frontiers | Frontiers in Political Science

TYPE Original Research
PUBLISHED 14 May 2025
DOI 10.3389/fpos.2025.1595345

Mapping AI's role in NSW governance: a socio-technical analysis of GenAI integration

Luis Lozano-Paredes*

Transdisciplinary School, University of Technology Sydney, Sydney, NSW, Australia

OPEN ACCESS

EDITED BY
Gianluca C. Misuraca,
Polytechnic of Milan, Italy

REVIEWED BY
John R. T. Bustard,
Ulster University, United Kingdom
Mark Esposito,
Hult International Business School,
United States

*CORRESPONDENCE
Luis Lozano-Paredes
✉ luishernando.lozanoparedes@uts.edu.au

RECEIVED 19 March 2025
ACCEPTED 30 April 2025
PUBLISHED 14 May 2025

CITATION
Lozano-Paredes L (2025) Mapping AI's role in NSW governance: a socio-technical analysis of GenAI integration.
Front. Polit. Sci. 7:1595345.
doi: 10.3389/fpos.2025.1595345

This paper examines the integration of Generative Artificial Intelligence (GenAI) in New South Wales (NSW) Government processes through a socio-technical lens using Actor-Network Theory. Rather than viewing GenAI as a passive tool, the study conceptualizes these systems as active actants that reshape governance networks, redistribute authority, and reconfigure democratic accountability mechanisms. Through document analysis, actor-network mapping, and scenario analysis of potential breakdowns, the research reveals that while NSW has established comprehensive AI governance frameworks, significant gaps remain in addressing the unique challenges posed by GenAI systems. Historical analysis of algorithmic failures in Australian public administration, including Revenue NSW's automated debt recovery system and the federal Robodebt scheme, demonstrates the consequences when technical systems undermine democratic principles. The paper proposes a regulatory sandbox approach to balance innovation with democratic safeguards, highlighting the need for governance frameworks that recognize GenAI's role in reshaping political relationships. The findings contribute to scholarly debates by demonstrating the insufficiency of purely technical or ethical frameworks that do not address the political dimensions of AI integration in public governance.

KEYWORDS

generative AI, actor-network theory, public governance, socio-technical systems, algorithmic governance

1 Introduction

The rapid advancement of Generative Artificial Intelligence (GenAI), particularly large language models (LLMs), presents transformative opportunities and significant global challenges for public governance systems. As governments increasingly deploy these technologies to enhance service delivery, streamline operations, and inform policy decisions, fundamental questions emerge regarding their implications for democratic accountability, institutional legitimacy, and citizen rights. This paper examines the integration of GenAI in the New South Wales (NSW) Government of Australia, offering a socio-technical analysis that conceptualizes these systems not merely as tools but as active actants in governance networks that reshape power relationships and decision-making processes.

Integrating GenAI into public administration represents a distinctive challenge that transcends conventional digital transformation initiatives. Unlike traditional automation technologies that execute predefined functions, GenAI and LLMs demonstrate emergent capabilities to generate human-like text, process unstructured information, and produce outputs that can directly influence policy formulation, regulatory enforcement, and citizen engagement. Their seeming authority, persuasiveness, and inherent opacity in their reasoning processes raise critical questions about how democratic oversight and judgment can be maintained in increasingly AI-mediated governance systems.

The NSW Government provides a particularly instructive case study. As one of Australia's largest public administrations, NSW has established comprehensive AI governance frameworks that include ethical considerations and democratic accountability. NSW has also created formal governance structures through its AI Strategy, Ethics Policy, and Assurance Framework to guide the responsible implementation of AI technologies across its departments and agencies. These frameworks provide a fertile ground for examining how theoretical aspirations for ethical AI translate into practical governance arrangements and the potential gaps that emerge between formal structures and operational realities.

This study employs Actor-Network Theory (ANT) as a methodological-theoretical framework to analyze the socio-technical dynamics of GenAI integration. Rather than treating technical systems as passive tools only subject to human direction, ANT recognizes both human and non-human entities as actants with agency that jointly constitute networks. This approach reveals how GenAI reshapes political relationships by mediating between citizens, officials, and policy frameworks, potentially redistributing authority in ways not fully captured by traditional governance models that presume exclusively human agency. The latter is particularly relevant in the context of LLMs being uncovered as having embedded political biases (Rozado, 2024).

The study also draws on concepts of breakdown and repair from infrastructure studies to anticipate potential vulnerabilities in GenAI implementation and the mechanisms required to address them. As Star (1999) and Jackson (2014) have argued, breakdowns—when systems fail to meet expectations—reveal underlying connections within networks that typically remain hidden during smooth operation. Similarly, repair processes represent opportunities for transformative reconfigurations of socio-technical and political relationships. By examining historical breakdowns in analogous algorithmic governance systems, this study identifies patterns that may inform more resilient approaches to GenAI governance.

The study addresses four central questions: First, how do LLMs and GenAI function as actants within NSW government networks, reshaping workflows, authority relationships, and decision-making processes? Second, what vulnerabilities emerge as GenAI becomes increasingly integrated into public governance, particularly regarding transparency, bias, and accountability? Third, how adequate are NSW's existing governance frameworks for addressing these vulnerabilities and maintaining democratic legitimacy? Finally, what additional governance mechanisms might be required to ensure that GenAI implementation enhances rather than undermines democratic values in public administration and public policy?

To address these questions, the study employs a mixed-methods approach combining document analysis of NSW Government policies and frameworks, actor-network mapping of key relationships in GenAI implementation, and scenario analysis to anticipate potential breakdown scenarios. This analysis contributes to scholarly debates on AI governance by demonstrating the insufficiency of purely technical or ethical frameworks that do not address the political dimensions of AI integration. As Jasanoff (2016) has argued, technologies are not merely tools but active participants in constituting social order and political relationships. This perspective suggests that effective AI governance requires technical safeguards or ethical principles and fundamentally reconsidering how agency, accountability, and democratic legitimacy operate in human-AI governance networks.

The study is structured as follows: Section II presents the theoretical framework, elaborating on Actor-Network Theory and concepts of breakdown and repair as analytical lenses for examining GenAI in governance. Section III details the methodology, outlining the document analysis, actor-network mapping, and scenario analysis approaches employed. Section IV examines the NSW Government's AI Strategy and workforce integration initiatives, while Section V critically assesses NSW's AI governance framework. Section VI analyzes historical examples of algorithmic governance breakdowns in Australia, extracting lessons for GenAI implementation. Section VII discusses the findings and offers recommendations for democratic governance of GenAI, including three anticipatory scenarios of potential breakdowns. Finally, Section VIII concludes by synthesizing key insights and implications for maintaining democratic legitimacy as GenAI becomes increasingly embedded in administrative processes.

2 Navigating networks and breakdowns

Drawing on Actor-Network Theory (ANT) (Latour, 2007; Latour, 1996), this paper explores how GenAI needs to be understood as a significant actant within governmental procedures, redistributing authority and reconfiguring democratic oversight mechanisms. This theoretical-methodological approach illuminates how GenAI integration reshapes governance structures by triggering renegotiations of political power, procedural rule, and accountability relations.

2.1 Actor-network theory and political governance

Actor-network theory, rooted in science and technology studies (STS), offers a distinctive lens for analysing AI in governance by refusing to separate technical systems from political structures (Latour, 2007; Latour, 1996; Papilloud, 2018). Rather than viewing GenAI as merely a tool deployed by human actors, ANT recognizes technological systems as actants with political significance that actively reshape governance networks. This approach reveals how GenAI systems like LLMs do not simply execute predetermined functions but can actively mediate political relations, reshaping how decisions are made, policies implemented, and public accountability maintained.

Three key concepts from ANT are particularly valuable for understanding AI's implications in governance:

First, recognizing both human and non-human actors and actants illuminates how LLMs can reshape democratic accountability by mediating between citizens, officials, and policy frameworks. When an LLM generates policy recommendations or drafts administrative decisions, it exercises a form of agency that complicates traditional governance accountability chains. The NSW Government's integration and promotion of GenAI as a co-actant (Digital.NSW, 2024c; Digital.NSW, 2024b) within administrative processes reconfigures who—or what—exercises authority in governance networks.

Second, the concept of translation reveals how AI systems transform policy intentions through their implementation. Translation in ANT refers to the process by which actors are brought into networks, with their interests realigned to fit (Latour, 2007). As LLMs or other algorithmic platforms interpret regulations, draft documents, or analyze citizen inputs, they do not simply transmit information but actively reshape it according to their logic. This translation process raises critical questions about democratic legitimacy when non-elected technical systems mediate policy intentions established through democratic processes.

Third, blackboxing—the process by which a technology's internal workings become opaque to users—presents fundamental challenges to democratic transparency. In government AI projects, blackboxing can obscure the logic behind algorithmic decisions, complicating efforts to ensure accountability to citizens (Gutiérrez, 2023).

2.2 ANT and GenAI in democratic governance

Recent scholarship has advanced the conversation on how AI integration reshapes political governance processes, offering both transformative potential and presenting challenges to democratic oversight (Cheong, 2024; Taeihagh, 2021). LLMs have become a focal point within this context due to their ability to generate text resembling human language, enabling automation of traditionally human political functions like policy analysis, regulatory drafting, and citizen communication.

Studies reveal that GenAI in public administration can assist in processing substantial data volumes, identifying trends, and generating insights and future scenarios that traditional methods might miss (Reid et al., 2023). However, some studies highlight the importance of tackling ethical and accountability concerns associated with GenAI's expanding role, addressing issues around transparency, potential bias, and the need for human oversight to prevent harm (Mergel et al., 2024).

