists have stated as early as 1992 that we will soon be able:

> [. . .] to communicate and interact with each other in a totally anonymous manner. Two persons may exchange messages, conduct business, and negotiate electronic contracts without ever knowing the True Name, or legal identity, of the other. Interactions over networks will be untraceable, via extensive re-routing of encrypted packets and tamper-proof boxes which implement cryptographic protocols with nearly perfect assurance against any tampering. [. . .] These developments will alter completely the nature of government regulation, the ability to tax and control economic interactions, [. . .] The State will of course try to slow or halt the spread of this technology, citing [. . .] fears of societal disintegration. Many of these concerns will be valid; [. . .] But this will not halt the spread of crypto anarchy.[77]

We can recognize an early idea of blockchain in these words, and governments are identified as sources of obstacles to blockchain's deployment. However, this is a shortsighted vision, which is built on the assumption that the state would not deploy blockchain for its purposes (purposes that include the administration of public services), and that is not the case, as we have seen in this article.

The right to good administration—an essential right to make public administration accountable to its citizens—could be violated by public administration should it implement the blockchain-based public services without creating an appropriate digital ecosystem for such services to be real; without adjusting the existing (or creating new) legal framework and procedural rules; and without creating mechanisms for citizens to ask questions, verify data, update information, and have other means to participate in these processes. So as to ensure these means of participation, citizens should have ways to interact and overcome the

[76] De Filippi & Loveluck, *supra* note 46, at 26.

[77] Timothy C. May, *The Crypto Anarchist Manifesto*, https://groups.csail.mit.edu/mac/classes/6.805/articles/crypto/cypherpunks/may-crypto-manifesto.html.

digital—in this particular case, blockchain—divide that is a pending issue to solve in much lesser (in terms of technological complexity) matters.

For example, if public administration is using blockchain to process and grant electricity bonuses, it could simply not update the blockchain where a citizen could see his or her application, and the citizen—not aware of deadlines and administrative procedures—might miss important dates or re-submission requirements or simply not understand what is going on. Then the question is who is supervising the public administration blockchains and how accountable should this supervising entity be: the public supervisor of the public administration, the Ombudsman and its Office might need additional technological, human, and financial resources to undertake this enormous task.

In particular, smart contracts (described in part III of this article) could be a serious obstacle to the right to be heard: if the public administration uses smart contracts, which execute themselves once certain conditions are met, then the citizen might not have time between the decision and execution to exercise his or her right to be heard before the decision affecting him or her negatively takes place. For example, if children of a large family become of age, certain welfare bonuses awarded to large families could be affected, for instance, the aforementioned electricity bonus. However, if this procedure was automated with the smart contract, the day a child becomes of age might become a condition triggering the non-application of electricity bonus, but that would not mean that the family stopped qualifying for the electricity bonus on different grounds, such as low monthly income of its members. This process would also mean that the family would not have time to explain their situation, but would probably need time to re-present the application for the entitlement to the electricity bonus because the contract is impossible to stop from executing itself. In the meantime, and for all the time that this application would be processed, the family would pay a full price of electricity, although legally entitled to bonus.

This is of course a speculation and a hypothetical situation as the social welfare has not (yet?) been subject to smart contracting nor blockchain, but it represents an emblematic situation when the right to be heard could be seriously compromised.

In addition to that, the complexity to update information on blockchain—in particular, if it is a nationwide blockchain with many nodes from different public administrations and millions of citizens accessing it—might slow down any procedure or processing of requests, entitlements, or remedies. Perhaps this scalability problem will be solved with time and once blockchain reaches higher maturity levels.

Similar reasoning applies to the right to information and the right to remedy as well: complexity of blockchain coupled with lack of control of how the public administration is managing the information on blockchain could make citizens more powerless and increase the sense of mistrust and disillusion. Indeed, blockchain could turn into the wall between public administrations and citizens, whereas it should be a bridge bringing the two parties closer and helping them understand each other better.

VI. CONCLUSION

The aim of this article was to look at the right to good administration on the one hand, and blockchain technology on the other: are they friends or foes? The article argues that they can be none and both, as we have examples to support both claims. It is obvious though, that so as to make really substantial claims on the matter, we need to carry out a more exhaustive research on how blockchain reflects social needs and entitlements in general and set its relationship with the right to good administration within a wider framework of inquiry. However, as limited as this research is, it still permits us to realize that blockchain in itself is not an answer to all the hurdles that citizens face while interacting with public administration but could be a part of a set of technological tools that citizens could benefit from in such interactions.

In relation to the above, blockchain is usually seen as a technology that should be taken as it is—decentralized, not-hierarchical, anonymous or pseudonymous, etc.—as if all these features were written in stone and could not be subject to modifications. Instead of thinking about what blockchain in its original sense permits, the public sector should use blockchain while thinking about what citizens might need and could get thanks to the blockchain technology applied by and to public administration. Furthermore, blockchain should not be used by the public administration to deprive people of their entitlements or "datify"—turn into a code—social fragility and dependence of many citizens. On the contrary, blockchain should be a tool for citizens to make sure that they get from the public administration what is due to them, and get it fast, and the public administration is transparent, acts legally, and controls its own actions.

Furthermore, while debating the possibilities of blockchain in the public sector, we should not forget that the public sector is different from the private one: for instance, the margin for anonymity within the public sector is much more limited—if possible at all—than it is in the private sector, and implies further and additional requirements in terms of safety, accessibility, accountability, and further legitimate and justified social expectations. Therefore, to talk about blockchain in the public sector without taking into account that the public sector is subject to higher standards and is by its very nature less flexible than the private sector, is to start with an erroneous presumption and, consequently, condemn the blockchain application to failure from the very beginning.

Moving towards the main object of this article—the interaction between the right to good administration and blockchain technology—the use of blockchain to guarantee the right to good administration is a part of the wider debate on transformation of the public sector: this transformation is a complicated endeavor and a continuous process. We know approximately when it started thanks to the advancements of information and communication technologies during the last century, but we see no end to it. In fact, on the one hand, newer technologies are emerging, and novel possibilities are taking shape, and, on the other hand, the public sector is so complex, multilayered, and dynamic that it is a never-ending task to digitally reshape and make compatible all the different ways in which the citizens, businesses, and public administrations interact with each other.

Furthermore, the discussion on blockchain and the right to good administration belongs to a broader discussion on the impact of (disruptive, emerging, new, or combined thereof) technologies on the legal systems globally, and in this particular case, on administrative law as such, which is also turning into global administrative law.[78] Within the framework of this global administrative law, such administrative tools as registries and other record-keeping mechanisms and systems could be supported by blockchain technologies by giving these registries, mechanisms, and systems internationally recognized legal solidity, recognition, and trustworthiness.

We need a common political commitment not only within the EU, where such a commitment already exists,[79] but also globally so as to build together governments, public administrations, public services, and digital skills that would empower people, meet their needs, and help

[78] Cassese & D'Alterio, *supra* note 13, at 2.

[79] *See, e.g.*, *Ministerial Declaration of eGovernment - the Tallinn Declaration*, EUROPEAN COMM'N (Oct. 6, 2017), https://digital-strategy.ec.europa.eu/en/news/ministerial-declaration-egovernment-tallinn-declaration; *Berlin Declaration on Digital Society and Value-Based Digital Government*, EUROPEAN COMM'N (Dec. 8, 2020), https://digital-strategy.ec.europa.eu/en/news/berlin-declaration-digital-society-and-value-based-digital-government.

them to live better lives as citizens of democratic societies.[80] These democratic societies, where human rights prosper, need to work together, and blockchain could be a powerful tool to achieve, fulfil, and protect some of these rights and ensure better public services, that with every year gains more international relevance and dimension.

There are many questions open for future research. What seems to be a promising line of research is, for example, the study of interplay of different technologies—AI, blockchain, etc.—within the public sector and how this interplay could reverberate on the quality and accessibility of public services and citizen empowerment.

In addition to that, further questions emerge, for instance, how inclusivity is guaranteed and how these technologies also affect the internal workings of public administrations that undergo a continuous re-organization in terms of financial, human, and technological resources. The right to good administration is a useful tool to guide these administrations in this never-ending, yet absolutely necessary, endeavor.

[80] *See* U.N. Secretary-General, *Road Map for Digital Cooperation: Implementation of the Recommendations of the High-Level Panel on Digital Cooperation*, U.N. Doc. A/74/821 (May 29, 2020) (providing an example of governmental collaboration).

The Digital Transformation of Tax Systems Progress, Pitfalls, and Protection in a Danish Context

LOUISE BLICHFELDT FJORD AND PETER KOERVER SCHMIDT[*]

ABSTRACT

The authors examine possibilities and challenges in using digital tools to obtain tax simplification and to improve tax assessment, collection, and transparency. Hence, the main objectives of the article are, from a legal perspective, to shed additional light on the relations between tax administrations and taxpayers in an increasingly digitalized world and to discuss how this development may influence taxpayers' rights and the overall efficiency of tax systems. In doing so, practical experiences—incurred in Denmark during its journey from a paper-based and manual tax administration process toward a more digitalized one—are analyzed. Against this background, it is concluded that many states around the world, including Denmark, have come a long way in making tax processes smoother and more efficient through the use of digital tools for the benefit of both taxpayers and tax administrations. However, at the same time, global as well as Danish experiences clearly show that states, in their pursuit to digitalize tax administrations further, need to take appropriate measures into consideration in order to ensure the legality and transparency of the digital tax administration processes.

I. INTRODUCTION

In a recent report released by the Organization for Economic Cooperation and Development (OECD), the continuous global trend toward digital transformation of tax administrations is highlighted. Moreover, it is stated that this trend has been accelerated by the

[*] Louise Blichfeldt Fjord, PhD, Assistant Professor, Copenhagen Business School, and Director, CORIT Advisory. E-mail: lbf.law@cbs.dk. Peter Koerver Schmidt, PhD, Professor (WSR), Copenhagen Business School, and Academic Advisor, CORIT Advisory. E-mail: pks.law@cbs.dk.

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

COVID-19 pandemic to such an extent that digital contact channels now dominate interactions between tax administrations and taxpayers.[1]

Even though the global pandemic may have accelerated the digital transformation of tax systems, policymakers, and tax administrations have been preoccupied with implementing solutions based on information technology (IT) for quite some time to make it easier for taxpayers to meet their tax obligations, to enhance compliance, and to increase efficiency. In addition, the digitalization of tax administrations—as well as of the global economy as such—has attracted the interest of tax scholars, and ever-growing literature deals with various issues related hereto. Accordingly, several different research streams may be identified within the field of *tax and technology*.[2]

In this article, however, our primary focus is on the possibilities for and challenges with using digital tools to improve tax simplification, tax assessment, tax collection, and tax transparency.[3] Hence, our main aims are—from a legal perspective—to shed additional light on the relations between the state (i.e., the tax jurisdiction) and its citizens (i.e., the taxpayers) in an increasingly digitalized world and to discuss how this development may influence taxpayers' rights and the overall efficiency of tax systems.[4]

[1] Org. for Econ. Co-operation and Dev. [OECD] FORUM ON TAX ADMIN., TAX ADMINISTRATION 2022 22–23 (2022). The report provides comparative information on fifty-eight advanced and emerging economies, accounting for around 90 percent of the global GDP.

[2] *See, e.g.*, Claudio Cipollini, *A Systemic Introduction to Tax and Technology*, IBFD Int'l Tax Stud. 3 (2022). The author identifies the following six research streams: taxation of the digital economy, technology and tax collection, technology and tax transparency, technology and tax simplification, technology and taxpayers' rights, and, finally, technology and taxation in developing countries.

[3] In this article, the term *tax simplification* is understood as endeavors to reduce the complexities of the tax system as such, including tax code complexity, structural complexity, policy complexity as well as administration and compliance complexities. For more on the dimensions of tax complexity, *see* Lynne Oates & Gregory Morris, *Tax Complexity and Symbolic Power*, *in* TAX SIMPLIFICATION 25-32 (Chris Evans et al. eds., 2015). The notions *tax assessment and collection* are used broadly, i.e. as the overall process of assessing taxpayers' income statements and actually collecting the taxes (including taking action against those who have not filed a return in time or paid their taxes when due). For more on these tax administration functions, *see Tax Administration 2022*, *supra* note 1, at 54, 122. The term *tax transparency* is used to refer to the transparency of taxpayers' affairs through the automatic exchange of information between states as well as through strengthened reporting and disclosure requirements for taxpayers and third parties. For more on the notion of tax transparency, *see* Johanna Hey, *General Report – The Notion and Concept of Tax Transparency*, *in* TAX TRANSPARENCY 3 (Funda Başaran Yavaşlar & Johanna Hey eds., 2019).

[4] In this article, the term *tax simplification* is understood as endeavors to reduce the complexities of the tax system as such, including tax code complexity, structural

In doing so, we will illustrate a number of practical examples incurred by Danish policymakers and authorities during Denmark's journey from a paper-based and manual tax framework toward a more digitalized one. The intention is that this approach will make the issues discussed more tangible and enable policymakers, officials, and scholars to learn from real-life examples provided in a Danish context.[5] Moreover, Denmark is considered a useful case for this purpose because Denmark has one of the most digitalized public administrations in the world and Denmark historically has been at the forefront of implementing digital solutions into their tax administrative framework.[6] Finally, some of Denmark's digital initiatives—in particular, the unsuccessful ones—have been subject to intense debate in the Danish media and scrutiny by national institutions, such as the National Audit Office and the Ombudsman.[7]

complexity, policy complexity as well as administration and compliance complexities. For more on the dimensions of tax complexity, *see* Lynne Oates & Gregory Morris, *Tax Complexity and Symbolic Power*, *in* TAX SIMPLIFICATION 25-32 (Chris Evans et al. eds., 2015). The notions *tax assessment and collection* are used broadly, i.e. as the overall process of assessing taxpayers' income statements and actually collecting the taxes (including taking action against those who have not filed a return in time or paid their taxes when due). For more on these tax administration functions, *see Tax Administration 2022*, *supra* note 1, at 54, 122. The term *tax transparency* is used to refer to the transparency of taxpayers' affairs through the automatic exchange of information between states as well as through strengthened reporting and disclosure requirements for taxpayers and third parties. For more on the notion of tax transparency, *see* Johanna Hey, *General Report – The Notion and Concept of Tax Transparency*, *in* TAX TRANSPARENCY 3 (Funda Başaran Yavaşlar & Johanna Hey eds., 2019).

[5] For a similar example-based approach used in a general administrative justice context *see e.g.*, Jennifer Raso, *Implementing Digitalisation in an Administrative Justice Context*, *in* THE OXFORD HANDBOOK OF ADMINISTRATIVE JUSTICE 521 (Marc Hertogh, 2021).

