ISSN 2281-5147

POLICING THE AI ADJUDICATOR: THE STUDY OF ALGORITHMIC ACCOUNTABILITY FOR AUTOMATED DECISION MAKING SYSTEMS IN THE PUBLIC SECTOR.

Mitisha Gaur*

Abstract

AI is ubiquitous in public and private sectors used for optimizing tasks through complex data analysis. While the technology is promising, its use in high-risk domains raises concerns about trust, fairness, and accountability. This chapter analyzes AI backed automated decision-making systems being used by public authorities and advocates for a strict governance framework based on meaningful transparency, risk management and algorithmic accountability practices focused on safeguarding fundamental rights and upholding the rule of law by adhering to the principles of natural justice.

*Mitisha is a Marie Skłodowska-Curie Action's fellow working as an Early Stage Researcher (Law) with the Legality Attentive Data Scientists (LeADS) Project funded under the EU's Horizon 2020 Framework. The primary focus of her research is the use and regulation of high-risk artificial intelligence systems deployed in adjudication environments such as those in courts, regulatory bodies, government departments etc.; during her research Mitisha is focused on studying transparency and legal viability vis-à-vis AI systems and the social impact these systems have when deployed in adjudication environments. She is based at the Lider Lab, Scuola Superiore Sant'Anna, Pisa (Italy).
Mitisha.Gaur@santannapisa.it
This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

Table of Contents

Keywords

Algorithmic Accountability – Meaningful Transparency – Automated Decision Making– Public Authorities – Artificial Intelligence

1. Introduction

Artificial intelligence has turned into a seemingly ubiquitous presence with its use spanning over multiple sectors such as energy, finance, education, healthcare, navigation and public administration. The central appeal of using AI based technology (AI Systems) lies in the purported public sentiment around its superintelligence. The primary driver of this superintelligence is associated with the ability of AI Systems to

ISSN 2281-5147

identify patterns within a dataset and generate insights by using analytical techniques rooted in statistical analyses, recognition of recurring patterns, mathematical computations etc. These techniques guide optimisation efforts for various activities across sectors. For example: the use of AI based prediction analytics can help in the optimisation of energy load across power grids by harnessing user data to decide where electricity is to be supplied in order to be compatible with the user requirements that vary across homes, industries and commercial buildings. Similarly, the use of these prediction based analytics fuelled by AI Systems has also permeated more dynamic and sensitive fields such as the financial sector where AI systems are used by banks for the assessment of credit applications, within public administrations to disburse government subsidies to persons eligible under welfare schemes and also by law enforcement authorities in order to decipher criminal activities in areas with high criminal activities. The governance framework applicable to an AI System is determined by the level of risk which may be associated with the AI System. The classifications for the levels of AI risk adopted by the European Union's (EU) AI Act which is the primary legislation governing AI systems across the EU, are divided in four broad categories, namely (1) Unacceptable Risks: AI systems marked for unacceptable risk are prohibited from being used and include AI Systems acting as social scoring systems used by financial institutions to evaluate candidates for their creditworthiness based on behavioural data regarding spending habits, credit history etc., AI Systems which aim to manipulate children or other vulnerable groups such as emotional manipulation through the use of virtual assistants, the use of AI Systems for real-time remote biometric processing such as emotion recognition of individuals in work spaces etc.; (2) High-Risk: The tasks performed by AI systems in circumstances which may have a significant and (potentially) harmful impact on the quality of life as well as the freedoms and liberties enjoyed by human beings are classified as high-risk tasks. Consequently, the AI Systems used to perform, augment or assist in the performance of any such high-risk tasks are termed as High-risk AI Systems. These include the use of AI systems for law enforcement functions such as those focused on evaluating the viability of evidence in the course of investigation or those used to evaluate the risk of a person becoming the victim of criminal offences etc., the performance of public administration functions such as to evaluate the eligibility of applicants for public benefits such as welfare benefits, healthcare assistance and associated services; (3) Limited Risk: These includes chatbots used in

ISSN 2281-5147

customer service and AI Systems with capabilities to create deepfakes,; and lastly (4) Minimal Risk: These include AI Systems that are used to perform low-risk functions such as AI systems acting as spam filters, writing and text editing tools etc.