As GenAI becomes more integral to government, policy and political processes, there is increasing attention on concerns like algorithmic transparency, bias, and accountability (Janssen and Kuk, 2016; Singhal et al., 2024; Taeihagh, 2021). Its deployment raises fundamental questions about delegating authority to GenAI, as these systems have the potential to shape or even determine policy outcomes (Veale and Brass, 2019) and are also embedded with a particular ideological view (Rozado, 2024).

2.3 Breakdown and repair in democratic AI governance

The ideas of breakdown and repair (Star, 1999; Jackson, 2014) provide valuable analytical tools for understanding democratic resilience in AI-enhanced governance systems. These concepts become especially important in implementing GenAI in the government sphere, where technological systems often face setbacks and require adjustments that have political implications.

'Breakdowns' (Star, 1999) describe moments when systems, infrastructures, or technologies fall short of expectations, exposing the underlying connections within a network that typically remain hidden during smooth operation. This concept draws on Latour (1992) work, which explores breakdowns through the lens of everyday artefacts and their role in guiding social order and behaviors. In political and policy contexts, breakdowns occur when systems fail to align with democratic values, expose accountability gaps, or produce outcomes that undermine public trust. When an AI system produces biased outcomes or operates in ways citizens cannot understand or contest, it can create democratic deficits that require intervention.

'Repair,' on the other hand, follows breakdown and encompasses the technological fixes, policy updates, and ethical recalibrations needed to realign these systems with public sector values. Jackson (2014) sees repair as not a simple return to the previous state but a productive process that can lead to transformation and improvement. The repair might mean reworking the relationships and roles within the actor network when integrating GenAI into public administration, creating opportunities for ethical reflection, policy updates, and systemic enhancement. In the NSW government, breakdowns could surface as technical setbacks or ethical issues that expose weaknesses in the socio-technical infrastructure. Repair becomes essential in addressing these issues, helping build public trust by openly tackling AI-related challenges.

By combining Actor-Network Theory with concepts of breakdown and repair, this paper develops a framework for critically examining how GenAI reshapes political accountability and democratic legitimacy in NSW governance. The following section outlines the methodological approach to mapping these socio-technical dynamics in NSW's emerging AI governance landscape.

3 Methodology

This study employs a framework combining document analysis, actor-network mapping, and scenario analysis to investigate the integration of GenAI in the NSW Government. Data were collected exclusively from authoritative public sources, including NSW government strategies, policies, training modules, and oversight reports, as the authorized implementation and use of GenAI is relatively recent and contacted government workers expressed discomfort or unawareness regarding the framework. This desktop research approach focuses on policy intent and documented practice, providing a foundation for examining NSW's AI governance structures.

3.1 Actor-network mapping

To systematically identify and analyze relationships among actors in NSW's GenAI ecosystem, a structured Actor-Network Mapping protocol was employed, aligned with established ANT methodological approaches (Justesen, 2020; Nimmo, 2011). The protocol consisted of four key phases:

1 Actor Identification: Systematic identification of human actors (public servants, policymakers, citizens), non-human technological actors (LLMs, data infrastructure), and institutional/regulatory actors (policies, frameworks) through document analysis.
2 Relationship Mapping: Documentation of connections between actors using a matrix-based approach that classified relationships according to direction, strength, nature, formality, and stability.

3 Translation Process Analysis: Examination of how actors are enrolled into networks through close reading of policy documents, analysing problematization, enrolment, and mobilization processes.
4 Network Visualization and Analysis: Creation of visual representations, including radial network diagrams illustrating centrality and connectivity and flow diagrams depicting information and influence paths.

This mapping revealed key actors and their roles in NSW's GenAI network (Table 1), highlighting the centrality of the AI Assurance Framework as a mediating structure between technical systems and governance mechanisms.

3.2 Document analysis

The document analysis followed a structured protocol, selecting materials based on their relevance to GenAI integration in NSW governance (Digital.NSW, 2024c; Digital.NSW, 2024b; Digital.NSW, 2024a; Digital.NSW, 2024d; Parliament NSW, 2024). A thematic coding approach was employed, with initial codes derived from the research questions and ANT framework, supplemented by inductive codes that emerged during analysis. Qualitative analysis software organized themes, ensuring systematic tracking of coding decisions. The analysis was enriched by examining evidence from relevant cases, such as the NSW Ombudsman's findings on AI-related failures and insights from the 'Robodebt' inquiry (Clarke et al., 2024; Podger, 2023; Rinta-Kahila et al., 2024). These insights will be expanded in a subsequent section of the paper.

3.3 Scenario analysis

Given that GenAI implementation in NSW is at an early stage with no significant operational failures yet documented, this study employed scenario analysis to complement document analysis and actor-network mapping. This approach enables the exploration of potential vulnerabilities without requiring direct observation of failures.

Three hypothetical scenarios were developed based on the socio-technical dynamics identified through ANT mapping. Each scenario examined how actors might realign or conflict in response to the breakdown, drawing on concepts of breakdown and repair to anticipate potential adaptations in AI governance.

3.4 Methodological rationale

Three key considerations drove the selection of the methodological approaches:

1 Capturing Socio-Technical Complexity: ANT provides a robust framework for analyzing the integration of generative AI into governance processes as it refuses the artificial separation between technical and social domain (Latour, 2007), treating both human and non-human entities as actants with agency.
2 Addressing Early-Stage Implementation: The relatively recent introduction of GenAI into NSW governance makes traditional empirical approaches such as impact assessments premature. ANT's focus on emergent networks provides analytical purchase even at this early stage, while scenario analysis helps anticipate potential vulnerabilities.
3 Navigating Data Access Constraints: The decision to rely on document analysis was shaped by both ethical considerations and practical constraints, as many government workers contacted were either unaware of the GenAI framework or uncomfortable discussing it, given its recent introduction.

The combination of ANT mapping with scenario analysis represents a methodological innovation that addresses the challenges of studying emerging technologies whose impacts are very recent due

TABLE 1 Key actors and their roles in NSW's GenAI network.

Actor category	Actor	Primary role	Key connections	Network position
Human	Mediating information flowsvants	End users of GenAI systems; interpreters of AI outputs	LLMs; AI Ethics Policy; Citizens	Interface between technical systems and governance structures
Human	Policymakers	Framework developers; strategic direction setters	AI Strategy; AI Assurance Framework; Oversight bodies	Decision-making nodes with high influence
Human	Citizens	Service recipients; subjects of AI-influenced decisions	Public servants; Government services	External stakeholders affected by network outcomes
Non-human technological	LLMs	Text generation; data analysis; decision support	Public servants; Data infrastructure; AI Assurance Framework	Central technological actants mediating information flows
Non-human technological	Data infrastructure	Data storage; processing; management	LLMs; AI systems; Security frameworks	The technical foundation supporting AI capabilities
Institutional/Regulatory	AI Ethics Policy	Ethical guidance; value alignment; risk mitigation	Public servants; Policymakers; AI Assurance Framework	Normative structure shaping actor behavior
Institutional/Regulatory	AI Assurance Framework	Governance mechanism; compliance structure; risk assessment	All actors	Central coordinating structure

to their early stages of implementation. It offers a framework other jurisdictions can adapt to ideating their GenAI integration initiatives.

Some limitations in these methods are worth noting, such as potential bias by the author in document selection, the subjective nature of thematic coding, and the challenge of fully capturing actor relationships in a static map. Additionally, reliance on public documents may not fully reflect the internal, current, or informal dynamics within the NSW Government, which, due to the early implementation and inception of these systems, are very difficult to obtain. The black box of governance is sometimes very closed, but this methodology aims to first articulate the topic at large. The following sections analyze the socio-technical dynamics of GenAI integration in NSW, exploring related ethical challenges and breakdown-repair concepts. The discussion will synthesize findings, examining how GenAI influences NSW Government operations and the associated ethical and political implications.

4 Context: NSW Government's AI strategy and workforce integration

This section examines the NSW Government's current GenAI initiatives through a political governance lens. It focuses on how its policy frameworks attempt to distribute authority, accountability, and oversight in AI-enabled governance. Particular attention is given to the NSW AI Strategy (Digital.NSW, 2024b), the NSW Ethics Policy and Assurance Framework (Digital.NSW, 2024a; Digital.NSW, 2024d), and workforce integration measures such as the 'Chatbot Prompt Essentials' learning module (Digital.NSW, 2024c). These initiatives reveal the governance architecture constructed to mediate human-AI relations in public administration and establish political legitimacy for algorithmic decision-making within democratic institutions.

4.1 NSW Government's AI governance architecture

The NSW Government has introduced a governance framework for integrating AI within its public sector operations to establish democratic legitimacy while capturing efficiency benefits. The framework establishes a political authority structure that positions elected officials and public servants as the ultimate decision-makers while incorporating GenAI systems as supporting actors in governance processes. This architecture consists of several interconnected components that together form what might be termed a socio-technical governance ecosystem:

A foundational element is the AI Ethics Policy (Digital.NSW, 2024a), which establishes normative principles guiding AI's use across government operations. The policy emphasizes transparency, accountability, fairness, and privacy, aligning AI tools with democratic values and societal expectations. By establishing these principles, the policy creates a governance mechanism to maintain democratic legitimacy while incorporating increasingly autonomous technical systems. This approach positions ethical frameworks as political instruments that mediate between technical capabilities and democratic requirements.

Complementing this normative framework, the AI Assurance Framework provides a structured review process for AI initiatives, mandating evaluations of AI projects to confirm their adherence to established ethical principles (Digital.NSW, 2024d). This framework represents an attempt to operationalize abstract ethical principles into concrete governance mechanisms. The review involves considering questions about public benefit, privacy, and security and establishing a procedural approach to maintaining human authority over algorithmic systems. Notably, this framework covers all AI projects except those involving commercial AI tools without customization, creating a potential governance gap in the oversight architecture.