[6] United Nations, E-Government Survey 11–14 (2020), and European Commission, Digital Economy and Society Index 66-76 (2021). *See also* Finansministeriet, Danmarks digitaliseringsstrategi – Sammen om den digitale udvikling [The Ministry of Finance, Denmark's Government's digitalization strategy – together for the digital development] (2022). The first national digitalization strategy was launched in 2001. For more, *see* Hanne Marie Motzfeldt & Azad Taheri Abkenar, DIGITAL FORVALTNING [DIGITAL MANAGEMENT] 20–22 (2019). Finally, *see* Digitaliseringsstyrelsen, Vejledning om digitaliseringsklar lovgivning [The Digitalization Agency, Guidelines for Digitalization Ready Legislation], Guideline nr. 9590 af 12.6.2018, which main aim is to ensure that digitalization is included in all of the preparatory work for new legislations from start to finish.

[7] The National Audit Office [*Rigsrevisionen*] is an independent institution placed under the Danish Parliament. Its main tasks are to determine whether public accounts are correct (financial audit) and to examine whether government-funded agencies and enterprises comply with current laws and regulations (compliance audit) as well as whether the administration has a sufficient focus on economy, efficiency, and effectiveness (performance audit). *See* Folketinget Rigsrevisionen, www.rigsrevisionen.dk (last visited

As the overall topic of the digital transformation of tax systems is extremely comprehensive, a number of delimitations had to be made to ensure a sufficient focus. Accordingly, we found it necessary to exclude explicit considerations on indirect taxes, duties, and tariffs. Moreover, we have not included considerations on developing countries' specific challenges with respect to the digitalization of their tax administration. While these areas are important, they exceed the scope of this paper and require further research.

It is part of the aim of this article to explicate relevant areas of Danish law as it stands (*de lege lata*) or as it stood. This explanation is done in accordance with the traditional Danish legal dogmatic method of interpretation and by relying on commonly accepted sources of law, including the wording of the tax provisions in question, statements in the *travaux préparatoires*, and Danish case law.[8] Furthermore, to facilitate a deeper understanding of the various processes that have led to the digital transformation of the Danish tax administration, historical, legal sources play a significant role in the article.[9]

Additionally, to provide a comprehensive insight into the international and Danish digital transformations of tax systems, broader considerations concerning good public administration are also included.[10] In this regard, a number of other sources are relied on as well, including reports from major international and Danish organizations and institutions, white papers, academic literature from

Sep. 8, 2022). The Danish Parliamentary Ombudsman's [*Folketingets ombudsmand*] main task is to help ensuring that the public administration acts in accordance with the law and good administrative practice, thus protecting citizens' rights vis-à-vis the administration. The Ombudsman investigates complaints and opens cases on his own initiative and carries out monitoring visits. *See* Velkommen til ombudsmanden, www.ombudsmanden.dk (last visited Sept. 8, 2022).

[8] For more on interpretation in Danish tax law *see e.g.*, Peter Koerver Schmidt, *Legal Pragmatism – A Useful and Adequate Explanatory Model for Danish Adjudication on Tax Avoidance?*, Nordic Tax J. 29 (2020).

[9] Hence, it may difficult or impossible to understand the present state of tax law and tax administration without knowing what led for current ills. *See e.g.*, Reuven-Avi Yonah, *Why Study Tax History?*, 48 INTERTAX 687, 687–89 (2020) (reviewing to it. In other words, to comprehend the current state-of-play one has to understand what the lawmakers were trying to achieve in the past. Further, solutions in tax tend to repeat themselves in cyclical fashion, and therefore studying the past can suggest remedies STUDIES IN THE HISTORY OF TAX LAW (Peter Harris & Dominic de Cogan eds., vol. 9 2021)).

[10] Broadly speaking the field of *public administration* is concerned with the institutional arrangements for the provision of public services and regulation of governmental activities, whereas administrative law examines these arrangements in terms of legal principles such as legality, fair procedure, and proportionate use of power. However, there is a close connection between the two disciplines. John S. Bell, *Comparative Administrative Law*, *in* The Oxford Handbook of Comparative Law 1254 (Mathias Reimann & Reinhard Zimmerman eds., 2019).

various research fields, statements from the Danish National Audit Office, and expositions from the Danish Ombudsman.

The remainder of this article is structured as follows: Section two includes a short introduction to the fields of taxation and tax administration. Section three contains an analysis and a discussion of the opportunities for simplifying tax administration through digitalization. Section four explores the digitalization of tax assessment and collection procedures. Section five analyzes and problematizes issues with tax transparency in a digital context. Section six considers the future possibilities for and challenges of further digitalization of the tax administration. Finally, Section seven presents and discusses the overall conclusions.

II. TAXATION AND TAX ADMINISTRATION IN A NUTSHELL

Even though the statutes of most states do not include an explicit definition of the notion of tax, the term is typically defined as a compulsory levy, which is imposed by an organ of government, for public purposes and without regard to the particular benefits received by a taxpayer (i.e., it is an unrequited payment).[11] There are three main goals of taxation: to raise revenue for necessary government functions and public purposes, to redistribute income, and to steer behavior.[12]

Taxation is typically a heavily regulated area, where an overwhelming amount of statutes and regulations prescribe how the taxable amount (i.e., the tax base) should be computed and how the tax payment should be calculated. This area of the law can be labeled material tax law. However, it is worth mentioning that tax law as a discipline normally is viewed as a subdiscipline within the field of administrative law and that formal tax law thus is concerned with broader questions concerning how a tax administration is authorized to work as well as which remedies that it has available.[13]

[11] Marjana Helminen, *General Report*, *in* The Notion of Tax and the Elimination of Double Taxation and Double Non-Taxation 17, 160–61 (Sdu Uitgevers, International Fiscal Association, 2016). The same understanding of the notion of tax applies in a Danish context. *See* Lars K. Terkilsen, *Denmark*, *in* The Notion of Tax and the Elimination of Double Taxation and Double Non-Taxation 297, 297–314, at 297 (Sdu Uitgevers, International Fiscal Association 2016).

[12] Reuven Avi-Yonah, *The Three Goals of Taxation*, 60 Tax L. Rev. 1, 3–4 (2006).

[13] At least in the continental European traditions, administrative law is concerned with the powers and organization of the executive organs of the state. *See* Bell, *supra* note 10, at 1252. Moreover, an accelerating diffusion of administrative principles among legal systems appears to take place. *See* Francesca Bignami, *Comparative Administrative Law*, *in* The Cambridge Companion to Comparative Law 167, 168–69 (Mauro Bussani & Ugo Mattei eds., 2012). *Material tax law* is sometimes also labeled *substantive tax law*, and

While a part of administrative law, tax law does have specific traits. Not only is the tax legislation often among states' lengthiest and most complex statutes, taxes also have the particular function of financing government operations and impact the vast majority of citizens in various ways.[14] Accordingly, the fact that taxation involves costly mass administration—as well as the fact that efficiently running a tax system is largely dependent on taxpayers' own reporting and self-assessments—should be kept in mind when analyzing and discussing the relations between tax administrations and taxpayers.[15] This also applies to the digital transformation of tax administrations.[16]

A. Tax Simplification and Digitalization

Even though simplicity is often highlighted as one of the central tenets of a good tax system,[17] tax legislation tends to be complex.[18] One explanation for this complexity is the reliance on income taxation, as the measurement of income inevitably contains difficult questions.

formal tax law is sometimes referred to as *procedural tax law*. *See* Pasquale Pistone, *General Report*, *in* Tax Procedures EATLP International Tax Series 18, 7–9 (2020).

[14] *See generally* Lawrence Zelenak, *Maybe Just a Little Bit Special – After All*, 63 Duke L.J. 1898 (2014) (this article is a response to the claim that tax law is no different from other areas of law; thereby, the article contributes to the longstanding discussion in the United States about so-called tax myopia or tax exceptionalism).

[15] The overall costs of running the tax system are often perceived to be high and may roughly be divided into three categories: 1) *distortion costs*, i.e., costs that arise when taxes affect taxpayers' decisions; 2) *administrative costs*, i.e., cost incurred by the tax administration in order to establish and operate systems to manage all aspects of taxation; and 3) *compliance costs*, i.e., costs incurred directly by taxpayers in order for them to comply with their tax-related obligations as well as for third parties involved in the process. *See* Jonathan Shaw et al., *Administration and Compliance*, *in* Dimensions of Tax Design – The Mirrlees Review 1100, 1105–06 (James A. Mirrlees & Stuart Adam eds., 2010).

[16] *See, e.g.*, the discussion in Benjamin Walker, *New Wave Technologies and Tax Justice*, *in* TAX JUSTICE AND TAX LAW 261 (Dominic de Cogan & Peter Harris eds., 2020).

[17] Dating all the way back to Adam Smith, *simplicity* has been hailed as a core principle of a good tax system, *see generally,* Adam Smith, An Inquiry into the Nature and Causes of the Wealth of Nations (Reprint by Elec. Book Co. 2000). In more recent times, *simplicity* was included in what has been referred to as the Ottawa Principles, i.e., a set of broad taxation principles that should apply to electronic commerce. S*ee* Committee on Fiscal Affairs, Electronic Commerce: Taxation Framework Conditions (1998) (presented to Ministers at the OECD Ministerial Conference, *A Borderless World: Realising the Potential of Electronic Commerce*). Years later, the importance of the Ottawa Principles was reaffirmed in the Final Report on Action 1 in the BEPS Project. S*ee* OECD, Addressing the Tax Challenges of the Digital Economy – Action 1: Final Report 134 (2015).

[18] *See generally* Joel Slemrod, *Why'd You Have to Go and Make Things So Complicated?*, *in* TAX SIMPLIFICATION 1 (Chris Evans et al., eds., 2015).

Accordingly, some of this complexity can be viewed as a necessary price to be paid in order to aid legislators in fine-tuning the income tax liability, that is, to personalize income taxation according to certain taxpayer characteristics, to obtain some desired level of redistribution in society (e.g., horizontal or vertical equity),[19] and to steer taxpayers' behavior in a certain direction (e.g., to invest more in the green transition or reduce pollution). However, it has been argued that part of this complexity is not for the purpose of contributing any social value but is merely caused by misguided legislative initiatives that have ended up distorting the economy.[20] The reasons behind such misguided attempts can be many, including the fact that legislators, tax officials, and taxpayers are all subject to cognitive limitations and that these limitations may be exploited.[21]

The need to protect tax systems against avoidance and evasion adds to this complexity and, more recently, the increasing mobility of taxpayers and the cross-border affairs of large, multinational enterprises have put further pressure on legislators and tax administrations.[22] As a consequence, many states have (often unsuccessfully) embarked on simplification reforms aimed at reducing the complexity of tax rules.[23] More successfully, technology to reduce the compliance burden for taxpayers—as well as for providing better and more reliable information to tax administrations—has played a significant role in many states. As further elaborated in the following sections, common examples are electronic filing of returns (often prefiled to some extent), online tax payments, and the delivery of online taxpayer assistance.[24]

[19] The concepts of horizontal or vertical equity have been subject to extensive debate in the legal literature. *Horizontal equity* means that taxpayers who are positioned identically relative to the tax base should pay equal tax, whereas the concept of *vertical equity* stipulates that taxpayers with different amounts of income or wealth should pay different amounts of tax. This latter concept is often reflected in states' use of progressive tax rates. For a discussion of the concepts, see, e.g., Ira K. Lindsay, *Tax Fairness by Convention: A Defense of Horizontal Equity*, 19 Fla. Tax Rev. 79 (2016).

[20] Slemrod, *supra* note 18, at 7.

[21] For more on the complexity and opacity of the United States' Federal Income Tax and a discussion of the consequences and possible technological solutions hereto, see, e.g., David I. Walker, *Tax Complexity and Technology*, 97 IND. L.J. 1095 (2022).

[22] *See generally* CONRAD TURLEY ET AL., *International Tax Administration Solutions in Major Countries*, *in* A NEW DAWN FOR THE INTERNATIONAL TAX SYSTEM (2017).

[23] At times, even well-meaning attempts to simplify the legislation have themselves created complexity. *See, e.g.*, Judith Freedman, *Managing Tax Complexity*, *in* TAX SIMPLIFICATION, *supra* note 20, at 253, 256.

[24] Turley et al., *supra* note 22. For a discussion of such possibilities in the context of the United States, see Joseph Bankman et al., *Using the "Smart Return" to Reduce*

B. The International Development

Initially, it should be recognized that, as countries differ in respect to their policy and legislative environment as well as administrative practices and culture, tax administrations face a varied environment within which to administer their taxation systems.[25] However, in general, most tax systems around the world operate with what may be described as a sequential process. This implies that taxpayers should be identified, and taxpayers are required to identify and report transactions and incomes as well as subject their income to the appropriate tax rules. On this basis, the tax obligation of each taxpayer should be calculated and paid. Subsequently, tax administrations should have the option to audit the tax assessment of each taxpayer and to enforce the taxation, and taxpayers should have the option to dispute the taxation.[26]

While obviously being country-specific, the general development of tax administration around the world has been characterized as evolving from Tax Administration 1.0 to 2.0, implying a digitalization of what was previously paper-based and manual sequential processes.[27] Further, the digitalization has created new opportunities for the data use and analytical tools by tax administrations to support the sequential tax administration processes.[28] Arguably, this development has resulted in efficiency gains and an increase in effectiveness of tax administration processes for taxpayers and the administration.[29]

Simplifying the sequential tax administration process for taxpayers through digitalization may be seen as a significant improvement in many ways. Notably, tax administrations have not only focused on reducing quantifiable costs from the administrative burden but also on costs associated with frustrations and anxiousness experienced by taxpayers uncertain of complex tax legislation and detailed reporting obligations.[30] Some of the digital initiatives, developed and implemented to support taxpayers in the sequential tax administration process, are so-called nudge techniques. These techniques aim to encourage and

Evasion and Simplify Tax Filing, 69 Tax L. Rev. 459 (2016), and Joseph Bankman, *Using Technology to Simplify Individual Tax Filing*, 61 Nat'l Tax J. 773 (2008).

[25] OECD, Using Third Party Information Reports to Assist Taxpayers Meet their Return Filing Obligations — Country Experiences with the Use of Pre-populated Personal Tax Returns 3 (2006).

[26] *See* OECD, Tax Administration 3.0: The Digital Transformation of Tax Administrations 10 (2020).

[27] *Id.* at 7, 76.

[28] *Id.* at 10.

[29] *See generally* Turley et al., *supra* note 22; OECD, *supra* note 26.

