JENRS

Received: 04 May 2026, Revised: 01 June 2026, Accepted: 02 June 2026, Online: 22 June 2026

DOI: https://doi.org/10.55708/js0506005

From ITIL to AIOps in Public Sector: A Systematic Literature Review

Catherine Ganduri*

Systems Analyst, Information Resources Division, Texas Commission on Environmental Quality, Austin, Texas, 78753, United States

Email(s): catherine.ganduri@tceq.texas.gov

*Corresponding author: Catherine Ganduri, Austin, Texas, United States, catherine.ganduri@tceq.texas.gov

ABSTRACT: Public-sector agencies rely on complex and highly regulated digital systems to deliver essential services. ITIL-based change and release management supports operational control, but many agencies still depend on manual approvals, fragmented operational data, and reactive monitoring. Artificial Intelligence for IT Operations (AIOps) and Machine Learning Operations (MLOps) can improve anomaly detection, failure prediction, release validation, and data-driven governance. However, the literature gives limited attention to how these technologies can be integrated into public-sector change and release workflows. A structured search of IEEE Xplore, ACM Digital Library, ScienceDirect, and Google Scholar identified 120 records. After duplicate removal, title/abstract screening, full-text assessment, and backward searching, 39 sources were retained for synthesis, including 25 primary studies and 14 complementary methodological, standards, and governance sources. The findings are organized into four themes: AI-driven monitoring and incident management, public-sector governance constraints, implementation challenges in legacy environments, and limited adoption of AI in change and release management. The review identifies gaps in explainability, empirical validation, AI-ready change records, and framework development, and proposes future directions for AI-supported operational governance in public-sector IT.

KEYWORDS: AIOps, Change Management, DevOps, ITIL, IT Service Management, MLOps, Public Sector IT, Release Management,

1. Introduction

Public-sector agencies increasingly depend on interconnected digital services to deliver regulatory, administrative, and citizen-facing functions. In these environments, service reliability is not only a technical requirement but also a governance responsibility. ITIL provides structured practices for incident management, change enablement, release management, service validation, and continual improvement, while the broader IT service management literature shows that these practices can improve service quality, accountability, and operational consistency [1-3].

At the same time, public-sector IT environments are becoming more complex because of cloud services, hybrid infrastructure, cybersecurity requirements, data integration, and modernization of older systems. Artificial Intelligence for IT Operations (AIOps) applies machine learning, event correlation, anomaly detection, predictive analytics, and automation to operational data such as logs, metrics, alerts, and incidents [4-6]. Related work on machine learning for computer systems and networking shows that data-driven techniques can support fault prediction, performance optimization, and adaptive operations [7]. MLOps extends these ideas by adding lifecycle controls for model deployment, monitoring, validation, and retraining [8].

Change and release management are central to IT service governance because they control when, how, and under what conditions system modifications are introduced into production. In traditional ITIL-oriented processes, change management evaluates need, risk, authorization, and stakeholder impact, while release management coordinates packaging, deployment, rollback planning, and post-implementation review [1-3]. These practices are especially important in public agencies because system changes may affect statutory reporting, public records, environmental monitoring, licensing, benefits delivery, or other essential services.

AIOps shifts IT operations from reactive monitoring toward predictive and data-driven operations. AIOps platforms can consolidate logs, alerts, traces, incidents,

and performance metrics, then use machine learning to identify anomalies, correlate events, recommend root causes, and support automated actions [4-6]. For change and release management, the same operational signals can be used to estimate deployment risk, compare current system behavior against historical baselines, and validate whether a release has introduced abnormal behavior.

MLOps is relevant because AIOps models themselves are production systems that require monitoring, retraining, versioning, and governance. In a public-sector context, AI automation should be auditable, explainable, and subject to human approval. Therefore, an AI-enabled change process should not replace ITIL controls; it should strengthen them through better evidence, earlier risk detection, and traceable decision support.

