Ethics and Development of Advanced Technology Systems in Public Administration

António da Costa Alexandre[1] and Luís Moniz Pereira[2(✉)]

[1] Lisbon, Portugal
antoniocalexndre@outlook.pt
[2] Departamento de Informática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Monte da Caparica, Lisbon, Portugal
lmp@fct.unl.pt

Abstract. This article analyses aspects related to the implementation of advanced technology systems in Public Administration, taking into account the scope of action of its entities and services, as well as its main distinction with the private sector, which lies in the pursuit of the public interest, understood as justification for the execution of public policies, but also in the broader scope of the human development index. The difficulties and resistance to the adoption and use of technology in Public Administration will be examined, also in the light of practical situations whose implementation proved to be inadequate, resulting in the analysis of ideas for the future. We approach the evolution of public management models that are emerging through the action of technology. We prepare the ground for machine ethics in Public Administration by framing ethics in public services in general terms, with reference to AI systems designed in line with mainstream ethics. Of the various ethical issues that arise in this domain, we pay attention to the issue of privacy and the balance that needs to be achieved so that the use of data can contribute to ethical, beneficial and reliable technologies. We sound alerts in the field of discrimination and prejudice that the bias of technology can show. The explainability and transparency of technological systems provide confidence to decision-makers and citizens, helping to clarify responsibilities in the decisions of Public Administration agents, and the researchers working for them .

Keywords: Ethics · Advanced Technology · Public Administration

1 Introduction

In this article we list some Public Administration (PA) definitions that allow delimiting the field of action of public services and entities [1, 2], which, due to the breadth of areas in which they operate, can be the main buyer of technology, without neglecting the regulatory functions that the PA is responsible for ensuring. The theoretical framework is complemented by data from interviews carried out in other academic investigations.

A. da Costa Alexandre—Independent Researcher.

E. González-Esteban et al. (Eds.): The ETHNA System Project, LNCS 13875, pp. 224–247, 2023.
https://doi.org/10.1007/978-3-031-33177-0_14

The public interest is the main distinguishing factor of PA vis-à-vis the private sector, whose main objective is to achieve economic benefits, which can justify the search for non-profit technologies, in favour of benefits and improvement of the human development index.

We frame the adoption of technology by public bodies and services by studying two well reported known cases: both regarding the implementation of fraud risk assessment systems in granting social support, in the Netherlands and in the United States [3]. Although with operational distinctions, beneficiaries are penalised by having subsidies refused in spite of the absence of grounds, and by the inability of the systems to provide the reasons for the refusals, so that the latter cannot be contested.

Technology is also causing an evolution in public management models of PA entities and services, with the emergence of new paradigms, closer to citizens through the technological systems that assist them in carrying out tasks and implementing public policies [4, 5].

The machine ethics approach will be preceded by the theoretical framework of ethics in PA, as a necessary condition to prepare an environment conducive to ethical principles that indicate solutions to questions and problems that may arise from the implementation of technology in general.

The process of conceiving, adopting and using technologies is analysed, in an attempt to identify facilitating conditions in terms of the research and organisational culture, leadership, plus the advantages that technology can bring to public services and entities in terms of effectiveness, efficiency, transparency, and trust of citizens.

Privacy is one of the key points for citizens to trust the technological solutions that support PA bodies, without neglecting the need to find a balance between data protection and necessary concessions that allow progress, facilitate security, or contribute to public health [6].

Discrimination caused by technology arising from technological solutions that give continuity to prejudices present in society, due to the use of non-representative, inaccurate or simply wrong training data, is also mentioned [7]. Examples are presented of systems for evaluating the risk of recidivism in criminal activity used by USA judges in the decision to grant parole, which penalise citizens of African origin.

The question of the explainability of technological solutions is analysed from a dual perspective, in terms of the information that citizens have the right to obtain on issues they raise before the PA, but also in terms of attributing responsibilities to service agents and public entities [8].

2 Public Administration, Public Interest and Human Development

The expression "Public Administration" (PA) can be understood according to two different meanings: (i) regarding the mode of organisation with the characteristics and specificities of the institutions that comprise it, with a view to achieving the public interest, in compliance with the law and rights, and (ii) regarding organisational action, object of study in the sociology of organisations [1].

Several versions of the PA definition were compiled by [9], with the aim of listing the PA concepts that many conceived to demonstrate the scope and nature of the discipline

of PA. The perception of the concept of PA is important, from the outset, to reach the meaning of the word "administration" as an effort that requires cooperation from the intervening parties in the pursuit of a common objective. The PA participates in the organisation of public policies and government programs [9].

According to [2] it is easy to define PA if one is satisfied with a simplistic definition. In that case, it will be the government in action managing public affairs and implementing public policies, in a public context—without which the PA does not exist—, plus affirming its role in the execution of governmental activity. These authors highlight the decisive role of the PA in the implementation of the public interest, defining it as the "universal label" that involves public policies and their various execution programs. The public interest is generally considered an asset by the community. However, less simplistically, it can be used both to promote public policies that represent an effective common good, and to obscure other policies whose acceptance by the community is not so evident. Walter Lippmann, cited by [2] affirmed "the public interest may be presumed to be what humans would choose if they saw clearly, thought rationally, and acted disinterestedly and benevolently" [10].

Haeberlin and Comim [11] refer that [12] considered that the replacement of religious beliefs, as a principle of government, with "belief in the public interest" was at the origin of the construction of modern societies. These authors defend the insufficiency of a legal approach *stricto sensu* of public interest, advocating a broader approach that contemplates human development.

Sen [13] in his studies on ethics and economics, argues that the criterion of utility should be replaced by the notion of substantive freedoms, considering the autonomy of free choice and the real conditions of a life that makes sense to value it. This author influenced the first United Nations document on human development.

The Human Development Index (HDI) has been established since 1990, viz. in [14]) issued by the United Nations Development Program (UNDP). In the various developments it has undergone, it takes on the role of contributing to correction of distortions in the model proposed by the HDI, in what concerns the role of poverty and of inequalities in development.

According to [15] "human development can provide the possibility of longer, healthier, and more creative lives, achieving goals that humans value, with commitments that shape development with a view to equality and sustainability of the planet we share."

3 Technologies and Public Administration

Digitization processes tend to replace existing processes by altering state governance and labour relations, intensifying development tensions, as a result of the growing trend to implement Artificial Intelligence (AI) [16].

Giest and Klievink [3] understand that because of digitization and the increasing use of systems based on algorithms and data processing, a new service provision regime is being created. Technologies such as AI alter technical fields, communication channels, decision-making functions and mechanisms, as well as their levels of intervention control [17] and the ethics thereof.

Giest and Klievink [3] studied the case of the State of Michigan where the Automated Fraud Detection System (MiDAS) was implemented, and the Dutch Childcare

Allowance case. In the US, the implementation of the AI system led to about a third of the Unemployment Agency staff being laid off. In the Dutch case, incremental changes affected the structure of the organisation, along with radical changes in the work of bureaucrats that had led to wrong decisions in the Dutch childcare subsidy.

Based on these observations, we see new organisational structures and human resource coordination in both of these two use cases, definable as administrative process innovation. In Michigan, we also see a conceptual overall breakthrough where there's a new way to approach the problem of fraud, through an AI system that can handle the process more accurately, faster, and more effectively than a human being. These observations speak of a more radical innovation in the latter case, and of only successive and incremental innovations in the Dutch case. This is in line with research that emphasises how the implementation of information technologies in public sector organisations is a complicated process and where change emerges incrementally [18], cited by [3].

In such scenarios, some envision new bureaucratic roles that correspond to digital processes. Bovens and Zouridis [19] suggest that there will be three groups of employees: (1) in the data processing process, such as system designers, lawyers, and system analysts; (2) managers in control of the production process; and (3) the "interfaces" between citizens and the information system, in the help desk, and also lawyers who deal with complaints. System designers, public policy implementers, and IT specialists are the new equivalents of bureaucrats [19]. However, some bureaucrats will be needed in the implementation of public policies and decision-making [20], and the ethical concerns involved. Thus, algorithms will just be additional decision support tools rather than becoming autonomous agents on their own, with their own ethics.