[30] *See generally* Walker, *supra* note 21.

promote correct taxpayer behavior and are based on behavioral insights into each individual taxpayer, online self-service tools, and targeted help. Examples of the latter are online live chats and virtual assistants who, based on artificial intelligence, provide information and to various extents fulfil assistance functions.[31]

Further, tax administrations use the traditional media and social media to communicate general information, deadlines, and updates to taxpayers.[32] Another increasing trend among tax administrations in their effort to simplify tax compliance is the use of mobile apps, which are becoming increasingly transactional. The most sophisticated apps offered by tax administrations are now a primary way for taxpayers to access information and personal tax accounts, to communicate with the tax administration, to submit information and tax returns, and to pay taxes.[33]

However, while some at the forefront of tax administration provide full-service mobile apps for specific parts of the taxation system, most tax administrations still rely on e-filing and e-payment channels. Accordingly, in a survey conducted by the OECD with respect to average e-filing, for the years 2018 to 2020, it was concluded that in the participating countries, more than 90 percent of business taxpayers submitted their tax returns electronically, whereas 85 percent of personal income tax returns were submitted electronically—both types of taxpayer returns had experienced an increase of approximately 19 percentage points since 2014. In assessing these figures, it should be noted that, for a number of tax administrations, a 100 percent e-filing rate has already become a reality.[34] As for e-payments rates, more than

[31] Alfredo Collosa, *Digitization of Tax Administrations, and Facilitation of Tax Compliance*, Inter-American Center of Tax Administrations (Oct. 4, 2021), https://www.ciat.org/ciatblog-digitalizacion-de-las-administraciones-tributarias-y-facilitacion-del-cumplimiento-tributario/?lang=en; OECD, *supra* note 26, at 10–11; *see also* Johanna Hey, *General Report – The Notion and Concept of Tax Transparency*, *in* Tax Transparency 3 (Funda Başaran Yavaşlar & Johanna Hey, eds., 2019), section, 1.2.3.3. (the role of digitalization).

[32] *See* OECD, *supra* note 26.

[33] OECD, Tax Administration 2021 Comparative Information on the OECD and other Advanced and Emerging Economies 87 (2021). As examples, Brazil's tax and customs "Normas" and Russia's special tax regime "Professional income tax" are discussed.

[34] OECD, Tax Administration 2022: Comparative Information on the OECD and other Advanced and Emerging Economies 55–56 (2022). The number of countries that were able to provide the average e-filing rates for the years 2018–2020 were 47 with respect to business income tax returns and 50 with respect to personal income tax returns. However, only 33 and 31 countries were able to provide information on the average e-filing rates for the years 2014 and 2020 with respect to business income tax returns and personal income tax returns respectively.

86 percent of tax payments measured by number and more than 88 percent measured by value were made electronically in 2020.[35] The slightly higher percentage of e-payments by value suggests that larger taxpayers particularly use e-payment.

While these figures all suggest an increased simplification in the tax administration process through options for e-filing and e-payment of taxes, a number of jurisdictions still experience a high volume of paper-based tax returns as well as payments through nonelectronic means, although this has been significantly reduced during the COVID-19 pandemic and is expected to decline further over time.[36]

A subsequent simplification step to e-filing and e-payment (although prior in the sequential tax administration process) is the prefilled tax return, where the tax administrations make a draft of the tax return available to taxpayers by populating the taxpayer's return with information typically provided from third parties.[37] A number of benefits from implementing prefilled tax returns have previously been discussed and may, inter alia, include a reduction in taxpayer compliance costs and system costs of the tax administration in the time taxpayers spend on the return and in the volume of involuntary errors by the taxpayers.[38]

A prerequisite for offering such prefilled tax returns is the construction of a comprehensive and reliable information system with large-scale agile information processing.[39] However, the complexities of the legal frameworks governing taxes are a challenge to more automated tax calculations, and, while machine-readable legislation can help automate the calculation process using algorithms, the capabilities of information systems in this respect have been and remain limited.

To account for the limitations of information systems, prefilled tax returns may initially be used for simple and frequent types of taxpayers (thereby decreasing requirements for the capabilities of the information

[35] *Id.* at 56. With respect to the average e-payment rates for the years 2018–2020, 47 countries were able to provide this information.

[36] *Id.*

[37] OECD, *supra* note 33, at 62–66. In the report, examples from China, New Zealand, Norway, Peru, Russia, and Spain are discussed. *See also* Alfredo Collosa, *Pre-prepared Tax Statements: An Instrument of Facilitation and Control*, Inter-American Center of Tax Administrations (May 28, 2021), https://www.ciat.org/pre-prepared-tax-statements-an-instrument-of-facilitation-and-control/?lang=en. The author argues that preprepared tax returns are very effective to achieve the goal of making it as easy as possible for taxpayers to comply with tax obligations in a simple way and without feeling doubt.

[38] *See also* Collosa, *supra* note 37 (This author also discusses the benefits of a reduction in postverification programs, improvements of the impression of the tax administration, as well as the perception that it is acting in real time as well, as increased collection.)

[39] *Using Third Party Information Reports*, *supra* note 25, at 10.

system), that is, within tax regimes that allow few deductions and credits and in places where the tax return can be verified with third-party data sources.[40] A typical example of suitable taxpayers to be offered prefilled tax returns is employees where the employer provides data to tax administrations.[41]

As already indicated, another prerequisite for offering prefilled tax returns is that tax administrations are able to collect all the relevant data. Therefore, obliging third parties to report information is regarded as a vital step in offering prefilled tax returns.[42] Accordingly, comprehensive systems requiring third parties to report income, for example, employment-related payments, such as wages, bonuses, and other fringe benefits, should typically be reported by the employer, whereas interests and dividends should typically be reported by financial intermediaries. Reporting obligations covering assets might include the sales and purchases of shares and bonds, which should typically be reported by financial intermediaries, and deduction-related information may be information on union fees, home mortgage interest, contributions to unemployment insurance and retirement savings plans, and childcare expenses typically reported by a number of intermediary third parties.[43] An inherent advantage of imposing reporting obligations on such large and institutionalized third parties is that they generally have the capacity to professionalize the processes, and they are likely to benefit from economies of scale.[44]

Some argue that the sum of information provided by the taxpayer and third parties offers tax administrations a high level of transparent tax data on each taxpayer, which is justified by principles of legal and equal taxation and the general public interest.[45] However, it is challenging to balance between the convenience of collecting information without the need for cooperation by (or knowledge of) the taxpayer and the risk of jeopardizing taxpayers' trust in the tax administration. Consequently, as is further discussed below, the process of collection and utilization of the collected information needs to be transparent for the taxpayer as well.[46]

Unsurprisingly, following the principle of garbage in garbage out,

[40] *See Tax Administration 2022*, *supra* note 1, at 57.

[41] *Using Third Party Information Reports*, *supra* note 25, at 10.

[42] *See Tax Administration 2022*, *supra* note 1, at 57–62.

[43] *See Tax Administration 2022*, *supra* note 1, at 57–62; *Using Third Party Information Reports*, *supra* note 25, at 10.

[44] *See Using Third Party Information Reports*, *supra* note 25, at 10.

[45] Roman Seer, *Purpose and Problems of Tax Transparency: The Legal Perspective, in* 17 TAX TRANSPARENCY 17, pg. 2 in online version (Funda Başaran Yavaşlar & Johanna Hey eds., 2019).

[46] Hey, *supra* note 3.

prefilled tax returns are only as good as the data received by the tax administrations. Accordingly, it has been argued that an important aspect of prefilled tax returns is taxpayers' self-assessment—even when the taxpayer merely confirms the proposal received.[47] This system may be implemented by requiring all taxpayers to respond, either by confirming that the return gave a complete and accurate picture of the taxpayer's tax affairs or by adjusting the information included in the prefilled tax return. Alternatively, a system of deemed acceptance can be adopted, that is, the taxpayer only has to react if the taxpayer has amendments or additions.[48] However, as is further discussed below regarding the practical experiences from Denmark, maintaining the self-assessment obligation upon taxpayers may imply challenges from a taxpayers' right perspective. As an example, it may be difficult for taxpayers to review their tax returns based on information that they have not provided themselves and combine this information with a sufficient understanding of the rules applicable to their tax affairs.

C. Danish Experiences

In many ways, the Danish development correlates with the international tendencies presented in the previous section. Hence, in 1995, Denmark made it possible for individuals to file their tax returns online. Accordingly, Denmark was among the first movers when it came to utilizing IT for such tax administration purposes, and, already in 2004, 68 percent of the tax returns of individuals were handled online, increasing to 96 percent in 2009. Moreover, it was made possible for corporations to file their tax returns online as of 2005.[49]

To a large extent, these tax returns are automatically prefilled using information received from various intermediaries, such as employers, banks, and pension funds. The approach includes a deemed acceptance of the prefilled tax return after the expiry of a notice period. For a significant part of the Danish individual tax base, complete online prefilled tax returns are thus being generated.[50]

[47] *See also* Collosa, *supra* note 37.

[48] *See Using Third Party Information Reports*, *supra* note 25, at 10.

[49] Jørgen G. Christensen & Peter B. Mortensen, Overmod og afmagt, 36–37 (2018); *see generally* Org. for Econ. Co-operation and Dev. [OECD], Tax Administration 2017: Comparative Information on OECD and Other Advanced and Emerging Economies (2017) (In 2014, it became mandatory for Danish corporations to file their tax returns online). For more about this requirement, *see* Mette B. Larsen and Asger L. Høj, DIAS – digitalisering af selskabsselvangivelsen, SR-Skat, 104 (2015) (Nowadays, in Denmark, all filing of tax returns for individuals as well as for corporations take place online.) *See also Tax Administration 2022*, *supra* note 1 at Table D13.

[50] *See Tax Administration 2022*, *supra* note 1, at Table A.46.

The Danish success in implementing e-filing and prefilled tax returns may partly be explained by the fact that Denmark already had introduced taxation at source for employees (pay-as-you-earn taxation or PAYE) in 1970.[51] Hence, a condition for the efficient operation of the PAYE system was an increased use of electronic data processing.[52]

To fully grasp the importance of this head start, the historical development of the Danish PAYE system will be explored further, as it illustrates how Denmark, as one of the first countries in the world, was able to lead its tax administration procedures in a digital direction.[53]

The preparation for the introduction of the Danish PAYE system had started back in the late 1950s when the Danish Ministry of Finance invited the International Business Machines Corporation (IBM) to assist in analyzing and defining the necessary preconditions and barriers for introducing such a system. Simultaneously, a joint venture was launched between the central government and the municipalities with the task of acquiring computers as well as the task of developing and operating the new PAYE tax system and related supporting systems.[54]

Among other things, these efforts led to the creation of the Central Personal Registration System in 1965 and to the creation of the Central Registration System for Companies and Employers in 1975.[55] The reason behind the creation of these registers was to ensure that each taxpayer, employer, and other information-providing third parties could be uniquely identified.[56]

In 1972, a report was published in which an appointed committee gave a number of recommendations on how to improve the efficiency of the PAYE tax system, including how to create a system that, as far as possible, made additional payments unnecessary and eased the burden of control.[57] Many of these suggested improvements (as well as other improvements) were implemented during the 1970s and 1980s.[58]

[51] Lov. nr. 100 af 31.03.1967 om Kildeskatteloven [Act on Taxation at Source], (Den.).

[52] Christensen & Mortensen, *supra* note 49, at 36–37.

[53] Lov. nr. 100 af 31.03.1967 om Kildeskatteloven [Act on Taxation at Source], (Den.).

[54] Søren D. Østergaard, *The Danish Tax System and the 'No Touch Strategy' in* History of Nordic Computing 58–64 (Christian Gram et al. eds., 2014).

[55] Lov nr. 239 af 10.06.1968 om folkeregistrering [Act on Personal Registration] (Den.); Lov nr. 151 af 24.04.1974 om erhvervsregistret [Act on Registration of Businesses] (Den.).

[56] *See generally* Østergaard, *supra* note 54.

[57] Betænkning fra udvalget til forbedring af kildeskatten [Report from the Commission on Improving Taxation at Source] Report no. 638 (1972) 6–7 (Den.).

[58] Østergaard, *supra* note 54 at 60–61.

This development paved the way for the launch of an integrated digital self-service strategy in the mid-1990s. The strategy contained a so-called no touch goal with an aim of ensuring that taxpayers interacted with the tax administration in the most cost-effective and time-effective way, that is, by enabling taxpayers to help themselves online through the website of the tax administration or by communicating online with tax officers.[59] In general, this strategy has proved to be a success, and, in 2004, the Danish tax administration was awarded the Danish eCommerce Prize (E-handelsprisen) for their digital self-service tax system called "TastSelv."[60]

Obviously, the journey toward a digitally based tax system has only been made possible through massive investments in IT, but, along the way, several legislative changes have also been made to facilitate the transition, to provide sufficient legal basis, and to clarify the responsibilities of taxpayers, third parties, and the tax administration.

Important examples of these initiatives are the adoption of the new Tax Control Act and the Act on the Reporting of Information to the Tax Administration in 2017.[61] The former contains rules concerning the obligations and responsibilities of taxpayers when providing information to the tax administration as well as rules on the powers available to the tax administration. The latter contains rules on the obligations for employers, banks, and others to report information about taxpayers to the Danish tax administration.

These two laws replaced a principal act dating back to 1972,[62] which partly was based on a principal act dating all the way back to 1946.[63] Accordingly, it was broadly agreed that there was an urgent need to replace the old legislation, which, through numerous smaller amendments over the years, had become an unsystematic patchwork.[64] Moreover, it was argued that the old legislation did not sufficiently take the digital transformation of the tax filing processes and tax control

[59] Org. for Econ. Co-operation and Dev. [OECD]., *SURVEY OF TRENDS AND DEVELOPMENTS IN THE USE OF ELECTRONIC SERVICES FOR TAX PAYER SERVICES DELIVERY* 16 (Mar. 2010), https://www.oecd.org/tax/administration/45035933.pdf.

[60] Lise Sønnichsen, *TestSelvhistorie*, 13 SKATTEREVISOREN 1, 100 (2010).

[61] Lov nr. 1535 af 19.12.2017 om Skattekontrolloven [Act on Tax Control] (consolidated act 283 of 03.02.2022) (Den.), and Lov. nr. 1536 af 19.12.2017 om Skatteindberetningsloven [Act on Reporting of Information to the Tax Administration] (consolidated act 1754 of 30.08.2021) (Den.).

[62] Lov nr. 568 af 16.11.1972 om selvangivelse af indkomst og formue [Act on self-assessment of income and property] (Den.).

[63] Lov nr. 392 af 12.07.1946 om selvangivelse af indkomst og formue [Act on self-assessment of income and property] (Den.).

[64] Jan Pedersen, *Ny skattekontrollov*, SR-Skat 33 (2018).

processes into account,[65] including the fact that most individual taxpayers no longer prepared and provided a tax return, but rather received an electronically generated yearly statement containing information provided by various third parties, which often could not be amended online by the taxpayer. In this context, it appeared reprehensible to still hold the taxpayer liable for the correctness of the information in the yearly statement generated and obtained in this way.[66]

Accordingly, in the new Tax Control Act, the requirement to prepare and provide a tax return was replaced by an obligation to disclose relevant information.[67] Moreover, it is now explicitly stated that this obligation does not comprise liability for information, which is or should have been provided by third parties, provided that the third party is independent of the taxpayer and that the information is to be used in the yearly statement.[68]

Despite these improvements, concerns still exist when it comes to the question of whether taxpayers' rights are sufficiently protected in the digital era.[69] As a consequence, the Danish Ombudsman launched a thorough investigation into the tax administration's digital procedures and IT systems in 2021.[70] One aim of this investigation is to examine a number of existing IT systems in order to assess whether the systems sufficiently support the Danish tax administration in complying with their obligations following from general administrative law, as set out in the in the Public Administration Act as well as in the General Data Protection Regulation (GDPR).[71] Another aim of the investigation is to

[65] Lovforslag nr. 13 2017/2018 [bill no. 13 2017/2018] (Den.).