These frameworks establish a hierarchical governance structure where human officials retain formal authority while AI systems are positioned as tools subject to human oversight. However, AI systems' increasing autonomy and authority in day-to-day administrative processes may challenge this formal structure in practice. The training initiatives explored below suggest a more complex agency distribution than the formal governance architecture might imply.

4.2 'Chatbot prompt essentials' and workforce integration

The Chatbot Prompt Essentials learning module (Digital.NSW, 2024c) represents a key component of NSW's preparation of its workforce for AI integration, revealing insights into how political authority may be redistributed in practice by GenAI. This module aligns with the infrastructure philosophy (Große, 2023) shaping the government's adoption of GenAI and AI more broadly (Parliament NSW, 2024).

The module is designed to upskill government employees in crafting effective prompts for GenAI systems, teaching public servants to create prompts that encourage a natural conversational flow with LLMs. By introducing prompt engineering (Bozkurt, 2024; Sahoo et al., 2024; Wang et al., 2024), the module establishes a new form of human-AI interaction where public servants learn -and are expected- to guide AI systems rather than simply oversee them. This approach positions public servants as prompt engineers who shape AI outputs rather than sovereign decision-makers who merely review AI recommendations, subtly shifting the distribution of agency and authority in practice.

This training initiative reveals tension within NSW's AI governance approach. While the formal governance architecture positions humans as the final authority over AI systems, the practical implementation suggests a more collaborative relationship where humans must learn to communicate with AI systems to achieve desired outcomes effectively. Moreover, they are expected to. This tension reflects broader challenges in establishing democratic governance of increasingly autonomous technical systems actively shaping political and administrative processes (see Figure 1).

4.3 Positioning NSW's approach to global AI governance

The NSW Government's approach to AI governance is distinctive in its breadth and proactive nature, setting a high benchmark for public sector AI frameworks globally. Compared to other jurisdictions, NSW's governance architecture reveals political priorities and

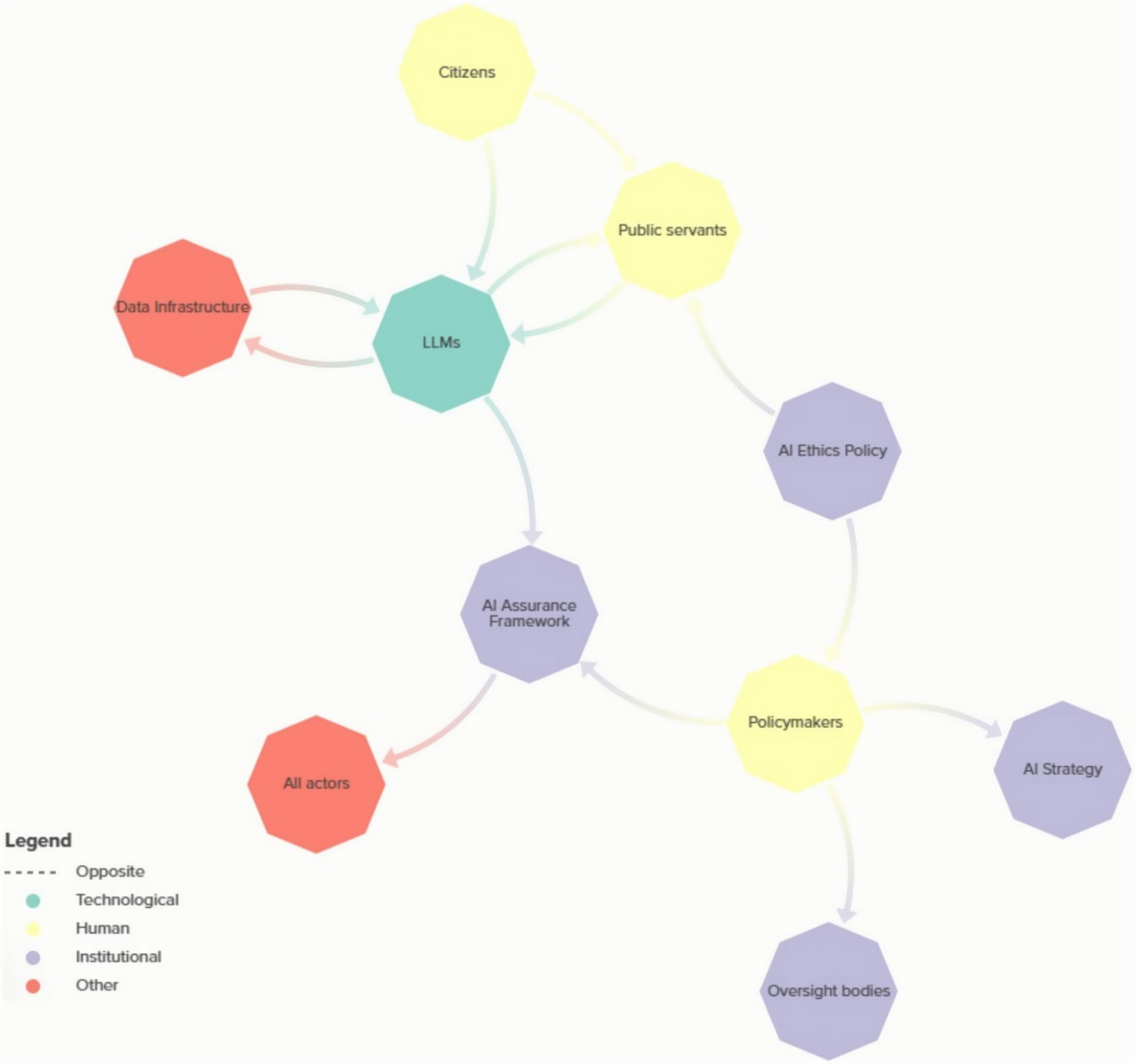

FIGURE 1
NSW GenAI governance architecture. A structural map of the core governance network for GenAI integration in NSW public administration. It illustrates the relationships between human actors (public servants, policymakers, citizens), technological systems (LLMs and data infrastructure), and institutional frameworks (AI Ethics Policy, AI Assurance Framework, and Strategy). Summary map elaborated by the author using Kumu.io.

assumptions about the relationship between human and artificial agents in democratic governance.

While other regions are developing their AI governance frameworks, NSW's strong focus on integrating ethical frameworks with workforce preparation and technical implementation distinguishes it from approaches in jurisdictions where AI governance may be more fragmented. For instance, the European Union's AI Act offers a comprehensive regulatory framework that categorizes AI risks (Wagner et al., 2023), while the United States issued an Executive Order on AI to guide federal agencies in creating assurance structures (Saheb and Saheb, 2024). Canada's Algorithmic Impact Assessment tool and Singapore's AI Governance Framework also emphasize responsible AI implementation but with different emphases on risk assessment and industry collaboration (Saheb and Saheb, 2024).

These global efforts underscore the increasing focus on establishing legitimate governance structures for AI systems. Still, the NSW Government's model is one of the most thorough in the public sector worldwide, particularly in its attention to practical workforce integration alongside formal governance frameworks. This comprehensive approach suggests an attempt to create a governance architecture that maintains democratic legitimacy while adapting to the redistribution of agency that AI systems inevitably introduce into governance processes.

The initiatives described above establish the formal governance architecture for GenAI in NSW public administration. The following sections will examine how these frameworks operate in practice, analysing their effectiveness in maintaining democratic legitimacy, accountability, and human oversight as AI systems become increasingly integrated into governance processes. Particular attention will be paid to how these formal structures address—or fail to address—the potential for breakdowns in AI-enabled governance systems and the mechanisms available for repair when such breakdowns occur.

5 Critical assessment of NSW's AI governance framework

This section critically assesses NSW's AI governance framework, examining its adequacy for maintaining democratic oversight, ensuring political legitimacy, and protecting citizen rights in the context of increasingly autonomous AI systems:

5.1 Ethical challenges as political governance challenges

Algorithmic models, including GenAI, raise essential considerations regarding how training data biases may influence public service outputs and potentially undermine democratic values of fairness and equality (Anthis et al., 2024; Brown, 2024; Gutiérrez, 2023). To counteract these democratic risks, the NSW Government has adopted AI Ethics principles and policies that stress fairness, inclusivity, and accessibility (Digital.NSW, 2024a). These principles require GenAI systems utilized by the government to incorporate input from diverse stakeholders and comply with anti-discrimination laws. However, in many senses, these principles represent normative aspirations rather than enforceable governance mechanisms, raising questions about their effectiveness in preventing outcomes that might be inherent to LLMs and GenAI systems (Anthis et al., 2024). The challenge is not merely ethical but fundamentally political: How can algorithmic systems respect democratic values and remain accountable to citizens rather than simply efficient at administrative tasks?

GenAI's reliance on extensive datasets that may include personal information also raises serious privacy concerns implicating citizen rights in democratic societies. The data-driven nature of AI can amplify privacy risks, mainly if sensitive data is processed without proper safeguards, potentially undermining citizens' privacy rights and autonomy. The Office of the Australian Information Commissioner (OAIC) has underscored privacy as a key concern in AI use. A 2023 OAIC survey revealed that many Australians feel uneasy about government agencies processing personal data with AI, highlighting the need for stringent privacy protections that respect democratic rights to informational self-determination (Pane, 2023).

5.2 Evaluating the NSW AI ethics policy: democratic governance mechanisms

While NSW's AI Ethics Policy and Assurance Framework represent significant steps toward responsible AI governance, their adequacy for maintaining democratic legitimacy in increasingly autonomous AI systems requires systematic assessment. The NSW AI Ethics Policy establishes eight core principles to guide AI development and use: transparency, fairness, accountability, privacy, security, reliability, contestability, and public benefit (Digital.NSW, 2024a). However, when explicitly evaluated for their capacity to ensure democratic governance of GenAI systems, several strengths but also limitations emerge:

5.2.1 Strengths of democratic governance mechanisms

The principle of contestability represents a significant democratic safeguard, as it explicitly requires that AI-driven decisions be subject to appeal and human review. This creates essential mechanisms for citizens to challenge algorithmic determinations, aligning with Cohen and Suzor's (2024) emphasis on public contestability and benefit as essential for democratic AI governance. By ensuring citizens can meaningfully contest AI-influenced decisions, this principle helps maintain democratic legitimacy in increasingly automated administrative processes.