The Michigan Integrated Data Automated System (MiDAS) [3] has also been widely discussed not only in the media but also in legal proceedings, such as a class action lawsuit and Auditor General Reports, and new legislation has been passed to accommodate changes in the way Michigan Unemployment deals with MiDAS regarding claims of fraud.

3.1 Michigan Automated Integrated Data System (MiDAS)

In the USA, in the State of Michigan, data mining techniques were implemented for the automatic detection of program fraud in support related to food stamps and unemployment benefits, with the aim of reducing operating costs and targeting fraud in insurance claims, the so-called Michigan Integrated Data Automated System (MiDAS). This system in a fully automated way identified 48,000 fraud claims of unemployment benefit recipients—a five-fold increase over the previous system [21]. Some Michigan residents on food assistance, for example, were automatically disqualified by the system [22]. Automatic fraud determinations also occurred if recipients did not respond to the questionnaire within 10 days or if the MiDAS system automatically considered their responses unsatisfactory [22]. Claims from beneficiaries were systematically denied by the system, which failed to provide evidence to support MiDAS' accusations of fraud [23]. This system was presented by [3] to highlight its flaws on detecting fraud in the referred support claims.

Therefore, both cases are very similar regarding the objectives linked to the implementation of both AI systems, which is to streamline the process – making it more

efficient and cost-effective – and to identify fraud more reliably. But, in the case of Michigan, this objective goes a step further by pointing to very limited or even non-existent human intervention in the process, to the point of dismissing employees for specific tasks and replacing them with an automated decision-making system, resulting in major changes in the way bureaucratic work is defined [3].

3.2 Dutch Childcare Subsidy Fraud Evaluator Algorithm

The Dutch Childcare Allowance case was widely publicised in the media and extensively documented and reviewed by supervisory bodies: the Data Protection Authority, the Auditor General, and the Dutch Parliament. This system was aimed at classifying the risk of systematic fraud, based on an algorithm that was fed with examples of correct and incorrect applications in order to learn how to 'recognize' those that pose a risk of fraud. This innovative system sought efficiency, but also with the political objective of repressing fraud. There was a strong organisational incentive to rapidly detect fraud, which, counterproductively, provided a disincentive to look critically at system results. It was set up in a hurry and the procedures, its work instructions poorly adjusted, and with little specialisation at various levels, to cite civil servants who worked at the agency. The focus on efficiency and effectiveness in combating fraud was not accompanied by an adequate organisational structure [3, 24]. As a consequence, the Dutch tax authorities used algorithms that wrongly identified fraud in the attribution of support for access to day care centres. The alleged offenders, several with an immigration background, were summoned regarding the financial support, which caused great financial and psychological difficulties for the families involved. The data protection authority would come to conclude that the data processing used by the AI system was discriminatory [3].

3.3 Public Management Technologies and Models

New Public Management (NPM) began in the 1980s in OECD countries, due to the realisation that Public Administrations represented an excessive burden and a tendency towards inefficiency. This model seeks to reverse the trend towards an increase in the number of workers controlling public expenditure, resorting to private services to carry out public functions that do not collide with the social function of the State. It also promotes automation, via information technologies [25]. According to [4] this model of public management is closely associated with ideas of the managerial style of the private sector, a strong orientation towards customer service, and the use of less hierarchical organisational control mechanisms. This paradigm is characterised by three characteristics, namely: disaggregation by dividing the hierarchies of large public organisations; competition allowing for multiple forms of provision developed between suppliers, and incentive characterised by specific pecuniary performance incentives for personnel.

Ojo et al. [4] refer to three main successors for NPM: Digital Age Governance (DAG), Public Value Management (PVM), and New Public Governance (NPG), all with a common aspect – the centrality of digital technologies. DAG highlights the influence and impact of information and digital technology developments on public sector management, with emphasis on 1) transparency; 2) the use of platforms to transmit information; and 3) creation of service shared centres. Within the scope of transparency, open access

to information in a specific area or program of public interest is promoted (the so-called targeted transparency). The digital platform aims to increase citizen engagement by fostering civic participation, better engaging citizens to crowdsource ideas about policies and programs, plus opening more channels for citizens to voice their opinions about government services and programs. Shared services seek to achieve greater effectiveness and efficiency [4, 5].

PVM has as its paradigm the achievement of public value, attained by deliberation of citizens' representatives and government staff, surpassing the sum of individual preferences of public service users. It is based on a strategy aimed at creating public value. It seeks the use of real-time data to protect entities and services, favouring autonomy and entrepreneurship, and values innovation through creativity and out-of-the-box ideas. The PVM seeks through political management to ensure legitimacy in the assumption of value propositions. It resorts to social networks to stimulate opportunities of public value; it also seeks the realisation of politically empowered citizenship; plus seeks to learn across multiple levels and audiences, cultivating a broader vision of well-being for all [4, 26].

The NPG public management model is based on institutional and network theory [27]. NPG is achieved through networking and collaboration, public-private partnerships, and citizen involvement in public decision-making. Co-production/co-creation with citizens is fostered by collaboration in networks. In public-private partnerships (PPP), both sectors share risks and resources to achieve value. Through PPP innovation is also attained, given the different origins of the partners. A summary of the characteristics of paradigms subsequent to New Public Management (NPM) is found in [27].

Disruptive digital technologies such as AI are enacted in public management models associated with the era of digital government and public value management. We found no case where AI was deployed in the context of NPG. This may be due to the nature of AI solutions currently being deployed [4].

The European Law Institute (ELI) conceived the "Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration" [28], considering that PA is confronted with specific challenges resulting from the implementation of AI algorithmic decision-making systems and machine learning. Use of these techniques poses specific problems relating to the principle of good administration. Furthermore, issues such as transparency, accountability, compliance, and non-discrimination are particularly relevant in the context of PA. This European project led to the development of such ELI Model Rules [28]. Although inspired to some extent by EU legislation, it is compatible not only with existing EU legislation but also with draft legal provisions, the ELI Model Rules have been designed in such a way as not to depend on EU law. In this sense, they can serve as inspiration for national legislators (even outside the EU), for governments and PA. Algorithmic decision making can be approached in several ways. The central idea behind the Model Rules [28] is an impact assessment, adopting an approach that distinguishes between high-risk systems that warrant an impact assessment (in its Annex 1) and low-risk systems (in its Annex 2) where such an assessment is not warranted, and even systems that cannot be easily classified *ex ante* as belonging to Annex 1 or Annex 2, in which case a risk assessment will be made if screening reveals that the system poses at least a substantial risk [28].

3.4 Public Administration and Technology Adoption

A crucial role is reserved for the PA in the adoption of AI, considering its double condition as a user of the enormous potential of these technologies, but also with regulatory functions of AI, defining procedures, rules and competences, and particularly with regard to the Ethics of AI, establishing the minimum conditions for the private sector and citizens to deploy and use it in accordance with ethical requirements, but this is still lacking, hence our call to attention. EU Member States have given greater emphasis to the regulatory role of PA or, at best, to a facilitating role, that is, as the entity that defines the basic conditions for the ethical use of AI by private entities and citizens, but not by the PA itself. PA's role of "first buyer" [29] and direct beneficiary of AI is somewhat neglected. Most AI literature views the government as a regulator. Discussion of the role of PA from the point of view of an AI user is scarce, although PA is increasingly becoming a significant user of AI [30]. [31] refers it is essential that this issue be taken on with greater accuracy, considering the various areas in which PA operates, from health to education, and other areas with a wide range of services and entities that represent an essential market for AI, not only as buyers, but as well as disseminators of these technologies in the economic and social sectors. In this regard, [32] understands that PA will be the area where the opportunities and challenges of AI will constitute a priority on the agenda, with repercussions felt by those administered. In the opinion of [29], the key to success is to gain the trust of users in the processing of their data. To reach this opinion, overall, 686 AI use cases in the public sector were collected and analysed as reported in [33, 34].

