Responsible AI beyond compliance: governing everyday use in the public sector

Transforming Government: People, Process and Policy

Gideon Mekonnen Jonathan
Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden

Received 6 June 2026
Revised 2 July 2026
22 July 2026
13 August 2026
26 August 2026
Accepted 27 August 2026

Abstract

Purpose – Artificial intelligence (AI) is increasingly embedded in public sector organisations, yet governing its use remains challenging due to tensions between innovation, accountability and institutional constraints. While prior research has largely focused on normative frameworks and ethical principles, there is limited empirical understanding of how AI governance is enacted in practice. This study aims to address this gap by examining how public sector organisations dynamically align AI use with governance and accountability requirements.

Design/methodology/approach – The study adopts an interpretive single-case study design within a public-sector organisation. Drawing on a socio-technical and process-oriented perspective, it analyses how everyday AI practices and organisational control mechanisms interact over time. Data were collected through semi-structured interviews across hierarchical levels and complemented by documentary analysis.

Findings – The findings reveal a persistent governance–practice gap, wherein employees adopt AI through informal, bottom-up practices that often diverge from formal governance structures. In response, governance evolves incrementally through layered and adaptive mechanisms, including technical controls, strengthened human oversight and informal managerial guidance. Rather than displacing accountability, AI use reinforces human judgement and intensifies verification practices. Governance is thus not enacted as a static system of compliance, but emerges through recursive interactions between practice and control.

Research limitations/implications – This study is based on a single interpretive case, which limits statistical generalisability and calls for further validation across contexts. The reliance on interview data may introduce bias, particularly regarding informal and sensitive practices such as shadow AI use. Future research should adopt comparative and longitudinal designs to examine how AI governance evolves over time and across institutional settings. The findings also highlight the need for theory development that integrates socio-technical dynamics with governance processes, encouraging further exploration of dynamic alignment in diverse organisational and technological environments.

Practical implications – The findings suggest that effective AI governance requires a shift from restrictive, compliance-focused approaches towards enabling and adaptive strategies. Public sector organisations should provide approved AI tools, clear usage guidelines and controlled environments (e.g. sandboxes) to support transparent and responsible use. Governance mechanisms should be embedded within everyday workflows and supported by strong human oversight, review practices and accountability structures. Importantly, organisations must recognise the increased effort required for verification and invest in training and organisational readiness. Treating governance as an ongoing capability rather than a static policy is essential for managing AI responsibly.

Social implications – The findings highlight important implications for accountability, transparency and public trust in AI-enabled government. By showing that AI use intensifies rather than replaces human judgment, the study underscores the continued centrality of human responsibility in public decision-making.

Transforming Government: People, Process and Policy
Emerald Publishing Limited
1750-6166
DOI 10.1108/TG-06-2026-0318

However, the persistence of informal and “shadow” AI practices raises concerns about uneven access, reduced transparency and potential risks to fairness and oversight. The results suggest that enabling and practice-informed governance approaches are essential to ensure that AI adoption in the public sector supports legitimacy, inclusiveness and responsible innovation in the service of broader societal values.

Originality/value – This study contributes to the information systems literature by conceptualising AI governance as a process of dynamic alignment between organisational practices and governance mechanisms. It extends existing work by providing an empirically grounded account of how governance co-evolves with AI use under conditions of institutional constraint and limited organisational readiness. The findings highlight the importance of socio-technical dynamics, human-in-the-loop oversight and adaptive governance approaches in achieving responsible AI in the public sector.

Keywords Artificial intelligence, AI governance, Responsible AI, Generative AI, Public sector, Digital government

Paper type Research paper

1. Introduction

1.1 Background

Artificial intelligence (AI) is increasingly embedded within public sector organisations, reshaping how information is processed, decisions are made and services are delivered. Across domains such as healthcare, taxation and urban governance, AI-enabled systems are deployed to augment decision-making, automate administrative tasks and enhance service efficiency (Al-Besher and Kumar, 2022). As part of broader digital transformation agendas, AI is positioned as a key driver of public value creation, with its effectiveness contingent on organisational capabilities such as data quality, institutional capacity and managerial commitment (Jonathan *et al.*, 2025).

However, the integration of AI into public-sector contexts introduces governance challenges that go beyond those typically encountered in private organisations. Public institutions operate under conditions of democratic accountability, legal scrutiny and normative expectations of fairness, transparency and legitimacy. Consequently, AI deployment is not merely a technical or operational concern but a matter of public responsibility and institutional trust (Wirtz *et al.*, 2021). These challenges are amplified by the opacity, adaptability and probabilistic nature of AI systems, particularly when their outputs inform decisions that have direct consequences for citizens.

In response, policymakers and scholars have advanced the notion of responsible AI, emphasising principles such as transparency, accountability and human oversight. However, the operationalisation of these principles remains contested. Regulatory approaches, including the European Union’s AI Act, tend to frame trustworthiness in terms of risk management and compliance, often overlooking the socially constructed and context-dependent nature of trust (Laux *et al.*, 2024).

At the organisational level, AI governance has emerged as the primary mechanism for translating ethical principles into practice. However, the existing literature remains conceptually fragmented and empirically limited. There is little consensus on how AI governance should be defined or structured, and only limited insight into how governance mechanisms are enacted *in situ* (Mäntymäki *et al.*, 2022; Birkstedt *et al.*, 2023).

This limitation is particularly consequential in light of the increasing accessibility of AI technologies, especially generative AI, which enables employees to incorporate these tools into everyday work practices, often beyond formal governance arrangements. This creates a fundamental tension between formal governance structures designed to ensure accountability and the emergent, practice-based appropriation of AI that enables flexibility and innovation.

1.2 Research aim and question

Addressing this gap, this study develops an empirically grounded, process-oriented account of AI governance in the public sector. Drawing on a socio-technical perspective, it conceptualises governance not as a static system of rules and controls, but as an emergent and evolving process shaped through recursive interactions between everyday AI use and organisational control mechanisms.

The following research question guides the study:

RQ1. How do public sector organisations dynamically align the use of AI with governance and accountability requirements as these technologies become embedded in everyday work practices?

This study makes three contributions to the IS literature. Firstly, it extends research on IT governance and digital innovation by developing a process-oriented, socio-technical account of AI governance that explains how governance is enacted and adapted in practice. Secondly, it develops and empirically grounds the concept of dynamic alignment as a recursive process linking everyday AI use and organisational control. Thirdly, it offers evidence-informed guidance for shifting public sector governance from restriction towards enabling, adaptive approaches.

2. Literature review and theoretical framework

2.1 The governance–practice gap in responsible AI

Efforts to institutionalise responsible AI in the public sector have largely centred on developing high-level ethical principles and regulatory frameworks that emphasise transparency, accountability and human oversight (De Sousa *et al.*, 2019). For example, Laux *et al.* (2024) argue that regulatory approaches such as the European Union's AI Act equate trustworthiness with acceptable levels of risk, thereby reducing trust to a compliance-oriented construct.