The public benefit principle establishes an explicit requirement that AI systems serve the public interest, creating a normative benchmark against which GenAI implementations can be evaluated regarding their contribution to democratic governance. This principle potentially guards against technology-driven implementations that prioritize efficiency over democratic values and citizen interests.

5.2.2 Limitations for democratic governance

The ethics policy, however, contains limited provisions for GenAI-specific risks such as hallucinations, the generation of plausible but factually incorrect content, and emergent capabilities that could undermine democratic accountability. Developed before the widespread availability of sophisticated LLMs, such as the latest versions developed by OpenAI or Anthropic and the potential of open-source models like DeepSeek and the Deep Research functionalities across models, the policy does not adequately address how citizens and oversight bodies can verify the accuracy of AI-generated content used in governance processes.

The framework provides inadequate guidance on human-AI collaboration in democratic decision-making processes. While the policy requires human oversight, it provides limited practical guidance on how public servants should evaluate, interpret, or potentially override AI-generated recommendations in different contexts. This gap is particularly concerning for democratic legitimacy when GenAI systems increasingly produce more natural language outputs that may appear authoritative despite limitations. Thus, structuring the possibility for government to delegate decision-making and action-building from the outputs of these models.

There is also insufficient attention to potential conflicts between algorithmic efficiency and democratic deliberation. The policy does not address balancing the speed and scale of AI-generated outputs with the time-intensive processes of democratic consultation, stakeholder engagement, and careful consideration that characterize legitimate public governance.

The policy also offers minimal guidance on managing AI-generated content in public communications that shape citizens' understanding of government processes. As NSW agencies potentially will use GenAI to draft communications, policy documents, or citizen responses, the ethics policy provides little specific guidance on ensuring transparency about AI authorship or distinguishing between human and AI-generated public communications.

5.3 Assessing the NSW AI Assurance Approach's democratic readiness

The NSW AI Assurance Approach (Digital.NSW, 2024d) is designed to operationalize the ethics policy through a structured assessment process for AI projects. However, its capacity to ensure

TABLE 2 Evaluation of NSW AI Assurance Framework's Democratic Governance Capacity.

Dimension	Assessment	Justification
Democratic oversight mechanisms	Partially adequate	The framework establishes review processes but lacks specific provisions for independent democratic oversight beyond departmental assessments.
Citizen contestability	Partially adequate	Somewhat strong emphasis on enabling citizens to challenge AI-influenced decisions, though implementation guidance is limited.
Political accountability	Inadequate	Insufficient clarity on how political responsibility is maintained when decisions incorporate AI-generated inputs.
Transparency requirements	Inadequate	While documentation is required, specific standards for explaining LLM outputs to citizens and democratic representatives are lacking.
Bias mitigation	Minimally adequate	The framework acknowledges bias concerns but provides limited guidance on identifying and addressing political or demographic biases.
Public participation	Largely absent	Limited provisions for citizen input into AI system design, deployment, or evaluation.

democratic oversight and political accountability varies significantly across different dimensions, as shown in Table 2.

This assessment reveals that while the NSW AI Assurance Framework provides a reasonable foundation for technical governance, significant gaps remain in its capacity to ensure democratic legitimacy and political accountability. Most notably, the framework lacks explicit provisions for citizen participation in AI governance, has limited mechanisms for independent oversight beyond departmental self-assessment, and has insufficient attention to how AI-influenced decisions remain politically accountable in democratic terms.

5.4 Comparative analysis with democratic AI governance models

When compared to emerging democratic AI governance models internationally, NSW's approach demonstrates both strengths and significant areas for improvement. The European Union's AI Act, for example, provides more robust democratic safeguards through a detailed risk classification system, stronger transparency requirements, and more precise political accountability mechanisms for high-risk AI systems (Wagner et al., 2023; Union, 2024). Similarly, the UK Government's Pro-Innovation Approach to AI Regulation addresses generative AI more explicitly, with specific provisions for maintaining democratic oversight and ensuring political accountability (GOV. UK, 2023).

Critics like Acemoglu have argued that effective democratic AI governance requires not just ethical principles but robust institutional mechanisms to ensure that AI development serves broader social welfare rather than narrow technical or bureaucratic interests (Acemoglu, 2021a; Acemoglu, 2021b). Against these critiques, NSW's framework does demonstrate a commitment to human oversight but lacks specific democratic institutions to ensure political accountability and citizen participation in AI governance.

Similarly, Marcus has emphasized the need for governance frameworks to specifically address the limitations of LLMs, including their propensity to generate plausible but false information that could undermine informed democratic deliberation (Marcus, 2023; Marcus, 2024). The NSW framework acknowledges these concerns but provides limited practical mechanisms for ensuring the accuracy of AI-generated content used in governance processes.

For NSW to maintain democratic legitimacy as GenAI becomes more deeply integrated into governance processes, its framework will need to evolve beyond technical and ethical considerations to address the political dimensions of AI governance more directly. This includes establishing clear lines of political accountability for AI-influenced decisions, creating meaningful opportunities for citizen participation in AI governance, and ensuring that democratic values of transparency, deliberation, and contestability are not sacrificed for algorithmic efficiency.

To effectively uphold democratic principles amid the growing influence of GenAI, governments must design systems that align with ethical standards and remain responsive to the political consequences of their deployment. As the following section illustrates, the moments when systems falter—whether through bias, opacity, or procedural disruption—often expose the fragility of existing governance frameworks. These breakdowns serve as critical entry points for understanding how socio-technical failures can undermine political legitimacy and where targeted repairs might reinforce democratic accountability (see Figure 2).

6 Breakdowns and repairs in GenAI implementation

Breakdowns in GenAI and other algorithmic models can appear in various forms—technical failures, ethical dilemmas, or moments when AI systems fall short of meeting public or operational expectations. Within a political governance context, such failures reveal technical shortcomings and fundamental challenges to democratic accountability, procedural fairness, and institutional legitimacy. This section thus analyzes historical examples of algorithmic governance breakdowns in Australia to extract lessons for NSW's emerging GenAI implementation, examining how socio-technical failures manifest as political governance challenges.

While NSW's current GenAI initiatives have not yet experienced significant documented failures, examining past AI implementation breakdowns provides crucial insights for anticipating and mitigating potential risks. Two prominent Australian cases—Revenue NSW's automated debt recovery system and the federal Robodebt scheme—offer relevant precedents illuminating potential governance vulnerabilities in algorithmically mediated administrative systems.

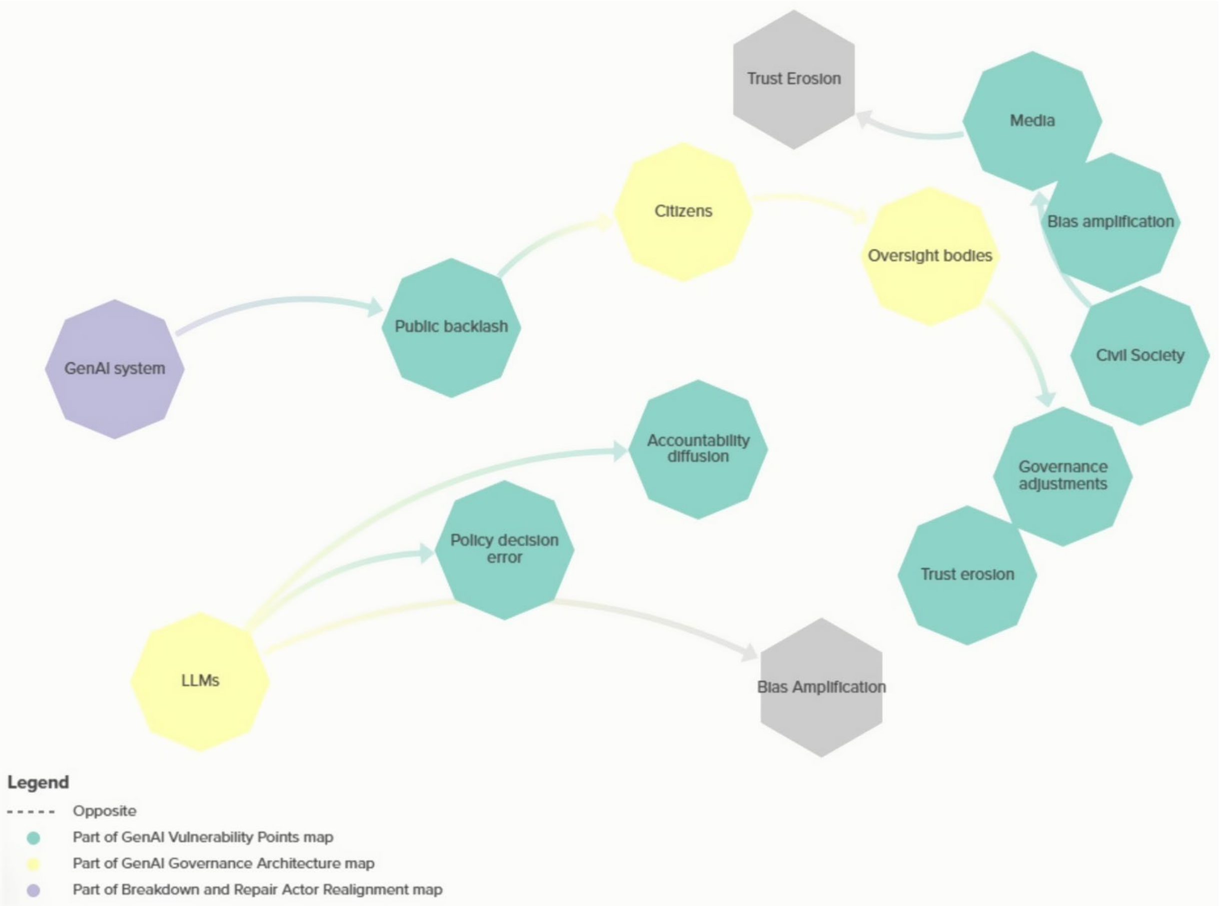

FIGURE 2
Vulnerability points in GenAI implementation. This diagram highlights potential governance vulnerabilities such as bias amplification, policy distortion, accountability gaps, and erosion of public trust. It visualizes how socio-technical tensions may emerge as GenAI becomes embedded in service delivery, design and decision-making. Summary map elaborated by the author using Kumu.io.