The adoption of AI in PA procedures has the potential to bring greater efficiency and effectiveness in the provision of services to companies and citizens, increasing the level of satisfaction and confidence in the quality of the public service [4, 29]. The main applications of AI in this context include process automation, virtual agents and speech analytics, predictive analytics for decision making, sentiment analysis, and document review [4, 35]. Madan and Ashok [36] highlighted two specific technologies: machine learning (ML) and natural language processing (NLP), considering that these two technologies characterise most AI applications in PA, with reference to the cross-case analysis of [36] and to the AI case study archive of the European Commission and Joint Research Centre.

In this perspective, PA can also be seen in its distinctive aspects to the viewpoint of the private sector. For the pursuit of public interest differs from the profit motive that presides over private entities. In the study and implementation of PA AI systems they must not just aim at economic benefits, in accord with what was previously exposed concerning the pursuit of the public interest, and its human development perspective.

4 Ethics in Public Administration

The functional environment of public organisations is dynamic, full of unpredictable and ambiguous international events, changes in power, domestic policy challenges, and technological changes that affect the functioning of PA entities. Currently, workers in public functions increasingly work in a world characterised by volatility, uncertainty, complexity, and ambiguity, resulting in multifaceted and contradictory requests that can

configure dilemmatic situations [37]. In this scenario, values, integrity, conduct and ethics can motivate and support state agents.

Quotes [38]: "Savater [39] assumes that ethics is a kind of moral synthesis. An imagined ideal of collective existence, a social community with a united and coherent character." Ethics is thus a natural concern in the discussion about the real role of PA decision-makers in exercising the discretion of the powers entrusted to them, as well as in the pursuit of the public interest by holders of political office [40] and striving to ensure that algorithms respect and embody such ethics [41].

Interest in ethics in PA has been increasing with the introduction of the principles of NPM [25] in public sector bodies and services [1, 42, 43], because of the adoption of notions such as efficiency, obtained results, and economy, sparse in ethical content and conducive to pressures in the public service, causing doubts/confusion about the application of ethical standards. In this way, it is important to question why these ends, also implying a reflection on the means and relationships that are established within entities [44]. The importance of ethics is revealed, from the outset, by the vision and perception of the reason for being of PA and how its action should be undertaken in terms of demands and adaptation to change. The existence of a current crisis of values is often mentioned in different environments and contexts. Indeed, the aim of a civilization of lightness means all but living lightly, at present. Although the weight of social norms has lightened, life feels heavier. In unemployment, precariousness, instability in couples, and duration of overwork [45].

Lipovetsky [45] highlights the importance of ethical values and ideas, realising that despite the findings of extreme individualism, we are not located at the zero degree of values. The sense of moral indignation has by no means been eradicated, as our societies reiterate a stable core of shared values: human rights, honesty, respect for children and rejection of violence and cruelty. Although it refers to the inconsideration of sacrifice in the old-fashioned way, it does not mean the disappearance of the spirit of responsibility and solidarity.

It is in a context that considers ethics in public entities within a general scope that the ethical issues of AI can be more accurately framed. Notwithstanding that their specificity might be seen from an autonomous perspective, the ethical tools that already exist in PA, or that may come to be contrived, will facilitate the dissemination and application of this new burgeoning area of ethics [6].

5 The Ethics of the Machine

In the framework of ethical issues in AI systems for PA, it is important to summarise the ethics of the machine in general terms to outline possible adaptations to the public sector. The ethics of the machine, an expression coined by [46], has as its main objective to conceive a machine oriented with compatible ethical principles in the scope of the decisions it takes and on the possible consequences of its action. Fundamental human rights complemented with codes of ethics are considered capable of mitigating the risk of AI to evolve in an opposite direction to human values, by identifying risks, priorities, vulnerabilities [41, 47–51]. The Universal Declaration of Human Rights (UDHR), although not binding, is part of the legal system of democratic states.

For [52] transmitting ethics to robots by observing humans faces a prior complication—the human difficulty in determining which human attitudes collide with ethical principles, and how to resolve our contradictions. The greatest benefit may lie not in what we teach the machines, but in what we learn from teaching them. [41, 50, 51] also mention the difficulty of the lack of a universal ethics. On the other hand, our moral conduct is essentially instinctive and only in more complex situations it is necessary to think about right or wrong in terms of specific difficulties or discomfort.

The issues of weaponry in general which uses AI and the dangers arising therefrom with ever more powerful and more efficient weapons are dealt with in some detail by [53], who warn of the central paradox that the greater the digital capacity of a society, the more vulnerable it becomes. Computers, communications systems, financial markets, electricity networks (and the digital command and control systems they depend on)—even the functioning of the democratic system—involve systems that are, to varying degrees, vulnerable to manipulation or cyberattack. In a conflict situation, the most extreme form of protection may involve cutting the network connections, taking the systems offline. Disconnection could become the ultimate form of defence. Consequently, ethical principles need to be considered for addressing issues and situations that the law still has difficulty in resolving.

The existence of areas of absence of legality; the tendency towards the universality of ethical postulates; the anticipation of legislators' responses; the assistance given to the interpretation of legal rules; the capacity to integrate loopholes in the law; the increased efficiency of bonds arising from voluntary adoption—are some of the foundations that support permanent ethical intervention [54].

Bostrom and Yudkowsky [55] warn of the necessary care with applied ethics in contexts that are very different from those of the human condition. Common normative precepts are subject to conditioning in the face of different empirical conditions, and it is also important to consider whether these precepts are suitable for future hypothetical cases in which their preconditions might be considered invalid.

What is relevant is that an agent, when faced with a problem, can have the ability to generate hypotheses and choose the solution that seems most appropriate to him/her, through a causality supported by their experiential memory, by previous decisions, and by accumulated preferences [41].

Hagendorff [56] and Jobin et al. [57] selected documents with references to ethical principles and guidelines in the field of AI, in the public and private sectors, immediately indicating that the ethical challenges of AI are transversal to society. These authors listed the main ethical issues they identified in their research: transparency; justice and equity; non-maleficence; non-discrimination; responsibility; and privacy, are the dominant ones. Jobin et al. [57] emphasise that the principle of transparency prevails with approaches related to explainability and interpretability, with a main focus on the use of data, the human interaction with automated decisions, as well as data system applications. Transparency is presented as a method to minimise harm, but also as a factor that fosters trust, as well as a promoter of freedom and autonomy. The above references have been the object of criticism that considers transparency as a pro-ethical principle, a condition that can affect or harm other practices or ethical principles, namely set forth by [58].

5.1 Contributions of Ethical Perspectives in the Implementation of AI Systems

In an investigation at Instituto Superior Técnico, University of Lisbon, utilitarian ethics were examined, and general egalitarianism found questionable in an AI system that does not confer a main advantage to its owner. It was found that either AI offers an individual advantage to the owner, or it will not be adopted. The consequent lack of interest in its purchase will have negative effects on production, restricting research and development [59]. Hence there is an onus to incorporate in AI systems ethical values proximate to their owners, though without disregarding general social values.

In this respect, the ideas of Rawls [60] have been defended for the field of AI, for example by [61]: "A Theory of Justice could help explore known problems of data bias, injustice, liability and privacy, in relation to machine learning and AI applications in government." [62], in his approach to the ethics of AI, considers that in Rawls's conception of justice, the veil of ignorance can be replaced by a much more natural condition of prudent selfishness in a finite world. The American contractualist Rawls is also cited by [63] in reference to a new social contract with the intervention of algorithms. Leben [64] describes a Rawlsian algorithm as an alternative to a utilitarian solution, which will assess the probability of survival for each person in an autonomous vehicle, for each action, calculating which action would obtain agreement starting from an original position with fair negotiation. Rawls [60] is also mentioned, for example, at the European Union level regarding the need to use AI in decision-making which can overcome human judgement. It is questionable whether a human will still contribute to a stronger protection of data subjects, or whether the better performance of machines even regarding the political and legal values at stake; for example, guaranteeing fair equality of opportunities, as defended by [60], thereby making human intervention redundant or dysfunctional [65].

Rule-based ethical theories, such as the Kantian one [66], are considered promising for machine ethics due to the structure of their judgments. According to a formalist interpretation of Kant's categorical imperative—"Act as if the maxim of your action should become, by your will, a universal law of nature" [66] —a machine might, for example, place prospective actions on top of traditional maxims (forbidden, permitted, obligatory), in the top-down approaches to the development of programming agents in which Kant's utilitarian ethics and deontological ethics are likely to be applied [65].