[66] *See also*, the criticism put forward by Borger- og Retssikkerhedschefen, *Redegørelse fra arbejdsgruppen ved skattekontrollovens ansvarsregler* [The Head of Tax Payers' Rights, Report from the Working Party on the Tax Control Law's Rules on Liabilities and Sanctions] (2011).

[67] Lov nr. 1535 af 19.12.2017 om Skattekontrolloven § 2(1) [Act on Tax Control] (Den.).

[68] *Id.* at § 2(2).

[69] For more on digitalization and general administrative law in a Danish context *see e.g.*, PER B. SØRENSEN, *FORVALTNINGSRET MED ET DIGITAL PERSPEKTIV* (2017).

[70] Folketingets Ombudsmand [The Danish Parliament's Ombudsman], News release of Mar. 17, 2021, Skattekontoret sætter fokus på digitalisering hos skattemyndighederne (2021) (Den.).

[71] Lov nr. 571 af 19.12.1985 om Forvaltningsloven [Public Administration Act] with later amendments (consolidated act no. 433 of 22.04 2014) (Den.), and 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, 2016 OJ (L 119) together with Lov nr. 502 af 23.05.2018 om Databeskyttelsesloven [Danish Act on Supplementing Provisions to the Regulation on the protection of natural persons with regard to the

look at a number of IT projects that have not been finalized yet to ensure that the administrative law requirements are sufficiently considered already in the developing phase. At the time of writing, the Ombudsman's thematic report on the tax administration's digital procedures and IT systems has not been published.

Over the years, the Ombudsman has closed a number of more limited and more narrowly scoped investigations concerning the use of digital tools and applications in the Danish tax administration. One of these investigations concerned the so-called One Tax Account system.[72] This system was meant to facilitate payments between businesses and the tax administration through one single account. However, when the new account was launched, the system contained an error, which entailed that no interests were levied on the outstanding payments for a major part of the businesses. The accumulated interest was subsequently collected manually with significant extra costs and inconvenience for the tax administration as well as the concerned taxpayers. Against this background, the Ombudsman criticized that the system was launched despite containing deficiencies that harmed a great number of taxpayers.[73]

Another of the Ombudsman's investigations concerned a new IT system that was meant to facilitate automatic transfers to the tax administration of information concerning transactions taking place on sharing platforms focused on renting out houses, apartments, summer cottages, and more (e.g., Airbnb). In connection to these transactions, the Ombudsman highlighted the fact that even though the new IT system only collected and transferred information—and thus did not directly generate any decisions itself—the information provided through the new system was also used by other agencies. These agencies then used the transferred information when making administrative decisions. Against this background, the Ombudsman emphasized that it is important that the agency responsible for the new IT system sufficiently takes account of the possible uses of the generated information by other agencies. Moreover, the Ombudsman considered it deplorable that the taxpayers' right to legal representation had not

processing of personal data and on the free movement of such data] (Den.). *See also*, paragraph 5.2 below.

[72] The legislation behind the One Tax Account system [Èn Skattekonto] was adopted in 2006. *See*, Lov nr. 513 af 07.06.2006 om opkrævning via én skattekonto [Act on Tax Collection Through One Account] (Den.). The rules took effect from Aug. 1, 2013. *See*, Ministerial Order nr. 577 of 30.05.2013 (Den.).

[73] Redegørelse fra Folketingets Ombudsmand [The Ombudsman's Report], case no. 17/03200 3–4 (Den.).

sufficiently been taken into account during the system development.[74]

Finally, the Ombudsman has also examined how the Danish tax administration communicates with taxpayers through social media.[75] In this respect, the Ombudsman has emphasized that the tax administration's opening and administration of a Facebook account should be considered an activity within the scope of public administration and that common administrative law rules should be adhered to. Accordingly, the Ombudsman highlighted that proper procedures should be in place for answering and archiving enquiries from taxpayers, for the preparation of full notes, and for the handling of confidential information.[76]

In conclusion, the Danish tax administration has come a long way in making tax processes smoother and more efficient through the use of digital tools. A prominent example is the fact that most individual taxpayers no longer need to prepare and file a tax return, as they automatically receive a prefilled, digitally prepared, yearly statement. However, the Danish development also shows that digitalization poses a number of challenges with respect to the protection of taxpayers' rights, including a need to recalibrate taxpayers' information obligations and liabilities as well as a need to alter internal tax administration procedures.

III. DIGITALIZATION OF THE TAX ASSESSMENT AND COLLECTION

The primary purpose of any tax administration is arguably the collection of tax revenue.[77] In doing so, the tax administration has to examine the completeness and correctness of tax returns, assess tax obligations, collect the taxes (sometimes by force), and provide guidance to taxpayers.[78]

Accordingly, tax administrations are faced with a number of heavy and costly tasks. In the aftermath of the financial crisis of 2008, many states have tried to stretch the resources allocated to their tax administrations further, for example, by carrying out reorganizations,

[74] *See* Redegørelse fra Folketingets Ombudsmand [The Ombudsman's Report], case no. 21/01499 (Den.).

[75] *See* Udtalelse fra Folketingets Ombudsmand [Ombudsman's Statement], case no. 18/03627 (Den.).

[76] For a thorough examination of the tax administration's (digital) communication channels *see also,* Borger- og Retssikkerhedschefen, *Undersøgelse af skatteforvaltningens kommunikations-kanaler* [The Head of Taxpayers' Rights, Examination of the Tax Administration's Communication Channels] 67–70. (2011).

[77] *See Tax Administration 2022*, *supra* note 1, at 32–33.

[78] Matthijis Alink & Victor van Kommer, *Core Business of Tax Administration*, *in* Handbook Of Tax Administration chap. 2 (2015).

auditing taxpayers based on information-driven risk profiling, and enhancing the digital processes and tools deployed.[79] Concerning the latter, it has, on the one hand, been argued that IT has the potential to cut the costs of processing taxpayer information, to reduce the risk of errors, and to help to expose noncompliance. However, on the other hand, it has been argued that new IT can be extremely difficult and costly to implement.[80] In the following subsections, we will explore these opportunities and challenges further.

A. The International Development

As stated, an important and time-consuming function of tax administrations is the assessment of accuracy and completeness of reported information. Generally, this assessment has happened and to a great extent still happens through different audit types, such as comprehensive or issue-oriented audits, inspections of books and records as well as in-depth investigations of suspected tax fraud, and potential visits to the taxpayer's premises.[81] However, accelerated as a result of the COVID-19 pandemic, advances in technology have led administrations to consider new ways of engaging with taxpayers during the audit process, including electronic submissions of audit-related documentation, increased use of automated electronic checks, and validations and matching of reported information.[82] As it is further discussed below, the transparency of these digitalized compliance actions is critical to supporting voluntary compliance, perceptions of fairness, and prophylactic effects.[83]

A significant part of more targeted and managed compliance is driven by the increased availability of data. Therefore, most tax administrations now apply data science techniques and analytical tools in audit case selection and analytics, including behavioral analysis, to build a more holistic understanding of compliance risks, behavioral patterns, and appropriate compliance interventions.[84] This approach allows tax administrations to better identify the tax returns, claims, or transactions, which may require further scrutiny. Furthermore, these

[79] *See generally* Turley et al., *supra* note 22.

[80] Shaw et al., *supra* note 15.

[81] *See Tax Administration 2022*, *supra* note 1, at 96.

[82] *See generally* Turley et al., *supra* note 22; *see Tax Administration 2022*, *supra* note 1.

[83] *See Tax Administration 2022*, *supra* note 1, at 96.

[84] *See generally* Org. for Econ. Co-operation and Dev. [OECD], *Advanced Analytics for Better Tax Administration: Putting Data to Work*, 20 (May 13, 2016), https://read.oecd-ilibrary.org/taxation/advanced-analytics-for-better-tax-administration_9789264256453-en; *see also Tax Administration 2022*, *supra* note 1, at 97, 102.

models—many of which can operate in real time—allow administrations to conduct automated electronic checks on all returns or on transactions of a particular type and allow administrations to use rule-based approaches to treat some defined risks. Therefore, these models provide tax administrations with more effective and efficient ways to undertake the assessment of taxpayer compliance.[85]

In terms of collecting taxes, the payment of taxes has, as explained above, largely been digitalized over the years with the majority of tax payments now being paid electronically. Further, the collection of taxes that have not been paid timely has been digitalized. The collection of outstanding tax payments is not only important for financing public spending,[86] but it is also important for maintaining high levels of voluntary compliance and citizens' trust in the overall tax system.[87] Accordingly, as the main goal of taxation is to collect revenue, the goal of debt management will arguably be to increase the net present value of outstanding tax debt while respecting legal principles and the perceived fairness of the tax system.[88]

The traditional approach for tax collection is a standard process that is applied uniformly to all debt until it is either paid or written off. A more effective approach, however, focuses on the debtor instead of the debt and uses advanced analytics and behavioral sciences to understand the driving mechanisms of the debtor's behaviour. Accordingly, instead of following a fixed order, predictive techniques may be used to identify taxpayers who are unlikely to meet their obligations but likely to respond to debt-management intervention, while prescriptive techniques may be applied to determine how to communicate most effectively with these segmented taxpayers.[89]

Accessing the data in the chain of the collection system and having the data properly structured are prerequisites for building risk models with predictive power that can be used to forecast payment behavior. For example, a risk model may predict payment behavior by distinguishing between cases where the tax debt is likely to be paid without further intervention and cases where early intervention is almost certainly needed. Based on such predictive models, tax administrations can target their nudge campaigns. For example, they

[85] *See Tax Administration 2022*, *supra* note 1, at 111–12.

[86] S*ee Tax Administration 2021*, *supra* note 33, at 44.

[87] *See Tax Administration 2022*, *supra* note 1, at 129–30.

[88] *See* Org. for Econ. Co-operation and Dev. [OECD], *Working Smarter in Tax Debt Management*, 29 (Oct. 24, 2014), https://read.oecd-ilibrary.org/taxation/working-smarter-in-tax-debt-management_9789264223257-en.

[89] *See Advanced Analytics for Better Tax Administration*, *supra* note 84, at 24, 26; *see also Working Smarter in Tax Debt Management*, *supra* note 88, at 19; *Tax Administration 2021*, *supra* note 33, at 136–37. *See generally* Turley et al., *supra* note 22.

can change how choices are presented without limiting the taxpayers' options or economic incentives.[90]

As discussed above, tax administrations are rather data-rich organizations in terms of information submitted by third parties. However, tax administrations also have information on the historical performance of taxpayers and their previous interactions with taxpayers.[91] Accordingly, one of the biggest challenges is integrating and understanding all of the available data to gain deep insights into taxpayers' behaviour and payment risks. This process is difficult because tax systems usually are built around individual taxes and tend to focus on individual debt claims rather than the taxpayer.[92] Unsurprisingly, risk modelling and analytics require expertise and competent people who understand both data analysis and business analysis and who can collaborate with tax administration experts. In other words, data scientists are a prerequisite for successfully developing analytical tools for collecting taxes.[93]

It has been argued that the application of advanced analytics is particularly suited for tax debt management, as the payment cycle is relatively short—either the debt is paid, or it is not. This system makes it easier to run trials supported by behavioural insights strategies for testing different wording in reminder letters and assessing the results in terms of payments. In this respect, it is important that tax administrations strike a balance between collecting the amounts due and assisting taxpayers to avoid distress. In other words, tax administrations should maintain a compliant attitude among taxpayers to avoid a reputation of being too lenient on the speed of recovery of tax debt and to ensure equal and consistent treatment of taxpayers.[94]

In practice, the debtor-oriented approach is usually based on segmentation.[95] Initially, segmentation will usually divide taxpayers

[90] *See* Org. for Econ. Co-operation and Dev. [OECD], *Successful Tax Debt Management: Measuring Maturity and Supporting Change*, 87 (Mar. 28, 2019), https://www.oecd.org/tax/forum-on-tax-administration/publications-and-products/successful-tax-debt-management-measuring-maturity-and-supporting-change.pdf.

[91] *See Working Smarter in Tax Debt Management*, *supra* note 88, at 23; *see also,* e.g., *Successful Tax Debt Management: Measuring Maturity and Supporting Change, supra* note 90, at 21.

[92] *See Working Smarter in Tax Debt Management*, *supra* note 88, at 24.

[93] *See id.* at 25.

[94] *See Tax Administration 2021*, *supra* note 33, at 130; *see also Working Smarter in Tax Debt Management*, *supra* note 88, at 24.

[95] *See* Antonio Faúndez-Ugalde et al., *Use of Artificial Intelligence by Tax Administrations: An Analysis regarding Taxpayers' Rights in Latin American countries*, 38 COMPUT. L. & SEC. REV. 1, 3–4, 6 (2020); *see also Advanced Analytics for Better Tax Administration*, *supra* note 84, at 25, 30. *See generally* Turley et al., *supra* note 22.

into large taxpayers, smaller and medium-sized enterprises, and individual taxpayers.[96] However, effective segmentation has been argued to require at least three additional levels of maturity:[97]

1. subsegmentation based on certain debt and taxpayer characteristics, e.g., debt size, debt age, and the business sector;

2. risk-based clustering by incorporating taxpayer behavior, which allows tax administrations to apply more targeted strategies to high-risk, noncompliant taxpayers and to apply strategies that are based on the value and complexity of collecting tax debt from other segments of taxpayers;

3. dynamic risk clustering aiming at improving the matching of treatments to each risk cluster and debt prevention must be dynamic by using a feedback loop to ensure continuous improvement.

Accordingly, based on the characteristics of both the debt itself and the debtor, a debt will then receive a risk classification. A group of taxpayers with the same classification can be clustered, enabling a segmented approach whereby the tax administration applies similar debt treatments across the group. This cluster also supports a level playing field by ensuring that taxpayers with similar characteristics are treated consistently.[98]

In the following section, Danish experiences with the digitalization of tax assessment and collection will be discussed, including a number of unsuccessful (to say the least) implementations of digital tools that aimed at harvesting rationalization and efficiency gains without appropriately accounting for the inherent risk in developing and applying a new IT system.

B. Danish Experiences

In terms of applying digital means to improve tax assessment procedures, Denmark has successfully implemented a number of the analytical tools discussed in the previous section, including audit case

[96] *See Working Smarter in Tax Debt Management*, *supra* note 88, at 25.

[97] *See id.* at 27–30.

