Governance in the Age of Artificial Intelligence: Exploring the Impact on Transparency, Accountability and Rule of Law

Shaza Arif*

Abstract

Artificial Intelligence (AI) has diverse potentials. The increasing use of AI makes it imperative to examine its impact on an important societal aspect — Governance. The paper uses three UN-proposed indicators of good governance: i. transparency, ii. accountability and iii. rule of law to assess the potential impact of AI in future governance. The research uses qualitative research design based on open-ended semi-structured interviews with AI experts to study the expert opinion on the subject. The data has been analysed through thematic analysis to capture the risks and opportunities associated with AI integration in governance. The paper's findings reveal that the employment of AI in governance is going to considerably impact transparency, accountability and the rule of law. This becomes increasingly concerning for non-democratic contexts. There is a consensus that Explainable AI (XAI) could, to a certain degree, mitigate the dangers associated with the technology in the domain of governance by yielding a positive impact on the concerned indicators of governance. The paper proposes that while AI integration may become imperative for policymakers in the future, it is equally important to prioritise transparency, accountability and rule of law to ensure good governance.

Keywords: Governance, Artificial Intelligence (AI), Transparency, Accountability, Rule of law.

* The author is Senior Research Associate, Centre for Aerospace & Security Studies (CASS). Email: shaza.arif@casstt.com

Introduction

Emerging technologies mark the hallmark of advancements and progress in modern world. Amongst the diverse list of emerging technologies, Artificial Intelligence (AI) stands as one of the prominent enablers, powering different technologies. While there is no widely accepted definition of AI, it generally refers to machines performing tasks that require complex decision making usually inherent associated human beings. [1] This ability and unique characteristic provides unprecedented efficiency across different domains.

The promises of AI are enhancing its adoption widely, across both public and private sectors.[2] The penetration of algorithms has marked the onset of a new technological era where data-driven decision-making has become increasingly prevalent in various sectors as well as different aspects of daily life.

Amongst the expanding scope of applications of AI, governance stands as a notable example. The technology holds promise to impact governance systems around the world significantly. Assuming a prominent role in governance, AI can facilitate the decision making process, refine efficiency and reduce costs in the concerned sectors. Governance in the age of AI is a complex and multifaceted element that touches upon several aspects of society, including ethics, morals, politics, administration, law etc. At its core, it refers to the decision making processes and the manner in which human exercise power vis-à-vis AI.[3] The phenomenon involves applications such as facial recognition systems, algorithmic analysis to facilitate decision making and intelligent systems for enhanced public services. With the rise of AI in governance, there is a need to ensure that the decision making processes align well with the indicators of good governance. As states and policymakers increasingly turn to algorithms to make decisions and manage complex systems, it is imperative to explore the potential impacts of AI on various

[1] Haroon Sheikh, "Mission AI: The New System Technology," (Springer Nature: Geneva, 2023), 410.
[2] Kevin C. Desouza and Rashmi Krishnamurthy, "Chatbots move public sector toward artificial intelligence," *Brookings*, 2017, https://www.brookings.edu/articles/chatbots-move-public-sector-towards-artificial-intelligence/.
[3] Xinyue Hao , Emrah Demir and Daniel Eyers , "Beyond human-in-the-loop: Sensemaking between Artificial Intelligence and Human Intelligence Collaboration," *Sustainable Futures* 10, (2025): 101152, https://doi.org/10.1016/j.sftr.2025.101152

aspects of governance. There are several indicators to evaluate good governance proposed by various international organisations. In this context, United Nations (UN) has listed eight indicators of good governance.[4] To keep the scope of the research paper focused, the paper will take up only three indicators, i.e., i. transparency, ii. accountability and iii. rule of law. The analysis will study the corresponding relation of AI with each of the indicators to draw the broader impact on governance. Furthermore, while the opportunities offered by AI are immense, there are equally significant challenges that also require due deliberation. The study aims to explore the conceptual understanding and practical applications, offering insights on how governance systems can evolve in future in light of technologies such as AI. The two research questions will be addressed in the article: a) How does the integration of AI in governance affect transparency, accountability and rule of law? b) What are the key implications of AI integration in governance? The paper analyses the role of AI in governance vis-à-vis each indicator separately. The paper then puts together the way forward. Lastly, the paper concludes by reflecting on the key takeaways.

The study uses qualitative design to conduct open-ended interviews. The sample has been selected via purposive sampling. The interviews were conducted in virtual/online mode. Ethical considerations were well considered during the process. There was no recording of the interview, rather note-taking was used to record the responses. The paper uses socio-technical theory, and the level of analysis is the state and governance systems. Data has been analysed using thematic analysis.