This critique reveals a broader limitation in existing approaches to AI governance: the implicit assumption that formal compliance mechanisms are sufficient to ensure responsible AI use.

As a result, AI adoption frequently outpaces formal governance arrangements, giving rise to informal or "shadow" use (Waters-Lynch *et al.*, 2025). We conceptualise this phenomenon as a governance–practice gap, defined as the persistent misalignment between formally articulated governance mechanisms and the situated enactment of AI in everyday organisational practice.

Structural constraints further reinforce this gap. Public sector organisations face technological, organisational and environmental barriers that hinder the systematic implementation of AI (Rjab *et al.*, 2023), as well as heightened exposure to ethical, legal and regulatory risks (Wirtz *et al.*, 2021).

2.2 Socio-technical dynamics of AI use and governance

To explain how the governance–practice gap emerges and evolves, this study adopts a socio-technical and practice-based perspective rooted in Information Systems research. Socio-technical systems theory conceptualises organisational outcomes as emerging from the recursive interaction between digital artefacts, human actors and organisational structures.

At the micro level, employees engage with AI in situated and pragmatic ways.

At the organisational level, actors responsible for governance respond by introducing and adapting control mechanisms, including formal policies, technical restrictions, training programmes and oversight procedures. Efforts to govern AI through traditional control structures, therefore, encounter inherent limitations.

Human oversight is widely proposed as the safeguard for responsible AI, but its limits are increasingly recognised. Automation complacency and bias can undermine human review, particularly under workload and are difficult to eliminate through training alone (Parasuraman and Manzey, 2010); oversight requirements may then create a false sense of security when reviewers cannot meaningfully scrutinise outputs (Green, 2022). AI also reconfigures administrative discretion and human-in-the-loop work, redistributing rather than removing judgement (Bullock, 2019; Grønsund and Aanestad, 2020). Verification effort is therefore not the same as effective oversight – a distinction we develop in the findings and discussion.

The interaction between micro-level practices and organisational control mechanisms gives rise to a dynamic tension between agency and control. Instead, governance must be understood as an emergent and evolving process shaped through recursive interactions between practice and control.

2.3 Legitimacy, accountability and public value

In the public sector, these socio-technical dynamics are embedded within broader institutional demands for legitimacy and public value. Public organisations operate under conditions of democratic accountability and are subject to legal, ethical and societal expectations that extend beyond efficiency and performance.

Existing research identifies multiple categories of risk associated with AI in the public sector. These risks are directly linked to legitimacy: when AI systems inform or automate decisions affecting citizens, opacity and complexity can undermine trust in public institutions (König and Wenzelburger, 2021).

From an information systems perspective, these legitimacy challenges are not external constraints but are enacted through socio-technical processes.

These dynamics implicate public value, which requires that organisations create value recognised as legitimate by citizens and authorising environments, not merely deliver efficient outputs (Moore, 1997); in digital government and particularly in Global South settings, realising public value is conditioned by institutional capacity, infrastructure and equity of access (Twizeyimana and Andersson, 2019). Accountability itself is best understood as a relational obligation to explain and justify conduct to a forum that can pass judgement (Bovens, 2007).

AI governance in the public sector, therefore, cannot be reduced to compliance or risk management.

2.4 Theoretical positioning

We define dynamic alignment as the provisional, practice-driven process through which an organisation continuously reconciles emergent AI use with its governance and accountability requirements. To avoid conflation, we distinguish four facets: as a process, it is the ongoing reconciliation of use and control; its mechanism is recursive feedback, whereby practice exposes the limits of controls and controls reshape practice; its provisional outcome is a temporary, never-settled state of "good enough" alignment in which formal and shadow practices coexist; and as an organisational capability, it is the capacity to sustain this reconciliation over time. Throughout, "dynamic alignment" refers to the process, with the other facets named where intended. The theorisation is abductive, moving iteratively between case patterns and established theory.

To aid precision, we distinguish levels of the framework: dynamic alignment is the central construct (the process); recursive feedback is its overarching mechanism; displacement, shadow normalisation and verification intensification are empirically observed sub-mechanisms; the five themes in Section 4 are the empirical manifestations; and the public-sector conditions in Figure 1 are the contextual boundary conditions.

Figure 1 presents the theoretical model developed in this study. At the micro level, employees engage in situated, experimental and sometimes informal uses of AI that may diverge from formal governance structures, generating a governance–practice gap. At the organisational level, governance actors respond by introducing and adapting control mechanisms, including policies, oversight structures and technical controls.

2.5 Distinguishing dynamic alignment from related constructs

The concept of dynamic alignment advanced here shares a family resemblance with several established constructs and its distinctiveness warrants clarification. Adaptive governance, developed primarily in the study of social-ecological systems, describes how institutions adjust to environmental change through experimentation, learning and polycentric decision-making (Folke *et al.*, 2005). While dynamic alignment likewise foregrounds learning and adjustment, it is concerned specifically with the recursive coupling between situated technology use and organisational control, rather than with the ecological resilience of institutions as such. Similarly, the dynamic capabilities perspective explains how organisations sense, seize and reconfigure resources to sustain advantage in changing environments (Teece, 2007). Dynamic capabilities are an organisational-level property oriented towards strategic renewal; dynamic alignment, by contrast, is a governance process oriented towards reconciling accountability with emergent use, enacted as much through frontline practice as through managerial strategy.

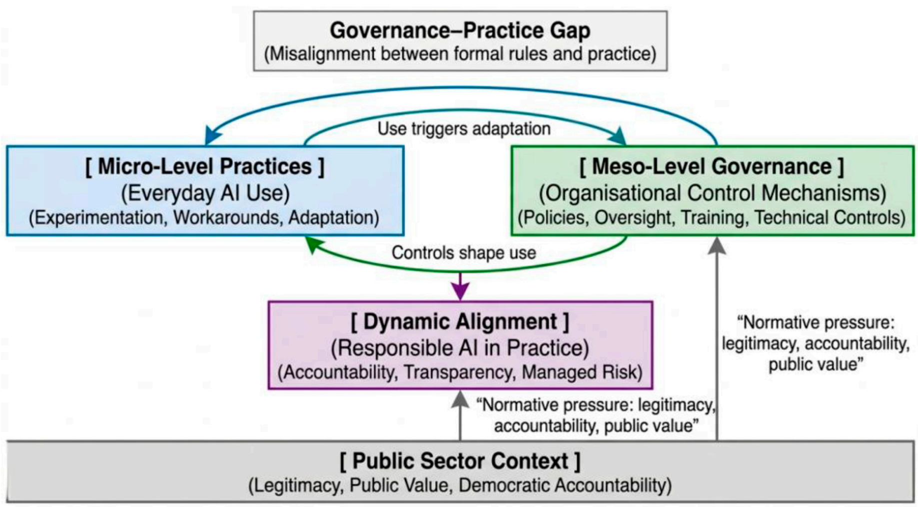

Figure 1. A socio-technical model of AI governance as dynamic alignment between practice and control