6.1 Revenue NSW automated system: undermining procedural fairness

The Revenue NSW automated debt recovery system, operational from 2016 to 2019, represents a significant case of digital algorithm-enabled administrative failure within NSW. The system was designed to automate the recovery of unpaid fines and debts through mechanisms including bank account garnishment and property seizures.

The system employed relatively simple automation rather than sophisticated AI, yet its implementation illuminates critical vulnerabilities relevant to more advanced GenAI applications in democratic governance:

a Algorithmic Inflexibility and Procedural Fairness: The system operated with rigid rule-based algorithms that could not adequately account for individual circumstances or vulnerabilities. Despite policy guidelines requiring consideration of financial hardship, the algorithmic implementation effectively overrode these considerations, resulting in automatic garnishment of bank accounts even for individuals experiencing severe financial distress.

b Diminished Human Oversight of Democratic Authority: Despite the significant consequences of its decisions, the system operated with minimal human review. The NSW Ombudsman found that in many cases, no meaningful human assessment occurred before automated enforcement actions were triggered (Jobberns and Guihot, 2024). This 'automation bias' led human operators to defer to the system's recommendations even in cases where intervention was warranted, effectively transferring democratic authority from accountable human officials to unaccountable technical systems.

c Democratic Transparency Deficits: Affected individuals received limited information about decisions or how they could contest them. The system's operations were essentially black-boxed from both the public and many frontline staff, making meaningful democratic contestation impossible. This opacity directly undermined citizens' ability to participate in decisions affecting their rights and interests—a fundamental requirement of democratic governance.

d Disproportionate Impact on Marginalized Communities: The system's effects were unequal. Analysis showed that Indigenous Australians, people with disabilities, and those experiencing homelessness or financial hardship were disproportionately impacted. This disproportionate impact reflected and potentially amplified existing power imbalances within democratic systems, raising fundamental questions about algorithmic systems' compatibility with democratic principles of equality and non-discrimination.

The Revenue NSW case demonstrates how algorithmic systems can undermine democratic governance when technical implementation fails to incorporate procedural fairness, transparency, and accountability. Although the current NSW AI Ethics Policy explicitly addresses many of these concerns, this example serves as a cautionary tale about the gap that can emerge between policy principles and technical implementation—a gap that could widen with more complex GenAI systems.

6.2 Robodebt: systemic failure through algorithmic governance

The federal 'Robodebt' scheme (officially the Online Compliance Intervention) represents one of Australia's most significant examples of algorithmic governance failure with profound political implications. The scheme operated from 2016 to 2019 and used an automated system to identify and recover alleged welfare overpayments by matching income data from the Australian Taxation Office with welfare payment records from Centrelink (Services Australia).

The Robodebt case illustrates how algorithmic systems can undermine fundamental democratic principles when implemented without appropriate political oversight:

a Flawed Algorithm Design and Democratic Decision-Making: The core breakdown in Robodebt stemmed from its reliance on income-averaging—a methodology that assumed consistent income across fortnightly periods based on annual tax data. This crude averaging approach produced systematic errors, creating false debt notices for individuals whose income fluctuated throughout the year (Carney, 2019). The decision to employ this flawed methodology revealed a privileging of technical efficiency over democratic principles of fairness and accuracy.

b Reversed Burden of Proof and Democratic Rights: The scheme shifted the burden of proof from the government to citizens, requiring individuals to disprove algorithmically generated debt calculations rather than requiring the government to verify debts before pursuing recovery (Clarke et al., 2024). This inversion of traditional administrative justice principles directly undermined democratic rights to procedural fairness and due process.

c Minimized Human Judgment in Democratic Governance: Despite dealing with complex welfare payment rules and individual circumstances, the system minimized human review. The Royal Commission into the Robodebt Scheme (Podger, 2023) found that human oversight was deliberately reduced over time, with staff pressured to process cases quickly rather than thoroughly assess them. This represented a conscious decision to transfer democratic authority from accountable human officials to unaccountable algorithmic systems.

d Opacity in Democratic Decision-Making: Affected individuals received debt notices with minimal explanation of how the amounts were calculated. The algorithmic process that determined debts was effectively black-boxed, making it difficult for recipients to understand or challenge determinations (Rinta-Kahila et al., 2024). This opacity directly undermined democratic principles of transparency and contestability.

e Institutional Resistance to Democratic Accountability: Perhaps most troublingly, the Robodebt scheme demonstrated remarkable resistance to feedback and correction. Despite early evidence of systematic errors, agency leaders defended the program and resisted calls for reform, highlighting how algorithmic systems can develop institutional momentum that resists democratic oversight and accountability (Podger, 2023)

The Royal Commission's findings highlighted the scheme's technical and political failures. The algorithm's flawed assumptions interacted with organizational culture, political priorities, and power imbalances to create a system that was resistant to democratic oversight and accountability (Podger, 2023).

6.3 Mapping historical failures to GenAI governance requirements

While Revenue NSW and Robodebt employed relatively simple algorithms compared to contemporary language models, their failures provide crucial insights for GenAI governance in democratic systems. Table 3 maps specific failure modes from these cases to potential GenAI risks and corresponding governance safeguards necessary to maintain democratic legitimacy.

This comparative analysis table reveals that while NSW's current AI governance framework addresses some of these risks in principle, significant gaps remain in practical implementation mechanisms. In particular, the framework provides limited guidance on:

1 How to conduct effective human oversight of increasingly sophisticated and persuasive LLM outputs that may appear more authoritative than human judgment

2 Specific explainability requirements for GenAI systems that operate through complex associations rather than explicit rules

3 Methodologies for detecting and mitigating biases in LLM-generated content that could systematically disadvantage specific communities

4 Protocols for independent assessment of GenAI systems' impacts on democratic rights and interests

5 Mechanisms for affected individuals to contest AI-influenced decisions that affect their rights and interests effectively (see Figure 3).

6.4 Articulation work and democratic repair in GenAI governance

After considering previous breakdowns and gaps in the frameworks, the concept of 'articulation work'—ongoing human efforts to adapt systems to real-world situations and contexts (Star,

TABLE 3 Mapping historical AI failures to GenAI Governance requirements.

Historical failure mode	Corresponding GenAI risk	Required democratic governance safeguard
Algorithmic inflexibility	LLMs may generate responses without considering individual contexts or exceptional circumstances	Mandatory human review of GenAI outputs in consequential decisions; documentation of context-specific considerations
Insufficient human oversight	More persuasive and authoritative-seeming LLM outputs may increase automation bias among human reviewers	Training in critical assessment of AI outputs; clear authority and responsibility for human override of AI recommendations
Transparency deficits	Increased complexity of LLMs may further obscure decision rationales from affected individuals	Explainability requirements tailored to generative AI; robust rights of access to information about AI use in decisions
Disproportionate impacts on vulnerable groups	LLMs trained on historical data may reproduce or amplify existing societal biases	Mandatory equity impact assessments; ongoing monitoring of disparate impacts by demographic groups
Reversed burden of proof	GenAI-generated analyses may be presumed correct without verification	Explicit policies requiring verification of AI-generated findings before consequential actions
Resistance to correction	Institutional investment in AI systems may create resistance to acknowledging limitations	Independent oversight mechanisms: channels for external experts and affected communities to flag concerns

1999)—provides a valuable framework for addressing the gaps identified earlier. In the context of GenAI governance, articulation work involves several critical dimensions that maintain democratic legitimacy in technically mediated governance systems:

1 Interpretive flexibility: Human operators must be empowered to interpret and contextualize LLM outputs rather than treating them as authoritative pronouncements. This requires both technical training and organizational cultures that value human judgment over algorithmic efficiency.
2 Boundary spanning: Effective articulation work requires individuals bridging technical, legal, ethical, and domain-specific expertise to evaluate GenAI outputs in context. NSW's training initiatives, like the Chatbot Prompt Essentials module (Digital.NSW, 2024c) represent initial steps in this direction but may need expansion to address the political dimensions of AI-mediated governance.
3 Vertical integration: Articulation work must occur at multiple levels—from frontline staff interpreting individual outputs to senior leaders evaluating system-wide patterns and impacts. NSW's current framework emphasizes frontline interpretation but provides less guidance on systematically evaluating structural impacts on democratic governance.
4 Reciprocal transparency: True articulation work requires explaining AI to humans and making human values and priorities legible to technical systems through thoughtful design and prompt engineering. This bidirectional transparency is essential for maintaining democratic values in systems that increasingly shape human governance decisions.