The principle of responsibility theorised by Hans Jonas was at the origin of the construction of a new understanding regarding the ethical aspects resulting from the position technology assumes in society. He considered responsibility as a basic ethical principle of human action, given the issues raised by technology, defending how the theory of responsibility can be seen as an ethics par excellence [67]. This author refers to an imperative for the new technological types of human action, that addresses the new type of object of action. Jonas [67] states: "Act in such a way that the effects of your action are compatible with the permanence of an authentically human life in society on Earth." The new imperative is addressed to public policy, as opposed to Kant's categorical imperative being addressed to the individual subject [67]. Alexandre [6] found that such Jonas's principle of responsibility is still not accepted within the scope of machine ethics. However, it seems to us this gap may be filled in the near future, as a result of his intergenerational approach, which is concerned with the legacy for the next generations.

6 European Union: An "Engine" of AI Ethics

The European Union (EU) has been looking for foundations for the implementation of an ethical AI, publishing several documents that justify an analysis not only for their pioneering spirit, but also for the repercussions that it may have in European countries. The EU's concerns with ethical aspects and fundamental human rights are evident, namely in the promotion of a human-centred AI defending human values as an "engine" of development and economic and social progress, seeking in particular to mitigate existing inequalities.

Alexandre and Pereira [68] have addressed the main documents produced by the European Union highlighting ethical concerns with the implementation of AI in PA. Of the documents analysed by these authors, we highlight three:

– In 2019, the High-Level Expert Group on Artificial Intelligence (AI HLEG), which was appointed by the European Commission [69], released the Ethical Guidelines for a Trusted AI, the first institutional document emphasising at planetary scale the importance of ethics in the development and implementation of AI. In this document, PA was also considered a beneficiary of improvements that AI systems present to the efficiency of public services, in their provision of public goods and services to society [68].

– In the diagnosis made by the "White Paper on AI—An European approach towards excellence and trust," published 19 February 2020, on the use of AI in the public sector, where the innovation capabilities and greater efficiency of services are also highlighted, with express reference to hospitals, transport services, public utility in general, financial supervisors, and other areas of public interest [7, 68].

– The Proposal for a Regulation, of the European Parliament and of the Council, from 21 April 2021, which establishes harmonised rules on Artificial Intelligence, known as the Artificial Intelligence Act (AIA) [69], constitutes the first institutional proposal that aims to regulate AI, openly challenging those who believe that the law should not regulate emerging technology, a very common stance in Silicon Valley [70]. This proposal for a community regulation highlights a change in the Commission's narrative: whereas in the White Paper one could identify the reversal of EU priorities, previously assuming a global competition neglecting fundamental rights, the "Proposal for a Regulation" advocates to ban AI practices which, via high-risk AI systems, may violate the European Union values and fundamental rights enshrined in the Charter of Fundamental Rights of the European Union, while proposing softer provisions for lower- and medium-risk AI systems. This proposed regulation foresees the future rules regulating AI to be supervised and monitored by national authorities. This may be an opportunity for member states not yet having created an entity with competences in the field of ethics in AI to create, as proposed in [41], a "National Ethics Commission for AI" to oversee the foreseen regulatory structures by Brussels, with a higher-up competence in the field of ethics [68].

6.1 European Strategies

According to [29, 34] the AI Strategies of EU Member States regarding the public sector show ethical concerns, with some expressing the intention to develop an ethical framework to guide the implementation of AI in the public sector, seen as a contribution to

establishing trust between workers in public functions and citizens, ensuring the quality and compliance with ethical values of the AI used by the PA. At the level of legal regimes, legal reforms are referred to with the aim of streamlining the development and use of AI. Some countries have noted difficulties in exchanging data between public entities and the private sector, and thus intend to create "regulatory structures" to facilitate this exchange. A significant number of these reforms will be specific to each sector, with emphasis on health [29].

In the European Union, in February 2022, national AI strategies had already been published in 23 Member States, also including Norway, establishing conditions for the development and acceptance of AI. The public sector's pioneering spirit is affirmed in the 2021 Review of the Coordinated Plan on Artificial Intelligence. It affirms the potential of AI in modernising the public sector itself, with the capacity to: "(i) automate simple and repetitive cognitive activities freeing-up labour time for more high-value activities; (ii) increase the predictive capabilities, enhancing data-driven decision-making; and (iii) support user-centric service personalization, increasing the effectiveness of public service delivery" [33].

7 Obstacles in Public Administration to the Implementation of AI

7.1 Organisational Culture

At the level of public services and entities, an organisational culture is decisive in the implementation of AI and its acceptance. In organisational dynamics, it is important to combat the inertia inherent in a routine rigidity that inhibits change and the development of new capabilities, but also the inertia that results from the constant change caused by electoral cycles, in democratic regimes that may lead to successive political changes, where concerns about re-election in the latter part of a term usually condition the decisions of political office holders. The influences of other government departments cannot be overlooked, as well as pressure from civil society and the media. As a result, PA, in addition to its internal difficulties, is also part of a turbulent and volatile external environment [36]. Inertia is considered a critical factor, due to the scarcity of resources for pilot and/or innovative projects, but also due to the difficulty in finding AI specialists. Inertia may also result from the rigidity of bureaucratic factors, centralised decisions, poorly trained workers, and too from attitudes contrary to the sharing of data within public entities and with outside departments [36].

At the PA's internal level, facilitating factors for the implementation of AI are an organisational culture with innovative and dynamic characteristics, favouring experimentation to face the risks of these technologies. Transformational leaders will motivate their workers to change and will seek to influence design as well as interaction with other agencies, services, and departments [30, 71, 72].

The Technology-Organization-Environment (TOE) structure of [73], as mentioned in [36], is indicated to explore the adoption of technology in different environments, due to the identical importance that it gives to both organisational and technological contexts in the implementation of technology in organisations.

The processes of diffusion and adoption of technological innovation according to [74], operate for long periods of time, can be understood in multiple analyses, and at

different levels of aggregation. They highlight the inseparability of human values, in the complex interactions between technological innovation, people, scientific concepts, aspirations, and consequences.

7.2 Contribution of Public Procurement

According to [6] the clear definition of AI ethics issues should be placed *ex ante*, in the acquisition and development of AI systems by the PA, with a clear specification in the terms of reference. The public entity may resort to preliminary market consultation, listening to potential suppliers as to the conditions they offer to guarantee ethical requirements. Thus, in the formulation of public policies with a view to implementing these technologies, the Government's purchasing power can play a key role in this matter, as referred in [7] and [29] with a determination of ethical criteria in the acquisition requirements that ensure private companies designing AI systems adequately meet public standards. This requirement can also contribute to the dissemination of ethical AI in the private sector.

Desouza et al. [75] propose in this context an agile acquisition process that allows for iterative development lifecycles through the acquisition of hardware and software in stages, ensuring that early access to knowledge of a sector focuses on problem definition, rather than developing detailed solution specifications.

7.3 Data – What Balance Between Security, Privacy, and Innovation?

The PA, in the performance of its activities, collects a huge amount of data from citizens, superior to most private entities [4], and data collected, for example, in the areas of taxation, health, education, and social security are particularly sensitive. Data privacy and security have become increasingly important, resulting in the publication of legislation of which the EU's General Data Protection Regulation (GDPR) is the highest exponent [76].

Madan and Ashok [36] refer that the accessibility of data and its use by governments for purposes other than those that were collected raise serious concerns related to privacy. On the one hand, the use of data can lead to superior public policy and service delivery about duty-oriented and service-oriented public values. However, at the same time, it undermines the social public value of privacy. [77, 78] state that these data provide information that may allow classifying citizens into clusters of micro populations. Thus, the accessibility of data and its use by governments for purposes other than those for which it was collected raise serious privacy concerns. There is a balance that needs to be achieved between the protection of personal, sensitive, and confidential data and, on the other hand, the potential of data collected by the public sector for the design and development of AI systems aimed at the public services, which manifest aptitude for the improvement of public services that will reflect on the citizens' own living conditions, thus contributing to progress.