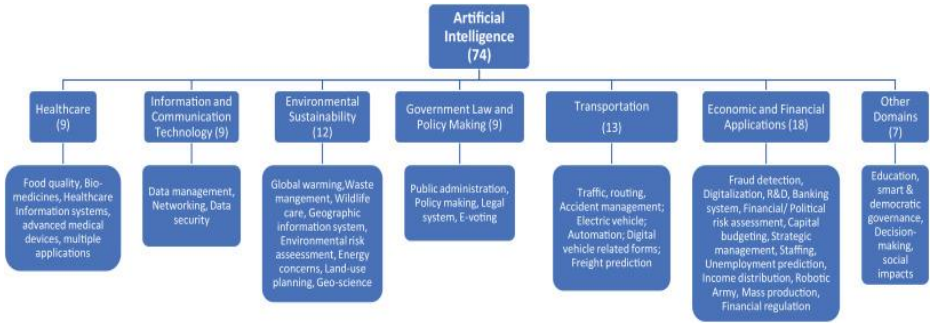
Literature Review

AI has led to innovative applications that profoundly impact our lives despite being barely 60 years old.[5] With the advent of the technological era, it has become inevitable for the government to adopt more efficient governance

[4] United Nations, "Global and National Leadership in Good Governance," https://www.un.org/en/chronicle/article/global-and-national-leadership-good. governance#:~:text=The%20most%20cited%20definition%20has,follows%20the%20rule%20of%20law

[5] Gagan Deep, Anshita Yadav and Ritika Chopra, "Artificial Intelligence and Effective Governance: A Review, Critique and Research Agenda," *Sustainable Futures* 2, no.3 (2020): 100004, https://doi.org/10.1016/j.sftr.2019.100004

mechanisms. As per the study of Deep, Yadav and Chopra, AI has the potential to bring transformative changes to various government sectors, as illustrated in the figure below.[6]



Source: *Gagan Deep, Anshita Yadav, and Ritika Chopra, "Artificial Intelligence and Effective Governance: A Review, Critique and Research Agenda," Sustainable Futures 2, no.3 (2020): 100004, https://doi.org/10.1016/j.sftr.2019.100004.*

Refinements in Machine Learning (ML) techniques, coupled with the supporting back-end data infrastructures, are likely to expand the role of algorithms in governance. The faster the progress, more tangible their impact on governance would be. The implications for transparency, accountability and the rule of law in governance have been discussed in the existing literature. It is noteworthy that accountability and rule of law emphasised by Francis Fukuyama in his book, *"Origin of Political Order"* are pivotal components of a modern political order and strengthening of various institutions.[7]

Transparency is recognised as an important aspect of good governance.[8] According to a study by Felzmann et.al, transparency is defined as openness, accessibility, visibility and interpretability of data/information.[9] Transparency is essential to ensure that the decisions

[7] Francis Fukuyama, *The Origins of Political Order: From Prehuman Times to the French Revolution* (New York: Farrar, Straus and Giroux, 2012), 22.
[8] Murat Jashari and Islam Pepaj,"The Role of the Principle of Transparency and Accountability in Public Administration," *Acta Universitatis Danubius* 10, no.1 (2018):60-70, https://core.ac.uk/download/pdf/229465497.pdf.
[9] Heike Felzmann, Fosch Villaronga, Catherine Lutz, and Tamò-Larrieux, "Robots and Transparency: The Multiple Dimensions of Transparency in the Context of Robot Technologies," *IEEE Robotics and Automation Magazine* 26, no.2 (2019): 71-78,

made by the policy makers are grounded in sound evidence rather than biased decisions.[10] According to a study by Nick Bostrom and Eliezer Yudkowsky, there is a widespread view that algorithmic decision making involving social functions should be predictable and transparent to those they govern.[11] In the United States (U.S.), as per Executive Order 13960, for governments transparency should embody three basic characteristics.[12] Information should be understandable, traceable and available to stakeholders. The challenge of transparency in the age of AI governance has been identified in various government documents, including "The Report on the National Artificial Intelligence Research and Development Strategic Plan: 2019."[13]

In the scholarly work of Chen & Wang, concern has been raised that AI could significantly reduce transparency in governance regarding data, data analysis, and decision making.[14] Coglianese and David Lehr echo similar concerns asserting that flexible nature of ML algorithms and their capacity to identify predictively useful patterns in large data sets can provide them with a considerable edge in decision making yet this comes with a cost to the transparency.[15] Further explaining the concept, two important concepts of

---

https://ieeexplore.ieee.org/abstract/document/8684252?casa_token=SYzA52F99TAAAAAA:REMYEPE8EfXFdnQ8W5k2bAmMPFvKKemep0Cm9vYnHLwF5s5BYY8PdE2qHvCLn8HtSJ3LWS5Xhg

[10] Stefan Larrson and Fredrik Heintz, "Transparency in Artificial Intelligence," *Internet Policy Review* 9, no.2 (2020): 1-16, https://policyreview.info/concepts/transparency-artificial-intelligence.

[11] Nick Bostrom and Eliezer Yudkowsky, "The Ethics of Artificial Intelligence," in *Artificial Intelligence Safety and Security*, ed., Roman V. Yampolskiy (London: Chapman and Hall/CRC, 2018),66.

[12] Federal Registrar, "Executive Order 13960: Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government," December 3, 2020, https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government.

[13] *The National Artificial Intelligence Research and Development Strategic Plan: 2019,* Report (Washington DC: Networking and Information Technology Research and Development, 2019), 25, https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf.