[98] *See id.*at 35. *See generally* Turley et al., *supra* note 22.

selection. However, as further elaborated below, Denmark has struggled to find a suitable digital solution for the collection of tax debt.

Nowadays, all parts of the public administration in Denmark use digital tools to support, to steer, and sometimes even to decide on cases. The Danish tax administration, among other agencies, makes use of such tools most noticeably.[99] Accordingly, robotic process automation is widely used to collect data, and, combined with the use of various digital templates and case-handling systems, many procedures within the public administration have become highly automated.[100]

One of the most prominent examples is the fully automated digital generation of yearly statements for most individual taxpayers.[101] However, as taxpayers still have the possibility to add information to the automatically generated yearly statement, control measures have to be put in place. Among these measures are digital blockers, which the Danish tax administration started using as of the income year 2017.[102] These digital blockers are able to prevent taxpayers from adding information to their yearly tax statement through the self-service IT system (TastSelv) if the added information does not conform to the information that the tax administration already has obtained about the taxpayer.[103]

Further, as of the income year 2018, the Danish tax administration started applying a broader and more advanced digital tool that is capable of going over all of the added information by taxpayers through the self-service IT system. The digital tool has the capacity to go

[99] Motzfeldt & Abkenar, *supra* note 6, at 32.

[100] Hanne M. Motzfeldt & Emilie Loiborg, *Digital sagsbehandling – om legalitet, styring og interne regler* [*Digital Case Processing – About Legality, Management, and Internal Rules*], FESTSKRIFT TIL BENT OLE GRAM MORTENSEN, at 262–64 (2022). Redegørelse fra Folketingets Ombudsmand [The Ombudsman's Report], Opgradering af Skatteforvaltningens ESDH-system [Upgrading the Tax Administration's ESDH System], Case No. 21/01501 (2022) (emphasizing that it is paramount that the responsible agency monitors whether such systems support correct application of relevant legislation and sufficiently takes account of the possible interaction with other agencies' case handling systems). *See* Udtalelse fra Folketingets Ombudsmand [The Ombudsman's Statement], Skatteforvaltningens registrering af og korrespondance med partsrepræsentanter i ældre it-systemer [The Tax Administration's Registration of and Correspondence with Party Representatives in Older IT Systems], Case No. 21/00385 (2021). This criticism has also been raised with respect to a number of other IT systems used by the tax administration.

[101] *See supra* Part 3.2.

[102] Press Release, Skattestyrelsen [Danish Tax Admin.], Digitalt kontrolværktøj stopper forkerte indtastninger af fradrag i TastSelv for knap 50 mio. kr. [Digital Tool Stops Incorrect Typing-in of Deductions for 50 Million Danish Kronor] (Mar. 5, 2020).

[103] *See also* Folketingets Skatteudvalg [Danish Parliament's Tax Comm.], SAU alm. del endeligt svar på spørgsmål 361, [Final Answer to Question 361] (2016-2017) (offering more about the background of the tax administration's use of digital blockers).

through approximately 1,000 yearly statements every minute and is able to point out which taxpayers have added information to the yearly statement that is atypical or, in other ways, diverges significantly from the average pattern.[104]

Moreover, with respect to the rather new and relatively popular phenomenon of trading in cryptocurrencies, the Danish tax administration has started using a digital control tool that can help decode the extensive amount of data related to such transactions and find out where the risk of incorrect filing is highest. At the same time, the tax administration has started using knowledge and information from other areas of control, such as information from banks on cross-border fiat money transfers, when auditing taxpayers with such activities. This information has given the tax administration a more accurate picture of what is actually going on with respect to taxpayers' trading in cryptocurrencies, and the efforts have already led to the amendment of several taxpayers' yearly statements.[105]

However, these tools are not the first examples of the tax administration using IT to select taxpayers for individual audit. For instance, back in 2006, the tax administration started using a digital tool to select businesses for tax audit.[106] The tool was developed against the backdrop of legislation enacted in 2004, which was intended to pave the way for digital submissions of information of financial accounts and for enabling the tax administration to move toward a more digital and risk-based audit approach.[107] On the basis of the data received digitally

[104] Danish Tax Administration, *supra* note 102 (in the very first year of application, the use of the new digital tool led to the manual investigation of 1300 taxpayers pin-pointed by the tool).

[105]. Press Release, Skattestyrelsen [Danish Tax Admin.], Kryptovaluta [Crypto Currencies] (Mar. 2020); *see* Skatterådet [The Tax Council], SKM2019.15.SKTST, Pålæg af oplysningspligt efter skattekontrollovens § 8 D, stk. 1- handel med virtuel valuta [Imposition of a Duty to Provide Information Pursuant to Section 8 D, subsection 1 of the Tax Control Act Trade in Virtual Currency] (2019). With the permission of the Danish Tax Council, the Danish Tax Administration has gained access to information from three Danish exchange services for transactions with cryptocurrencies. *See infra* Part 5.2 (discussing more on Denmark's exchange of information with other states). Moreover, the Danish tax administration has received information about the transactions of Danish citizens from a Finnish exchange service for transactions with cryptocurrencies.

[106]. Folketingets skatteudvalg [Danish Parliament's Tax Comm.], Vedrørende lov nr. 1441 af 22. december 2004 om digitalisering af regnskabsoplysninger mv. (L 31, folketingsåret 2004/05, 1. samling) [Regarding Act No. 1441 of 22 December 2004 on Digitization on Accounting Information etc. (L 31, Parliamentary Year 2004/05, 1st session)], SAU alm. del, annex 36 (2006-2007).

[107]. Lov nr. 1441 of 22.12.2004 *Lov om digitalisering af regnskabsoplysninger, ophævelse af virksomheders underretningspligt og afskaffelse af kildeskattebøderne* [Act on Digitalisation of Financial Information, Removal of Businesses' Information Duties and Abolition of Withholding Tax Fines] (Den.).

and annually from the businesses as well as other data already collected by the tax administration, the new tool should thus assist the Danish tax administration in selecting businesses for further manual audit more effectively.[108]

The legislation enacted in 2004 coincided with the beginning of a major renovation and renewal of the Danish tax administration's IT systems. One reason for the administration to embark on this ambitious project was the fact that the number and complexity of its IT systems had grown significantly over the years and that the entire system posed a technological risk.[109] Accordingly, it was decided to grant the tax administration funds to invest in and develop a new IT structure in the following years.[110]

Besides mitigating the technological risk, the aim of the major update of the IT structure was to make the tax administration more effective and, thereby, be able to cut down massively on staff.[111] Accordingly, the number of staff within the entire Danish tax

[108] Lovforslag L 31 af 17.12.2004 forslag til lov om ændring af skattekontrolloven og lov om opkrævning af skatter og afgifter m.v. [Proposal for an Act on Amendments to The Tax Control Act and Act on Collection Taxes and Duties etc.], § 3 (Den.); *see* Skatteudvalget [The Tax Comm.], *Gennemsigtighedsrapport SKATs kontrolarbejde* [Transparency Report on the Tax Administration's Audit Activities], SAU Alm.del Spørgsmål 281 (2017), 6 (showing that as part of the risk-based audit approach, the tax administration divides taxpayers into various segments containing different characteristics and tax challenges); *see* Rigsrevisionen [The Nat'l Audit Off.], Beretning til statsrevisorerne om ToldSkats indsats mod sort økonomi [Report to the State Auditors on the Tax Administration's Activities Targeting the Black Market], RB A303/05 (2005) Accordingly, 2004 appears to mark an important turning point in the tax administration's audit approach, seeing (roughly speaking) a change from a broad and uniform approach toward an approach based on segmentation and risk-profiling of taxpayers.

[109] *See* Rigsrevisionen [The Nat'l Audit Off.], Udvidet notat til statsrevisorerne om ToldSkats IT-systemer [Extended Memo to the State Auditors on the Tax Administration's IT Systems], RN D101/04 (2004). Hence, the tax administration's overall IT structure was described as a "spaghetti pot," comprising seventy-one IT systems that were interconnected through 453 connections. Moreover, the tax administration's own systems had 312 additional connections to external IT systems.

[110] *See* Finansudvalget [The Parliamentary Fin. Comm.], Aktstykke [Committee Appropriation] 157 (2004) (giving the first of two steps planned for the project); *see* Finansudvalget [The Parliamentary Fin. Comm.], Aktstykke [Committee Appropriation] 151 (2006) (giving the second step).

[111] *See* Christensen & Mortensen, *supra* note 49, at 155 (arguing that the massive update of the tax administration's IT structure combined with an extensive reorganization of the entire Danish tax administration carried on from 2005 an onwards (in essence, a merger of the municipal and state-based tax departments into one single and centrally managed organizational unit) should be perceived as one big cost-cutting exercise); *see, e.g.*, Org. for Econ. and Co-Operation and Dev., Ctr. for Tax Pol'y and Admin., TAX POLICY REFORMS IN DENMARK 12 (2015) (providing more information on the amendment of the entire organizational structure of the Danish tax administration in these years).

administration was reduced significantly from more than 10,000 in 2004 to just over 6,000 in 2015.[112]

One IT project that was contemplated to contribute extensively to a more effective tax administration—in line with the international best practices for debtor-oriented and risk-based debt management—was the so-called Common Debt Collection System (*Ét Fælles Indrivelsessystem* [EFI]). The aim of EFI was to gather all public debt collection within one system and thus allow for the administration to replace a vast part of the manual debt collection activities with automated digital debt collection.[113] Initially, the plan was to put EFI into use in the second half of 2007, but the inauguration had to be postponed several times. Finally, in 2013, EFI was put into use, but the application of the system had to be stopped again a few months later, and, in 2015, it was decided to scratch EFI completely. The reason for this cancellation was that the Ministry of Taxation believed that mending the many defects in EFI would be too costly and risky.[114]

Among EFI's many defects were certain legality issues. Hence, an investigation carried out by the legal advisor to the Danish State had shown that various functionalities in EFI contributed to the collection of debt claims that were actually obsolete.[115] Moreover, the data quality was, in many instances, found to be so poor that correct debt collection would not be possible. Accordingly, even in the relatively short period in which EFI actually had been applied, several taxpayers had been subject to unlawful debt collection. On top of that, EFI posed problems with respect to the General Data Protection Regulation (GDPR), as the system did not delete the taxpayers' information after the debt was collected and actually continued to gather new information on the taxpayers, including information about their spouses and children.[116]

Another problem with this system was that the tax administration had continued to cut down on staff in the debt collection unit despite the fact that the launch of EFI was postponed several times. In other words, the Danish tax administration had tried to harvest the rationalization

[112] *See* Christensen & Mortensen, *supra* note 49, at 30.

[113] *See* Finansudvalget [The Parliamentary Fin. Comm.], Aktstykke [Committee Appropriation] 151 (2006). Accordingly, it was expected that EFI would enable the public administration to cut approximately 200 man-years.

[114] Skatteministeriet [Ministry of Tax'n], REDEGØRELSE OM ÉT FÆLLES INDDRIVELSESSYSTEM [Report on one common debt-collection system], SAU alm.del, annex 48 (2014/2015), 51–52.

[115] Kammeradvokaten [Legal Advisor to the Danish State], RAPPORT OM LEGALITETSANALYSE AF EFI-DELSYSTEMFUNKTIONALITETER [Report on legality analysis of the EFI-system functionality] (2015), 5–6.

[116] Michael Tell, *Denmark*, *in* TAX TRANSPARENCY 473, 473–89 (Funda Basaran Yavaslar et al. eds., 2019); *see also infra* Part 5.2. (expanding on GDPR-related issues).

gains upfront despite the risk inherent in developing and applying a new IT system.[117] As a consequence, the amount of uncollected debt increased substantially and a vast amount ended up becoming obsolete and thus was never collected. In 2014 and 2015, debt claims amounting to 900 million DKK and 1.3 billion DKK respectively were forfeited.[118] While obviously being a disaster from a fiscal point of view, the writing down of debt claims also conflicted with the principle of equal treatment, as it arbitrarily benefitted a number of noncompliant taxpayers that potentially could have paid what they owed.[119]

Another example of legality problems from the digitalization of the Danish tax administration concerned the rules for municipal real estate taxation, which basically is a wealth tax on land.[120] Application of such a wealth tax requires valuations, and the task of preparing all these valuations is demanding. Consequently, the Danish tax administration has for years applied various IT systems to overcome this task.

In 2003, a number of technical amendments were made to the rules on municipal real estate taxes.[121] One of these amendments gave access to a deduction in the basis for taxation for certain improvements made on taxpayers' owned land.[122] However, when the new rules were subsequently applied by the tax administration, it appeared that the basis for taxation—calculated automatically by an IT system—became extremely low in certain situations. A local tax official discovered these odd results and informed his superiors who facilitated a change in 2005

[117] Christensen & Mortensen, *supra* note 49, at 155.

[118] Skatteministeriet, *supra* note 114, at 14.

[119] *See* Skatteministeriet, *Skatteministeriet indgår forlig med EFI-leverandører* [*The Ministry of Taxation settles claims against the suppliers of EFI*], Skatteministeriet, Mar. 29, 2019 (explaining that new staff resources were given to the debt-collection unit, that large efforts were made to restore an effective debt-collection function, and that the Ministry of Taxation also sued some of the suppliers for damages, which resulted in a settlement agreement in 2019); *see* Gældsstyrelsen, *Nyt inddrivelsessystem lønindeholder mere end 1 mia. kr. af danskernes gæld* [*New debt collection system withhold more than 1 billion DKK in Danes' salary payments*], Gældsstyrelsen, Dec. 18, 2020 (detailing a new digital debt-collection tool called PSRM that was launched in 2020.); *see* Rigsrevisionen, *Beretning afgivet til Folketinget med Statsrevisorernes bemærkninger* [*Report to the Parliament with the remarks of the State auditors*], Rigsrevisionen, 7-14, 18-26 (2020) (noting that the tax administration's project steering and application of PSRM has been subject to criticism).

[120] Lov nr. 34 af 02.18.1961 Lov om kommunal ejendomsskat [Act on Municipal Real Estate Tax] as changed by Act no. 1463 of 06.10.2020 (Den.).

[121] Lov nr. 1047 af 12.17.2002 Lov om om ændring af lov om beskatning til kommunerne af faste ejendomme og lov om vurdering af landets faste ejendomme [Act amending the Act on taxation of immovable property to municipalities and the Act on the valuation of the country's immovable property] (Den.).

[122] Lbkg. no. 1463 of 06.10.2020 Promulgation of the Municipal Property Tax Act at § 1(5).

to the IT system as the odd results were perceived to be in conflict with the intentions behind the legislation. The change made to the IT system was carried out in direct cooperation between the Danish tax administration and the supplier of the IT system.[123]

In the following years, discussions took place internally in the Danish tax administration about the legality of the changes made to the IT system, as no simultaneous amendments had been made to the wording of the law itself. However, it was not until 2009 when taxpayers had started making complaints that the Danish tax administration's top management asked the legal advisor to the Danish State to investigate whether the changes made to the IT system constituted an unlawful change of practice. In 2010, the legal advisor concluded that this activity was in fact unlawful in the case.[124]

Subsequently, the tax administration's decisions and behavior were subject to severe criticism from the state auditors.[125] Hence, it was sharply criticized that the Danish tax administration had decided to change the IT system without carrying out a full legal analysis of the possible need for a change in the wording of the law itself. In addition, the state auditors criticized that the tax administration had not reacted appropriately to earlier warnings about the unlawfulness of the change in practice and that the tax administration had not stopped the unlawful practice at an earlier stage.