A particularly sensitive issue is predictive policing, that is to police individual behaviour in order to attempt to predict rule infringement by the individual. The main ethical issues gravitating around data selection and machine biases, such as "visualisation and interpretation of forecasts, transparency and accountability, time and effectiveness

as well as the problem of stigmatisation of individuals, environments and community areas" [79].

7.3.1 Privacy and Data Collection

The concept of privacy has evolved since the dawn of the fourth industrial revolution, in the permissions that users grant to electronic platforms and in the exposure that users make of their lives on social networks. Even so, privacy is an aspect that inspires concern in the implementation and use of AI in public services. Ensuring respect for citizens' privacy is considered one of the essential aspects for winning trust in AI and for it to be successful [6].

In the interviews carried out by [6] a specialist's reference was made to the balance that must be achieved between data privacy and the need to compromise in terms of solutions that are adequate to guarantee public health and, ultimately, save human lives. It was mentioned that in a pandemic situation, like the recent pandemic situation due to the spread of the Covid 19 virus, several rights can be limited. The right to privacy is almost considered evangelical, seen as if it were a greater right than the right to life itself. Therefore, it is necessary to establish a discussion regarding the limitation of the right to privacy in a pandemic context. The specialist cited the example of South Korea, which had a lot of cases in the first SARS. To prevent the spread of Covid 19, it used data such as economic transactions with credit cards, mobile phone data, and Bluetooth communication so as to test people and quarantine them before transmitting the virus. This decrease in privacy has occurred despite South Korea having a democratic regime. For this interviewee, the issue of privacy was identified as the most important in the PA, and he also referred that although research on data from people with cancer is regulated, preventing the identification of the patient, a detailed analysis of the data allows this identification.

8 Discrimination and Prejudice

Prejudice, discrimination and racism, or biases related to gender equality, are risks involved in any economic or social activity. Human decision-making is not immune to error and bias. However, the same prejudice in the field of AI and particularly in PA can have a much greater effect [6]. Within the scope of AI, discrimination is identified "on grounds of sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation" [7]. Berryhill et al. [80] refer that in the very conception of AI tools, the lack of diversity is reflected in the technology industry. For example, only about 19% of research related to AI is authored by women; on the other hand, the number of AI publications authored by women has stagnated since the 1990s.

The models of recidivism risk introduced in the US judicial system, designed by data-fed algorithms, demonstrate racial discrimination against African-Americans [81–83]. They make it possible to assess the danger of recidivism that each convict represents, with the conviction that this assessment is more accurate than a judge's off the cuff guess [84]. The LSI–R questionnaire focuses on ten different topics, addressing questions about the inmate's birth, education, family, neighbours, and friends, classified with different

indices depending on their importance, as mentioned by [85], in the publication "LSIR – What is the Level of Service Inventory-Revised Risk Assessment?", edited by LLC, Trial Lawyers. For [82], the questionnaire does not directly identify the race, as this would be illegal, but with the quality of the details that each prisoner provides, this illegal question becomes almost superfluous [86].

An independent investigation carried out by ProPublica (a non-profit entity doing investigative journalism), analysed COMPAS—Correctional Offender Management Profiling Alternative Sanctions—concluded that defendants of African descent were much more likely to be incorrectly assessed as having a higher risk of recidivism compared to Caucasian individuals, as for the latter was found an incorrect tendency towards low-risk flagging.[1]

In the interviews carried out by [6] of specialists in the world of AI, from university professors, civil society agents, to public administration leaders, the following were highlighted:

Ethical issues related to discrimination that AI systems are still unable to avoid, resulting from the bias of algorithms that produce discriminatory decisions, were the main issues identified in this context.
Neural networks were mentioned due to the lack of language concepts capable of explaining the decisions they make.
The biases result from the actions of the humans that are at the origin of the data which will contribute to discriminatory actions. If the methods used cannot avoid this bias, then the mistakes of the past will be repeated and reinforced.
The need for communication of the possible biases, by those responsible for the design, to the end adopters of the systems.

In machine learning, the under- or over-representation of data for some groups may occur, so the careless use of data in AI training can cause or perpetuate discrimination and inequalities that already exist in society. AI itself can contribute to sexism and gender stereotypes through tools that enhance such behaviours, e.g., most virtual assistants are given female names and personalities associated with sexist/stereotypical "female reactions" and some hardier physical robots (e.g., rescue robots) were given male forms [87].

Discrimination within the scope of AI must also be seen in terms of unequal access, considering citizens with insufficient resources and/or knowledge to establish a technological connection with Public Administration services and bodies, as they are in a situation of digital exclusion. Electronic administration cannot serve as a factor of discrimination between citizens with digital capabilities and digital illiterates. The obligation, in the most complicated periods of the pandemic, to hold basic, secondary and higher education classes through the Zoom and Teams platforms, faced with the difficulty of connecting to the internet due to the lack of necessary equipment, either by students or by some teachers [88, 89]. In an interview conducted by [6] it was mentioned that, of

[1] Interested readers can consult the work and methodology of this entity at: https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis.

course, knowing how to send an email or use WhatsApp is not enough for a full-fledged insertion in the world of technology.

Bias in AI is verified when machine learning can lead to discrimination against people, specific groups, generally marginalised by gender, social class, sexual orientation, race, or religion. It may originate from prejudiced attitudes in the very design of the AI system, and/or be due to the use of non-representative, inaccurate or simply wrong training data. These are cases of discrimination, legally defined as the unfair or unequal treatment of an individual (or group) based on certain protected characteristics (also known as protected attributes) such as income, education, gender, or ethnicity [24, 90].

Service providers, both public and private, must consciously address issues of prejudice and discrimination, ensuring that behaviours, experiences, and views that represent the diversity of the population have been considered, cf. Committee on Standards in Public Life [91].

9 Explainability/Transparency, Transparency and Responsibility

In terms of explainability, the opacity of the algorithms that constitute autonomous AI systems that make decisions, but having difficulty in explaining the respective grounds, is a major concern. Users expect to understand or else be explained the grounds for those decisions. In the absence of these grounds, recipients are limited in any claim they wish to present. [92] refer that at present machines are not "good storytellers", meaning that they cannot explain the gist and flow of their reasoning and conclusions in a language readily understandable by human users.

According to [8], opacity can be: (i) intentional to protect intellectual property [93]; (ii) illiterate, where a system is only understandable to those with technical skills; and (iii) intrinsic, the complexity of the system makes the understanding of its decisions difficult for any human being.

Olsen et al. [94] conceive that the implementation of AI in PA raises some concerns that may cloud legal measures, delaying the implementation of these systems. The first concern is the loss of control over systems and therefore a clear link to accountability when decisions are made. In the exercise of discretionary power, the agent assumes responsibility for the decision. The fear is present even when the AI system is used in conjunction with the human or supervised by the human, yet the deference to the machine creates a vague sense of responsibility for the decision. The second fear is the loss of human dignity, with the reduction of humans to mere "cogs in the machine". Taking away the ability to understand and communicate freely with another human being can easily lead to alienation and loss of human dignity. Finally, the difficulty of using data that will lead to false and discriminatory decisions was also identified.

A hybrid system was proposed by [94]: the algorithm using machine learning in PA would be used to produce drafts of decisions that can increase efficiency in services or organisations without reducing quality, provided that the data submitted to machine learning presents a sufficiently high volume and learning has a point of reference in well-founded decisions. The authors state that trainees also learn from previous decisions and that humans differ from algorithms because these tend to be more rigorous than humans.

Actions can be filed in the courts against procedures with automated decision-making, due to the opacity of the systems preventing the understanding of the reasons

for the decision [8]. These systems, in particular when opacity is intrinsic, need to be adapted to the need to justify acts in force in PA, so that their implementation in the public sector does not result in the annulment of acts whose foundations are unintelligible.