The regulatory matrix under the AI Act varies across the 4 risk levels namely- (1) the AI systems exhibiting unacceptable risk are prohibited from being used; (2) the ones exhibiting high-risk are bound by a comprehensive set of legal obligations which include periodic and event-based compliances that are associated with both the technical and organisational requirements focused on use of high-risk AI systems such as risk assessment and mitigation, issuance of instructions of use, fundamental right impact assessment, conformity assessment, technical documentation, record-keeping etc.; (3) AI systems with low risk are bound by minimal reporting requirements and finally, (4) the AI Act exempts the use of AI systems with minimal risk from its purview, however with the increase in the use of generative AI tools, this may change.

The two-fold regulatory obligations, namely: technical and organisational, that are imposed on relevant stakeholders engaged with high-risk AI Systems, which include providers of AI Systems i.e. entities that develop and subsequently license a high-risk AI System and a deployer who may be a natural or a legal person such as an organisation, company, public authority, that uses an AI System to perform functions.

This chapter is focused on the use of AI Systems by government departments such as taxation authorities, family and child welfare departments etc as well as by judicial authorities such as courts, tribunals etc (collectively referred to as Public Authorities). The use of AI systems by Public Authorities has a direct impact on the health, safety and fundamental rights of the decision-subjects. The acknowledgement of the risk associated with the use of AI systems in this domain is also reflected in the AI Act's classification of an AI system used by Public Authorities to assist in performing sensitive tasks such as the dispensing of public welfare, assist judges in researching and interpreting facts etc. as a high-risk AI system.

ISSN 2281-5147

2. Automated Decision Making Systems and Public Authorities: Preserving the Adjudicatory Fabric.

There are multiple applications of AI Systems within Public Authorities, however for the purposes of this chapter, the central focus lies on the use of AI Systems in their capacity as automated decision-making systems (ADMS). These ADMS may be machine learning based statistical tools which provide quantifiable indications to the user such as rate of successful resolution of a legal dispute (whether in favour of the petitioner or the defendant) based on a given set of facts or provide a risk based scoring associated with the applications they process such as the application to request public welfare funds.

These types of inputs by the ADMS have a material impact on the manner in which the user of the ADMS views the applications presented to them. Another popular ADMS tool is the newer generative AI Systems (GenAI) such as the famous large language models ChatGPT and Gemini, that are backed by natural language processing technology and may be designed to provide answers to the questions a user may pose to it. The mimicking of human behaviour by GenAI may lead to increased trust between the deployer and the AI System, however, numerous investigations have observed flaws within the GenAI system which has been observed to produced fictitious answers to queries posed to it. This phenomenon has been termed as *hallucinations*. A prominent example is when ChatGPT constructed a fictional caselaw to support its answer to a question placed before it.

Against this backdrop, the efforts to govern the development and use of High-risk ADMS by Public Authorities, a crucial factor to consider is the methodology of use associated with such an ADMS.

The decision making process across Public Authorities is divided into four (4) key stages: (1) acquisition of information based on which a decision has to be made; (2) the analysis of the information; (3) selection of decision based on the analysis of information and lastly; (4) the implementation of the selected decision.

ISSN 2281-5147

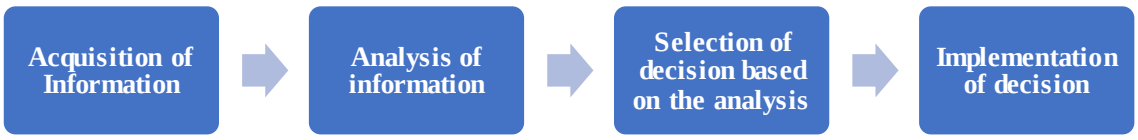


Figure 1: Stages Of Decision Making Within A Public Authority

The manner in which ADMS is used by Public Authorities is materially affected by the stage of decision making within which such ADMS is deployed, as the instructions of use, associated risk as well as transparency requirements will differ. Prior to delving into the technical and organisational constraints attached to the ADMS, it is crucial to understand the context within which the ADMS is deployed by the Public beyond the simplistic reduction of "to assist in decision making". This assistance can be understood as a plethora of tasks and ranges in the level of automation associated with it. This can be at the performance of simple tasks such as the streamlining of information focused on expediting tasks (low level of automation), the assistance in writing a judgement (moderate level of automation) or the choosing of a decision based on historical data on behalf of the Public Authority (high level of automation).