Dynamic alignment is also distinct from the tradition of strategic and IT governance alignment, which examines the fit between business strategy, IT strategy and organisational infrastructure (Henderson and Venkatraman, 1993). That literature treats alignment largely as a desired end-state to be engineered through planning and structural design. Our formulation departs from this in two respects. Firstly, alignment is treated not as an achievable equilibrium but as a provisional and continuously renegotiated accomplishment. Secondly, alignment is driven from the bottom up by the everyday appropriation of technology, consistent with practice-based and technology-in-practice perspectives (Orlikowski, 2000; Leonardi, 2011; Faraj *et al.*, 2018), rather than being cascaded from strategic intent. It also resonates with recent calls to treat the management of AI as an ongoing organisational undertaking rather than a one-off design problem (Berente *et al.*, 2021). Dynamic alignment thus integrates the learning orientation of adaptive governance, the reconfiguration logic of dynamic capabilities and the fit concern of alignment research, but reframes them as an ongoing, practice-driven process of governing responsible AI use.

Generative AI alters the alignment problem relative to traditional IT in three ways. Firstly, provisioning: generative tools are adopted from outside official channels at near-zero cost, so use precedes and bypasses formal acquisition, whereas traditional systems are centrally procured and configured. Secondly, output character: generative outputs are probabilistic and plausibly wrong, shifting effort from operating a system to verifying its content. Thirdly, visibility: because use occurs on personal devices and accounts, it is largely invisible to the organisation. Together, these features make the governance–practice gap structurally wider and faster-moving than in conventional IT settings.

3. Research methodology

This study adopts an interpretive qualitative design to examine how AI governance is enacted and evolves within a public sector organisation. Given its focus on processes, interactions and contextual dynamics, a qualitative approach is suited to capturing the complexity of socio-technical phenomena and the meanings actors attribute to them. The interpretive stance emphasises contextualised understanding and acknowledges the researcher’s role in co-constructing meaning with participants.

3.1 Research strategy

The study uses a single-case design to enable an in-depth, contextually grounded investigation of AI governance in practice. Case study research is well-suited to examining contemporary phenomena in real-life settings, especially where the boundaries between the phenomenon and its context are unclear (Yin, 2018). Here, AI governance is treated not as an isolated construct but as an embedded organisational process shaped by technological, institutional and human factors.

The selected public sector organisation in Kenya offers such an opportunity, allowing detailed examination of how AI use emerges and how governance mechanisms respond. Denscombe (2017) similarly notes that case studies suit complex social situations that require a holistic understanding rather than broad generalisation. The aim is analytical, not statistical, generalisation – linking empirical insight to theoretical constructs. By focusing on one context over time, the study traces interactions between everyday AI use and governance mechanisms, capturing the alignment, adaptation and control dynamics central to the research question.

The case organisation is a national one-stop public service delivery programme in Kenya. The programme integrates multiple government services under a single service point and is designed to improve citizens’ access to administrative services through physical service centres and digital platforms. It was selected as an information-rich case (Denscombe, 2017) because it combines two conditions that make the dynamics of AI governance especially observable: knowledge-intensive work involving document drafting, summarisation and citizen communication – tasks for which generative AI is readily applicable – and a public sector accountability regime characterised by legal liability, data-protection obligations and procurement constraints. This combination renders visible the tension between the pull of AI-enabled efficiency and the institutional demands of legitimacy and accountability, making the organisation an information-rich setting for examining how governance is enacted in practice.

The unit of analysis is the practice–control interaction – the recurring episodes in which everyday AI use meets, evades or provokes organisational controls – rather than the organisation as a whole or discrete decisions; this focus follows from the process-oriented research question.

Situating the case within its national context is important for interpreting these dynamics. Kenya has positioned itself as a regional digital-government leader, articulated through its Digital Economy Blueprint and successive e-government initiatives and the one-stop service model examined here reflects this ambition to consolidate citizen-facing services. At the same time, the institutional environment imposes distinctive constraints. The Data Protection Act (2019) and the Office of the Data Protection Commissioner establish statutory obligations for handling citizen data, heightening the liability associated with any exposure of personal information to third-party AI services. Public procurement is tightly regulated, which slows the formal acquisition of approved AI tools and licences. Politically, public administration operates under strong expectations of accountability and public scrutiny; socio-economically, uneven connectivity, constrained ICT budgets and disparities in access to devices shape both how employees reach AI tools and how citizens experience digitally mediated services. These institutional, political and socio-economic realities constitute the "public sector context" depicted in Figure 1 and condition the governance–practice dynamics reported below.

3.2 Data collection

Data were collected between November 2025 and April 2026 using semi-structured interviews as the primary method. A purposive sampling strategy ensured participants possessed relevant knowledge of AI use and governance, spanning hierarchical levels to capture both micro-level practice and macro-level governance.

Twenty-eight interviews were conducted: one chief information officer (CIO), eleven managers and 16 frontline staff. Participants are identified throughout the findings by unique codes – the CIO, managers (M1–M11) and frontline staff (F1–F16) – to enable traceable attribution while preserving confidentiality. Interviews with frontline staff focused on everyday use – experimentation, workarounds and the integration of AI into routine tasks – while interviews with managerial actors explored organisational responses such as policy development, oversight, training and accountability. Sample size was guided by data saturation; following Francis *et al.* (2010), collection was deemed sufficient once additional interviews yielded no new themes and the observed thematic repetition indicated saturation had been reached. Table 1 summarises the participants and the analytical focus of each role group. Table 1 also reports, at the role-group level, the principal functional areas represented and participants' AI-use experience, without compromising anonymity.

Interviews were conducted between November 2025 and April 2026, on average lasted approximately 65–80 min and were conducted online via Zoom. With participants' consent, interviews were audio-recorded and transcribed *verbatim* and an interview guide structured

Table 1. Overview of interview participants

Role group	No.	Principal functional areas	AI-use experience	Primary analytical focus
Chief information officer	1	ICT governance, information security	Regular	Governance strategy, risk posture, accountability
Managers (M1–M11)	11	Policy, legal drafting, service delivery, ICT	Occasional–regular	Oversight, policy adaptation, informal control
Frontline staff (F1–F16)	16	Legal drafting, customer service, document processing	Occasional–daily	Everyday use, workarounds, verification practices
Total	28			

around AI use, perceived risks, governance mechanisms and accountability was used flexibly to pursue emergent themes. The guide was piloted with five purposively selected early interviews (included in the 28); refinement was limited to rewording and reordering questions and no analytical categories were altered. The full interview guide – structured around AI use, perceived risks, governance mechanisms and accountability – is provided as supplementary material (Appendix).