Although AI adoption in the public sector is increasing, government environments face constraints that are less visible in private-sector implementations. Public agencies must address transparency, accountability, explainability, procurement constraints, legal compliance, stakeholder oversight, and public trust [9], [10]. Research on AI-enabled change management in public administration also emphasizes organizational readiness, cultural alignment, human oversight, and the need to manage institutional change rather than treating AI as a purely technical tool [11], [12].

Existing research has made substantial progress on AIOps for monitoring and failure management, and recent work has begun to connect AIOps with ITSM and ITIL-oriented automation [13]. However, less attention has been given to how AI can support formal change and release management decisions, including change risk scoring, deployment scheduling, change advisory board (CAB) support, rollback planning, and post-release validation. This is important because production ML systems can create hidden technical debt and reliability risks if monitoring, validation, and model governance are weak [14], [15]. Explainable AI is also necessary where automated recommendations affect high-risk public-sector decisions [16].

This review addresses four research questions:

RQ1: How have artificial intelligence techniques been applied in IT operations management and service governance?

RQ2: What role do AIOps and MLOps technologies play in supporting release and change management within complex IT environments?

RQ3: What challenges and limitations exist when implementing AI-driven operational automation in public-sector IT systems?

RQ4: What research gaps remain in the application of AI to government IT infrastructures?

The contribution of this paper is fourfold. First, it synthesizes literature across ITSM, AIOps, MLOps, DevOps, and public-sector AI governance. Second, it explains why the current evidence base is mature for operational monitoring but less mature for formal change approval and release governance. Third, it identifies technical, organizational, and governance barriers that affect AI adoption in public-sector change and release management. Fourth, it proposes a compact AI-enabled ITIL framework and maps its components to the reviewed sources so that readers can trace the framework back to the evidence base.

2. Methodology

A systematic literature review was conducted using structured search, screening, and synthesis process. The review was guided by PRISMA reporting principles, established guidance for standalone literature reviews, and thematic analysis procedures [17-19]. Although the topic contains scoping-review characteristics because direct public-sector change/release studies remain limited, a systematic review label was retained because the study used predefined databases, Boolean search strings, inclusion and exclusion criteria, quality assessment, and a documented screening trail. The aim was not to map every possible source but to synthesize bounded evidence base relevant to AI-enabled IT operations and governance.

The review focused on peer-reviewed literature and high-quality standards or governance sources relevant to AI-enabled IT operations, change management, release management, MLOps, DevOps, and public-sector IT governance. The search covered literature published from 2015 to 2025. The final audit of the search log, retained sources, and coding decisions was completed on 4 May 2026.

Table 1: Selection criteria used for the literature review

Criteria	Description
Publication type	Peer-reviewed journal articles, conference papers, standards, and public-sector governance reports
Databases	IEEE Xplore, ACM Digital Library, ScienceDirect, Google Scholar, and backward reference searching
Search window	2015-2025; final audit completed on 4 May 2026
Search terms	AIOps; AI in IT operations; MLOps; ITIL; ITSM; change management; release management; DevOps; public-sector AI; AI governance
Inclusion criteria	Sources addressing AI-enabled IT operations, change/release governance, MLOps, DevOps deployment, public-sector AI adoption, or responsible AI controls

Criteria	Description
Exclusion criteria	Non-English sources, non-technical opinion pieces, duplicated studies, and publications unrelated to operational governance

Table 2: Boolean search strings, search fields, and audit details

Database	Fields searched	Boolean search string
IEEE Xplore	Title, abstract, keywords, metadata; full text when available	("AIOps" OR "artificial intelligence for IT operations" OR "machine learning for IT operations") AND ("IT service management" OR ITSM OR ITIL OR "change management" OR "release management" OR DevOps OR "continuous delivery")
ACM Digital Library	Title, abstract, keywords, metadata	("AIOps" OR "failure management" OR "anomaly detection" OR "root cause analysis") AND (logs OR metrics OR incidents OR "IT operations")
ScienceDirect	Title, abstract, keywords	("artificial intelligence" OR AI OR "machine learning") AND ("public governance" OR "public sector" OR "public administration" OR "government IT") AND ("service management" OR "change management" OR "release management")
Google Scholar	Title and full-text metadata returned by the search engine	(AIOps OR MLOps OR "AI in IT operations") AND (ITIL OR ITSM OR "change management" OR "release management") AND ("public sector" OR government OR governance)