[14] Yu-Che Chen, Michael J. Ahn, and Yi-Fan Wang, "Artificial Intelligence and Public Values: Value Impacts and Governance in the Public Sector," *Sustainability* 15, no. 6 (2023): 4796-4812, https://www.mdpi.com/2071-1050/15/6/4796.

[15] Cary Coglianese and David Lehr, "Transparency And Algorithmic Governance," *Administrative Law Review* 71, no. 1 (2019): 1-56, https://www.jstor.org/stable/pdf/27170531.pdf?casa_token=fTZx4e68qo0AAAAA:urFP

transparency are highlighted. *Fishbowl Transparency* explains "What is happening" and *Reasoned Transparency* reveals "why something is happening." Governance using AI will fail to address both types of transparency.[16] According to Bahrevar & Khorasani AI is used in a way that promotes transparency in governance requires a multi-faceted approach.[17]

Several measures including data selection, preparation procedures and active and passive monitoring have been proposed to enhance transparency across different sectors.[18] Felzmann, Villaronga, Lutz and Aurelia further emphasise that clear guidelines and regulations should be established by governments for the employment of AI in decision making processes.[19] Algorithmic accountability, guidelines for data privacy and transparency in decision making was also identified as important steps in this regard. Daniel et al's study further suggests that human oversight and feedback loops and use of understandable algorithms is highly recommended.[20] Felzmann et al also add that public participation and engagement in the development and use of AI should be prioritised by the government.[21] This may comprise engagement with civil society for feedback and accumulation of diverse perspectives for the AI systems.

Accountability is also an important indicator of good governance. According to Hohma et al., accountability is defined as the relation between

6wGJXuRj1Dc_1e-RpkKpiXcIfLKSe9F8-MZbtgryURfRt4fVzVLYnxSlhVjz1MUkQONHGFjT7PxqxolJJjRi39RvSNdsSOPEUwV0eDr1w3VC88c.

[16] Coglianese and Lehr, "Transparency And Algorithmic Governance."

[17] Concordia University, "Accountability and Transparency in AI Systems: A Public Policy Perspective," November 2021, https://www.concordia.ca/content/dam/ginacody/research/spnet/Documents/BriefingNotes/AI/BN-105-The-role-oF-AI-Nov2021.pdf

[18] "Accountability and Transparency in AI Systems."

[19] Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz and Aurelia Tamò-Larrieux, "Towards Transparency by Design for Artificial Intelligence," *Science and Engineering Ethics* 26, (2020): 3333-3361, https://link.springer.com/article/10.1007/s11948-020-00276-4

[20] Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, et al., "The Ethics of Algorithms: Mapping the Debate," *Big Data & Society* 3, no. 2 (2016): 2053951716679679, https://journals.sagepub.com/doi/full/10.1177/2053951716679679.

[21] Felzmannm, Villaronga, Lutz and Larrieux, "Towards Transparency by Design for Artificial Intelligence," 3351.

a particular actor and the group to which the actor holds the responsibility to justify its actions.[22] This factor acts as a performance evaluator of the concerned stakeholder. As per Konigs, accountability is seen as one of the important gaps existing in AI and its integration into society.[23] De Sio & Mecacci also agree to the preceding assertion and stress that a complex combination of human agents and technical systems may lead to serious accountability gaps.[24] These accountability gaps could be identified as moral accountability gaps, public accountability gaps and active accountability gaps. The issue requires timely measures to address the accountability challenges. Regardless of how advanced the AI systems become; they would always be subjected to accountability challenges. Cussin points out that training data, learning models and source codes are all distinctive and unique elements that would be influencing algorithmic decision making and would not be clear to the public.[25] The development of algorithms by private companies can further add to the existing complications.[26] In their book, Stahl Schroeder and Rodrigues refer to the concept of "many hands" where involvement of different entities lead to complications.[27] The view was also emphasised that accountability lies not only at the point of execution but also at the point of inception. The level of human involvement in the accountability loop also remains a towering challenge for the decision makers. A common finding among the studies is that the automation of several tasks could reduce the administrative burden on human resource. However, the utility will be greatly diminished when

[22] Ellen Hohma, Auxane Boch, Rainer Trauth and Christoph Lutge, "Investigating Accountability for Artificial Intelligence through Risk Governance: A Workshop-Based Exploratory Study," *Frontiers in Psychology* 14 (2023): 1-17, https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1073686/full#ref6.

[23] Peter Konig, "Artificial Intelligence and Responsibility Gaps: What is the Problem?," *Ethics and Information Technology* 24, no. 36 (2022):24-36, https://doi.org/10.1007/s10676-022-09643-0.

[24] Fillipo Santoni de Sio and Giulio Mecacci, "Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them," *Philosophy & Technology* 34, no.4 (2021):1057-1084, https://link.springer.com/article/10.1007/s13347-021-00450-x#Fn1.