Regan et al. [78] conceive that the opacity of AI systems in whatever PA service challenges the traditional responsibilities of administrators, regardless of substantive policy, or whether decisions directly involve citizens. Therefore, when considering the use of AI systems, the manager should take into account:

The likely effects of an AI system on its oversight, as well as on unintelligible material in its decision making.
Determine the regime of responsibility resulting from the decisions of such systems.
Determine acceptance levels of AI autonomy and the extent to which they should be changed.

Transparency and explainability in AI-based decisions can elicit greater trust from both PA officials and citizens. However, the downside of increased transparency is the ability to manipulate the system for private reasons [4, 95].

The implementation of AI and respective ethical guidelines is a multidisciplinary process requiring contributions, among others, from technology, ethics, statistics, law, social sciences, legislators, journalists, including politicians and the population at large [43, 80]. Depending on the respective application, contributions from sociologists, psychologists, doctors, or others with experience in the respective area will also be necessary [96].

10 Conclusions

Public Administration (PA) services and entities as advanced technology users, and namely AI, represent a particularly important position in the market, given the number of areas that comprise it and the organisations that gravitate in its sphere, thus being considered as the main promoter and buyer of technological systems. The PA is also assigned the role of regulator, responsible for protecting users in terms of security, but also for defining the basic conditions, in particular for the growing preoccupations with ethical use of AI systems by private entities and citizens.

The public interest that characterises the public administrative function, as the main differentiating factor from private activity, must justify the search for technological solutions that contribute to improving the human development index (HDI), allowing for longer, healthier, and more creative lives, in line with the human-centricity aims of Industry 5.0 and Society 5.0.

Technology, namely AI, is producing changes in technical areas, in communication, and in decision-making. Public services have been endowed with greater effectiveness, efficiency, and transparency.

However, the implementation of AI in PA as a user has received less attention from political power, compared to the regulatory role of AI in terms of security and the conditions for the implementation of AI ethics.

Some models for evaluating the risk of fraud in the attribution of social benefits, such as the subsidy for day-care centres in the Netherlands, or the unemployment fund in the

State of Michigan, in the United States, were poorly implemented, with flaws in the assessments carried out, and in the inability to present adequate grounds in identifying fraud risks.

The interest in ethics in PA has been increasing with the application of New Public Management principles such as efficiency, results, and economy, which are however sparse in ethical content and conducive to pressures on the service. The importance of ethics is revealed by the vision and perception of the PA's reason for existing and how its actions should be enacted in terms of diligent demand and adaptation to change.

It is in a context that considers ethics in public entities in a more general context that the ethical issues of AI can be framed with greater accuracy, albeit their specificity may be envisaged from an autonomous perspective. The AI ethical tools that already exist in the PA, or that come to be conceived in it, will facilitate the dissemination and application of this new area of ethics.

The ethics of the machine is faced *ex ante* with the difficulty arising from the non-existence of a universally accepted ethics, therefore it being necessary to define which human attitudes collide with ethical principles, and how to resolve our own contradictions. The ethics of AI has also been designed to fill in an absence of legal rules, aiming to anticipate responses from the legislator, but also contributing to the interpretation of legal rules, the integration of loopholes in the law, plus the increased efficiency provided by voluntary adherence bindings. Some AI systems have been designed respecting principles of ethical currents such as utilitarianism, contractualism, deontology, or virtue ethics. The new imperative defended in [67]: "Act in such a way that the effects of your action are compatible with the permanence of an authentically human life on Earth," addresses public policies, contrary to Kant's categorical imperative directed only to the individual, yet still forgotten by this new branch of ethics can positively contribute to the realisation of a more universal machine ethics.

References

1. Bilhim, J.: Teoria Organizacional – Estruturas e Pessoas, 7th edn. Instituto Superior de Ciências Sociais e Políticas, Lisbon (2013)
2. Shafritz, Jay, M., Russell, E.W., Borick, Chrisopher, P., Hyde Albert, C.: Introducing Public Administration. Ninth Edition, Routledge (2017) https://himia.umj.ac.id/wp-content/uploads/2021/08/Introducing-Public-Administration-by-Jay-M.-Shafritz-et-al.-z-lib.org_.pdf
3. Sarah, N.G., Bram, K.: More than a digital system: how AI is changing the role of bureaucrats in different organizational contexts, Public Management Review. https://www.tandfonline.com/doi/full/10.1080/14719037.2022.2095001 (2022)
4. Ojo, A., Mellouli, S., Ahmadi Zeleti, F: A Realist Perspective on AI-era Public Management – An Analysis of Mechanisms, Outcomes and Challenges of AI Solutions in the Public Sector. In: 20th Annual International Conference on Digital Government Research, Association for Computing Machinery. https://dl.acm.org/doi/abs/10.1145/3325112.3325261 (2019)
5. Tan, E., Crompvoets, J.: A new era of digital governance. KU Leuven Public Governance Institute. Wageningen Academic Publishers (2022). https://www.wageningenacademic.com/doi/pdf/10.3920/978-90-8686-930-5_1
6. Alexandre, A. C.: Novos Desafios da Administração Pública: Inteligência Artificial e Ética nos Sistemas Inteligentes com Autonomia. Masters dissertation, Instituto Superior de Ciências Sociais e Políticas, Lisbon (2022) https://www.repository.utl.pt/handle/10400.5/25585

7. European Commission: White Paper on Artificial Intelligence: A European approach to excellence and trust https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (2020)
8. Cobbe, J.: Administrative law and the machines of government: judicial review of automated public-sector decision-making. Legal Studies, vol. 39, pp. 636–655, Cambridge University Press, Cambridge MA, USA. (2019) https://scholar.google.com/citations?view_op=view_citation&hl=en&user=TsHYXWoAAAAJ&alert_preview_top_rm=2&citation_for_view=TsHYXWoAAAAJ:9yKSN-GCB0IC
9. Nwanisob, Beluchi C., Christopher, I. C.: Definition of Public Administration. Am. Int. J. Bus. Management (AIJBM) **3**(9), (2020). https://www.aijbm.com/wp-content/uploads/2020/09/I395661.pdf (2020)
10. Lippmann, W.: Public Opinion. Transaction Publishers, New Brunswick and London, Second printing (1998)
11. Haeberlin, M., Comim, F.: Todos e cada um de nós: o interesse público como critério de desenvolvimento humano. Revista Brasileira de Políticas Públicas, vol. 10. (2020)
12. Coulanges, F.: La Cité Antique. Librairie Hachette, Paris (1900)
13. Sen, A.: Development as freedom. Oxford University Press. (1999)
14. United Nations Development Programme: Human Development Report 1990: Concept and Measurement of Human Development. UN, New York. https://hdr.undp.org/content/human-development-report-1990 (1990)
15. United Nations Development Programme: Human Development Report 2010: The Real Wealth of Nations: Pathways to Human Development. Palgrave Macmillan, New York. https://hdr.undp.org/system/files/documents/human-development-report-2010-complete-english.human-development-report-2010-complete-english (2010)
16. Petropoulos, G., Marcus, J. S., Moes, N., Bergamini, E.: Digitalisation and European Welfare States. Bruegel, Brussels. https://www.bruegel.org/sites/default/files/wp_attachments/Bruegel_Blueprint_30_ONLINE.pdf (2019)
17. Erdurmazli, E.: Effects of Information Technologies on Organizational Culture: A Discussion Based on the Key Role of Organizational Structure. In: Süleyman Davut Göker (Ed.) A Closer Look at Organizational Culture in Action. https://www.intechopen.com/chapters/72534 (2020)
18. Meijer, A., Bekkers, V.: A Metatheory of E-Government: Creating Some Order in a Fragmented Research Field. Government Information Quarterly. https://www.sciencedirect.com/science/article/abs/pii/S0740624X15000568 (2015)
19. Bovens, M., Zouridis, S.: From Street-Level to System-Level Bureaucracies: How Information and Communication Technology is Transforming Administrative Discretion and Constitutional Control. Public Administration Review 62. https://onlinelibrary.wiley.com/doi/10.1111/0033-3352.00168 (2002)
20. Keiser, L.: Understanding Street-Level Bureaucrats' Decision Making: Determining Eligibility in the Social Security Disability Program. Public Administration Review 70(2):247 – 257. DOI:https://doi.org/10.1111/j.1540-6210.2010.02131.x (2010)
21. Gilman, M.: States Increasingly Turn to Machine Learning and Algorithms to Detect Fraud, U.S. News & World Report, Feb. 14. (2020) https://www.usnews.com/news/best-states/articles/2020-02-14/ai-algorithms-intended-to-detect-welfare-fraud-often-punish-the-poor-instead
22. Richardson, R., Schultz, J.M., Crawford, K.: Dirty Data, Bad Predictions: How Civil Rights Violations Im pact Police Data, Predictive Policing Systems, and Justice, New York University Review Online, Vol. 94:15 (2019)
23. De la Garza, A.: States' Automated Systems are Trapping Citizens in Bureaucratic Nightmares with their Lives on the Line. Time, May 28 (2020) https://time.com/5840609/algorithm-unemployment/