Drawing on Jackson (2014) framing of repair as a creative and transformative process, several potential repair mechanisms emerge as particularly relevant for maintaining democratic legitimacy in NSW's GenAI implementation:

1 Rapid response protocols: Establishing clear procedures for identifying, escalating, and addressing potential GenAI failures, with designated responsibility and authority for democratic intervention.
2 Feedback integration systems: Creating structured mechanisms to capture, analyse, and respond to patterns in GenAI system outputs, particularly identifying systematic errors or biases that affect democratic rights.
3 Collaborative repair forums: Establishing multi-stakeholder processes for addressing significant failures, including technical experts, policy specialists, affected communities, and oversight bodies to maintain democratic legitimacy through inclusive deliberation.
4 Public transparency about repair: Documenting and publicly reporting on system failures and repair processes, building trust through openness about limitations and improvements to maintain democratic legitimacy through transparency.
5 Continuous learning mechanisms: Systematically capturing insights from breakdown-repair cycles to inform both technical refinements and governance improvements, creating an adaptive learning system that can maintain democratic legitimacy in rapidly evolving technical environments.

NSW's current AI governance framework addresses some of these elements, particularly through its emphasis on monitoring and evaluation. However, the specific mechanisms for capturing, analyzing, and learning from GenAI failures remain underdeveloped. As the government advances its GenAI implementation, strengthening these repair mechanisms will be crucial for building resilient and trustworthy AI systems that maintain democratic legitimacy.

The lessons from Revenue NSW, Robodebt, and international cases underscore that technical systems cannot be separated from their social, organizational, and political contexts. Effective GenAI governance requires attention to technical specifications and the socio-technical networks in which these systems operate. As NSW continues to integrate GenAI into governance processes, maintaining this socio-technical perspective will be essential for anticipating vulnerabilities and developing effective repair mechanisms when breakdowns occur in ways that maintain democratic accountability, procedural fairness, and political legitimacy. The next section of this paper will engage with discussions and policy recommendations based on this broad analysis.

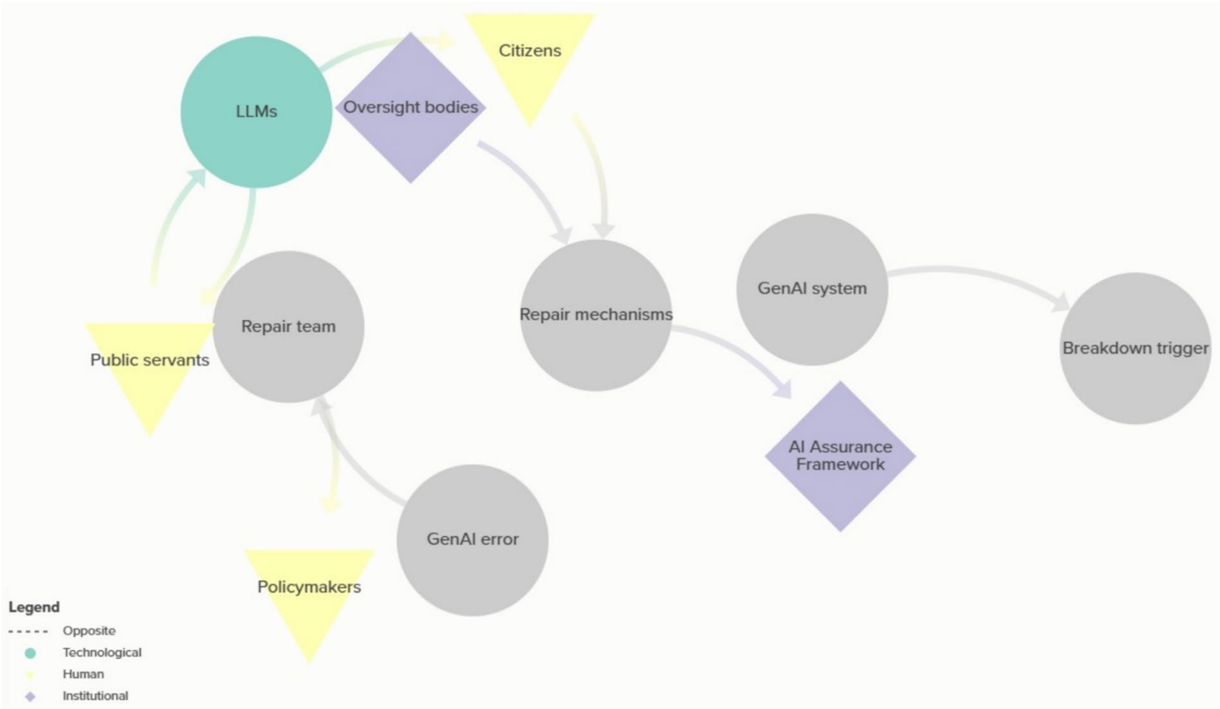

FIGURE 3
Breakdown and repair actor realignment. This map captures the reconfiguration of governance networks in response to socio-technical failures. It illustrates how repair actors, oversight bodies, and frontline staff respond to GenAI breakdowns by negotiating authority, trust, and accountability through new alignments. Summary map elaborated by the author using Kumu.io.

7 Recommendations: scenarios of GenAI as a political actant in NSW governance

To complement the document analysis, actor-network mapping, and the different identified processes, articulations and gaps, this section will address three scenarios that anticipate potential GenAI implementation breakdown and explore how governance networks might respond. These scenarios were developed based on the socio-technical dynamics identified through ANT mapping and informed by patterns observed in historical algorithmic failures while accounting for the unique characteristics of GenAI systems. By exploring these potential futures, vulnerabilities and governance challenges that might remain hidden until real-world breakdowns occur can be identified, this section will build recommendations for developing anticipatory governance approaches.

7.1 GenAI scenarios: anticipating breakdown and repair

Given that GenAI implementation in NSW is at an early stage with limited operational history, these scenarios provide a structured approach to anticipating how breakdowns might manifest and how governance networks might respond. The scenarios were developed based on the socio-technical dynamics identified through ANT mapping, drawing on historical patterns of algorithmic failure while accounting for the unique characteristics of GenAI systems.

Each scenario explores a different type of potential breakdown, examining how actors within NSW's governance network might realign or conflict in response and identifying potential repair mechanisms that could maintain democratic legitimacy. These scenarios are not predictions but analytical tools that illuminate potential vulnerabilities and governance challenges that might otherwise remain hidden until real-world breakdowns occur.

7.1.1 Scenario 1: LLM-generated error in policy development

In this scenario, a NSW department uses a GenAI system to analyse public submissions on a proposed policy change, generating a summary report that informs the final policy decision. However, the LLM introduces subtle but significant errors in its interpretation of citizen input—mischaracterizing opposition to specific measures as support, aggregating responses in ways that obscure key concerns from marginalized communities, and hallucinating patterns that align with its training data rather than the actual submissions. These errors influence the policy direction, leading to decisions that do not accurately reflect public sentiment.

This scenario highlights several critical vulnerabilities in GenAI integration:

1 Epistemic challenges: GenAI systems can present persuasive but inaccurate interpretations of data, particularly when processing unstructured information like public submissions. The black-box nature of LLMs makes it challenging to verify

the accuracy of these interpretations without labor-intensive human verification.

2 Democratic representation risks: When AI mediates between citizen input and policy decisions, it can inadvertently filter or distort citizen voices, potentially undermining representative democracy. Marginalized communities whose language patterns or concerns differ from dominant training data may be particularly affected.

3 Accountability diffusion: When errors are discovered, responsibility becomes diffused across the network—the GenAI system, the public servants who prompted it, the policymakers who relied on its outputs, and the technical team implementing it. This diffusion creates challenges for democratic accountability and remedy.

The ANT mapping reveals how such a breakdown would stress connections between key actors in the network. The relationship between public servants and citizens becomes mediated through the GenAI system, introducing new opacity forms. The formal authority of policymakers remains intact, but their exercise of that authority is shaped by the LLM's interpretation of citizen input. The AI Ethics Policy and Assurance Framework provide general principles but lack specific guidance for mitigating these risks in policy development contexts.

Potential repair mechanisms include:

- Implementing structured verification processes where human reviewers sample and verify AI interpretations against original citizen submissions
- Creating technical guardrails that flag potential hallucinations or biased interpretations
- Establishing clear protocols for citizen contestation of AI-generated summaries of public input
- Developing explicit role definitions that maintain human responsibility for accurate representation of citizen views

7.1.2 Scenario 2: public backlash to AI-driven service

In this scenario, a NSW agency implements a GenAI-powered chatbot as the primary interface for a public service, reducing wait times and increasing accessibility. However, citizens begin reporting that the system provides inconsistent information, appears to treat different demographic groups differently in its responses, and makes it challenging to reach human representatives when issues arise. Civil society organizations investigate and publish findings suggesting systematic bias in how the system interprets and responds to different communication styles, disadvantaging specific communities. The media amplifies these concerns, creating a public backlash that undermines trust in both the particular service and broader government AI initiatives.

This scenario illuminates several socio-technical vulnerabilities:

1 Visibility of bias: While human bias often remains implicit and challenging to detect systematically, algorithmic bias can become visible through patterns across many interactions, creating focal points for public criticism and undermining trust.

2 Accountability expectations: Citizens may hold government AI systems to higher standards than either private AI systems or traditional government services, expecting both the efficiency of automation and the flexibility of human judgment.

3 Remediation challenges: Once trust is broken, technical fixes alone may be insufficient to restore public confidence, requiring broader governance reforms and meaningful public engagement.

Citizens who previously interacted directly with public servants now navigate an AI interface that mediates access to services. Civil society organizations emerge as essential actors in the network, exercising informal oversight functions that highlight gaps in formal governance structures. The breakdown reveals the limitations of the AI Assurance Framework in anticipating and addressing bias in citizen-facing applications.

Potential repair mechanisms include:

- Creating transparent monitoring systems that track service outcomes across demographic groups.
- Establishing clear pathways for citizens to reach human representatives when the AI system fails to address their needs.
- Implementing formal channels for civil society organizations to flag potential biases or failures.
- Developing participatory processes for affected communities to help redesign the system to address their needs better.