[25] Jessica Cussin, "Towards AI Security," (Paper, Centre for long-term Cybersecurity, 2019), https://cltc.berkeley.edu/wp-content/uploads/2019/02/CLTC_Cussins_Toward_AI_Security.pdf

[26] Cussin, "Towards AI Security."

[27] Bernd Carsten Stahl, Doris Schroeder and Rowena Rodrigues, *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges* (New York: Springer Briefs, 2023), 69.

there is an absence of public trust in them and a responsibility vacuum in case things do not go as they were meant to be. Hence, there is a mounting pressure on government agencies to address the accountability gaps in their respective systems. The literature suggests that the absence of transparency is not the sole challenge for accountability. Loi & Matthias Spielkamp stress that the absence of legal or de facto entitlement also impedes fruitful results. [28] One of the most proposed solutions to overcome the accountability issues in AI and governance, besides review boards and stakeholder involvement, is meaningful human control, as mentioned by a study conducted by Lane.[29]

There is a need to differentiate between the applicability of transparency and accountability given that the concepts are slightly different vis-à-vis governance and AI. In governance, transparency refers to open government practices and the lack of information gaps between the public and government.[30] It refers to the openness of the government's decisions and policies. With regards to AI, the concept has different applicability. In AI, transparency refers to the openness of AI algorithms.[31] It refers to how decisions are made by AI systems, to enable better understanding of the internal workings.[32] Likewise, in governance, the concept of accountability entails the government to be answerable to the citizens for its decisions and policy making.[33] On the other hand,

[28] Algorithm Watch, "Towards Accountability in the Use of Artificial Intelligence for Public Administrations," November 2021, https://algorithmwatch.org/en/wp-content/uploads/2021/05/Accountability-in-the-use-of-AI-for-Public-Administrations-AlgorithmWatch-2021.pdf.

[29] Lotti Lane, "Artificial Intelligence and Human Rights: Corporate Responsibility in AI Governance Initiatives," *Nordic Journal of Human Rights* (2023):1-22, https://www.tandfonline.com/doi/full/10.1080/18918131.2022.2137288.

[30] Redeemer Dornudo Yao Krah and Gerard Mertens, "Transparency in Local Governments: Patterns and Practices of Twenty-first Century," *State and Local Government Review* 52, no.3 (2020): 200-213, https://doi.org/10.1177/0160323X20970245

[31] Weimin Ouyang, "Research on the Role of Algorithm Transparency in Algorithm Accountability," (Paper presented at Advances in Social Science, Education and Humanities Research Conference, Singapore, November 29-30, 2019).

[32] Ouyang, "Research on the Role of Algorithm Transparency in Algorithm Accountability."

[33] Muhammad Khotami, "The Concept of Accountability in Good Governance," (Paper presented at Advances in Social Science, Education and Humanities Research Conference, Singapore, November 23-25, 2017).

accountability in AI refers to the reasons, explanations and justifications regarding responsibility of decisions made by AI systems.[34]

The third indicator of good governance considered in the research paper is the 'rule of law.' The concept is defined by the United Nations as follows: "The rule of law is a principle of governance by which all persons, institutions and entities, public and private, including the state itself, are accountable to laws that are publicly promulgated, equally enforced, independently adjudicated and consistent with international human rights norms and standards".[35]

According to the study conducted by Greenstein, it has been argued that this phenomenon is based on the two-pillar transparency principle. In this framework the process should be accessible to the public/citizens through political representation, and procedural safeguards should be present in the form of the ability to contest the decisions.[36] However, as per Rosengrun, the technological advancements and their role in digital decision making pose a serious looming threat to the rule of law.[37] The underlying complexities of the AI systems not only impact various governance aspects but also undermine the traditional notions underpinning the 'rule of law.' Several research papers have mentioned the assistance/support provided to judges in making decisions using AI. In a study by Liu, Lin, and Chen, it is mentioned that the U.S. has been using the criminal justice system with AI integration to mediate between the state and accused.[38] It is becoming

[34] Reuben Binns, "Algorithmic Accountability and Public Reason," *Philosophy and Technology* 31 (2018): 543-556, https://link.springer.com/article/10.1007/s00146-023-01635-y#citeas

[35] United Nations, "What is Rule of Law," https://www.un.org/ruleoflaw/what-is-the-rule-of-law/#:~:text=For%20the%20United%20Nations%20(UN,and%20which%20are%20consistent%20with.

[36] Stanley Greenstein, "Preserving the Rule of Law in the Era of Artificial Intelligence (AI)," *Artificial Intelligence and Law* 30, no.2 (2022):291-323, https://doi.org/10.1007/s10506-021-09294-4https://link.springer.com/article/10.1007/s10506-021-092944

[37] Sebastian Rosengrun, "Why AI is a Threat to the Rule of Law," *Digital Society* 1, no.2 (2022): 1-10, https://link.springer.com/article/10.1007/s44206-022-00011-5#:~:text=By%20challenging%20the%20rule%20of,least%20on%20the%20surface%20level.