The outcome of this criticism was that the parliament, in late December 2010, changed the wording of the law in order to reflect the, until then, unlawful practice that had been conducted by the tax administration. However, as the amendment would not have retroactive effects, it did not change the fact that a significant number of taxpayers had been subject to unlawful taxation for years. Accordingly, a large number of taxpayer complaints had to be dealt with subsequently.[126]

All in all, it must be acknowledged that the Danish tax administration, with some success, has implemented various digitalized tools to enhance taxpayer compliance, including the use of digital blockers and the more automatized selection of taxpayers for individual audit. However, the failed attempt to digitalize the debt collection processes, as well as the unlawful changes made to the IT system generating valuations for real estate tax purposes, clearly illustrate some of the difficulties that the Danish tax administration has run into

[123] Christensen & Mortensen, *supra* note 49, at 170–71.

[124] *See* Kommunaludvalget [Parliament's Municipalities Committee], L 65 Svar på Spørgsmål 13 Offentligt [Public Answer to Question 13 of L 65] (2010/2011).

[125] Statsrevisorerne, *Beretning om ulovlig opkrævning af ejendomsskatter* [*Report on unlawful collection of real estate taxes*], 3–4 (2011).

[126] Christensen & Mortensen, *supra* note 49, at 174.

along the way. Moreover, these experiences clearly show that it is of utmost importance to ensure that new IT systems are carefully assessed both from a technical and from a legal perspective.

IV. TAX TRANSPARENCY AND DIGITALIZATION

The term *tax transparency* is not clearly defined and covers both the affairs of taxpayers as well as the activity of tax administrations.[127] In an international tax context, the core of the term relates to the exchange of information between tax administrations in different countries, but, nowadays, it covers a much broader range of topics.[128] Despite this development, we mainly focus on the increasing transparency of taxpayers' affairs through the automatic exchange of information between states as well as through the strengthened reporting and disclosure requirements for taxpayers and third parties.[129]

It has been argued that transparency in connection to taxpayers' tax-relevant affairs is a precondition for a just and equal application of tax legislation. Further, it has been stated that transparency has a so-called deterrence effect, that is, an effect causing taxpayers to neither evade nor avoid taxes due to a perceived higher risk of detection.[130] In this context, digitalization plays an important role, as it can be used as a forceful enabler of tax transparency.[131] At the same time, however, it has to be ensured that the extent and use of data collection is proportional to its purpose and that taxpayer rights are appropriately taken into consideration.[132]

A. The International Development

Tax administrations need information about taxpayers to impose a tax assessment and collect taxes, and, in this respect, the fundamental problem of information asymmetry between the tax administration and the taxpayer is a difficult challenge. On the one hand, taxpayers have

[127] Alessandro Turina, *"Visible, Though Not Visible in Itself": Transparency at the Crossroads of International Financial Regulation and International Taxation*, 8 WORLD TAX J. 378, 380 (2016).

[128] Hey, *supra* note 3, at 3.

[129] *See, e.g.,* Tatiana Falcão & Armando L. Yaffar, *General Report: Exchange of Information: Issues, Use and Collaboration*, 105B CAHIERS DE DROIT FISCAL INT'L 197, 197 (2020) (summarizing the instruments for exchange of information in international tax matters).

[130] Hey, *supra* note 3, at 8.

[131] *Id.* at 12–13.

[132] Xiaoqing Huang, *Ensuring Taxpayer Rights in the Era of Automatic Exchange of Information: EU Data Protection Rules and Cases*, 46 INTERTAX 225, 229 (2018).

more and better information about the relevant facts, circumstances, and implemented tax structures. On the other hand, tax administrations generally have more knowledge and information about the content, interpretation, and application of the law.[133]

As already discussed above, tax administrations have invested significantly to reduce the asymmetry of digital initiatives developed and implemented to support taxpayers in the tax administration process. Further, with respect to reducing the information asymmetry of taxpayers' tax affairs, as already discussed, the last decade has witnessed an unprecedented increase in information collected on taxpayers, which is exchanged between tax administrations. The exchange of information regimes currently applicable are generally based on three sets of legal norms:

> 1. The tax treaty network with exchange of information provisions modelled after Article 26 of the OECD Model Tax Convention on Income and on Capital (2005 and subsequent updates).
>
> 2. The tax information exchange agreement network, modelled after the OECD Model Agreement on Exchange of Information in Tax Matters of 2002, with Model Protocol of 2015.
>
> 3. The multilateral Convention on Multilateral Administrative Assistance in Tax Matters originally signed in 1988 but rebranded as the most comprehensive and widely signed instrument on exchange of information since 2009.[134]

In respect of collecting information, the tax administrations can use different approaches. These have previously been categorized as[135]:

> - Voluntary disclosure, for example, publicly available tax strategies implemented by corporate taxpayers versus a legal obligation to disclose, such as mandatory disclosure rules for certain cross-border

[133] Stan Stevens, *Thematic Report: Cutting-Edge Techniques to Collect Information from Taxpayers, in* TAX TRANSPARENCY: 2018 EATLP CONGRESS ZURICH 7–9 JUNE 2018 97, 97 (2019).

[134] Falcão & Yaffar, *supra* note 129, at 201.

[135] *See* Stevens, *supra* note 133*; see, e.g.*, Falcão & Yaffar, *supra* note 129, at 202 (providing a comparative analysis of the various instruments and reporting obligations).

arrangements.[136]

- National reporting and exchange of information versus international exchange of information on request—spontaneously or automatically. The exchange may, inter alia, include facts, circumstances, financial information, advance tax rulings, advance pricing agreements, and country-by-country reporting (CbCR). Under the CbCR rules, a company is obliged to file detailed information and explanation on, inter alia, its group structure, revenue from related parties and third parties, total revenue, profit before tax, corporate income tax and withholding taxes, current-year accrued corporate income taxes, stated capital, accumulated earnings, tangible fixed assets and employees' internal transactions, a list of all constituent entities of the multinational enterprise, and a description of the nature of the activities of each constituent entity. All this listed information is exchanged between the tax administration of the jurisdictions in which the company is active.[137]

- Reporting of information to the tax authorities versus the public reporting implemented as a reputation mechanism, such as the ultimate beneficial owner register.[138]

[136] Council Directive 2011/16, O.J. (L 64) 1 (EU) (demonstrating that at the European Union level, member states have backed the implementation of the OECD's Base Erosion and Profit Shifting project through the adoption of several directives inter alia on administrative cooperation in the field of direct taxation); *see also* Turley et. al., *supra* note 22, § 16.2.2 (discussing regimes in inter alia the US, UK, and Canada).

[137] As of October 2022, there are over 3300 bilateral exchange relationships activated with respect to jurisdictions committed to exchanging CbCR, and the first automatic exchanges of CbCR took place in June 2018. *See Activated Exchange Relationships for Country-by-Country Reporting,* OECD, https://www.oecd.org/tax/beps/country-by-country-exchange-relationships.htm, (June 2022).

[138] Council Directive 2015/849, 2015 O.J. (L 141) 73 (EU). (On May 20, 2015, the European Union adopted the 4th Anti-Money Laundering Directive (Directive (EU) 2015/849 of the European Parliament and of the Council of 20 May 2015 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Regulation (EU) No. 648/2012 of the European Parliament and of the Council, and repealing Directive 2005/60/EC of the European Parliament and of the Council and Commission Directive 2006/70/EC, OJ L 141). According to this Directive,

All of these initiatives, which have been put in place with the aim of reducing information asymmetry and thereby increasing tax transparency, have been facilitated by digitalization to a large extent, which, on one side of the coin, arguably represents a chance for fairer and more efficient tax enforcement but, on the flip side, increases the risk of data abuse and encroachment of taxpayers' privacy.[139] The new technical means of data tracking imply that the range of data known by the tax administration is significant and may include family status, religious affiliations, health conditions, and business secrets.[140] While taxpayers who are privileged by living in democratic rule of law states may not need to be too concerned about data abuse, the risk should not be neglected, and tax administrations need to be careful to avoid risks of breaching taxpayer confidentiality as a consequence of exchanging information.[141] In this respect, it should be noted that information received under agreements on exchange of information is typically treated as confidential and given the same level of protection as information provided under domestic law. This protection will typically imply that information may only be disclosed to persons or authorities, including courts and administrative bodies, whose role is to deal with the assessment and collection of taxes or the prosecution of claims.[142] However, in this respect, it has previously been problematized that international agreements on the exchange of information often do not mention the protection of personal data but instead refer to domestic law.[143]

Consequently, although digitalization and "datafication" may be beneficial to fight tax evasion and improve equality in tax collection, the protection of taxpayers through guarantees of confidentiality and

Member States are obliged to implement rules requiring companies to disclose their ultimate beneficial owners and beneficiaries of trusts.).

[139] Hey, *supra* note 3, at 12.

[140] Anna-Maria Hambre, *Tax Confidentiality: A Legislative Proposal at National Level*, 9 WORLD TAX J. 2, sec. 1. (2017).

[141] Turley et al., *supra* note 22, § 16.2.2.

[142] Organisation for Economic Co-operation and Development [OECD], *Multilateral Convention on Multilateral Administrative Assistance in Tax Matters*, art. 22 (2011); *see, e.g.* Organisation for Economic Co-operation and Development [OECD], *OECD Model Tax Convention on Income and on Capital*, art. 26, ¶ 2 (2019).

[143] *See, e.g.,* Falcão & Yaffar, *supra* note 129, at 233 (problematizing that Article 22 (1) of the Multilateral Convention on Multilateral Administrative Assistance in Tax Matters is the only provision amongst the three treaties on the subject that mentions the protection of personal data; according to Article 22 (1), safeguards may be put in place by the supplying party in accordance with its domestic laws to ensure the protection of personal data).

principles of data protection are gaining support.[144] In line with these observations, many countries have implemented far-reaching legal reforms of their tax procedures as a result of the digitalization and automatization of the tax administration. However, it has been argued that data protection laws need to be adapted further to cope with this process of digitalization.[145] At the same time, support for a mandatory public disclosure policy of some taxpayers' tax return information is increasing, and voluntary compliance is encouraged by relying on the public eye to promote or even demand that taxpayers act responsibly.[146]

Another challenge with datafication, the use of algorithms, and the production of tax assessments automatically, is the lack of transparency during this process—not only for the taxpayers but also for the tax administrations themselves, for the courts overseeing the decisions, and for governments.[147] This lack of transparency in the used algorithms is also challenged by the fact that data analytics codifies the past, that is, it is based on the past and the assumption that patterns will be repeated.[148] Accordingly, the assumptions behind the decisions that determine the selection are unclear and hardly to be scored for fairness. Further, as predicative models are based on correlations that do not imply causality, there is a risk of stigmatization and discrimination when segmenting taxpayers based on shared characteristics. Arguably, predicative models, to some extent, sacrifice fairness for efficiency and, accordingly, predicative models require continuous feedback based on careful evaluations and assessments of the results provided by the

[144] Hey, *supra* note 3, at 14–15 (citing Philip Baker & Pasquale Pistone, *General Report*, *in* 100b THE PRACTICAL PROTECTION OF TAXPAYERS' FUNDAMENTAL RIGHTS § 3 (2015); Kimberly A. Houser & Debra Sanders, *The Use of Big Data Analytics by the IRS: Efficient Solutions or the End of Privacy as We Know It?*, 19 VAND. J. ENT. and TECH. 817, 866 (2017); *see* Kay Blaufus et al., *The Effect of Tax Privacy on Tax Compliance – An Experimental Investigation*, 26 EUR. ACCT. REV. 561, 562 (2017). *Contra* Laura C.B. Altafulla, *The Line between Tax Secrecy and Tax Transparency*, in TAX POLICY CHALLANGES IN THE 21st CENTURY 423, 445 (Raffaele Petruzzi & Karoline Spies eds., 2014).

[145] Hey, *supra* note 3, at 13.

[146] Hey, *supra* note 3, at 15 (citing Marjorie E. Kornhauser, Doing the Full Monty: *Will Publicizing Tax Information Increase Compliance?*, 18 CAN. J. L. & JURIS. 1, 101–03 (2005)); Joshua Blank, *Reconsidering Corporate Tax Privacy, Proceedings of the Annual Conference on Taxation*, 11 N.Y.U. J.L. & BUS. 31, 49 (2014); Reuven S. Avi-Yonah & Ariel Siman, *The 1 Percent Solution: Corporate Tax Returns Should Be Public (and How to Get There)*, 73 TAX NOTES INT'L. 627, 627 (2014); Lee A. Sheppard, *Should Corporate Tax Returns Be Disclosed?,* 142 TAX NOTES 1381, 1381, 1383 (2014); *see* Allison Christians, *Do We Need to Know More About Our Public Companies?*, 66 TAX NOTES INT'L 843, 843 (2012).

[147] Hey, *supra* note 3, at 13.

[148] *See* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 8 (2016).

model.[149]

Despite these legal concerns, the digitalization of the Danish tax administration does not seem to have resulted in severe debate or scrutiny by institutions or the public media. Actually, and as further explained in the paragraph just below, the Danish tax administration's comprehensive transfer of taxpayer information appears to be broadly accepted by Danish taxpayers.

B. Danish Experiences

As already described, the Danish tax administration receives information about Danish taxpayers from a long list of domestic third parties. However, in line with the general international development, the Danish tax administration is also heavily involved in cross-border exchange of information. Accordingly, the Danish tax administration continually receives vast amounts of information about Danish taxpayers from tax administrations around the world and transfers an abundance of information to the same foreign tax administrations in return.[150]

The Danish tax administration's wide possibilities for collecting and transferring taxpayer information are not subject to widespread concern. The fact that Denmark is a quite homogeneous welfare state with strong democratic traditions and a transparent public administration probably explains why the administration's activities are not subject to widespread concern. Accordingly, the Danish state's institutions, including the tax administration, generally command high levels of trust from Danish citizens.[151]

In other words, it appears to be broadly accepted that the Danish tax administration needs wide access to collect, store, and transfer information about taxpayers to perform its tasks appropriately and efficiently. This comprehensive access to taxpayer information is facilitated by a strong tradition for international cooperation combined with a comprehensive domestic legal framework.[152]

However, it is worth noting that the Danish tax administration's

[149] Stevens, *supra* note 133, at 152.

[150] *See also* Søren L. Nielsen & Bent Bertelsen, *Denmark*, *in* 105B CAHIERS DE DROIT FISCAL INTERNATIONAL, 307, 307 (2020).