The endeavour of decision making by a Public Authority is guided by the balancing of many crucial duties and associated responsibilities shouldered by such Public Authorities. These include the duty to uphold and protect the rights of citizens, the responsibility to exercise the rule of law, and the duty to adhere to the principle of natural justice. These principles of natural justice are the very fulcrum of robust judicial systems (such as courts and tribunals) and quasi-judicial systems (such as

ISSN 2281-5147

government department and boards providing licenses and administrative rulings based on legal statutes) across the world. These principles of natural justice are as follows- (1) The adjudicating authority must not be biased whether in favour of or against the persons seeking legal recourse; (2) Pronouncing of a reasoned order by the adjudication authority; (3) Absence of unjustifiable delay in adjudication; (4) Ability of a person to make legal representation in front of the adjudication authority and; (5) Adequate notice to be provided to a person to prepare for the legal proceedings initiated against them. Consequently, the material impact awarded by the principles of natural justice upon the decision making processes by Public Authorities is two-fold: (1) allows Public Authorities to build precedent and; (2) the adherence to the principles of natural justice allows for examination of the judgements of Public Authorities by supervising authorities such as superior courts with appropriate jurisdiction on the subject matter.

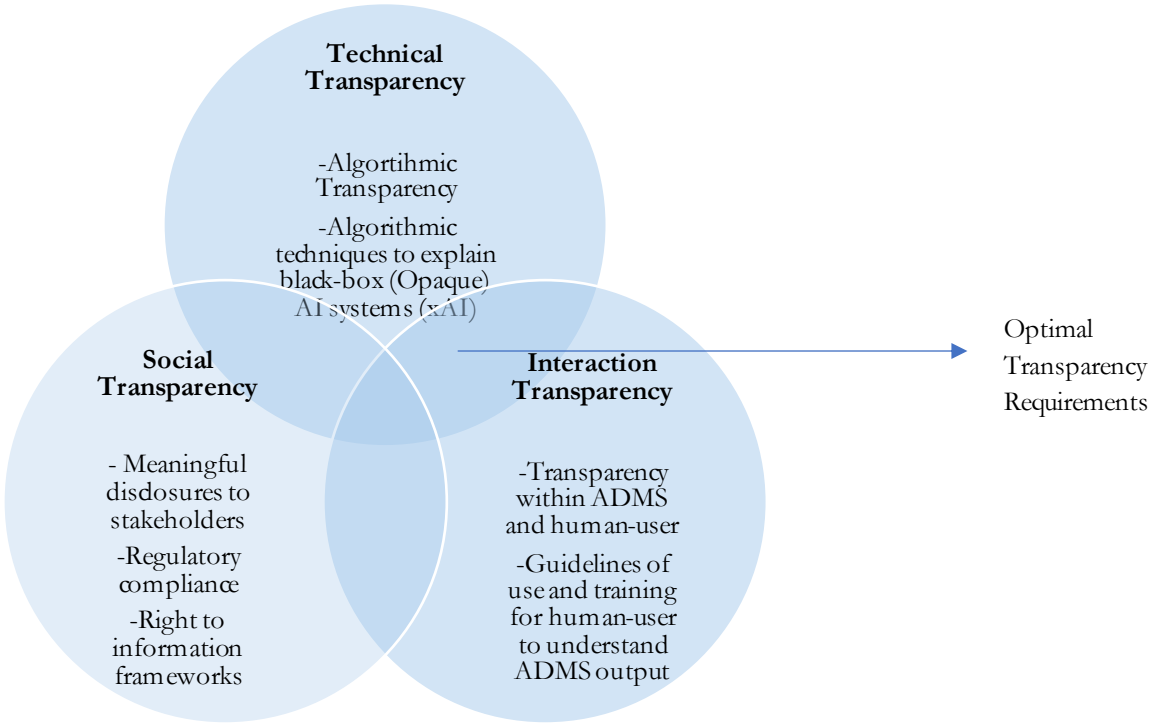
Against this backdrop, this chapter focuses on three crucial issues associated with the use of ADMS by Public Authorities, namely (1) How to develop an ADMS which can be safely deployed within a Public Authority to assist in carrying out judicial and quasi-judicial functions?; (2) How to ensure that the ADMS is deployed safely within a Public Authority and is being used in the correct context? and lastly; (3) How to protect the persons subjected to these decisions from adverse effects of the ADMS use by Public Authorities?