[38] Han-Wei Liu, Ching-Fu Lin and Yu-Jie Chen, "Beyond State v Loomis: Artificial Intelligence, Government Algorithmisation and Accountability," *International*

commonplace that 'risk assessment algorithms' are being consulted not only in procedures of deciding prison duration, and determining guilt and innocence but also in pretrial bails and post-trial sentencing.[39] Sidhu cited the famous "State v. Loomis case", in his research paper, where a defendant was charged six years in prison for a criminal act by a trial court. The punishment sentence was determined by 'algorithmic risk assessment' and the suspect was wrongfully denied his right to challenge the trial. Huq reveals that the formal and procedural conception of legal issues would be subjected to empirical contingencies with the integration of ML.[40] Apart from the vulnerabilities such as bias and errors in the system, the most critical aspect that was identified remains the absence of human factors.

The literature suggests that transparency, accountability and rule of law are at risk with the advent of AI in governance. The prevailing sentiment is that all three factors need to be considered crucial when integrating autonomy in various governance domains. It is pertinent to note that the available literature lacks expert opinions and is largely generic. This study adds experts-based insights via qualitative interviews.

Theoretical Framework

The socio-technical theory posits that the joint interaction of social systems and technical subsystems can lead to optimised governance outcomes. The theory rejects the idea that the technological outcomes are purely technical. In contrast, it argues that it is the interaction of technology with human systems which produces the desired impact on aspects such as governance. This aspect faces challenges and failure only when the interaction between the two systems is weak.

Journal of Law and Information Technology 27, no. 2 (2019): 122-141, https://academic.oup.com/ijlit/article-abstract/27/2/122/5316430?redirectedFrom=fulltext.

[39] Dawinder S. Sidhu, "Moneyball Sentencing," *Boston College Law Review* 671 (2015):2014-2026, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2463876.

[40] Aziz Huq, "Artificial Intelligence and the Rule of Law" (Paper, University of Chicago, Public Law working Paper No.764, 2021), https://ssrn.com/abstract=3794777.

In the research paper, the three factors under consideration i.e. Transparency, Accountability and Rule of Law constituted the social systems. On the other hand, integration of AI, ML and autonomous decision making constitutes the technical aspect.

The advancements in AI will not automatically translate into efficient governance sytems.AI is an enabler; its impact will come purely from its optimal utilisation. As suggested by the expert opinions, adapting it too quickly without the needed care and consideration will lead to adverse results and an overall negative impact. Hence, these technical advancements must be complemented by social mechanisms which include regulations, legal frameworks, and standards etc. In parallel, it is equally important that the technical systems also evolve as per the requirement of social systems.

Different concepts explained in the paper such as black-box nature of AI, stochastic models, data-driven subjectivity show discrepancies where technical systems are integrated into social systems without ensuring their compatibility with each other.

AI and Transparency

Black Box Nature

There was unanimous agreement among the respondents that AI models are inherently 'black boxes'. These models rely on deep learning utilising multiple layers of neural networks to process data and make decisions. Artificial neural networks are inherently complex and self-adaptive. Consequently, inner workings of the systems' are considerably complex, making the decision making notably opaque, especially when compared to human decision-making. The development, testing and deployment of the algorithms remain ambiguous. This lack of clarity makes justifying the AI system's decisions difficult, adversely impacting the transparency in decision making.

One respondent revealed that achieving transparency in governance was not impossible but it will take considerable time before the current algorithms become explainable. It was emphasised that publishing

algorithms will not make the governance transparent. Transparency can only be achieved when the decision making mechanisms in each layer of the artificial neural networks are deconstructed.

It was asserted that this complexity has the potential to impact transparency across a range of governance sectors. For instance, in the financial sector, AI models are employed to make decisions about credit scores, loan approvals, and other important financial matters. Without transparency in how these decisions are made, consumers may find it difficult to understand why they were denied a loan or why their credit scores decreased.

Similarly, in the criminal justice system, where machine learning techniques are extensively used, AI models make decisions about parole, sentencing, and even the likelihood of someone committing a future crime. Lack of transparency in the inner workings of these models can raise concerns about fairness.

For instance, a criminal justice system may advise a US$1000 fine for an individual X, from a minority community on a theft charge. The high charge might be associated with the prior high probability of the crime committed from the respective community. The system might issue a much severe punishment to deter the individual from future crimes based on his race. On the other hand, another individual Y, belonging from a majority community might be issued a standard US$600 fine for the same crime. Although the AI system might have used data points to come at the two different decisions, such scenarios raise questions regarding the transparency of the systems, owing to the presence of intentional or unintentional bias in training data.

Open-source Models

The participants also mentioned the threats to transparency concerning open-source models. They believed that open-source AI models were becoming increasingly user-friendly with time, making it likely that individuals, particularly in the private sector, may use them frequently. The use of open-source models could further impair transparency because

they are more vulnerable to attacks, theft, hacking, may perpetuate biases and may lack quality control and oversight.