2.1 *Research Design*

The initial database search identified 120 records. After removing 20 duplicates, 100 records were screened by title, abstract, and keyword relevance. Forty-five records were excluded at screening because they did not address AI-enabled IT operations, service governance, operational automation, or public-sector relevance. Fifty-five full-text articles and sources were then assessed for eligibility; 16 were removed because they lacked methodological clarity, technical depth, or relevance to change/release governance. The final synthesis included 39 sources, consisting of 25 primary studies and 14 supplementary methodological, standards, and governance sources. Figure 1 summarizes the study-selection process and allows readers to verify the flow from identification to final inclusion.

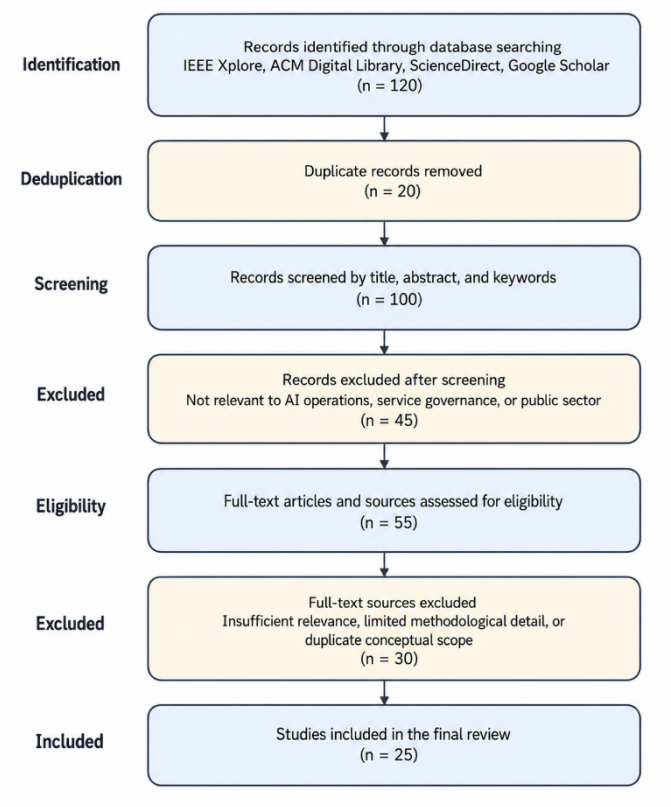

Figure 1: PRISMA-based study selection process.

2.2 *Data Extraction, Bias Mitigation, and Analysis*

For each selected source, data were extracted on publication year, application domain, AI technique, operational process, governance issue, empirical setting, and relevance to public-sector change and release management. The analysis used thematic coding to identify recurring patterns across the literature, followed by cross-checking themes for coherence and relevance to the research questions. The coding process generated four synthesis themes: AI-driven monitoring and incident management, public-sector governance and compliance, legacy-system and data integration barriers, and AI support for change and release management.

Because this review was conducted by a single independent author, bias mitigation relied on a documented audit trail rather than inter-reviewer reliability statistics. The audit trail recorded search strings, source databases, inclusion/exclusion reasons, quality-assessment notes, and theme-coding decisions. To reduce screening bias, borderline sources were rechecked against the research questions after full-text reading, and each retained source was classified as direct evidence or adjacent evidence. Direct evidence referred to sources that explicitly discussed AIOps, MLOps, ITSM, change management, release governance, or public-sector AI. Adjacent evidence referred to sources on ML production readiness, explainability, fairness, datasets, risk management, and AI management systems. This distinction prevented overstatement of the maturity of

AI-enabled change approval while still incorporating controls needed for responsible implementation.