The study observed established ethical procedures. Participation was voluntary and based on informed consent and participants were free to withdraw at any point. To protect confidentiality, all participants and the organisation have been anonymised: individuals are identified only by role group and any details that could identify the organisation or its staff have been removed or generalised. Given that the study concerns informal and unsanctioned ("shadow") practices, particular care was taken to ensure that no participant could be linked to specific disclosures.

Consistent with the interpretive stance, reflexivity was maintained throughout. The researcher remained attentive to the influence of their own assumptions on data generation and interpretation, recognising that meaning was co-constructed with participants. Analytical memos were used to surface and interrogate emerging interpretations, supporting a critical rather than purely descriptive engagement with the data.

To complement and triangulate the interviews, documentary analysis examined internal policy documents, ethical guidelines and AI-related governance artefacts, allowing formal structures to be compared with reported practice. This triangulation strengthened the credibility and confirmability of the analysis through cross-verification across sources.

The documentary corpus comprised organisational AI and data-handling guidance, ICT and security policies and internal communications relevant to AI use during the study period; documents were selected where they pertained to AI use or its control and were used to compare formally articulated rules with reported practice, rather than as a standalone data set. Documents were selected on three criteria – relevance to AI use or its governance, currency during the study window and organisational authority – and were coded with the same coding frame as the interviews, so that formal provisions could be set directly against reported practice; where the two diverged, the divergence itself became an analytical focus.

3.3 Data analysis

Data were analysed using thematic analysis following Braun and Clarke (2006), a flexible method for identifying and interpreting patterns in qualitative data and well suited to exploratory, interpretive work. The analysis was iterative and reflexive: transcripts were read and re-read for familiarisation; initial codes captured meaningful segments relating to AI use, governance practices and organisational responses; and codes were organised into broader themes reflecting recurring patterns. Coding decisions and theme development were documented to support transparency and auditability.

Consistent with the socio-technical framework, the analysis addressed three interrelated dimensions: the technical affordances of AI tools and how they were used; organisational control mechanisms, including their integration with existing IT and data governance; and issues of legitimacy, transparency and accountability in the justification and regulation of AI use. Particular attention was paid to critical incidents in which everyday use diverged from formal arrangements. These moments of misalignment were analysed as key sites of interaction between practice and governance, revealing how the organisation responded to emerging challenges and adjusted its controls over time.

Analytically, first-order codes grounded in participants' own language (for example, "using personal phones to access AI", "double-checking citations", "blocking sites drives use underground") were progressively abstracted into second-order themes (such as shadow normalisation, intensified verification and counterproductive control) and ultimately into the aggregate dimensions that structure the findings: everyday AI use, adaptive control, reinforced accountability, institutional constraint and co-evolution. This movement from informant-centric to theory-centric categories provided a transparent chain of evidence linking raw data to the conceptual model.

The approach is best characterised as an abductive, Gioia-informed thematic analysis: thematic analysis describes the coding logic and the Gioia data structure (Table 2) displays it transparently, the two being complementary rather than competing.

The trustworthiness of the analysis was addressed through four established criteria. Credibility was supported by triangulation across role groups and between interview and documentary data, as well as by the use of rich, *verbatim* quotations. Transferability is enabled through a thick description of the case context, allowing readers to judge the applicability of the findings to other settings. Dependability was strengthened by documenting coding decisions and maintaining an audit trail of analytical memos and confirmability by grounding interpretations explicitly in the data and cross-verifying claims against multiple sources. Together, these measures support the rigour of an interpretive account that does not seek statistical generalisation but analytical resonance with theory.

Table 2. Data structure linking first-order codes to aggregate dimensions

First-order codes (illustrative)	Second-order themes	Aggregate dimension
"A lifeline"; "an open secret"; personal phones used	Shadow normalisation of AI use	Everyday AI use
Blocking "drives it underground"; "treat as a first draft"; peer review	Counterproductive vs practice-embedded control	Adaptive control
"Double-check every citation"; "rests with the human"	Intensified verification; non-delegable responsibility	Reinforced accountability
"No official safe pathways"; "do-it-yourself"	Absence of sanctioned pathways	Institutional constraint
"The rule changed because of practice"	Reactive governance adjustment	Co-evolution (provisional)

4. Findings

The findings reveal a persistent governance–practice gap between formal AI governance frameworks and the realities of everyday organisational use. Rather than being implemented as a predefined system, responsible AI governance emerges as an adaptive socio-technical process shaped through ongoing interaction between practice and control.

4.1 From policy silence to shadow normalisation

AI adoption originated informally, driven less by organisational strategy than by immediate operational pressures. Frontline staff and knowledge workers consistently described generative AI as a practical resource for managing workload intensity, enhancing efficiency and overcoming task-related challenges. A frontline legal drafter (F3) explained that AI had become "a lifeline…a way to overcome writer's block and save hours of work", while a customer service officer (F7) described it as "a helper…for a practical problem". These accounts reflect a strong alignment between perceived task demands and the affordances of AI tools.

In contrast, formal governance initially adopted a cautious and risk-oriented stance. The CIO emphasised concerns related to data protection and institutional liability, stating that "sensitive citizen data is never exposed through public AI tools". This position led to restrictive measures, such as blocking AI platforms on official systems. However, these measures did not eliminate use; rather, they displaced it. As one manager (M4) observed, "blocking these sites isn't stopping the use; it's just driving it underground" and a frontline employee (F9) confirmed, "we just use our phones…it's an open secret".

Documentary analysis corroborated this policy silence: the ICT and data-protection policies in force during the study window contained no provisions specific to generative AI and no approved tool list or usage guidance was located. In practice, shadow use followed a recognisable routine – staff drafted or summarised on a personal device, transferred the output into official systems and reworked it to fit house style and legal form, leaving no trace of the AI step in the record.

One frontline employee (F4) gave a characteristic account: "I had to draft a complex response to a parliamentary query on a tight deadline. I started by feeding the key case files and the query into the AI on my personal laptop, asking it to generate a preliminary structure and a first draft of the legal arguments. I then copied that text into our secure system. I spent the next two hours reworking every paragraph, checking every cited precedent against our official law library and adjusting the tone to match our office's formal style. By the time I signed off, the final document bore little resemblance to the AI's output, but it had saved me the better part of a day's work."