Apparent Transparency

i) Automated Decision Making

Several respondents provided examples of certain sectors where AI could apparently foster transparency, such as automated decision making. For instance, if machine learning algorithms are used for approving loan requests in the banking sector, the procedure could be more transparent if the system is blinded to the applicants' race. Similarly, different permits, like building permits, could be issued through automated decision making. The building permit application involves reviewing plans and specifications to ensure compliance with zoning regulations, building codes, and safety standards. Using AI, government agencies can develop systems that automate the review process. An AI system could scan the plans and specifications for specific key criteria and flag any areas that do not comply with regulations or standards. This can speed up the review process, reduce the workload on government employees and foster transparency given that a predefined set of factors would make the decisions. However, the respondents were convinced that although it may appear to increase transparency, the underlying complexity, loopholes, and the probability of error will negatively impact transparency. The complex inter-variable relationships make the inner workings and predictions considerably difficult. Hence, there was a concern that such "deceptive transparency" may lead to further adoption of AI in various government sectors.

ii) Predictive Analytics

The concept of deceptive transparency was also visible in predictive analytics. Algorithms could be used to analyse large amounts of data from public sources and government agencies to identify patterns and trends leading to more rational decision making and playing a role in preventing corruption or other abuses of power. For example, in tax collection, algorithms could be employed to identify discrepancies in tax returns and cross check them against other data sources to flag potential instances of

tax fraud. Similarly, in social welfare programs, algorithms could be used to analyse beneficiary data to detect anomalies, such as multiple claims for the same individual or claims for deceased individuals, indicating fraudulent activity. AI could be used to predict future trends and outcomes based on historical data helping government agencies identify potential problems before they occur and take steps to prevent them.

Although predictive analysis may make these processes more efficient yet they undermine the concept of transparency. Hence, it puts the decision-makers in a dilemma to choose one over the other. Moreover, it was also identified that transparency would always remain a dynamic concept in AI governance. What might appear transparent to the AI developer may not be transparent to government officials or citizens, and vice versa.

AI and Accountability

Stochastic by Design

The interview respondents believed that accountability is directly linked with transparency and both must be studied together. The respondents identified that using AI in decision-making could introduce new levels of complexity and opacity, making it difficult to scrutinise and hold decision-makers accountable.

There was a consensus that the inner workings of an AI model are stochastic by design implying that the output models can vary even when the inputs are the same. Accountability is challenging to assign since the method used for a certain decision came from a stochastic rather than a deterministic process, and it is strenuous to identify the cause of the particular outcome or decision. As a result, it can be challenging to determine when an AI system is making wrong decisions or when it is making decisions that are not in line with the public interest and whom to hold responsible for the particular act.

Responsibility Gaps

In governance aspects, there can be multiple sectors in which AI systems would operate in a decentralised manner; it can be challenging to identify who is responsible for their outcomes, creating responsibility gaps. For instance, in a potential scenario of managing public transportation of City X, several actors would be involved such as transport authority, technology providers and private companies, etc. In the likelihood of the AI systems not yielding the desired outcome, responsibility gaps could emerge as to who should be held accountable for such an act. Such scenarios lead to confusion regarding the responsibility of the outcomes. Greater penetration of AI in governance could adversely impact the human role in decision making. With so much algorithmic activity, humans may find themselves in a position where they are less inclined to hold themselves accountable for the errors committed by the machines. Moreover, there are concerns that algorithmic decision making could lead to the redistribution of power. As AI systems become more powerful, power could be concentrated and consolidated in the hands of a few individuals or institutions.

Such developments become even more complex in sensitive sectors such as the health care setups. In a potential scenario where the system arrives at the wrong diagnosis, it is possible that the human operator may be reluctant to take the direct responsibility on the premises that the system was programmed by the AI expert. Hence, such scenarios could lead to redistribution of responsibility amongst developers, implementers and users of autonomous systems.

The responsibility gaps are relevant for international collaborations or public-private partnerships given that they involve multiple actors. For example, a private company develops and operates a airport security system powered by AI for a country X. In that case, it can be challenging to determine who is responsible for the system's outcomes, particularly if the system's development and operation are subject to contractual agreements. In a potential scenario of the security systems failing repeatedly, the government can question the obligation of the private company in maintaining security of the airport via AI-enabled systems. On the other hand, the private companies could blame the government for flawed specifications or wrong data. It would remain in dilemma whether

the responsibility is wholly on the government, on the private company or shared responsibility on the two.

Ultimately, impact of AI on governance accountability will depend on a range of factors, including how AI systems are designed and implemented, the level of transparency and accountability integrated into these systems, and the regulatory and legal frameworks in place to govern their use.

AI and Rule of Law

Objectivity

There was a consensus that AI in governance holds the potential to remove human bias and ensure that the objectivity of law when derived from a well-trained AI system. However, this, may seem positive initially, it could lead to harsher outcomes for edge cases where crimes are committed out of necessity rather than bad intent. An example given by one respondent was the classic scenario of 'stealing bread to feed a hungry child'. While such an act may be legally considered an offense, it is also understandable and potentially justifiable, given the circumstances. It could undermine human judgement and discretion because AI algorithms designed to make decisions based solely on objective criteria, may overlook subjective or contextual factors that are crucial for ensuring fair and just decisions. This reduces complex scenarios to mere data points correlated and statistically weighted against each other.