2.3 *Quality Assessment and Coding Controls*

Each source was assessed against a compact quality checklist. The checklist examined whether the source clearly stated its problem context, described the AI or governance method, provided transferable evidence for IT operations, and addressed risks such as data quality, explainability, organizational readiness, or auditability. Sources that were strong in AIOps but weak in public-sector context were retained only when their findings were directly transferable to change, release, monitoring, or operational-risk management. Standards and governance sources were used as complementary evidence rather than as primary empirical studies.

Table 3: Quality assessment checklist for retained literature

Assessment area	Question applied during screening	Use in synthesis
Topical relevance	Does the source address AI-enabled operations, ITSM, MLOps, DevOps release control, or public-sector AI governance?	Determined whether the source contributed to one or more research questions.
Method clarity	Does the source explain its review method, empirical setting, framework logic, or technical approach?	Reduced reliance on unclear opinion-based material.
Operational evidence	Does the source discuss logs, incidents, deployment records, change data, service dependencies, or monitoring outcomes?	Supported links between AIOps evidence and ITIL change/release workflows.
Governance relevance	Does the source address transparency, accountability, auditability, risk management, or human oversight?	Informed public-sector requirements for accountable AI adoption.
Transferability	Can findings from enterprise or cloud settings be adapted to regulated government IT environments?	Separated directly applicable findings from contextual limitations.

Table 4.a and 4.b summarizes the quality profile of the 25 primary studies used as the main analytic evidence base. Ratings are reported as H = high, M = moderate, and L = low. The table is intended to make the evidence base transparent rather than to exclude all studies that scored low on one dimension; for example, a technical AIOps study could score low on governance relevance but still provide strong operational evidence.

Table 4a: Quality profile of technical and operational studies used for thematic synthesis

Ref.	Focus	T	M	O	G	Role in review
[3]	ITSM review	H	H	M	M	ITSM foundation
[4]	AIOps failure management	H	H	H	M	AIOps methods
[5]	AIOps real-world challenges	H	M	H	M	Implementation barriers
[6]	Cloud AIOps	H	H	H	M	Cloud observability
[7]	ML for systems/networking	H	H	H	L	Technical evidence
[8]	MLOps architecture	H	H	M	M	ML lifecycle control
[13]	Agentic AIOps framework	H	M	H	M	ITIL-aligned automation
[14]	ML technical debt	H	H	M	M	ML governance risk
[15]	ML production readiness	H	H	M	M	Model validation control
[20]	Software engineering for ML	H	H	M	M	ML engineering practice
[21]	ML systems challenges	H	H	M	M	Deployment challenges
[22]	ML deployment survey	H	H	M	M	Operational ML risks
[23]	Anomaly and RCA survey	H	H	H	L	Incident analysis
[24]	Log anomaly detection	H	H	H	L	Log-based monitoring
[25]	Time-series anomaly detection	H	H	H	L	Telemetry monitoring
[26]	LLMs for anomaly detection	M	H	H	L	Emerging AI monitoring
[27]	CI/CD systematic review	H	H	H	M	Release pipeline practices
[28]	DevOps adoption case study	H	H	M	M	Organizational barriers

T = topical relevance; M = methodological clarity; O = operational evidence; G = governance relevance. H = high, M = moderate, L = low.

Table 4b: Quality profile of public-sector and governance studies used for thematic synthesis

Ref.	Focus	T	M	O	G	Role in review
[9]	AI in public governance	H	H	L	H	Public-sector governance
[10]	AI and public sector	H	H	L	H	Public-sector challenges
[11]	AI change in public administration	H	M	L	H	Organizational readiness
[12]	AI-enabled organizational change	M	H	L	M	Change adoption lens
[29]	AI adoption in public administration	H	H	L	H	Government adoption evidence
[30]	AI capability in government	H	H	L	H	Public-sector AI capability
[31]	AI public management framework	H	M	L	H	Governance framework

T = topical relevance; M = methodological clarity; O = operational evidence; G = governance relevance. H = high, M = moderate, L = low.