3. The Transparency Triad: Informing ADMS within a Public Authority

The common thread across these three challenges (as discussed in Section 2) is that by virtue of the expectation of transparency from Public Authorities, the decisions of Public Authorities as well as any associated information which aides and assist such decision making is subject to explanation under the mechanisms of the access to information framework, through which an individual can seek specific information from Public Authorities. The combined reading of the legal requirements, duties and responsibilities as well as the explanation requirements associated with Public Authorities, transparency associated with decision making emerges as a central theme. Therefore, it is evident that ADMS deployed within Public Authorities must also comply with necessary transparency requirements. The transparency mandate

ISSN 2281-5147

associated with the use of ADMS within an Public Authority is tri-fold and comprises of (1) Technical transparency: This form of transparency is associated with the inner workings of the ADMS and the ability of the ADMS to provide a meaningful explanation about the output it produces; (2) Interaction Transparency: This form of transparency is associated with the ability of the human-user of an ADMS to adequately understand the inner workings of the ADMS and make an informed decision as to whether or not the output produced by the ADMS must be relied upon; and finally (3) Social Transparency: This pertains to the sharing of information (such as the underlying technology, the trustworthiness and safety) vis-à-vis the ADMS by the Public Authority with relevant stakeholders such as citizens, persons subjected to the decision in which an ADMS was involved, regulatory bodies etc. The triad of these three types of transparency related requirements creates the optimal transparency requirements for a Public Authority deploying ADMS.



ISSN 2281-5147

Figure 2: Tri-fold Transparency Mandate For Public Authorities Deploying ADMS

4. The Ever-Shifting Landscape: Context and Public Authority in ADMS

The deployment of an ADMS within a Public Authority raises crucial questions, particularly concerning the role of context of use and how it informs decision making. This also informs the manner in which the ADMS may be used in a safe and trustworthy manner, while protecting the interests of the developers, deployers as well as decision-subjects.

Contextual clarity during the development and the deployment of ADMS in a Public authority is crucial. This context is far ranging from the stage of decision making within which an ADMS is deployed to the decision-subjects and whether they are minors or members of a vulnerable class, the level of technical expertise showcased by the human-user relying upon the computations of the ADMS such as the level of AI literacy which dictates their ability to adequately comprehend and rely on the decision-outcome of an ADMS. Another crucial constraint is that context within an ADMS is ever changing and the technical infrastructure of the ADMS must evolve accordingly to accommodate it. For example: Changes in legal regulation or social norms may directly impact the context within which an ADMS must be deployed or relied upon.

Another crucial layer of contextual clarity within an ADMS is a defined purpose for which the ADMS is being used by an organisation. Is it designed to automate routine tasks like eligibility checks or delve into complex areas like parole decisions? The level of automation and the associated degree of human oversight may vary significantly depending on this background. For instance, an ADMS flagging the possibility of fraudulent tax returns might have a lower human oversight threshold compared to a system assessing child custody disputes.

The impact of an ADMS decision directly depends on who it affects, therefore a crucial consideration is that any decisions which directly or indirectly impact minors or vulnerable populations such as migrants and refugees, people with disabilities, racial and ethnic minorities etc. necessitate a more nuanced understanding of context.

ISSN 2281-5147

Public authorities rely on qualified users to interpret and implement ADMS outputs. The level of technical expertise these individuals possess forms another crucial layer of context, as noted previously. Therefore, comprehensive training becomes paramount to ensure that the users of the ADMS can understand the limitations of the ADMS such as hallucinations, presence of bias in data which leads to algorithmic discrimination, lack of transparency, improper data quality and diversity (which may again circle back to the problem of bias within the dataset used to develop AI systems) and can critically analyse its recommendations before implementing the same. This is also pertinent to combat cases of automation-bias within human-users where users are observed to overly rely on the decision outcome produced by an ADMS. In some cases, additional data or context not captured by the system might be crucial for the final decision, therefore it is crucial that in the absence of the same the user of the ADMS possesses adequate levels of AI literacy to spot the challenges associated with the ADMS and take necessary steps. Therefore, it has been noted that owing to these possible shortcomings, high-risk AI systems should not be deployed without meaningful human oversight.