In a potential scenario, an individual may be accused of breaking a traffic signal. The AI enabled system might punish the individual as per the standard practice. However, the defendant may appeal that the system has ignored the context where the action was taken in case of a severe emergency. The defendant may also hold that his previous clear record has also been set aside and he has been given a considerably harsher punishment by the system.

The cases of more sensitive nature, such as issues pertaining to family law vis-a-vis custody of children may be even more complicated. In such

scenarios, the factors such as empathy may not be considered by AI systems.

Another potential challenge is that the integration of AI in governance could lead to new legal challenges not adequately addressed by existing legal frameworks. For instance, how can an individual sentenced for a certain crime challenge the decision made by the AI systems.

Need for Well-informed Dataset

According to the responses, ensuring that AI models used in the legal system are transparent and auditable is crucial to mitigate the above mentioned risks This means that the underlying data, algorithms, and decision making processes are open to scrutiny and review. A well-formed dataset is essential to ensure that the AI system is trained to capture all of the intricacies of the law. Once such a system is established, it should undergo A/B testing against human decisions for the same cases to ensure that the AI system aligns with the intent of the law in addition to adhering to the letter of the law. A/B testing refers to developing two models "A" and "B" and comparing them with a control variable to determine which model performs better.

This is a challenging task requiring considerable improvements in AI techniques, research and human resource. Moreover, the challenging part of such a system would be enabling it to understand the 'grey area' of the law where everything is not black or white.

Opportunities

Some of the opportunities were highlighted by a respondent who believed that AI could help improve the efficiency of legal processes including contract review, and case management and contribute to legal research. AI has the potential to identify trends in legal data, facilitating better decision-making and more effective law enforcement. Through analysis of the past cases and legal precedents, AI models could predict the likelihood of success in a particular case empowering lawyers and litigants to make more informed decisions about whether to pursue a legal case thus

reducing the number of frivolous lawsuits in courts. Additionally, AI has the potential to reduce corruption by automating administrative tasks and creating more efficient systems for public services. By automating routine tasks AI can reduce costs and increase efficiency in legal processes, saving time and resources for both individuals and the government. However, it was emphasised that human decision making will always have an edge over the machines and therefore relying solely on algorithms for matters of critical importance may not be a rational choice.

Overall, AI can both augment and undermine governance subjected to how well it is integrated and regulated. It is expected that future AI models will be of hybrid nature comprising of AI powering human-decision making. The ultimate objective is to enhance efficiency, speed and data-processing power.

The findings of the literature review are in line with the expert opinions. This underscores that while algorithms may make governance more efficient in certain administrative tasks yet they also give rise to multiple challenges for the concept of governance itself. It is concerning that despite the concerns raised from certain quarters, policymakers worldwide are inclined to implement AI in diverse governance structures.

Way Forward – Explainable AI (XAI)

The respondents were of the view that existing literature addresses the enhancement in the aforementioned governance indicators. However, they believed those recommendations are contingent upon the underlying premise of explainable AI (XAI). They contend that until the explainable AI reaches a certain threshold, the recommendations regarding monitoring and formulating new regulations hold little practical value.

Explainable AI is a field of research that aims to develop AI models that are more transparent and interpretable. Respondents were of the view that XAI was an active area of research aimed at establishing a mechanism that demonstrates the causal relationship between input data and the output of the AI system. XAI algorithms are designed to provide explanations for their decisions, simplifying human comprehension of the rationale behind a specific decision.

For instance, in the context of autonomous vehicles, XAI could provide clear explanations regarding how the vehicle reached a particular decision. For example, if the vehicle is programmed to stop at a red light, the XAI system could explain the specific data points employed to detect the red light and explain how the decision to stop was made. This could enhance passengers' safety and trust and enable engineers to better identify and address potential system issues more effectively.

XAI Techniques

While the stochastic nature poses multiple challenges, the participants emphasised on specific XAI techniques that could contribute to governance. The respondents focused on five techniques and asserted that improvements in these techniques could ultimately enhance transparency, accountability and the rule of law.

i. Feature analysis is a technique used to identify features or inputs essential for the decision making process of an AI model. This technique can help identify factors driving the final output of the AI system and validate the decision making process.

ii. Output distribution mapping is a ML technique employed to adjust a model's output to match a desired distribution. In other words, it is a way to modify the predicted output of a ML model to meet specific requirements or constraints.

iii. Pre-classification latent feature analysis is a technique used to identify the underlying factors influencing the decision-making process of an AI system. This technique could be used to identify any hidden biases or inconsistencies in the decision-making process.

iv. Analysing Gradient Activations For a computer vision-based model (an AI system designed to analyse images or videos), examining the gradient activations of inner convolutional layers is an accurate measure for explaining what the AI system 'saw' when making the classification design. The gradient activation of inner convolution layers reflects the changes that the system undergoes as it analyses

the image. During the classification decision, visualising the gradient activations of the inner convolutional layers reveals which features in the input image were most important in the decision making process. This analysis of gradient activations enables a better understanding of the image regions that were the focus of AI when making the classification decision. In a potential scenario, a Convulsion Neural Network may classify the images of either cats or dogs. When presented with a cat's image, the gradient activations of the inner layer could demonstrate that the system focused on specific features such as the cat's ear or the fur's texture to classify it as a cat. By contrast, when presented with an image of a dog, the gradient activations could reveal that the AI system focused on features such as the shape of the dog's snout or the pattern of its coat.

v. Uncertainty Quantification can also help ensure that the decision made by AI systems are countersigned by the amount of uncertainty attached to the output. This assigns a confidence score to the output and gives an idea about the reliability of the output, enabling decision makers to make more informed decisions.