Overall, the quality assessment shows strong topical relevance across the selected studies. Operational evidence is strongest in AIOps, anomaly detection, DevOps, and observability studies, while governance relevance is strongest in public-sector AI and responsible-AI literature. This distribution supports the review's central finding that technical AIOps research is comparatively mature, whereas governance-oriented applications in public-sector change and release management remain less developed

3. Overview of Selected Publications

The reviewed literature shows a strong concentration of work on AIOps for operational monitoring, anomaly detection, and incident response. A smaller body of work addresses public-sector AI governance, organizational readiness, and explainability. Very few studies directly examine how AI can be embedded in ITIL change and release management. Table 4 summarizes representative studies used to build the thematic synthesis.

Additional ML engineering studies make clear that AI-enabled software lifecycle management requires workflows that differ from traditional software development: data collection, model training, evaluation, deployment, monitoring, and feedback loops must be coordinated [20-22]. AIOps literature also shows that log analysis, anomaly detection, root-cause analysis, and multivariate time-series monitoring are mature research areas, but they are typically evaluated for incident response rather than formal change approval [23-26]. DevOps and continuous deployment research provide a process foundation for release governance, emphasizing automated testing, pipeline visibility, security, and cross-functional collaboration [27], [28]. Public-administration studies add constraints: AI adoption depends on public value tensions, absorptive capacity, organizational capabilities, and risk-aware governance [29-33].

Table 5: Representative primary and complementary studies

Ref(s)	Focus area	Key contribution to this review
[4], [5]	AIOps failure management	Anomaly detection, event correlation, and real-world implementation challenges.
[6], [23]	Cloud and microservice AIOps	Incident detection, failure prediction, root-cause analysis, and cloud observability.
[8], [20]-[22]	MLOps and ML engineering	Lifecycle control, deployment readiness, monitoring, collaboration, and production maintenance.
[24]-[26]	Logs and anomaly detection	Deep log analysis, multivariate time-series detection, and emerging LLM-based methods.
[27], [28]	DevOps and release pipelines	CI/CD practices, automated testing, pipeline visibility, and cultural barriers.
[9]-[12], [29]-[33]	Public-sector AI governance	Transparency, accountability, institutional readiness, and organizational change.
[34]-[39]	Responsible AI controls	Model cards, dataset documentation, fairness, risk management, and AI management-system requirements.

4. Findings and Discussion

The thematic analysis shows that AI has strong operational value in monitoring and incident response, but weaker integration with formal governance processes. Figure 2 summarizes the main classification of AI in IT operations used in this review. The figure separates technical AIOps capabilities from governance-oriented ITIL touchpoints so that the reader can see where current research is mature and where change/release management remains underdeveloped.

4.1 AI-Driven Monitoring and Incident Management

The most mature area of AI-enabled IT operations is monitoring and incident management. AIOps studies demonstrate how machine learning can process logs, alerts, traces, and performance metrics to identify anomalies, correlate related events, and recommend likely root causes [4-6], [23-25]. These capabilities reduce

manual triage and support a shift from reactive operations toward predictive service management. For change and release management, the same telemetry can provide baseline behavior before deployment, early detection after deployment, and evidence for rollback or remediation decisions.