The final, and perhaps most critical, aspect of context is its dynamic nature. Public policies, demographics, and technology evolve constantly, this is relevant more so when the subject matter is the use of AI Systems within Public Authorities. An ADMS designed for efficient distribution of unemployment benefits during an economic downturn might need adjustments during a period of low unemployment. Regular reviews and updates are essential to ensure the ADMS adapts its algorithms and data sets to reflect the ever-changing environment such that the ADMS may produce results relevant to the current societal norms and use based requirements.

Public authorities face a complex challenge in deploying ADMS effectively. Striking a balance between automation and human oversight, ensuring fairness for all decision-subjects, and continuously adapting the system to a dynamic environment requires a nuanced understanding of context. Therefore, as discussed previously, transparency and an adequate level of AI literacy are key here. Public authorities need to be transparent about how they are using ADMS and put in place mechanisms for individuals to challenge automated decisions.

ISSN 2281-5147

5. Public Authorities and Algorithmic Accountability: The Final Piece Of The ADMS Puzzle

The final crucial piece of the puzzle which focuses on the developing and deploying an ADMS within a Public Authority, is algorithmic accountability. Algorithmic accountability is the practice of holding the deployers of algorithms responsible for its effects. This inclusion of responsibility through algorithmic accountability has a direct impact on the manner in which the ADMS is not only developed and deployed within a Public Authority to augment decision making but also how it is perceived socially.

The ADMS is a technical component or a tool which is deployed within a social and organisational environment, therefore this interaction between the technical components as well as the social and organizations components creates an interdependent ecosystem referred to as a sociotechnical system (STS). The theory of STS is essentially an organisational development approach that focuses on the complexities associated with workflow within an organisation based on the interaction between social (such as persons, levels of education and technical skills), organizational (such as processes, timelines and task flows) and technical elements (such as hardware and software components) within an organisation (collectively referred to as the STS Stakeholders). The characteristics of an STS, which vary greatly from one organisation to another, impacts the interaction that a technical component such as an ADMS has in the face of contextual information which are driven by the social and organisational factors within an STS. To simplify it further, imagine a big system within the Public Authority, like a machine with many parts. To make all the parts work well together, they need clear rules. These rules cover how different stakeholders interact with the system and how the system itself works. Therefore, for optimal functioning these rules should be transparent, meaning not only must they be easy to understand but also that each stakeholder must know the rules applicable to itself as well as its counterparts and the information pertaining to whether such rules have been followed or not should be readily available. Also, there needs to be a dedicated person in charge, making sure everything runs smoothly (human oversight). Things get even more complex if this system works with other similar ones, like connecting different machines. In these cases, it's crucial to have clear rules and a clear chain of command to avoid confusion or problems.

ISSN 2281-5147

The concept of algorithmic accountability is closely associated with the adherence to the rules that govern an STS. Therefore, in order to truly ensure algorithmic accountability, it is crucial to divide the accountability frameworks based on subject matter. In keeping with this, the algorithmic accountability framework is divided into 4 main parts, namely (1) Technical Accountability; (2) Organisational Accountability; (3) Social Accountability and; (4) Regulatory Accountability.

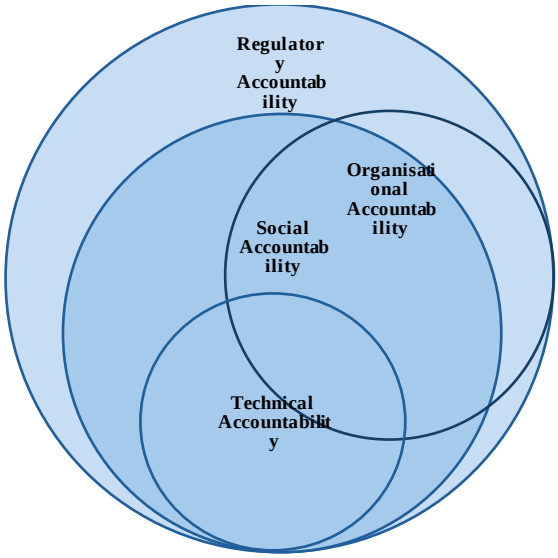


Figure 3: Interplay Between The Components Of Algorithmic Accountability

To ensure a fair and ethical application of ADMS within a Public Authority, a multi-pronged approach to accountability is crucial. Imagine a complex STS tasked with making critical choices such as the ADMS used within a Public Authority to investigate possible fraudulent activity vis-à-vis tax benefits and is required to function with transparency and fairness. This can be achieved through a layered framework encompassing each: technical, organizational, social, and regulatory accountability measures.