Over time, these informal practices became normalised. What began as isolated experimentation evolved into routine integration across core tasks such as drafting, summarisation and communication. A policy manager (M2) noted that AI had shifted from "a future topic" to "a tool that many of my staff use daily". Governance responses lagged behind this shift, moving slowly from policy silence to reactive and partial guidance. This temporal mismatch constitutes a central manifestation of the governance–practice gap.

Crucially, normalisation was not the result of defiance or non-compliance, but of the mismatch between restrictive governance approaches and the practical utility of AI tools. The high perceived value of AI, combined with low barriers to access, meant that prohibitive controls did not suppress use but merely relocated it beyond organisational visibility. As such, shadow normalisation reflects a structural response to governance conditions rather than individual deviation.

4.2 Layered and adaptive control mechanisms

Governance mechanisms did not appear as a unified or coherent system, but as a layered and evolving configuration of controls that were introduced, adjusted and combined over time. Three broad categories of control were identified: technical controls, formal human oversight mechanisms and informal managerial controls.

Technical controls, primarily in the form of network restrictions and platform blocking, were the most formalised yet least effective mechanisms. As one frontline employee (F1) noted, “the constraint doesn’t prevent the behaviour, it just makes it more inconvenient”. In practice, employees circumvented these controls with ease, using personal devices and external networks.

Formal human oversight mechanisms, such as peer review and managerial approval, were pre-existing organisational routines that became increasingly important in the context of AI use. A policy manager (M5) emphasised that “all documents must go through peer review and senior sign-off”, indicating that established accountability structures absorbed and stabilised the risks introduced by AI.

The documents reinforced this layering: pre-existing peer-review and sign-off requirements appeared in standing procedures, whereas no document authorised or restricted specific AI tools, so technical blocking functioned as an informal IT measure rather than a codified policy.

Informal managerial controls emerged as particularly significant. These included verbal guidance, implicit expectations regarding responsible use and explicit instructions to verify AI outputs. As one manager (M7) stated, AI-generated content “must be treated as a first draft…and rigorously fact-checked”. These controls were adaptable, context-sensitive and embedded within everyday practice.

Governance evolved incrementally in response to concrete incidents. A notable example involved a draft citing a non-existent law (described in detail in Section 4.3), which triggered immediate corrective action and prompted stricter review practices. Such incidents functioned as catalysts for governance adaptation, highlighting that control mechanisms were shaped by experiential learning rather than predefined design.

Notably, more formalised controls were, in this case, less influential on behaviour than practice-embedded ones, suggesting that governance effectiveness depends on how far mechanisms are aligned with organisational routines and professional judgement. This is a case-specific pattern rather than a general law: there were also contrary instances – technical blocking did deter use on official desktop systems, displacing rather than merely failing to constrain it and some staff declined to use AI for sensitive matters in the absence of an approved tool. Formal controls were thus selectively, rather than uniformly, ineffective.

4.3 Accountability and the reinforcement of human oversight

A central and consistent finding is the persistence and reinforcement of human accountability. Across all organisational levels, responsibility for outputs remained explicitly assigned to individuals rather than delegated to technology. The CIO stated that “accountability is crystal clear…it rests with the human and the office, not the tool”, while a policy manager (M2) emphasised that “the accountable person is the officer whose name is on the document. Period.”

Rather than diminishing accountability, AI use intensified it. Employees described increased vigilance, particularly in relation to verifying AI-generated outputs. A legal drafter (F3) noted, “I double-check every single citation…the stakes are too high” and a customer service officer (F7) explained, “I read it twice before sending…I never just copy and paste”.

These practices indicate that AI introduces a new layer of epistemic uncertainty that must be managed through human judgement.

A senior manager (M8) elaborated on a critical incident: “One of my juniors submitted a policy brief for approval. It was well written and well argued, but it cited a specific statutory provision that did not exist. When I challenged it, he admitted he had used a public AI tool for the research. The AI had fabricated the reference – it looked perfectly legitimate, with a plausible act number and year, but was wholly non-existent. That incident was a wake-up call. We now require that any AI-generated content be treated as a ‘draft for development’ and we have increased the intensity of our verification checks, particularly on citations and factual claims.”

As one manager (M4) observed, reviewers now actively question whether outputs “look plausible but might be completely fabricated”. In this sense, AI shifts effort from content production to validation and verification.

In highly accountable organisational contexts, this redistribution of effort is a defining feature of AI integration. We interpret this as an intensified verification effort rather than demonstrated accountability effectiveness: the evidence establishes that participants report more checking, not that oversight reliably catches errors and the limits of human oversight (Parasuraman and Manzey, 2010; Green, 2022) caution against equating effort with efficacy.

4.4 The absence of safe pathways and the persistence of informality

A recurring theme is the absence of formalised pathways for safe and sanctioned AI use. The organisation lacked approved tools, structured training or controlled environments for experimentation. As a policy manager (M5) stated, “there are no official safe pathways…it’s all do-it-yourself”, while a frontline employee (F5) expressed a desire for “an official tool…that I don’t have to hide”.

Institutional constraints, including procurement requirements, data-protection obligations and risk-averse organisational logics, shape this absence, limiting the organisation’s ability to integrate AI even as its use expands informally.

The consequences are multifaceted. Firstly, shadow practices persist, with employees relying on personal devices and accounts. Secondly, organisational learning is fragmented, as knowledge about AI use is shared informally through peer networks rather than formal channels; one manager (M9) described this as “a quiet, peer-to-peer network of innovation…completely outside any official oversight”. Thirdly, inequities emerge, as access to AI depends on individual resources such as personal devices and connectivity.

Participants recognised that restrictive approaches may be counterproductive, with one manager (M4) noting that punitive measures would “just drive it further underground”, while another (M9) emphasised the need for “a safe sandbox to bring this into the light”. This reflects an emerging recognition that governance must balance control with enablement.

4.5 Co-evolution of practice and governance

The findings suggest a co-evolutionary dynamic between AI use and governance. Participants’ retrospective accounts suggested that, over roughly the preceding year, AI adoption had expanded from isolated experimentation towards widespread informal integration. As one respondent (F9) observed, “it’s gone from a secret thing…to something everyone does”. Governance evolved along a slower trajectory, moving from “no policy” to “reactive policy” and gradually towards more proactive consideration.

This temporal lag produces a state in which informal practices and formal controls coexist in tension. Governance does not precede practice but follows it, often reacting to unintended consequences: employee practices expose the limitations of existing controls, prompting organisational responses that, in turn, reshape practice. As one manager (M2) explained, governance rules were adjusted because of how AI was being used in practice – "the rule changed because of…practice".

Overall, responsible AI governance is enacted as a process of dynamic alignment, in which shadow practices, adaptive control mechanisms and reinforced human oversight interact under conditions of institutional constraint and normative pressure. Alignment is therefore not a fixed endpoint, but an ongoing organisational accomplishment.