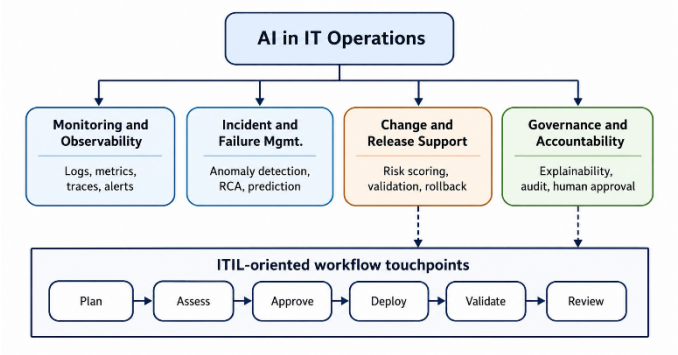

Figure 2: Classification of AI capabilities relevant to IT operations and ITIL workflows

4.2 *Change and Release Management as a Governance Bottleneck*

Change and release workflows remain less automated than monitoring workflows. Traditional ITIL processes rely on predefined approval paths and human assessment of change risk. DevOps and CI/CD research show that automated testing, deployment pipelines, and continuous feedback can improve release frequency and reliability [27], [28]. However, public-sector environments often require additional approval controls because releases may affect regulated services, public records, security controls, or statutory reporting. The review therefore indicates that AI should be used as decision support for risk scoring, schedule optimization, dependency analysis, rollback readiness, and post-release validation rather than as an unchecked autonomous approver.

4.3 *Public-Sector Governance, Compliance, and Accountability*

Public-sector AI adoption is shaped by accountability, fairness, transparency, privacy, legal compliance, procurement rules, and public trust [9-12], [29-33]. These constraints are directly relevant to AI-enabled change management because the output of a risk-scoring model may influence whether a system change is approved, delayed, or escalated. Governance therefore requires evidence that a model was trained on appropriate data, evaluated under relevant conditions, monitored after deployment, and reviewed by accountable personnel.

Responsible-AI literature provides mechanisms for making operational AI more auditable. Model cards document intended use, evaluation conditions, limitations, and ethical considerations [34]. Dataset datasheets document data provenance, composition, collection processes, and recommended use [35]. Bias and fairness studies show that model behavior can be affected by data quality and hidden assumptions, while fairness toolkits provide techniques for detection and mitigation [36], [37]. Governance frameworks such as the NIST AI Risk Management Framework and ISO/IEC 42001 translate these principles into lifecycle controls for risk management, monitoring, documentation, and continual improvement [38], [39].

4.4 *Legacy Systems, Data Quality, and Integration Barriers*

A recurring barrier across AIOps and MLOps literature is the quality and integration of operational data. AIOps depend on reliable logs, metrics, alerts, incidents, configuration data, deployment records, and service dependencies [5], [6]. Public-sector agencies often operate a mixture of legacy applications, custom platforms, cloud services, and vendor systems, which can make data normalization and real-time correlation difficult. MLOps literature reinforces that AI models require monitoring, validation, and retraining after deployment [8], [20-22]. Without these controls, models used for change-risk prediction may become inaccurate as infrastructure, data patterns, and service dependencies change.

4.5 *AI-Enabled ITIL Change and Release Framework*

Based on the synthesis, this review proposes a framework that links operational data, AIOps analytics, MLOps governance, ITIL change/release processes, and public-sector oversight. The framework is not intended to remove human control. Instead, it shows where AI can strengthen evidence-based decision-making while keeping accountability, explainability, and compliance within the governance process. Figure 3 presents the proposed framework.

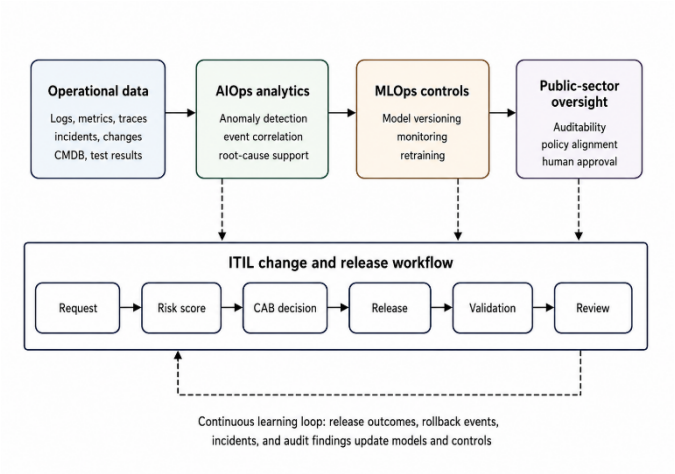

Figure 3: Proposed AI-enabled ITIL change and release management framework

The framework can be operationalized through lifecycle controls that connect planning, assessment, approval, deployment, validation, and review. Figure 4 shows how these controls distribute across the AI-enabled release process. It also clarifies that governance controls do not occur only at approval time; they appear before, during, and after deployment.