24. European Union Agency for Fundamental Rights: Bias in Algorithms –Artificial Intelligence and Discrimination. (2022), https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf
25. Carvalho, E.R.: Reengenharia de Processos na Administração Pública. Universidade de Lisboa, Instituto Superior de Ciências Sociais e Políticas (2001)
26. Sedlačko, M.: The five tasks of public value management: public value as a programme of administrative and societal democratisation. J. Public Adm. Policy **XIII**(2), Winter 2020/2021. (2020) https://sciendo.com/article/10.2478/nispa-2020-0019
27. Patapas, A., Raipa, A., Smalskys, A.: New public governance: the tracks of changes. Int. J. Bus. Social Res. **4**(5), (2014) https://econpapers.repec.org/article/lrclarijb/v_3a4_3ay_3a2014_3ai_3a5_3ap_3a25-32.htm
28. European Law Institute: Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration. (2022), https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Model_Rules_on_Impact_Assessment_of_ADMSs_Used_by_Public_Administration.pdf
29. Misuraca, G., van Noordt, C.: Overview of the use and impact of AI in public services in the EU, European Union (2020) https://publications.jrc.ec.europa.eu/repository/handle/JRC120399
30. Kuziemski, M., Misuraca, G.: AI governance in the public sector: three tales from the frontiers of automated decision-making in democratic settings. Telecomm Policy **44**(6) (2020) https://pubmed.ncbi.nlm.nih.gov/32313360/
31. Alexandre, A., Costa: Porquê a Ética na Inteligência Artificial? Jornal i, 14/10/2022. (2022) https://ionline.sapo.pt/artigo/783201/porqu-a-etica-na-intelig-ncia-artificial-?seccao=Opiniao_i
32. Aydın, A.: Transformation of Public Administration in the Impact of Artificial Intelligence Systems (2020), https://medium.com/@sibervatandas/transformation-of-public-administration-in-the-impact-of-artificial-intelligence-systems-d564bfc40764
33. Ricart, J., Van Roy, R., Rossetti, F.V., Tangi, L: AI Watch - National strategies on Artificial Intelligence: A European perspective. EUR 31083 EN, Publications Office of the European Union, Luxembourg (2022). https://ai-watch.ec.europa.eu/publications/ai-watch-national-strategies-artificial-intelligence-european-perspective-2022-edition_en
34. Tangi, L., van Noordt, C., Combetto, M., Gattwinkel, D., Pingnatelli, F.: AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector. EUR 31088 EN, Publications Office of the European Union, Luxembourg. (2022) https://publications.jrc.ec.europa.eu/repository/handle/JRC129301
35. Wirtz, B.W., Weyerer, J.C., Geyer, C.: Artificial intelligence and the public sector—applications and challenges. Int. J. Public Adm. **42**(7), 596–615 (2018)
36. Madan, R., Ashok, M.: AI adoption and diffusion in public administration: a systematic literature review and future research agenda. Govern. Inform. Quart. **40**(1), 101774 January (2023) https://www.sciencedirect.com/science/article/pii/S0740624X22001101
37. Wal, Z.: Trends and drivers of public administration in the twenty-first century, Handbook on Corruption, Ethics and Integrity in Public Administration -, edited Graycar, Adam, Edward Elgar Publishing Limited, (2020) https://book4you.org/book/5679954/f5fc1b
38. Maia, A.: Ética e Integridade na Governação Pública – A corrupção e a fraude, e sua prevenção. In: Maia, A. (Ed.) Ética e Integridade na Vida Pública, Almedina. Lisbon. (2020)
39. Savater, F.: Convite à Ética. Fim de Século, Lisbon (2008)
40. Amundsen, I.: Introduction to Public Sector Ethics: In: Public Sector Ethics, Amundsen, Inge & Andrade, Vicente P. (Eds.). Universidade Católica de Angola, ResearchGate. (2009), https://www.researchgate.net/publication/275968636_Introduction_to_Public_Sector_Ethics_Introducao_a_Etica_no_Sector_Publico

41. Pereira, L.M., Lopes, A.B.: Machine Ethics: From Machine Morals to the Machinery of Morality. SAPERE vol. 53, Springer Nature AG, Cham, Switzerland. (2020)
42. Mozzicafreddo, J.: O papel dos Cidadãos na Administração Pública. In: Reforma do Estado e da Administração Pública. Gestionária, Instituto Superior de Ciências Sociais e Humanas, Lisbon. (2000)
43. OECD: Hello, World: Artificial Intelligence and its Use – the Public Sector. OECD Working Papers on Public Governance (2019) https://doi.org/10.1787/19934351
44. Soares, L.: M P: A Ética na Administração Pública. Instituto Superior de Ciências Sociais e Políticas, Lisbon (2014)
45. Lipovetsk, G.: Da leveza para uma civilização do ligeiro. Edições 70, Lisbon (2017)
46. Anderson, M., Anderson, S.: Introduction: Machine Ethics, IEEE Intelligent Systems (July/August), IEEE-org. (2006), https://ieeexplore.ieee.org/abstract/document/1667946?casa_token=NTsFmOCQ5xgAAAAA:a8eMjkNh6GM8e6Xz3UCJ0OSo5CXuhJtIb0IlG3MAVVs5zzAKRUp8XY4k1fflW6Z5y5KKqWG1ig
47. Organization Economic Cooperation Development: Recommendation of the Council on OECD Legal Instruments on AI (2019), https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449
48. Organization Economic Cooperation Development: Artificial Intelligence in Society, OECD (2019) https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm
49. Tavani, H.: Ethics & Technology – Controversies, Questions and Strategies for Ethical Computing. 4th edition, John Wiley (2013), https://futuresinitiative.org/futureofwork/wp-content/uploads/sites/216/2018/07/Ethics_and_-Technology-Controversies-Questions-and-Strategies.pdf
50. Nankervis, A., Connell, J., Montague, A., Burgess, J. (eds.): Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-1614-3
51. Schawb, K.: Shaping the Fourth Industrial Revolution. Cologny, Geneva (2018)
52. Domingos, P.: The Master Algorithm, Penguin Books (2017)
53. Kissinger, H.: Schmidt, Eric & Daniel, Huttenlocker: The Age of AI And Our Human Future. Little Brown, London (2021)
54. Antunes, H.S.: Direito e Inteligência Artificial. Universidade Católica Editora, Lisbon (2020)
55. Bostrom, N., Yudkowsky, E.: The Ethics of Artificial Intelligence, Cambridge Handbook of Artificial Intelligence, edited by Keith Frankish and William Ramsey, Cambridge University Press (2011) https://doi.org/10.1017/CBO9781139046855
56. Hagendorff, T.: The ethics of ai ethics: an evaluation of guidelines. Mind. Mach. **30**(1), 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8
57. Jobin, A., Ienca, M., Vayena, E.: Artificial Intelligence: the global landscape of ethics guidelines. Health Ethics & Policy Lab, ETH Zurich (2019), https://export.arxiv.org/ftp/arxiv/papers/1906/1906.11668.pdf
58. Turilli, M., Floridi, L.: The ethics of information transparency. Ethics Inf. Technol.**11**, 105–112 (2009) https://link.springer.com/article/10.1007/s10676-009-9187-9
59. Fernandes, P., Santos, F., Lopes, M.: Norms for Beneficial A.I.: A Computational Analysis of the Societal Value Alignment Problem. INESC-ID, Instituto Superior Técnico, Universidade de Lisboa. https://doi.org/10.3233/AIC-201502 (2019)
60. Rawls, J.: A Theory of Justice. Revised edition, Harvard University Press and Google Books. (1971)
61. Grace, J., Bamford, R.: AI Theory of Justice: Using Rawlsian Approaches to Better Legislate on Machine Learning, Government, School of Advanced Study – University of London (2020), https://journals.sas.ac.uk/amicus/article/view/5161
62. Thomsen, K.: Ethics for Artificial Intelligence, Ethics for All. J. Behav. Robot. 359–363 (2019). https://www.degruyter.com/document/doi/10.1515/pjbr-2019-0029/html