These techniques when integrated into governance systems could contribute positively. The respondents were convinced they would make the systems more explainable. Hence, this interpretability could aid factors such as transparency, accountability and the rule of law. However, they by no means should be seen as problem-solving. The challenges mentioned would remain on the table and XAI could only contribute to making AI relatively more interpretable.

Policy Recommendations

Furthermore, each country needs to adopt actionable policy recommendations tailored to the respective country's context vis-a-vis enhancing transparency, accountability, and adherence to the rule of law in AI systems. Some important policy recommendation in this regard are as follows:

- There is a need for development of national standards for AI systems used in governance. These standards should reflect the minimum requirements for systems to be considered transparent. In this context AI experts together with legal experts should jointly draft these standards to ensure that the systems are interpretable.

- Specific legislation could be introduced that conditions all AI systems to be used in various government sectors to abide by the established standards.

- The established standards and legislations must be reviewed and updated regularly to align with the pace of new technological developments.

- Participation of states in international forums can also serve as an effective measure in learning best practices from each other and devising effective standards.

- There is a need to establish verification mechanisms that assess the level of explainability in AI systems. In this context, independent bodies could be established to keep a check on the AI systems.

- As AI becomes more integrated in governance systems, there must be timelines or a transition period established by the policymakers to be followed by the concerned sectors.

- Regular training of concerned government officials should also be conducted for smooth governance using AI systems.

- All AI-related projects to be used in governance should undergo AI risk assessments for better performance and risk mitigation.

- There is a need for allocation of specific funds for research in XAI techniques such as feature importance, output distribution mapping, pre-classification latent feature and analysis of gradient activations.

- The curriculum taught in universities should include the broader understanding of AI, importance of AI techniques and their interplay

with governance. Education institutes should also partner with private tech companies to facilitate training programs and workshops in this regard.

- Research on the above-mentioned XAI techniques should be encouraged by promoting grants and incentives to universities who are making advancements in XAI techniques.

- Collaboration between government, academia, and industry should be promoted to foster innovation and advancements in XAI.

- Feedback mechanisms should be developed to ensure that the diverse perspectives are considered from the civil society.

- Governments should also launch awareness programs for masses to be better informed about the interplay of AI and governance. The content should be accessible and understandable for the non-technical audience as well.

- A diligent approach with an adequately tested AI system will enable the automation of redundant and time-consuming parts of governance thus leading to an overall net positive impact.

- Deterministic systems or procedures have a defined set of inputs and outputs, and the outcome of the process can be predicted based on these inputs. They involve rules and procedures that could be codified into an AI system. Hence, data-driven decision-making that relies more on statistical analysis should be readily automated by AI. The sectors where the stakes are high should not be automated with AI for the time being.

- In the concerned sector, a detailed testing regime needs to be designed where the potential pitfalls of an AI system can be identified. This can be done by identifying edge cases that might be of concern to policymakers.

- A sandbox setting would be needed where the AI system is allowed to make decisions and a causal link between the input and output can be established. Once the performance (‘behaviour’) of such a

system has been established, potential pitfalls can then be addressed.

- By investing in XAI systems, it becomes relatively easier to interpret the systems. Hence, the states/organisations that are inclined towards integrating AI in governance structure, need to actively pursue research in this domain.

- Concepts such as transparency, accountability and rule of law need to be reviewed when considering algorithmic decision-making. The same concepts applied to human decision making may not readily be applicable for AI enabled governance. Hence, there is a need for more debate on these aspects and to define these concepts vis-à-vis governance in the age of AI.

Conclusion

As with varied roles that AI plays across sectors, its prominence in governance is undeniable. While there are clear indications that the advent of AI-enabled governance is likely to increase the efficiency of various governance tasks, it is also poised to introduce novel challenges. These challenges could adversely impact transparency, accountability and the rule of law. Future governance in the context of AI cannot afford to ignore these challenges, and subsequent actions need to be considered accordingly.

Explainable AI stands out as a crucial means to ensure that future governance is aligned with the principles of good governance. Hence, there is a pressing need to invest resources and research in this regard and explore various techniques that make the systems more interpretable, adaptable in governance structures, and have wider public acceptance. Furthermore, the governments should take the necessary actions in the form of standards, legislations, establishing timelines, learning best practices, training, feedback mechanisms, funding and public awareness for the smooth integration of this technology in various governance sectors. The responsible integration of AI can help overcome the existing challenges and prove beneficial for governance.