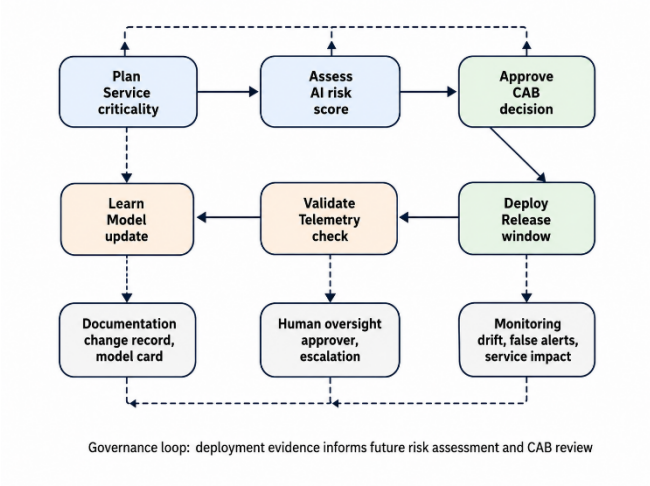

Figure 4: Governance controls across the AI-enabled release lifecycle

4.6 *Design Requirements for AI-Assisted CAB Decisions*

The synthesis indicates that an AI-assisted CAB should be designed as an accountable decision-support capability rather than a replacement for review boards. The system should combine operational telemetry with change records, but it must also expose why a recommendation was made and whether the supporting evidence is complete. Table 6 translates the literature into design requirements for public-sector use.

Table 6: Design requirements for AI-assisted change advisory decisions

Requirement	Rationale	Implementation controls
Explainable risk score	CAB members need to understand why change is low, medium, or high risk.	Show risk drivers, affected services, historical incidents, dependency confidence, and uncertainty.
Data provenance	Operational data may be fragmented across legacy systems and vendor platforms.	Track source systems, time windows, missing values, normalization logic, and dataset limits.
Human approval	Public-sector accountability cannot be delegated fully to an algorithm.	Require named approver, escalation path, and policy-based exception handling.
Release evidence	Approval should be connected to testing and deployment readiness.	Link test results, security checks, rollback plan, change window, and service-owner sign-off.
Post-release learning	The model should improve from release outcomes without hiding errors.	Feed incidents, rollback events, performance shifts, and user-impact data into review cycles.
Model governance	AIOps models can drift or become misaligned with policy requirements.	Use monitoring, retraining triggers, model cards, dataset documentation, and periodic audit review.

4.7 *Illustrative Public-Sector Workflow Simulation*

To make the proposed framework more concrete, consider an illustrative public-sector portal release. A change manager submits a release request for an online regulatory portal. The AI-assisted workflow first assembles evidence from the change record, impacted configuration items, prior incidents, test results, and service criticality. An AIOps model then generates a risk profile that identifies likely service dependencies, unusual historical outage patterns, and confidence limits in the available data. The CAB does not automatically accept the recommendation; instead, it uses the risk profile to ask targeted questions about rollback readiness, affected services, and compliance risk. During deployment, telemetry is monitored against historical baselines. After deployment, incidents, rollbacks, or abnormal performance shifts feed into the post-implementation review and future model retraining. Figure 5 summarizes this scenario as a decision-support workflow.