63. Rahwan, I.: Society-in-the-Loop: Programming the Algorithmic Social Contract. Ethics and Information Technology 20 (2018), https://link.springer.com/article/10.1007/s10676-017-9430-8
64. Leben, D.: A Rawlsian algorithm for autonomous vehicles. Ethics Inf. Technol. **19**(2), 107–115 (2017). https://doi.org/10.1007/s10676-017-9419-3
65. European Parliament: The ethics of artificial intelligence: Issues and initiatives, European Union. (2020), https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf
66. Kant, I.: Fundamentação da Metafísica dos Costumes, Edições 70, Lisbon. (2007)
67. Jonas, H.: El Princípio de Responsabilidad – Ensayo de una Ética para la Civilización, Herder. (1995)
68. Alexandre, A.C., Pereira, L.: European Union Fosters Ethical AI in the Public Administration. In: de la Iglesia, D.H., de Paz Santana, J.F., López Rivero, A.J. (Eds.) New Trends in Disruptive Technologies, Tech Ethics and Artificial Intelligence. DiTTEt Advances in Intelligent Systems and Computing, vol. 1430, Springer. (2022). https://doi.org/10.1007/978-3-031-14859-0_9
69. 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 (2021). https://eur-lex.europa.eu/legal content/EN/TXT/?uri=CELEX:52021PC0206
70. MacCarthy, M., Propp, K.: LAWFARE, Home Hard National Security Choices, Law Fare (2021). https://www.lawfareblog.com/machines-learn-brussels-writes-rules-eus-new-ai-regulation
71. van Noordt, C., Misuraca, G.: Evaluating the impact of artificial intelligence technologies in public services: Towards an assessment framework.In: ACM International Conference Proceedings Series (2020), https://doi.org/10.1145/3428502.3428504
72. van Noordt, C., Misuraca, G.: Exploratory insights on artificial intelligence for government in Europe. Social Sci. Comput. Rev. 40(2), 426–444, (2020) https://journals.sagepub.com/doi/abs/10.1177/0894439320980449
73. Tornatzky, L.G.,Fleischer, M.: The Processes of Technological Innovation. Lexington Books (1990). https://archive.org/details/processesoftechn0000torn/mode/2up
74. Eveland, J., Tornatzky, L.G.: Technological Innovation as a Process. The Processes of Technological Innovation. Lexington Books (1990). https://www.researchgate.net/publication/291824703_Technological_Innovation_as_a_Process
75. Desouza, K.C., Dawson, G., Chenok, D.: Designing, developing, and deploying artificial intelligence systems: lessons from and for the public sector Bus. Horizons **63** (2), 205–213 (2020)
76. European Parliament: Regulation (EU) 2016/679 of 27 April 2016, General Data Protection Regulation. (2016) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679
77. Pencheva I., Esteve, M., Mikhaylov, S.J: Big data and AI - a transformational shift for government: So, what next for research?. Public Policy Adm. **35**(1), 24−44 (2020) https://doi.org/10.1177/0952076718780537
78. Regan, P.M., Maschino, K.: Artificial intelligence and its impact in public administration. In: Shark, Alan R. (Ed.), National Academy of Public Administration. (2019) https://napawash.org/uploads/Academy_Studies/9781733887106.pdf
79. Gstrein, O. J., Bunnik, A., Zwitter, A.: Ethical, Legal and Social Challenges of Predictive Policing, Catholic Law Review, University Medical Center, Groningen (2019), https://pure.rug.nl/ws/portalfiles/portal/117795047/SSRN_id3447158.pdf
80. Berryhill, J., Heang, K.K., Clogher, R., McBride, K.: Hello, World: Artificial Intelligence and its Use – the Public Sector. OECD Working Papers on Public Governance (2019) https://doi.org/10.1787/19934351

81. Watkins, I.: The utility of level of service inventory - revised (LSI-R) assessments within NSW Correctional environments. (2011). https://www.semanticscholar.org/paper/The-utility-of-level-of-service-inventory-revised-Watkins/550f2eed8f440b4ba0acc459b5558f99086fe666
82. O'Neil, C.: Weapons of Math Destruction. Penguin Books, London (2019)
83. Fry, H.: Hello World: Being Human in the Age of Algorithms Hardcover – September, Transworld Publishers Ltd. (2018)
84. Andrews, D., Bonta, J.: The Psychology of Criminal Conduct. Routledge, 6th edition. (2018)
85. Roth Davies: What is the effect of a defendant's refusal to participate the LSI-R? (2023), https://www.rothdavies.com/criminal-defense/frequently-asked-questions-about-criminal-defense/sentencing-outcomes/what-is-the-effect-of-a-defendants-refusal-to-participate-in-the-lsi-r/
86. Alexandre, A.C.: A Inteligência Artificial é discriminatória? (II Parte), O preconceito racial presente em decisões judiciais coadjuvadas pelos sistemas artificiais "inteligentes", Jornal de Monchique, edition 470, 28th of January (2023)
87. Leslie, D.: Understanding artificial intelligence ethics and safety A guide to AI ethics, including responsible design and implementation of AI systems in the public sector. The Alan Turing Institute, London. (2019). https://www.turing.ac.uk/sites/default/files/2019-08/understanding_artificial_intelligence_ethics_and_safety.pdf
88. Otero, P.: Manual de Direito Administrativo, vol. 1. Editora Almedina, Lisbon (2013)
89. Council of Europe: Ad Hoc Committee on Artificial Intelligence Policy Development Group, Artificial Intelligence in Public Sector (2021). https://rm.coe.int/cahai-pdg-2021-06-2779-3226-6755-v-1/1680a29927
90. Ferrer, X., van Nuenen, T., Such, J.M., Coté, M. Criado, N.: Bias and Discrimination in AI: a cross-disciplinary perspective. IEEE Technol. Soc. Mag. **40**(2), (2021), https://ieeexplore.ieee.org/document/9445793
91. Committee on Standards in Public Life: Artificial Intelligence and Public Standards Life Artificial Intelligence and Public Standards - A Review by the Committee on Standards in Public Life, UK (2020), https://www.assets.publishing.service.gov.uk/government/uploads/system/uploads/attachmentdata/file/868284/Web_Version_AI_and_Public_Standards.PDF
92. Brynjolfsson, E., McAfee, A.: O negócio da inteligência artificial, Inteligência Artificial, Harvard Business Review. Actual Editora, Lisbon (2019)
93. Harrison, T.M., Luna-Reyes, L.F.: Cultivating trustworthy artificial intelligence in digital government. Social Sci. Comput. Rev. **40**(2), 494–511. (2022), https://journals.sagepub.com/doi/abs/10.1177/0894439320980122
94. Olsen, H.P., Slosser, J.L., Hildebrandt, T.T., Wiesener, C.: What's in the Box? The Legal Requirement of Explainability. In: Computationally Aided Decision-Making in Public Administration. Faculty of Law, University of Copenhagen, Elsevier (2019). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3402974
95. Janssen, M., Brous, P., Estevez, E., Barbosa, L.S., Janowski, T.: Data governance: Organizing data for trustworthy artificial intelligence Government Information Quarterly, **37**(3), 101493 (2020), https://pure.tudelft.nl/ws/portalfiles/portal/85607215/1_s2.0_S0740624X20302719_main.pdf
96. Whittaker, M., et al.: AI Now Report 2018. AI Now Institute, New York University. (2018). https://ainowinstitute.org/AI_Now_2018_Report.pdf