This co-evolutionary reading should be treated cautiously. Only one practice-to-policy sequence – the "non-existent law" incident described in Section 4.3 – was independently documented. However, participants described a broader tightening of expectations. The retrospective, single-window design means organisation-wide co-evolution is best regarded as a plausible interpretation that requires longitudinal confirmation rather than as a demonstrated finding. Accordingly, we present co-evolution as an analytically plausible but empirically under-determined process: we make no claim to a generalisable longitudinal pattern from this single case and treat the term as a directed hypothesis for future longitudinal testing. Table 3 maps the five empirical themes to the constructs of the conceptual model, providing the empirical grounding for the dynamic alignment process developed in the Discussion.

5. Discussion

This study examined how public sector organisations dynamically align the use of AI with governance and accountability requirements as these technologies become embedded in everyday work practices.

5.1 From governance gap to generative mechanism

The findings extend existing research on the governance–practice gap by demonstrating that the gap is not merely a temporary lag between policy and implementation, but a generative feature of socio-technical systems characterised by rapid technological diffusion and institutional constraint. Prior work has largely conceptualised this gap as a deficiency in organisational readiness or a failure in governance design (Wirtz *et al.*, 2021; Jonathan *et al.*, 2025, 2026). In contrast, the present study shows that the gap emerges endogenously through the interaction between accessible, high-utility technologies and governance approaches that rely on restriction and formal compliance.

Table 3. Mapping of empirical themes to conceptual model constructs

Empirical theme	Model construct	Representative evidence
Policy silence to shadow normalisation	Micro-level practice; governance–practice gap	"It's an open secret"
Layered and adaptive control	Meso-level governance	Technical, formal and informal controls
Reinforced human oversight	Accountability (verification effort)	"It rests with the human and the office"
Absence of safe pathways	Institutional constraint	"It's all do-it-yourself"
Co-evolution (provisional)	Recursive feedback (illustrated)	"The rule changed because of… practice"

When AI tools are readily available and offer immediate performance benefits, employees adopt them pragmatically to meet task demands. This displacement reduces the organisation's ability to monitor, understand and guide AI use, thereby expanding rather than closing the governance gap.

5.2 Reconfiguring governance: from control systems to situated practice

The extant literature conceptualises governance as a system of formal controls, comprising rules, policies and technical restrictions that seek to shape behaviour externally to organisational practice. Consistent with Jonathan *et al.* (2026), our findings challenge this conception by demonstrating that governance effectiveness depends less on formal authority than on its integration into the routines, relationships and norms through which work is accomplished. Within this case, the inverse relationship between formalisation and behavioural influence reveals a fundamental misalignment: highly formalised controls, such as technical blocking mechanisms, proved least effective because they remained detached from the workflows they sought to regulate. By contrast, informal controls, including managerial expectations, peer review and professional judgement, exerted greater influence precisely because they were embedded within everyday practice. Governance, therefore, is not simply imposed; it is enacted through practice.

5.3 AI and the transformation of accountability work

Contrary to dominant narratives that portray AI as automating decision-making and attenuating human responsibility, the findings indicate that accountability remains firmly vested in human actors and is, if anything, amplified through AI adoption (Jonathan *et al.*, 2026). Accountability consequently shifts from the production of content to its validation, interpretation and legitimisation. Moreover, accountability remains non-delegable, positioning human-in-the-loop oversight not as a discretionary safeguard but as a constitutive source of organisational legitimacy (Jonathan and Han, 2025).

5.4 Legitimacy, institutional constraint and public value

These observations map directly onto the "public sector context" block of the framework in Figure 1. In the Kenyan setting, the normative pressure exerted by that context is concrete rather than abstract: statutory data-protection duties under the Data Protection Act (2019), procurement rules that delay the provision of sanctioned tools and politically salient expectations of accountability jointly explain why the organisation was slow to formalise AI use even as demand grew. Institutional theory helps interpret this pattern: organisations conform to regulative and normative expectations to secure legitimacy (DiMaggio and Powell, 1983; Jonathan and Han, 2025; Suchman, 1995), and here the pursuit of legitimacy under tight institutional constraints both restrained formal adoption and, paradoxically, displaced use into unobservable channels.

5.5 Boundary conditions and generalisability

These conditions suggest the mechanisms are most pronounced where AI is accessible outside organisational control, work is knowledge-intensive and accountability is strong – as in this resource-constrained, procurement-bound setting. Theorising the variation clarifies the concept's scope. In resource-rich organisations that can rapidly procure sanctioned enterprise tools, the provisioning gap should narrow and shadow use decline, shifting the problem from visibility to configuration and monitoring. In highly regulated domains (e.g. health, justice), stronger sanctions may suppress overt use but intensify displacement, widening the very gap they target. Dynamic alignment thus describes a family of context-dependent trajectories rather than a single path, and a comparative study across these conditions is a priority.

The effectiveness of these oversight mechanisms is itself bounded. Practice-embedded controls and verification are most likely to be effective where review is routine, adequately resourced and backed by clear individual accountability and where reviewers retain the domain expertise to detect plausible-but-wrong outputs. Their effectiveness is likely to diminish over time, particularly in high-workload contexts where automation complacency takes hold, the absence of sanctioned tools limits visibility into use and accountability is diffuse (Green, 2022; Jonathan *et al.*, 2026; Parasuraman and Manzey, 2010). The intensified verification reported here should therefore be read as an effort whose efficacy is contingent on these conditions rather than as a guarantee of reliable oversight.

5.6 Citizen-level implications

The organisational dynamics carry consequences for citizens, which, in the absence of direct citizen-level data, we frame as evidence-informed implications rather than demonstrated outcomes. Entering work materials into public tools on personal devices, outside any data-processing agreement, is at odds with statutory obligations and exposes citizens to unobservable privacy risks (Janssen and Kuk, 2016; Haag and Eckhardt, 2017). Where access depends on personal devices and connectivity, service quality may vary with officers' private means rather than a deliberate standard (Sun and Medaglia, 2019; Twizeyimana and Andersson, 2019). Moreover, opaque AI mediation can foster technocratic trust, in which accountability rests on expertise rather than on public understanding (Sieber *et al.*, 2025; Bovens, 2007). These implications warrant a direct study of citizens' experiences.

To support future empirical work, Table 4 sketches a preliminary operationalisation of dynamic alignment, identifying core dimensions, observable indicators and possible data sources; these are offered as a starting point for measurement rather than a validated instrument.