[151] Tell, *supra* note 116, at 473–74; *see, e.g.,* Jacob G. Nielsen, Peter K. Schmidt & Helle Vogt, *Denmark*, *in* HIST TAX'N 243, 262 (2021).

[152] Preben B. Hansen & Lasse E. Christensen, *Denmark*, *in* 98b CAHIERS DE DROIT FISCAL INTERNATIONAL [STUDIES ON INTERNATIONAL FISCAL LAW], 249–74 (2020); *see also* Peter K. Schmidt, *The Emergence of Denmark's Tax Treaty Network: A Historical View*, 1 NORDIC TAX J. 49, 49–63 (2018) (detailing the historical development of Denmark's network of tax treaties and exchange of information agreements).

wide access to taxpayer information is balanced against legislation aiming at protecting taxpayers. Accordingly, the legal protection of taxpayers' rights follows from provisions in various statutory laws and regulations. For example, section 27 of the Public Administration Act stipulates that any person employed by or acting on behalf of the public administration is subject to a duty of confidentiality. Further, section 17(1) of the Tax Administration Act specifies that the tax administration has to treat all taxpayer information on economic, business, and personal matters as confidential.[153] Among other things, this process entails that the Danish tax administration may only transfer information on taxpayers to foreign tax administrations if the taxpayer explicitly permits such a transfer or if a clear legal basis for exchanging the information can be found elsewhere.[154]

With respect to the exchange of information, such a provision can be found in section 66 of the Tax Control Act. The provision thus contains the statutory legal basis for Denmark's exchange of taxpayer information with other tax administrations around the world.[155] Thus, the provision lists the legal bases on which the exchange can take place. These fall into four broad groups[156]:

> 1. EU directives;
>
> 2. tax treaties, bilateral and multilateral (more than seventy comprehensive bilateral tax treaties plus an international multilateral instrument and a Nordic multilateral treaty);
>
> 3. administratively concluded agreements on mutual assistance in tax matters, inter alia, Tax Information Exchange Agreements or TIEAs (around fifty of such agreements);

[153] Lov nr. 1441 af 22.12.204 *Lov om digitalisering af regnskabsoplysninger, ophævelse af virksomheders underretningspligt og afskaffelse af kildeskattebøderne* [Act on Digitalisation of Financial Information, Removal of Businesses' Information Duties and Abolition of Withholding Tax Fines].

[154] Nielsen & Bertelsen, *supra* note 150, at 316.

[155] Lov nr. 2612 af 28.12.2021 bekendtgørelse af skattekontrolloven § 61 [Act on Tax Control] (Den.) (providing a broad legal basis for the Danish tax administration to acquire information from third parties); Lov nr. 2612 af 28.12.2021 Bekendtgørelse af skatteindberetningsloven § 22 [Act on provision of information to the tax authorities] (Den.) (containing the legal basis for Denmark's FATCA-agreement with the United States Internal Revenue Service and other agreements based on the OECD's Common Reporting Standard).

[156] Nielsen & Bertelsen, *supra* note 150, at 307.

> 4. any other international agreement or convention relating to administrative assistance on tax matters to which Denmark has acceded.

Denmark's actual exchange of information with other tax administrations can take place spontaneously, by request, and automatically. Incoming and outgoing requests from other EU countries are mainly serviced through the eForm Central Application system, which is a digital platform based on a Common Communication Network mail system. Exchange of information with Nordic countries that are not EU Member States mainly takes place via tunnel-encrypted emails and information requests to non-EU and non-Nordic states are mainly made through an internal server-based, VPN-secured system or, if this is not possible, through ordinary mail.[157]

All automatically received material coming from abroad is treated automatically and forwarded to the relevant audit teams upon evaluation of the data quality. Moreover, these data are automatically implemented into the taxpayers' digitally generated yearly tax statements.[158]

Information is also exchanged with foreign tax administrations as a result of the international agreement on CbCR. The legal basis for applying these rules in a Danish context is found in sections 47 through 52 of the Tax Control Act.[159] The CbCR data received by the Danish tax administration is transferred to and stored in a data warehouse by the Danish tax administration. This enables the Danish tax administration to carry out data searches, make extracts, and prepare specific reports based on the stored CbCR data. These available opportunities assist the tax administration's transfer pricing specialists in the making of risk assessments of specific multinational enterprises and their group entities.[160]

As a public agency, the Danish tax administration has to take the GDPR into account with respect to all of its activities. As a main rule, the tax administration is allowed to process taxpayer information, pursuant to article 6(1)(e), which stipulates that the processing of information is lawful if it is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. However, when processing taxpayer information, the tax administration has to act in accordance with the

[157] *Id.* at 308. The VPN system is called *bluewhale*. The Danish tax administration is not able to send anything encrypted through the normal email system.

[158] Nielsen & Bertelsen, *supra* note 150, at 331.

[159] *See also* Bek nr. 1304 af 14.11.2018 Bekendtgørelse om land for landrapportering.

[160] Nielsen & Bertelsen, *supra* note 150, at 307–09.

general GDPR principles on legitimacy, proportionality, empowerment, accountability, and security.[161]

Overall, the Danish tax administration thus has wide access to collect, store, and transfer information about taxpayers domestically as well as across borders. This system is needed for the tax administration to perform its tasks appropriately and efficiently but has to take place in a way that sufficiently ensures taxpayers' rights. In this regard, it is reassuring that such protection, for instance, follows from various provisions found in the Public Administration Act as well as the Tax Administration Act, and that the tax administration is only allowed to transfer information cross-border if a clear legal basis for exchanging the information exists.

V. THE FUTURE OF HIGHLY DIGITALIZED TAX ADMINISTRATION

New emerging technologies and improvements of existing technologies provide huge opportunities for tax administrations in various areas and may enable tax administrations to expand the insights derived from available data significantly.[162] Moreover, some scholars see great opportunities for tax administrations in the use of blockchain technology.[163] Altogether, it has been argued that the use of such new technologies represents the beginning of a whole new era for tax administrations.[164] However, also in this context, taxpayers' rights are at risk of being jeopardized.[165] These opportunities and risks are further discussed below in an international as well as in a Danish context.

A. International Outlook

While digitalization arguably has significantly increased the efficiency and effectiveness of tax administration as well as helped to

[161] *See* Jan Trzaskowski & Max G. Sørensen, GDPR COMPLIANCE: UNDERSTANDING THE GENERAL DATA PROTECTION REGULATION 48–60 (2019).

[162] Walker, *supra* note 16, at 263–71 (emphasizing big data, computer power, and data analytics).

[163] Dennis Post, DIGITAL TAX ADMINISTRATION 4.0: TOWARDS A DISTRIBUTED TAX SYSTEM, IN TAX LAW AND DIGITALIZATION: THE NEW FRONTIER FOR GOVERNMENT AND BUSINESS, 39–59 (2021).

[164] *See* Gianluca Mazzoni, *(Re)defining the Balance between Tax Transparency and Tax Privacy in Big Data Analytics*, 72 BULL. FOR INT'L TAX'N 656, 663 (2018).

[165] Arthur J. Cockfield, *Cross-Border Big Data Flows and Taxpayer Privacy*, *in* ETHICS AND TAXATION 379, 379-396 (2020); Faúndez-Ugalde et al., *supra* note 4, at 6; *see, e.g.*, Duncan Bentley, *Taxpayer Rights and Protections in a Digital Global Environment*, *in* ETHICS AND TAXATION 251, 251–94 (2020).

reduce burdens for different taxpayer segments, it has been argued that some countries may be reaching the end of their ability to further reduce the tax gap or the administrative burdens in any significant way.[166] What seems most apparent is the fact that the digitalization of the tax administration has yet to move away from sequential taxpayer-facing processes and be integrated into taxpayers' daily lives and their operational systems. This move is Tax Administration 3.0.[167] A number of legal scholars have suggested that blockchain technology is the technical means to take tax administrations to the next level.[168] The OECD has argued—without being technology specific—that the core elements of Tax Administration 3.0 could include[169]:

> - that paying taxes will become a more seamless experience, as taxpayers' behavior and systems will be the starting point to facilitate compliance by design so that noncompliance will require deliberate and burdensome activities;
>
> - that digital platforms will become agents of tax administrations and carry out tax administration processes within their systems, implying that tax administration is conducted within a network of interacting trusted actors and no single point of failure, although the tax administrations ensure the quality, robustness, and reliability of the outputs;
>
> - that tax administration processes will become increasingly real time to stay synchronized with taxpayers' transactions and incorporate artificial intelligence to support characterizations and

[166] OECD, *supra* note 26, at 19.

[167] *Id.* at 12.

[168] *See* Charléne A. Herbain, *Fighting VAT Fraud and Enhancing VAT Collection in a Digitalized Environment,* 46 INTERTAX 6/7 579, 579–83 (2018); Sunny K. Bilaney, *From Value Chain to Blockchain: Transfer Pricing 2.0,* 25 INT'L TRANSFER PRICING J. 294, 296 (2018); Alicja Majdanska & Karol Dziwinski, *The Potential of a Standard Audit File: Tax in the European Union: A Chance for Coordinated VAT Administration?,* 72 BULL. FOR INT'L TAX'N 10 582, 593 (2018); Christina Dimitropoulou et al., *Applying Modern, Disruptive Technologies to Improve the Effectiveness of Tax Treaty Dispute Resolution: Part 1*, 46 INTERTAX 11 856, 868–70 (2018); Aleksandra Bal, *Taxation, Virtual Currency and Blockchain*, 68 INT'L TAX'N 19, 19–27 (2019); *see, e.g.*, Claudio Cipollini, *Blockchain and Smart Contracts: A Look at the Future of Transfer Pricing Control*, 49 INTERTAX 4 (2021).

[169] OECD, *supra* note 26, at 12–14.

> assessments of tax claims as well as balancing mechanisms where taxation cannot be settled on a transactional basis;
>
> - that taxpayers get the opportunity to check and question tax assessments that have been made based on automated and human decision-making;
>
> - that every taxpayer gets one digital identity, which will support a seamless connection between other government services and private actors;
>
> - that skilled employees will be supplemented with advanced analytics and decision-supporting tools to reduce the number of areas where compliance choices remain and in order to detect anomalies, leakages, and flaws in the tax system.

Such a transformation requires many things coming together and will, of course, be easier where the tax affairs of taxpayers are less complex.[170] However, the need for financing the increased public spending as well as changes in work patterns and new business models impose risks and difficulties for tax administrations and may be expected to grow over the coming years with the increasing digitalization of the economy.

In terms of changes in work patterns and business models, the rapid growth of the sharing economy through online platforms has led taxpayers to shift status from employees—where salaries are subject to withholding in the compliance by design system (PAYE)—to self-employment. While access to relevant information is addressed to some extent by domestic legal reporting requirements and cross-border exchanges of information, it does increase the complexity of tax administration and opportunities for noncompliance.[171] Another challenge is the increase in what may be referred to as "flexible workplaces." This challenge relates to mobile professionals who perform their work remotely from anywhere in the world, taking advantage of digital technologies. While this has previously mostly been a

[170] *Id.* at 12.

[171] Organisation for Economic Co-operation and Development, THE SHARING AND GIG ECONOMY: EFFECTIVE TAXATION OF PLATFORM SELLERS FORUM ON TAX ADMINISTRATION 25 (2019); Organisation for Economic Co-operation and Development, MODEL RULES FOR REPORTING BY PLATFORM OPERATORS WITH RESPECT TO SELLERS IN THE SHARING AND GIG ECONOMY 3 (2020).

characteristic of personnel of highly digitalized business models—that is, programmers and data scientists—the COVID-19 pandemic has been mainstreaming remote work. Therefore, companies deploying more traditional business models may find themselves increasingly looking for employees across national borders and allowing for a plus or minus two-hour time zone difference or even allowing for so-called digital nomads, who travel the world while working a full-time job remotely.[172]

As the global economy becomes increasingly interconnected and digitalized, it has been subject to intense debate that some businesses have been able to generate profits through participation in a significant and sustained way in the economy of a country without a local, physical presence creating a taxable presence.[173] Current international discussions may entail—if implemented—that tax administrations of multiple jurisdictions should have access to highly complex, large, and geographically distributed information on multinational businesses with complex supply chains and financial arrangements.[174] In this respect, the optimal system might, as stated above, be an increased reliance on

[172] Svetislav V. Kostić, *In Search of the Digital Nomad: Rethinking the Taxation of Employment Income under Tax Treaties,* 11 WORLD TAX J. 2 184, 203–06 (2019); *see* Stjepan Gadzo, *Croatia: A New (Tax-Free) Promised Land for Digital Nomads? (Part II),* KLUWER INT'L TAX BLOG (Sept. 8, 2022), http://kluwertaxblog.com/2022/02/28/croatia-a-new-tax-free-promised-land-for-digital-nomads-part-ii/.

[173] Louise F. Kjærsgaard, *The Ability to Pay and Economic Allegiance: Justifying Additional Allocation of Taxing Rights to Market States,* 49 INTERTAX 8, 648 (2021); Ana P. Dourado, *The OECD Report on Pillar One Blueprint and Article 12B in the UN Report,* 49 INTERTAX 3, 3–6 (2021); Thomas Greil & Stefan Eisgruber, *Taxing the Digital Economy: A Case Study on the Unified Approach*, 49 INTERTAX 53, 54, 57, 66 (2021); Xiaorong Li, *A Potential Legal Rationale for Taxing Rights of Market Jurisdictions*, 13 WORLD TAX J. 15, 26–36 (2021); Jinyan Li, *The Legal Challenges of Creating a Global Tax Regime with the OECD Pillar One Blueprint*, 72 BULL. FOR INT'L TAX'N 84, 88, 93 (2021); Hans Van den Hurk, *OECD's Pillar One and the Return of the Pencil!,* KLUWER INT'L TAX BLOG (Feb. 22, 2021), http://kluwertaxblog.com/2021/02/22/oecds-pillar-one-and-the-return-of-the-pencil/; Gino Sparidis et al., *Digital Economy Taxation Developments: A Marker for the Future of Taxes (Part 2),* KLUWER INT'L TAX BLOG (Feb. 5, 2021), http://kluwertaxblog.com/2021/02/05/digital-economy-taxation-developments-a-marker-for-the-future-of-taxes-part-2/; Daniele Frescurato & Velio A. Moretti, *The Carve-out of Financial Services from Pillar One: Good Times for a Step Further?*, KLUWER INT'L TAX BLOG (Nov. 23, 2020), http://kluwertaxblog.com/2020/11/23/the-carve-out-of-financial-services-from-pillar-one-good-times-for-a-step-further/; Vikram Chand & Damiano Canapa, *Pillar I of the Digital Debate: Its Consistency with the Value Creation Standard as Well as the Way Forward*, KLUWER INT'L TAX BLOG (Nov. 24, 2020), http://kluwertaxblog.com/2020/11/24/pillar-i-of-the-digital-debate-its-consistency-with-the-value-creation-standard-as-well-as-the-way-forward/; William Byrnes*, Recommendations for the Pillar One and Pillar Two Blueprints*, KLUWER INT'L TAX BLOG (Dec. 18, 2021), http://kluwertaxblog.com/2020/12/18/recommendations-for-the-pillar-one-and-pillar-two-blueprints/.