7.1.3 Scenario 3: governance conflict over AI-generated administrative decisions

In this scenario, a NSW agency uses GenAI to draft administrative decisions in a high-volume regulatory context, with public servants reviewing and approving the drafts before finalization. Over time, the agency increases efficiency targets, reducing the time allocated for human review. When a controversial decision is challenged through formal appeal processes, investigation reveals that the GenAI system had incorporated subtle reasoning errors that the human reviewer missed under time pressure. The incident triggers a conflict between oversight bodies that emphasize procedural fairness and agency leadership focused on operational efficiency, raising fundamental questions about the appropriate balance between human and algorithmic authority in administrative decision-making.

This scenario highlights several governance challenges:

1 Automation bias: The persuasive language and authoritative tone of GenAI outputs can lead human reviewers to defer to algorithmic recommendations even when they should exercise independent judgment, particularly under institutional pressure to improve efficiency.

2 Governance tensions: Different actors within the governance network may hold competing priorities—efficiency versus procedural fairness, innovation versus caution—creating tensions that surface when breakdowns occur.

3 Statutory interpretation: Existing administrative law frameworks may provide insufficient guidance for determining appropriate human oversight of GenAI systems, creating legal uncertainties about validity and review rights.

Agency leadership and oversight bodies that previously operated with a shared understanding of proper administrative process now conflict over the appropriate role of GenAI. The AI Ethics Policy provides normative principles supporting human oversight, but practical implementation pressures push against these principles. The breakdown exposes gaps in the governance architecture where technical implementation decisions interact with fundamental questions of administrative legitimacy.

Potential repair mechanisms include:

- Establishing clear minimum standards for human review of AI-generated administrative decisions
- Creating monitoring systems that track review time and modification rates to identify potential automation bias
- Implementing periodic audits that test the quality of human oversight through the deliberate introduction of errors
- Developing specific training for administrative decision-makers on identifying and correcting potential errors in GenAI outputs

These scenarios illuminate how breakdowns in GenAI implementation could manifest in NSW governance, revealing potential vulnerabilities that might not be apparent through document analysis alone. By mapping how actors would realign in response to these breakdowns, the study identifies specific gaps in current governance frameworks and potential repair mechanisms that could maintain democratic legitimacy in the face of GenAI-related challenges.

The scenarios also demonstrate the value of an anticipatory governance approach that prepares for potential failures before they occur. By identifying vulnerabilities early and developing repair mechanisms in advance, NSW can build resilience into its GenAI governance frameworks, enabling faster and more effective responses when real-world breakdowns inevitably occur. This approach aligns with the concepts of breakdown and repair from infrastructure studies, viewing failures not as endpoints but as opportunities for transformative reconfiguration of socio-technical relationships.

In navigating these tensions, NSW has an opportunity to pioneer governance approaches that maintain democratic legitimacy while capturing the benefits of AI-enhanced administration. By developing mechanisms that address the specific challenges posed by GenAI as an actant in governance networks, NSW can establish models that other jurisdictions can adapt to their contexts. The recommendations outlined below provide a starting point for this journey toward democratic AI governance that enhances rather than undermines the public's trust in government.

7.2 Required governance mechanisms

Drawing from the three scenarios of potential GenAI breakdown in NSW governance, a regulatory sandbox (Allen, 2019; Zetzsche et al., 2017) approach emerges as a promising framework for balancing innovation with democratic safeguards. This approach builds directly on the insights gained from analyzing how LLM-generated errors in policy development, public backlash to AI-driven services, and governance conflicts over AI-generated administrative decisions might manifest in practice.

The regulatory sandbox creates controlled testing environments where innovative GenAI applications undergo time-limited implementation with modified regulatory requirements while maintaining core democratic protections. In response to the scenario of LLM-generated errors in policy analysis, the sandbox implements assessment criteria that specifically evaluate how GenAI systems interpret and aggregate citizen input, with verification protocols that compare AI-generated summaries against original submissions. This addresses the epistemic challenges and representation risks identified when AI mediates between citizen voices and policy decisions.

Transparency mechanisms within the sandbox framework directly respond to public backlash by establishing visibility into GenAI systems' operations. The framework requires documentation of training data sources and limitations, mandates clear disclosure when citizens interact with AI systems, and emphasizes accessible explanations of AI-influenced decisions. These measures create the conditions for early identification of potential bias patterns before they trigger the cascading trust failures depicted in the scenario analysis.

Human oversight protocols address the vulnerabilities in the administrative decision-making scenario, particularly the risk of automation bias under efficiency pressures. The sandbox establishes minimum standards for human review of AI-generated content, creates monitoring systems that track review time and modification rates, and implements periodic audits that test the quality of oversight through deliberate error introduction. These measures maintain the balance between efficiency gains and procedural fairness requirements.

Democratic accountability structures within the sandbox respond to the diffusion of responsibility identified across all three scenarios. An independent oversight body reviews implementations within the sandbox environment, regular public reporting creates transparency into the experimentation process, and statutory review rights ensure citizens can contest AI-influenced decisions. These mechanisms prevent the accountability gaps that emerged when responsibility became distributed across technical systems and human operators.

The sandbox approach fundamentally depends on the participatory mechanisms that were notably absent in the scenario breakdowns. It incorporates citizen consultation on significant AI deployments, equity impact assessments to evaluate potential disproportionate effects on marginalized communities, and feedback channels for reporting concerns. These participation structures ensure that repair mechanisms are built into the sandbox design rather than improvised after breakdowns occur.

NSW can create controlled conditions for testing GenAI applications that anticipate and mitigate the specific vulnerabilities identified in the scenario analysis through this regulatory sandbox framework. This approach enables the validation of governance mechanisms before permanent implementation, generating valuable evidence about effective oversight approaches while maintaining democratic legitimacy during the innovation process.

8 Conclusion: towards democratic governance of generative AI

This paper has examined the integration of Generative Artificial Intelligence within NSW Government processes through a socio-technical lens, conceptualizing GenAI systems as passive tools and active actants within governance networks. The Actor-Network Theory framework has illuminated how these systems reshape political relationships, redistribute authority, and reconfigure accountability mechanisms in ways that challenge traditional governance models predicated on exclusively human agency.

The research reveals that while NSW has developed a comprehensive AI governance architecture through its AI Strategy, Ethics Policy, and Assurance Framework, significant gaps remain in addressing the unique challenges posed by GenAI systems. These sophisticated language models function as potent mediators that transform workflows and decision-making processes in ways that formal governance structures may not fully capture. As demonstrated through the analysis of training initiatives like the Chatbot Prompt Essentials module, the practical implementation of GenAI suggests a more collaborative relationship where public servants shape AI outputs rather than sovereign decision-makers merely reviewing AI recommendations.

This shifting dynamic, however, creates several vulnerabilities in democratic governance. Transparency deficits emerge from the black-box nature of LLMs, creating barriers to understanding how these systems translate policy intentions into administrative practices. Bias amplification risks reproducing or intensifying existing social inequalities in more subtle and persuasive forms than earlier algorithmic systems. Accountability challenges arise when decisions incorporate inputs from GenAI systems, complicating the determination of responsibility for outcomes. Procedural fairness risks surface when systems operate with implicit assumptions that may disadvantage specific communities, particularly when their training data does not adequately represent the diversity of citizen experiences.

The historical analysis of algorithmic failures in Australian public administration—including Revenue NSW's automated debt recovery system and the federal Robodebt scheme—demonstrates the profound consequences when technical systems undermine democratic principles of procedural fairness, transparency, and contestability. While these earlier systems employed relatively simple algorithms compared to contemporary language models, their failures provide crucial insights for GenAI governance in democratic systems.

The scenario analysis explored potential breakdowns in GenAI implementation—including LLM-generated errors in policy development, public backlash to AI-driven services, and governance conflicts over AI-generated administrative decisions. These scenarios illuminate how GenAI vulnerabilities might manifest in practice and how governance networks might respond. They underscore the value of an anticipatory governance approach that prepares for potential failures before they occur, viewing breakdowns not as endpoints but as opportunities for transformative reconfiguration of socio-technical relationships.

The concept of 'articulation work'—ongoing human efforts to adapt systems to real-world contexts—provides a valuable framework for maintaining democratic legitimacy in technically mediated governance systems. Practical articulation work requires interpretive flexibility that empowers human operators to contextualize LLM outputs rather than treating them as authoritative pronouncements. It demands boundary spanning that bridges technical, legal, ethical, and domain-specific expertise to evaluate GenAI outputs in context. It necessitates vertical integration of oversight across organizational levels and reciprocal transparency that both explains AI to humans and makes human values legible to technical systems.

For NSW to maintain democratic legitimacy as GenAI becomes more deeply integrated into governance processes, its framework must evolve beyond technical and ethical considerations to address political dimensions more directly. This includes establishing clear lines of political accountability for AI-influenced decisions, achieved by recognizing and nominating GenAI as an actant, and creating meaningful opportunities for citizen participation in AI governance, and ensuring that democratic values of transparency, deliberation, and contestability are not sacrificed for algorithmic efficiency.

The research contributes to scholarly debates on AI governance by demonstrating the insufficiency of purely technical or ethical frameworks that do not address the political dimensions of AI integration. Effective AI governance requires technical safeguards, ethical principles and a fundamental reconsideration of how agency, accountability, and democratic legitimacy operate in human-AI governance networks. As Jasanoff (2016) has argued, technologies are not merely tools but active participants in constituting social order and political relationships.

As GenAI systems become increasingly embedded in administrative processes, maintaining democratic oversight, human judgment, and public contestability becomes more crucial. The NSW Government has an opportunity to pioneer governance approaches that capture the benefits of AI-enhanced administration while preserving the democratic values that legitimize public governance. By developing mechanisms that address the specific challenges posed by GenAI as an actant in governance networks, NSW can establish models that other jurisdictions can adapt to their own contexts.