Table 4. A preliminary operationalisation of dynamic alignment for future research

Dimension of dynamic alignment	Observable indicator(s)	Possible data source
Governance–practice gap	Extent of unsanctioned vs sanctioned AI use; visibility of use to the organisation	Interviews; system logs; document review
Emergent practice (shadow normalisation)	Routinisation of AI across core tasks; reliance on personal devices and workarounds	Interviews; observation
Adaptive control	Balance of technical vs practice-embedded controls; lag between practice change and control change	Document review; interviews
Recursive feedback (mechanism)	Documented sequences of practice prompting control change over time	Longitudinal or document tracing
Reinforced accountability	Verification effort (checking behaviours) versus verification effectiveness (errors detected)	Interviews; review records; audits
Provisional alignment (outcome)	Coexistence of formal and shadow practice; size of residual gap	Repeated or longitudinal assessment

6. Conclusion

This study examined how a public sector organisation aligns its use of AI with governance and accountability requirements as these technologies become embedded in everyday work. The answer is that alignment is achieved not through predefined frameworks but through dynamic alignment: a provisional, recursive process in which bottom-up AI use and incremental control adjustments are continuously reconciled. The overarching mechanism is recursive feedback between practice and control; within it, displacement – restriction relocating rather than removing use – is the most visible empirically observed dynamic, alongside the reinforcement of human verification work. The contribution is integrative and process-oriented: it consolidates practice-based and governance perspectives into an account of governance-in-practice. It specifies how generative AI reshapes the alignment problem relative to traditional IT.

6.1 Implications

For research, the study advances a process-oriented understanding of AI governance by conceptualising it as dynamic alignment, highlighting how governance emerges through the continuous interaction between organisational practices and control mechanisms.

For practice, the findings support evidence-informed guidance rather than tested prescriptions: a staged pathway of interim usage rules and non-punitive disclosure, a governed sandbox and verification training and – where feasible – procurement of enterprise tools with data-processing safeguards. Because sandboxes, approved tools and procurement reform were not evaluated here, these are propositions and research priorities; in resource- and procurement-constrained organisations, realistic first steps are low-cost measures (interim guidance, disclosure channels, strengthened review) rather than immediate enterprise procurement.

For policy, the evidence cautions against prohibition and favours safe provisioning, transparency and avenues for citizens to contest AI-assisted decisions. For teaching, the case offers material on how governance evolves in practice rather than in the abstract.

6.2 Limitations and future research

This study has several limitations that point to avenues for future research. As a single retrospective case covering a relatively short period, it can only illustrate, rather than definitively establish, processes of co-evolution. The temporal dynamics would benefit from longitudinal and multi-wave investigation. The findings are also based on self-reported perceptions rather than system-generated data, preventing independent verification of the prevalence of shadow use or the effectiveness of oversight mechanisms. Although methodological safeguards were used, social desirability bias cannot be entirely ruled out. Future comparative and observational research across diverse regulatory and resource environments is needed to test the boundary conditions identified here and to advance dynamic alignment from an analytical lens to a more fully elaborated theory of responsible AI governance in the public sector. Finally, coding was undertaken by a single researcher. However, reflexive memos, an audit trail and cross-checking against documentary data were maintained to support interpretive consistency; single-coder analysis remains a limitation with respect to researcher subjectivity and independent or multiple coding would strengthen future work.

References

Al-Besher, A. and Kumar, K. (2022), "Use of artificial intelligence to enhance e-government services", *Measurement: Sensors*, Vol. 24, p. 100484.

Berente, N., Gu, B., Recker, J. and Santhanam, R. (2021), "Managing artificial intelligence", *MIS Quarterly*, Vol. 45 No. 3, pp. 1433-1450.

Birkstedt, T., Minkkinen, M., Tandon, A. and Mäntymäki, M. (2023), "AI governance: themes, knowledge gaps and future agendas", *Internet Research*, Vol. 33 No. 7, pp. 133-167.

Bovens, M. (2007), "Analysing and assessing accountability: a conceptual framework 1", *European Law Journal*, Vol. 13 No. 4, pp. 447-468.

Braun, V. and Clarke, V. (2006), "Using thematic analysis in psychology", *Qualitative Research in Psychology*, Vol. 3 No. 2, pp. 77-101.

Bullock, J.B. (2019), "Artificial intelligence, discretion, and bureaucracy", *The American Review of Public Administration*, Vol. 49 No. 7, pp. 751-761.

De Sousa, W.G., de Melo, E.R.P., Bermejo, P.H.D.S., Farias, R.A.S. and Gomes, A.O. (2019), "How and where is artificial intelligence in the public sector going? A literature review and research agenda", *Government Information Quarterly*, Vol. 36 No. 4, p. 101392.

Denscombe, M. (2017), *EBOOK: The Good Research Guide: For Small-Scale Social Research Projects*, McGraw-Hill Education (UK).

DiMaggio, P.J. and Powell, W.W. (1983), "The iron cage revisited: institutional isomorphism and collective rationality in organizational fields", *American Sociological Review*, Vol. 48 No. 2, pp. 147-160.

Faraj, S., Pachidi, S. and Sayegh, K. (2018), "Working and organizing in the age of the learning algorithm", *Information and Organization*, Vol. 28 No. 1, pp. 62-70.

Folke, C., Hahn, T., Olsson, P. and Norberg, J. (2005), "Adaptive governance of social-ecological systems", *Annual Review of Environment and Resources*, Vol. 30 No. 1, pp. 441-473.

Francis, J.J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M.P. and Grimshaw, J.M. (2010), "What is an adequate sample size? Operationalising data saturation for theory-based interview studies", *Psychology and Health*, Vol. 25 No. 10, pp. 1229-1245.

Green, B. (2022), "The flaws of policies requiring human oversight of government algorithms", *Computer Law and Security Review*, Vol. 45, p. 105681.

Grønsund, T. and Aanestad, M. (2020), "Augmenting the algorithm: emerging human-in-the-loop work configurations", *The Journal of Strategic Information Systems*, Vol. 29 No. 2, p. 101614.

Haag, S. and Eckhardt, A. (2017), "Shadow IT", *Business and Information Systems Engineering*, Vol. 59 No. 6, pp. 469-473.

Henderson, J.C. and Venkatraman, N. (1993), "Strategic alignment: leveraging information technology for transforming organisations", *IBM Systems Journal*, Vol. 32 No. 1, pp. 4-16.

Janssen, M. and Kuk, G. (2016), "The challenges and limits of big data algorithms in technocratic governance", *Government Information Quarterly*, Vol. 33 No. 3, pp. 371-377.

Jonathan, G.M. and Han, S. (2025), "GenAI in public sector transformation: balancing promise and prudence", *International Conference on Electronic Government and the Information Systems Perspective*, Springer Nature Switzerland, Cham, pp. 105-121.

Jonathan, G.M., Kuika Watat, J., Aasi, P. and Ayeni, F. (2026), "Bridging the governance-practice gap: a socio-technical study of AI co-evolution", *Proceedings of the Americas Conference on Information Systems (AMCIS 2026), AIS*, p. 22.