Technical accountability focuses on the inner workings of the ADMS, like ensuring the system is built well and uses reliable data. Meaningful and multi-faceted transparency, as explained in the previous sections, is key. These include deliberations such as "Can the users understand how decisions are made?" The presence of clear and meaningful explanations, even for those without technical expertise, are essential. Further, rigorous testing and audits are crucial to identify and eliminate bias before deployment of the ADMS within the Public Authority. This prevents the system from

ISSN 2281-5147

perpetuating social inequalities through skewed datasets based on which the ADMS may have been trained. Consequently, data governance becomes vital as well, focusing on pertinent questions such as "Where does the data come from?" and "How is the data used?" A robust technical accountability framework ensures that the data is accurate, complete, and collected ethically, with individuals having control over their personal information, is used by the ADMS. Finally, security and explainability are important. Strong cybersecurity measures protect the system from tampering, and the ability to meaningfully explain decisions helps identify errors and promotes fairness.

Organizational accountability focuses on responsibility within the STS itself, such as the assignment of clear roles within a complex environment. Defining roles and responsibilities ensures everyone working with the STS understands their part and more importantly, can be held accountable, in case of an adverse event such as ADMS led bias propagation. While ADMS helps to automate decision making tasks, human oversight, as noted previously, remains crucial. Human involvement through oversight allows for questioning decisions, considering the presence of contextual factors that the algorithm might not take into account, and maintaining alignment with ethical and legal principles associated with the use of high risk AI by Public Authorities such as ADMS. The ADMS user training provided by Public Authorities in conjunction with the developers of the ADMS, empowers those human-users interacting with the ADMS to understand its limitations and capabilities, as well as providing thorough instructions of use to the ADMS users which include risk escalation mechanisms, adverse output mechanisms etc. This training should be focused on empowering the human-user to analyse the ADMS outputs critically, identifying potential biases or errors. Risk management in these scenarios also remains crucial, as organizations implementing ADMS need a well-defined plan to identify and mitigate potential risks.

Further, social accountability empowers the public to hold institutions using ADMS accountable. Here, the focus also lies on meaningful transparency and encouraging public engagement and discussion. Public awareness ensures they understand how ADMS are used and its potential impacts on decision subject as well as the wider impact population that the societal fabric is comprised of. The core tenet of social accountability is transparency in communication, which focuses on the right to explanation vis-à-vis persons and their right to know how ADMS are used by Public

ISSN 2281-5147

Authorities which will consequently affect their lives. Additionally, it is to be noted that social accountability fuelled by public participation through public meetings such as townhouses, public consultations and conducting public polls allows for meaningful public involvement in the development and deployment stages of an ADMS. This has the ability to help in identifying potentially material issues before they arise and ensuring the system is designed ethically, in compliance with legal regulations and with social good in mind. Independent audits (by algorithmic watchdog organisations and citizen's rights groups) and reviews provide valuable insights and identify areas for improvement. Finally, grievance redressal mechanisms are essential for fairness and building public trust. Persons who are subjected to ADMS by the Public Authorities should have clear and accessible ways to challenge unfair or discriminatory decisions made by ADMS, which must be resolved within a stipulated timeframe.

Lastly, regulatory accountability sets the ground rules for the operation of an ADMS by a Public Authority to aide its efforts to perform material public functions, through establishing rules and regulations. Regulatory frameworks (such as the EU's AI Act) may be focused on defining clear expectations for fairness, transparency, and accountability, which may be imposed on the developers and deployers of the ADMS by means of regulatory compliance. Another useful solution may be the establishment of independent regulatory bodies, focused on overseeing the use of ADMS by Public Authorities, monitoring adherence to compliance and investigating potential breaches. The practice of impact assessments (which includes both algorithmic impact assessment (focused on the technical robustness of the ADMS) as well as the fundamental rights impact assessment (focused on the impact of the ADMS on the fundamental rights of the decision subjects) is a welcome and crucial tool, focused on the evaluation of potential risks ex-ante deployment, allowing for mitigation strategies to be put in place. These forms of ex-ante requirements are preferred over the enforcement of ex-post sanctions, in the case of high-risk AI such as the use of ADMS by Public Authorities since the degree of harm caused by a biased or faulty ADMS may cause material harm to the decision-subject which may not be mitigated through sanctions or monetary compensation. This brings us to sanctions and enforcement mechanisms such as operational injunctions enforced on ADMS until the algorithmic shortcomings are tackled, that are used to ensure accountability and deter misuse of the ADMS by Public Authorities.