This research underscored that the path toward democratic governance of GenAI is neither purely technical nor exclusively political but fundamentally socio-technical. It requires governance frameworks that recognize the active role of these systems in reshaping political relationships and decision-making processes. By anticipating potential breakdowns, developing robust repair mechanisms, and creating governance structures that maintain democratic legitimacy in the face of technological change, NSW can ensure that GenAI enhances rather than undermines public trust in government.

The wicked problems (Head, 2019; Head and Alford, 2015; Rittel and Webber, 1974) of governance cannot be solved through computational power alone. Rather, maintaining democratic legitimacy in AI-augmented governance requires ongoing articulation work—the human effort of contextualizing, interpreting, and evaluating algorithmic outputs within broader social and political values. As GenAI systems become increasingly embedded in administrative processes, the regulatory sandbox approach proposed in this research offers a promising framework for balancing innovation with democratic safeguards, ensuring that technological efficiency does not come at the cost of transparency, accountability, and the contestability essential to democratic governance.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LL-P: Validation, Data curation, Conceptualization, Project administration, Methodology, Formal analysis, Investigation, Writing – review & editing, Software, Supervision, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript. During the preparation of this work, the author used Anthropic's Claude 3.5 and 3.7 Sonnet to help with organising manuscript sections and improving readability of complex concepts. The author reviewed and edited all AI-generated content, verified all sources independently, and takes full responsibility for the published content.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acemoglu, D. (2021a). Harms of AI: National Bureau of Economic Research. Cambridge, MA, USA.

Acemoglu, D. (2021b). Redesigning AI: MIT Press. Cambridge, MA, USA: The MIT Press.

Allen, H. J. (2019). Regulatory sandboxes. *Geo. Wash. L. Rev.* 87:579. doi: 10.2139/ssrn.3056993

Anthis, J., Lum, K., Ekstrand, M., Feller, A., D'amour, A., and Tan, C. (2024). The impossibility of fair LLMS. Arxiv[Preprint]. doi: 10.48550/arXiv.2406.03198

Bozkurt, A. (2024). Tell me your prompts and I will make them true: the alchemy of prompt engineering and generative AI. *Int. Council Open Dist. Educ.* 16, 111–118. doi: 10.55982/openpraxis.16.2.661

Brown, N. B. (2024). Enhancing Trust in LLMS: algorithms for comparing and interpreting LLMS. Arxiv [Preprint]. doi: 10.48550/arXiv.2406.01943

Carney, T. (2019). Robo-debt illegality: the seven veils of failed guarantees of the rule of law? *Alternat. Law J.* 44, 4–10. doi: 10.1177/1037969X18815913

Cheong, B. C. (2024). Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Front. Human Dynam.* 6:1421273. doi: 10.3389/fhumd.2024.1421273

Clarke, R., Michael, K., and Abbas, R. (2024). Robodebt: a socio-technical case study of public sector information systems failure. *Australas. J. Inf. Syst.* 28:4681. doi: 10.3127/ajis.v28.4681

Cohen, T., and Suzor, N. (2024). Contesting the public interest in AI governance. *Internet Policy Rev.* 13:1794. doi: 10.14763/2024.3.1794

Digital.NSW. (2024a). Artificial intelligence ethics policy [online]. Available online at: https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-ethics-policy (Accessed February 17, 2025).

Digital.NSW. (2024b). Artificial intelligence strategy [online]. Available online at: https://www.digital.nsw.gov.au/policy/artificial-intelligence/artificial-intelligence-strategy (Accessed February 17, 2025).

Digital.NSW. (2024c). Chatbot prompt essentials [online]. Available online at: https://www.digital.nsw.gov.au/policy/artificial-intelligence/chatbot-prompt-essentials (Accessed February 17, 2025).

Digital.NSW. (2024d). NSW artificial intelligence assurance approach [online]. Available online at: https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assessment-framework (Accessed February 17, 2025).

GOV.UK (2023). A pro-innovation approach to AI regulation. Department for Science, Innovation and Technology and Office for Artificial Intelligence. Government of the United Kingdom.

Große, C. (2023). A review of the foundations of systems, infrastructure and governance. *Saf. Sci.* 160:106060. doi: 10.1016/j.ssci.2023.106060

Gutiérrez, J. L. M. (2023). On actor-network theory and algorithms: Chatgpt and the new power relationships in the age of AI. *AI Ethics* 4, 1071–1084. doi: 10.1007/s43681-023-00314-4

Head, B. W. (2019). Forty years of wicked problems literature: forging closer links to policy studies. *Polic. Soc.* 38, 180–197. doi: 10.1080/14494035.2018.1488797

Head, B. W., and Alford, J. (2015). Wicked problems: implications for public policy and management. *Admin. Soc.* 47, 711–739. doi: 10.1177/0095399713481601

Jackson, S. J. (2014). "Rethinking repair" in Media technologies: Essays on communication. eds. T. Gillespie, P. J. Boczkowski and K. A. Foot (Cambridge, MA, USA: MIT Press).

Janssen, M., and Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Elsevier* 33, 371–377. doi: 10.1016/j.giq.2016.08.011

Jasanoff, S. (2016). The ethics of invention: Technology and the human future. New York, NY, USA: WW Norton & Company.

Jobberns, H., and Guihot, M. (2024). Digital governance and neoliberalism: the evolution of machine learning in Australian public policy. *Law Technol. Hum.* 6, 29–52. doi: 10.5204/lthj.3408

Justesen, L. (2020). "Actor-network theory as analytical approach" in Qualitative analysis: Eight approaches for the social sciences. eds. M. Järvinen and N. Mik-Meyer (Thousand Oaks, CA: Sage), 327–244.

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. *Shap. Technol.* 1, 225–258.

Latour, B. (1996). On actor-network theory: a few clarifications. *Soziale Welt*, 47:369–381.

Latour, B. (2007). Reassembling the social: An introduction to actor-network-theory. Oxford: Oxford University Press.

Marcus, G. (2023). Controlling AI. *Commun. ACM* 66, 6–7. doi: 10.1145/3613250

Marcus, G. F. (2024). Taming Silicon Valley: How we can ensure that AI works for us. Cambridge, MA, USA MIT Press.

Mergel, I., Dickinson, H., Stenvall, J., and Gasco, M. (2024). Implementing AI in the public sector. *Public Manag. Rev.*, 26:1–14. doi: 10.1080/14719037.2023.2231950

Nimmo, R. (2011). Actor-network theory and methodology: social research in a more-than-human world. *Methodol. Innov.* 6, 108–119. doi: 10.4256/mio.2011.010

Pane, J. (2023). 2023 Australian community attitudes to privacy survey highlights 'rapidly shifting' landscape. Portsmouth, NH, USA: IAPP.

Papilloud, C. (2018). "Bruno Latour and relational sociology" in The Palgrave handbook of relational sociology. ed. F. Dépelteau, Palgrave Macmillan, Cham. 183–197.

Parliament NSW (2024). Artificial intelligence in New South Wales. Sydney, NSW, Australia: Parliament NSW.

Podger, A. (2023). Report to the Royal Commission into the Robodebt scheme.

Reid, A., O'callaghan, S., and Lu, Y. (2023). Implementing Australia's AI ethics principles: A selection of responsible AI practices and resources: CSIRO. Gradient Institute and Commonwealth Scientific and Industrial Research Organisation (CSIRO).

Rinta-Kahila, T., Someh, I., Gillespie, N., Indulska, M., and Gregor, S. (2024). Managing unintended consequences of algorithmic decision-making: the case of Robodebt. *J. Inf. Technol. Teach. Cases* 14, 165–171. doi: 10.1177/20438869231165538

Rittel, H. W., and Webber, M. M. (1974). Wicked problems: Man made. *Futures* 26, 272–280.

Rozado, D. (2024). The political preferences of LLMS. *PLoS One* 19:e0306621. doi: 10.1371/journal.pone.0306621

Saheb, T., and Saheb, T. (2024). Mapping ethical artificial intelligence policy landscape: a mixed method analysis. *Sci. Eng. Ethics* 30:9. doi: 10.1007/s11948-024-00472-6

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., and Chadha, A. (2024). A systematic survey of prompt engineering in large language models: techniques and applications. Arxiv [Preprint]. doi: 10.48550/arXiv.2402.07927

Singhal, A., Neveditsin, N., Tanveer, H., and Mago, V. (2024). Toward fairness, accountability, transparency, and ethics in AI for social media and health care: scoping review. *JMIR Med. Inform.* 12:e50048. doi: 10.2196/50048

Star, S. L. (1999). The ethnography of infrastructure. *Am. Behav. Sci.* 43, 377–391. doi: 10.1177/00027649921955326

Taeihagh, A. (2021). Governance of artificial intelligence. *Polic. Soc.* 40, 137–157. doi: 10.1080/14494035.2021.1928377

Union (2024). Regulation (EU) 2024/1689 of the European Parliament and of the council of 13 June 2024 laying down harmonised rules on artificial intelligence (Eur-Lex, European Union: artificial intelligence act).

Veale, M., and Brass, I. (2019). Administration by algorithm? Public management meets public sector machine learning. *Public Manag. Mach. Learn.* 121–149. doi: 10.1093/oso/9780198838494.003.0006

Wagner, M., Borg, M., and Runeson, P. (2023). Navigating the upcoming European Union AI act. *IEEE Softw.* 41, 19–24.

Wang, L., Chen, X., Deng, X., Wen, H., You, M., Liu, W., et al. (2024). Prompt engineering in consistency and reliability with the evidence-based guideline for LLMS. *Npj Digit. Med.* 7:41. doi: 10.1038/s41746-024-01029-4

Zetzsche, D. A., Buckley, R. P., Barberis, J. N., and Arner, D. W. (2017). Regulating a revolution: from regulatory sandboxes to smart regulation. *Fordham J. Corp. Fin. L.* 23:31.