Jonathan, G.M., Yalew, S.D., Gebremeskel, B.K. and Watat, J.K. (2025), "AI adoption in the public sector: organizational readiness and the pursuit of public value", *Complex Systems Informatics and Modeling Quarterly*, Vol. 45 No. 45, pp. 43-70.

König, P.D. and Wenzelburger, G. (2021), "The legitimacy gap of algorithmic decision-making in the public sector: why it arises and how to address it", *Technology in Society*, Vol. 67, p. 101688.

Laux, J., Wachter, S. and Mittelstadt, B. (2024), "Trustworthy artificial intelligence and the European Union AI act: on the conflation of trustworthiness and acceptability of risk", *Regulation and Governance*, Vol. 18 No. 1, pp. 3-32.

Leonardi, P.M. (2011), "When flexible routines meet flexible technologies: affordance, constraint, and the imbrication of human and material agencies1", *MIS Quarterly*, Vol. 35 No. 1, pp. 147-167.

Mäntymäki, M., Minkkinen, M., Birkstedt, T. and Viljanen, M. (2022), "Defining organizational AI governance", *AI and Ethics*, Vol. 2 No. 4, pp. 603-609.

Moore, M.H. (1997), *Creating Public Value: Strategic Management in Government*, Harvard University Press.

National Council for Law Reporting (2019), Data Protection Act, No. 24 of 2019, Government of Kenya, available at: www.kentrade.go.ke/wpontent/uploads/2022/09/Data-Protection-Act-1.pdf (accessed 15 May 2026).

Orlikowski, W.J. (2000), "Using technology and constituting structures: a practice lens for studying technology in organisations", *Organization Science*, Vol. 11 No. 4, pp. 404-428.

Parasuraman, R. and Manzey, D.H. (2010), "Complacency and bias in human use of automation: an attentional integration", *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 52 No. 3, pp. 381-410.

Rjab, A.B., Mellouli, S. and Corbett, J. (2023), "Barriers to artificial intelligence adoption in smart cities: a systematic literature review and research agenda", *Government Information Quarterly*, Vol. 40 No. 3, p. 101814.

Sieber, R., Brandusescu, A., Sangiambut, S. and Adu-Daako, A. (2025), "What is civic participation in artificial intelligence?", *Environment and Planning B: Urban Analytics and City Science*, Vol. 52 No. 6, pp. 1388-1406.

Suchman, M.C. (1995), "Managing legitimacy: strategic and institutional approaches", *The Academy of Management Review*, Vol. 20 No. 3, pp. 571-610.

Sun, T.Q. and Medaglia, R. (2019), "Mapping the challenges of artificial intelligence in the public sector: evidence from public healthcare", *Government Information Quarterly*, Vol. 36 No. 2, pp. 368-383.

Teece, D.J. (2007), "Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance", *Strategic Management Journal*, Vol. 28 No. 13, pp. 1319-1350.

Twizeyimana, J.D. and Andersson, A. (2019), "The public value of e-government – a literature review", *Government Information Quarterly*, Vol. 36 No. 2, pp. 167-178.

Waters-Lynch, J., Allen, D.W., Potts, J. and Berg, C. (2025), "Shadow user innovation: governing covert generative-AI use for dynamic-capability renewal", *Innovation*, pp. 1-17.

Wirtz, B.W., Langer, P.F. and Fenner, C. (2021), "Artificial intelligence in the public sector – a research agenda", *International Journal of Public Administration*, Vol. 44 No. 13, pp. 1103-1128.

Yin, R.K. (2018), *Case Study Research and Applications*, Vol. 6 Sage, Thousand Oaks, CA.

Further reading

Jobin, A., Ienca, M. and Vayena, E. (2019), "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, Vol. 1 No. 9, pp. 389-399.

Kellogg, K.C., Valentine, M.A. and Christin, A. (2020), "Algorithms at work: the new contested terrain of control", *Academy of Management Annals*, Vol. 14 No. 1, pp. 366-410.

Kuika Watat, J., Jonathan, G.M. and Zhang, L. (2025), "Why generative artificial intelligence does not survive first contact with bureaucracy", *Proceedings of the International Conference on Information Systems (ICIS 2025)*, *AIS*, pp. 1-8.

Margetts, H. and Dorobantu, C. (2019), "Rethink government with AI", *Nature*, Vol. 568 No. 7751, pp. 163-165.

Mittelstadt, B. (2019), "Principles alone cannot guarantee ethical AI", *Nature Machine Intelligence*, Vol. 1 No. 11, pp. 501-507.

Shneiderman, B. (2020), "Human-centered artificial intelligence: reliable, safe and trustworthy", *International Journal of Human–Computer Interaction*, Vol. 36 No. 6, pp. 495-504.

Zuiderwijk, A., Chen, Y.C. and Salem, F. (2021), "Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda", *Government Information Quarterly*, Vol. 38 No. 3, p. 101577.

Appendix

Semi-structured interview guide

Manuscript TG-06-2026-0318 – "Responsible AI Beyond Compliance: Governing Everyday Use in the Public Sector".

A. Introduction and consent.

- Purpose of the study; voluntary participation; confidentiality and anonymisation; consent to audio-record; right to withdraw.
- Could you briefly describe your role and main responsibilities?

B. Everyday AI use (micro-level practice).

- In your day-to-day work, where and how do you use AI tools (e.g. drafting, summarising, communication)? Please walk me through a recent example.
- Which tools do you use, and how did you come to use them? Are they provided by the organisation or accessed another way?
- When AI is not permitted or available on official systems, what do you do instead?
- How do you decide when AI is appropriate for a task and when it is not?

C. Perceived risks and data handling.

- What concerns, if any, do you have about using AI for your work (accuracy, confidentiality, citizen data)?
- How do you handle sensitive or personal information when using AI tools?

D. Governance mechanisms (meso-level control).

- What rules, policies or guidance govern AI use in your area? How did you become aware of them?
- What technical controls exist (e.g. restrictions, blocking), and how do they affect your work?
- What role do managers, peer review or sign-off play in relation to AI-assisted work?
- Have any guidelines or practices changed in response to something that happened? Please describe.

E. Accountability and verification.

- Who is accountable for the outputs when AI has been used? How is that understood in practice?
- How do you check or verify AI-generated content before it is used or released?
- Has AI changed how much checking or oversight your work requires?

F. Institutional context and safe pathways.

- Are there approved tools, training or “safe” environments for experimenting with AI? What is missing?
- How do procurement, data-protection or other institutional requirements shape what is possible?

G. Closing.

- Is there anything important about how AI is used or governed here that we have not discussed?

Corresponding author
Gideon Mekonnen Jonathan can be contacted at: gideon@dsv.su.se