ISSN 2281-5147

6. Conclusions

These four pillars of accountability are interconnected and work best when implemented in harmony. Technical measures ensure the fairness and transparency of the ADMS itself, while organizational measures establish clear roles and responsibilities for those deploying and using the system. Social accountability empowers the public through awareness, participation, and grievance redressal, and regulatory accountability sets the ground rules through regulation, independent oversight, and regulatory enforcement. By weaving these pillars together, we can ensure that ADMS are used responsibly, ethically, and with the public good at the forefront. This multifaceted approach allows us to build trust in the complex world of automated decision-making, ensuring it serves society effectively and fairly.

The primary roadblock in the investigations pertaining to the use of ADMS by Public Authorities is the seemingly opaque algorithms which are often found to be in use and consequently rupture the requirement of algorithmic transparency and the ability of users to perceive or explain (to the decision subjects) the inner workings of the ADMS on which the Public Authority relies. This opacity, in turn, while creates distrust in the minds of the decision subjects vis-à-vis the ADMS also provides a leeway to Public Administrations to dodge questions regarding inner mechanisms of the algorithms in use. Further, there has been an observed lack of internal training which leads to either the misuse, over-reliance or the unreserved mistrust regarding the use of the ADMS. Additionally, the observed pattern of the development and deployment of an ADMS for use by a Public Authority is that the same institution dons the hat for both the deployer as well as the developer, therefore there is no internal accountability or balance of powers within these two roles, which may lead to the unchecked perpetration of bias and lack of social as well as regulatory accountability in case of a misadventure at the hands of the Public Authority relying on an ADMS.

The benefits of using a high-risk AI System such as an ADMS cannot be considered in insolation with the responsibilities associated with the use of such tools. This becomes even more crucial in light of the fact that the deployer and user of the ADMS is a Public Authority, an organisation that wields immense power and its actions or

ISSN 2281-5147

inactions have a significant impact on the lives of people. Therefore, a holistic approach is required that spans across all STS Stakeholders (organisational, social and technical) while developing, deploying and using such an AI Systems, rooted in upholding meaningful transparency, promoting meaningful human oversight and keeping the two in check through use of algorithmic accountability measures.

7. Selected Readings

(1) Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. Regulation & Governance, 18(1), 3-32.

(2) Wang, A., Kapoor, S., Barocas, S., & Narayanan, A. (2024). Against Predictive Optimization: On the Legitimacy of Decision-making Algorithms That Optimize Predictive Accuracy. ACM Journal on Responsible Computing, 1(1), 1–45.

(3) Roehl, U. B. U. (2023). Automated decision-making and good administration: Views from inside the government machinery. Government Information Quarterly, 40(4), 101864.

(4) Mökander, J., & Axente, M. (2023). Ethics-based auditing of automated decision-making systems: Intervention points and policy implications. AI & SOCIETY, 38(1), 153–171.

(5) Chopra, A. K., & Singh, M. P. (2018). Sociotechnical Systems and Ethics in the Large. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 48–53.

(6) Ruckenstein, M., Lomborg, S., & Hansen, S. S. (2020, November). Re-humanising automated decision-making. In Workshop report from the ADM: Nordic Perspectives research network.

(7) Malgieri, G., & Pasquale, F. A. (2022). From Transparency to Justification: Toward Ex Ante Accountability for AI. SSRN Electronic Journal.

ISSN 2281-5147

(8) Diakopoulos, N. (2020). Accountability, transparency, and algorithms. The Oxford handbook of ethics of AI, 17(4), 197.

(9) Olsen, H. P., Hildebrandt, T. T., Wiesener, C., Larsen, M. S., & Flügge, A. W. A. (2024). The Right to Transparency in Public Governance: Freedom of Information and the Use of Artificial Intelligence by Public Agencies. *Digital Government: Research and Practice*, *5*(1), 1–15.

(10) Kaminski, M. E. (2023). Regulating the Risks of AI. Forthcoming, Boston University Law Review, 103, 22-21.

(11) Novelli, C., Casolari, F., Rotolo, A., Taddeo, M., & Floridi, L. (2024). AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act. Digital Society, 3, 13.

ISSN 2281-5147