[174] OECD, *supra* note 26, at 21.

the integration of tax rules into the different business accounting systems used by various businesses. However, it is important to justify why data needs to be collected rather than potential alternatives.[175] In order for the regime not to collect excessive unrequired documentation, arguably the rules have to be reasonably and carefully targeted at those transactions that are most likely to involve tax avoidance according to the rule of law.[176] Further, while digital recordings of payments, record keeping, and identity present many opportunities for tax administrations to increase transparency and to prompt compliance, digitalization may also produce transparency holes, that is, through the use of virtual currencies, cryptocurrencies, and opaque digital assets.[177]

To mitigate underreporting or no reporting of taxable income from cryptocurrencies and to promote harmonized rules, the EU Commission initiated a process intended to lead to the eighth update of the Directive on Administrative Cooperation.[178] More recently, the council presidency and the European Parliament reached a provisional agreement on the markets in crypto assets.[179] According to the press release, actors in the crypto-assets market will be required to declare information on their environmental and climate footprint. Further, the European Banking Authority will be tasked with maintaining a public register of noncompliant crypto-asset service providers and crypto-asset service providers with a parent company located in countries listed on the EU list of third countries (considered at high risk for anti-money laundering activities) and/or on the EU list of noncooperative jurisdictions for tax purposes. These protocols will be required to implement enhanced checks in line with the EU Anti-Money Laundering framework. Further, issuers of asset-referenced tokens need to have a registered office in the EU to ensure proper supervision and monitoring. Finally, crypto-asset service providers will need an authorization to operate within the EU, and national authorities will regularly transmit relevant information on the largest crypto-asset service providers to the European Securities and Markets Authority.[180]

Along the same line, the OECD published a report in which the need for greater transparency in the field of crypto-assets was highlighted,[181]

[175] *Id.* at 24.

[176] TURLEY ET AL., *supra* note 22, at chap. 16.

[177] OECD, *supra* note 26, at 22.

[178] See Proposal for a Council Directive amending Directive 2011/16/EU on administrative cooperation in the field of taxation COM/2022/707 final.

[179] Council of the European Union Press Release, Digital finance: agreement reached on European crypto-assets regulation (MiCA) (June 30, 2022).

[180] *Id.*

[181] ORG. FOR ECON. COOP. AND DEV. [OECD], TAXING VIRTUAL CURRENCIES: AN OVERVIEW OF TAX TREATMENTS AND EMERGING TAX POLICY ISSUES (2020).

and the OECD is developing a proposal to ensure sufficient reporting and exchange of information. The Committee on Fiscal Affairs has also approved a work plan to review the Standard for Automatic Exchange of Financial Account Information in Tax Matters.[182]

Accordingly, despite challenges and legal concerns, it seems safe to conclude that datafication and requirements for tax transparency on the international scene are here to stay.

B. Danish Outlook

In recent years, Denmark has also taken steps to facilitate a transition toward Tax Administration 3.0. Accordingly, on 1 July 2018, a new agency called the IT and Development Agency (Udviklings- og forenklingsstyrelsen) was established in connection to a major reorganization of the Danish tax administration.[183] The idea was that the new agency should support the development of a reliable and future-proof tax administration. Accordingly, the agency was given as its core task to maintain existing IT systems, ensure stable operations, and develop modern and future-proof IT solutions to the Danish tax administration.

In order to further support the successful development of such modern and future-proof IT solutions, the Danish Parliament, in December 2021, adopted a bill containing a number of changes to the Danish Tax Control Act.[184] The main aim of this bill was to pave the way for a more efficient and intelligent risk-based tax control.[185] In this context, it was acknowledged that the development of new IT solutions for tax control purposes required that the IT and Development Agency was allowed to collect and process various forms of information. Hence, the intention was to introduce a clear and unambiguous legal basis for

[182] ORG. FOR ECON. COOP. AND DEV. [OECD], REPORT ON THE IMPLEMENTATION OF THE RECOMMENDATION OF THE COUNCIL ON THE STANDARD FOR AUTOMATIC EXCHANGE OF FINANCIAL ACCOUNT INFORMATION IN TAX MATTERS, 320–21 (2nd ed. 2020); *see also* Luisa Scarcella, *Exchange of Information on Crypto-Assets at the Dawn of DAC8*, KLUWER INT'L TAX BLOG (Mar. 29, 2021), http://kluwertaxblog.com/2021/03/29/exchange-of-information-on-crypto-assets-at-the-dawn-of-dac8/.

[183] Danish Tax Administration, *A.A.1.1.1 Skatteministeriets organization Indhold [Contents of the Ministry of Taxation]*, *in* DEN JURIDISKE VEJLEDNING (2022). The IT and Development Agency is an agency under the Ministry of Taxation and forms part of the Danish Customs and Tax Administration. The agency has around 1,700 employees.

[184] Lov nr. 2612 af 28. 12. 2021 Lov om ændring af lov om et indkomstregister, skatteindberetningsloven og skattekontrolloven [Act on amendment of the income register act, the provision of tax information act, and the tax control act] (Den.).

[185] Lov nr. 73 af 21. 12. 2021 Lov om ændring af lov om et indkomstregister, skatteindberetningsloven og skattekontrolloven [Act on amendment of the income register act, the provision of tax information act, and the tax control act] (Den.).

the agency to collect, process, and combine data in order to develop tools based on data analytics, including machine learning tools.

Pursuant to section 68(1)–(3) of the Tax Control Act, the tax administration already had the right to pool data contained in the tax administrations' own IT systems in the course of its tax enforcement tasks. Further, for the purpose of assessing and collecting taxes, the tax administration already had access to necessary information about individuals and legal entities' economic and business affairs contained in the common national income register as well as in other public agencies' IT systems.[186]

However, these rules did not provide the necessary legal basis for collecting, processing, and combining data in the course of developing new IT tools. As the tax administration, to a limited extent, had already started to utilize and pool data from various registers when developing new IT tools, it was thus found to be of great importance to ensure a sufficient legal basis for these activities.

While it is commendable that legislative changes were made in 2021, to provide the necessary legal basis, it is reprehensible that the tax administration had already started to pool and use data for these purposes before the changes were adopted. In addition, as the new legal basis provided in 2021 is rather broad and generic, it is crucial that procedures are put in place to ensure sufficient protection of taxpayers' rights.[187]

In particular, concerns arise with respect to pooling and using data contained in various registers for the use of developing tools for the profiling of taxpayers.[188] However, at least according to the travaux préparatoires, the Danish legislature appears to be aware of these concerns. Hence, it is explicitly stated that [authors' own translation][189]:

> Within the framework of the proposed provisions, it will be ensured that appropriate mathematical or statistical procedures are used for the profiling. Technical and organizational measures will also be implemented, which will, in particular, ensure that factors that result

[186] Lov nr. 403 af 8.5.2006 [The Income Register Act] (consolidated act 284 of 3.2.2022) (Den.).

[187] *See e.g.,* Lov nr. 73 af 21. 12. 2021 Lov om ændring af lov om et indkomstregister, skatteindberetningsloven og skattekontrolloven, annex 1, 12–15, [Act on amendment of the income register act, the provision of tax information act, and the tax control act] (Den.).

[188] General Data Protection Regulation 95/46, art. 4(4), 2016 O.J. (L 119) 1.

[189] Lov nr. 73 af 21. 12. 2021 Lov om ændring af lov om et indkomstregister, skatteindberetningsloven og skattekontrolloven, 12. [Act on amendment of the income register act, the provision of tax information act, and the tax control act] (Den.).

> in inaccurate personal information are corrected and that the risk of errors is minimized. The measures must also secure personal data in a way that takes into account the potential risks to the interests and rights of the data subject and ensures that no differential treatment of taxpayers takes place on the basis of race, ethnic origin, political, religious, or philosophical convictions, labor union affiliation, genetic status, state of health, or sexual orientation. . . . [190]

Further, the travaux préparatoires contain a specific assessment of GDPR-related issues. The legislator is thus fully aware that basic GDPR rules have to be respected. It is also pointed out that Article 23(1)(e) of the GDPR allows that national legislation, to which the data controller or processor is subject, may restrict the scope of some of the obligations and rights provided for in the GDPR. This is allowed when such a restriction respects the essence of the fundamental rights and freedoms and is a necessary and proportionate measure in a democratic society to safeguard important objectives of general public interest, in particular an important economic or financial interest of the EU or of a member state, including monetary, budgetary, and taxation matters, public health, and social security.

This legislation is of particular relevance with respect to the so-called purpose limitation principle, that is, the requirement that personal data may only be collected for specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes.[191] The reason is that the new section 67(a) of the Danish Tax Control Act provides for pooling and utilization of register data for a broader purpose, that is, the development of new IT tools. Hence, this broad use of data may infringe upon the purpose limitation principle. However, as the pooling and utilization of register data arguably safeguards important objectives of general public interest, in particular the Danish state's important economic or financial interest, this infringement must probably be considered permissible.[192] At the end of the day, however, it all depends on how the development of the new IT tools, as well as the accompanying procedures, actually play out.

[190] Lov nr. 73 af 21. 12. 2021 Lov om ændring af lov om et indkomstregister, skatteindberetningsloven og skattekontrolloven, 14. [Act on amendment of the income register act, the provision of tax information act, and the tax control act] (Den.).

[191] Trzaskowski & Sørense, *supra* note 161, at 81–82.

[192] Lov nr. 73 af 21. 12. 2021 Lov om ændring af lov om et indkomstregister, skatteindberetningsloven og skattekontrollovenm, 3–8 (See comments provided by the Danish Data Protection Agency [*Datatilsynet*]).

Altogether, the Danish tax administration has taken important steps toward facilitating a transition toward Tax Administration 3.0, in a way where the development and use of new digital tools can be balanced against the need for protection of taxpayers. However, the steps currently taken still do not deviate from the traditional sequential taxpayer processes, and further steps are thus needed in order to move to the next level, that is, processes where taxation is fully integrated into the daily lives of taxpayers and their operational systems.

VI. CONCLUSIONS

To benefit both taxpayers and tax administrations, many states around the world have already come a long way in making tax processes smoother and more efficient through the use of digital tools. Prominent examples include the use of e-filing of partially or fully prefilled returns, e-payment and provision of online taxpayer assistance. This development, often referred to as the transition from Tax Administration 1.0 to 2.0, can also be observed in Denmark.

Despite the obvious benefits of this development, it is important to be aware of taxpayers' rights, including the risk of jeopardizing taxpayers' trust in the tax administration if such rights are not sufficiently protected. Consequently, the processes of collecting and utilizing taxpayer information needs to be transparent. Moreover, a sufficient legal basis and a clarification of the responsibilities for taxpayers, third parties, and the tax administration have to be ensured.

A good example of such an exercise is the amendment made in Denmark through the adoption of a new Tax Control Act in 2017. However, despite these improvements, concerns still remain when it comes to the question of whether Danish taxpayers' rights are sufficiently protected in the digital era. It is thus commendable that the Danish Ombudsman has launched a more thorough investigation into the Danish tax administration's digital procedures and IT systems to assess whether the tax administration's IT systems take sufficient account of obligations following from general administrative law and GDPR.

Advances in technology have also enabled tax administrations to digitalize audit processes through the increased use of automated electronic checks and validations, matching of reported information, and profiling of taxpayers. Accordingly, most tax administrations now apply data science techniques and analytical tools in audit case selection. This application allows tax administrations to better identify the tax returns, claims, or transactions, which may require further scrutiny. However, ensuring nondiscrimination, accountability, and transparency of such digitalized compliance actions is critical to facilitate voluntary compliance, protect taxpayers appropriately, sustain perceptions of fairness, and create prophylactic effects.

For some time, the Danish tax administration has also used various digitalized tools to ameliorate audit processes. Useful experiences thus have been made with respect to the use of digital blockers and automatized selection of taxpayers for individual audit. However, in other areas negative experiences have been made. Hence, the attempt to digitalize the debt-collection processes suffered from poor data quality, legality problems, GDPR deficiencies, and last but not least, a too optimistic view on when expected efficiency gains could be reaped (e.g., from cutting down staff within the tax administration). Further, the unlawful changes made to the system generating valuations for real estate tax purposes clearly showed that it is of utmost importance to ensure that the application of such digital tools are continually assessed—both from a technical and from a legal perspective.[193]

Tax administrations around the world have also embarked on initiatives to increase taxpayer transparency through digital means, for example, through the use of automatic exchange of information with other states. This initiative, on one side of the coin, arguably represents a chance for creating a fairer and more efficient tax enforcement, but, on the flip side, it also increases the risk of data abuse and encroachment of taxpayers' privacy. In a Danish context, the comprehensive transfer of information on Danish taxpayers to and from other states appears to be broadly accepted. This acceptance is probably caused by Denmark's strong tradition and preference for international cooperation in such matters combined with a thorough legal protection of taxpayers following from provisions in various statutory laws and regulations.

Looking ahead, most tax administrations will probably try to gradually move away from traditional sequential taxpayer processes and instead try to integrate taxation into taxpayers' daily lives and their operational systems—that is, realize what has been called Tax Administration 3.0. While this development obviously will pose new challenges and legal concerns, it seems safe to conclude that the digitalization and datafication of tax processes is here to stay.

Accordingly, it seems appropriate that Denmark, as an important first step, has prioritized creating a new agency dedicated to developing modern and future-proof IT solutions to the Danish tax administration. Further, it is commendable that the Danish Parliament, finally, has provided a clear legal basis for these endeavors. Hence, only a

[193] *See e.g.,* Paul Henman, *Administrative Justice in a Digital World: Challenges and Solutions*, *in* THE OXFORD HANDBOOK OF ADMINISTRATIVE JUSTICE 459, 459–80 (Marc Hertogh et al. eds., 2021).

determined, consistent, and combined focus on both the technological opportunities and the legal challenges can move the Danish tax system closer to realizing Tax Administration 3.0 and, at the same time, ensure sufficient protection of taxpayers' rights.[194]

[194] We thus share the general view presented by Steven M. Appel & Cary Coglianese, *Algorithmic Administrative Justice, in* THE OXFORD HANDBOOK OF ADMINISTRATIVE JUSTICE 481, 481–502 (Marc Hertogh et al. eds., 2021). The authors argue that although machine-learning algorithms and other automated tools present important challenges for governments, existing legal principles should prove to be no intrinsic or insurmountable obstacle to the responsible deployment of artificial intelligence. However, as the authors simultaneously stress, public administrators, elected officials, and concerned citizens must remain vigilant in their use of such digital tools, in order to help ensuring that artificial intelligence is used responsibly.