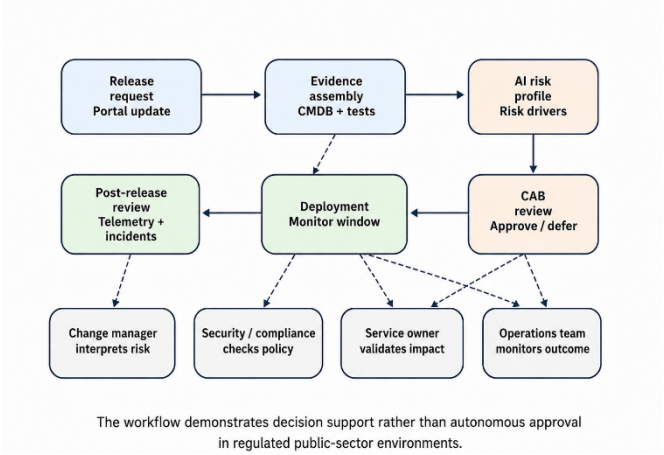

Figure 5: Illustrative public-sector workflow simulation for a portal release

5. Research gaps and Future Directions

The literature provides strong foundations for AIOps, MLOps, DevOps, and responsible AI, but several gaps remain when these areas are applied to public-sector

change and release management. Table 7 summarizes the main gaps and future research directions.

Table 7: Research gaps and future directions

Gap	Problem	Future direction
Limited empirical evidence	Few studies evaluate AI in real public-sector change/release workflows.	Conduct case studies using operational change records, incidents, release outcomes, and audit findings.
Weak explainability for CAB decisions	Risk scores alone are insufficient for accountable approval.	Develop explanation methods that show risk drivers, data confidence, affected services, and rollback readiness.
AI-ready change records	Change tickets often lack structured features needed for prediction.	Standardize change metadata: affected CI, dependency, test result, outage history, approval path, and service criticality.
Governance-aware MLOps	AIOps models are rarely connected to public-sector policy controls.	Map model monitoring, retraining, validation, access control, and documentation to ITIL and AI governance requirements.
Human-AI collaboration	Automation can reduce effort but may reduce accountability if poorly designed.	Evaluate decision-support interfaces for CABs, release managers, service owners, and compliance reviewers.
Transferability across agencies	Public agencies differ in infrastructure, law, data maturity, and risk tolerance.	Build adaptable frameworks with maturity levels rather than one-size-fits-all automation models.

Future research should move from general AIOps capability descriptions toward process-specific evaluation. Public-sector studies should examine where AI recommendations can safely enter the change workflow, which decisions remain strictly human-controlled, and how agencies can measure improvement in release risk, incident reduction, audit readiness, and service continuity. Research should also examine how MLOps controls can be embedded into ITSM so that operational AI models remain reliable after infrastructure, data, and policy conditions change.

A practical research direction is the development of an AI-assisted CAB decision-support model. Such a model would combine change history, dependency maps, incident records, test results, service criticality, and compliance requirements to produce an explainable change-risk profile. The profile should support, not replace, human decision-making by presenting risk factors, uncertainty, relevant precedent, and recommended mitigation actions. For public agencies, this design is preferable to black-box automation because it preserves accountability while improving evidence quality.

Another research direction concerns evaluation of metrics. Existing AIOps studies often evaluate detection accuracy or alert reduction, while public-sector change governance also requires measures such as auditability, approval transparency, incident avoidance, service-continuity protection, and stakeholder confidence. Future evaluations should therefore combine technical measures with organizational and governance outcomes.

6. Practical Implications for Public-Sector IT

For public-sector agencies, the findings suggest that AI-enabled change and release management should be introduced through a staged maturity path. Agencies should first improve the quality of change records, service maps, incident histories, and release metrics before relying on predictive models. Without these foundations, AI outputs may appear precise while reflecting incomplete data. The practical starting point is therefore not automation of approvals; it is improving operational observability and creating reliable evidence for human review.

A second implication is that AI should be embedded into existing ITIL and governance roles rather than deployed as a separate analytics experiment. Change managers, release managers, service owners, security teams, data stewards, and compliance reviewers should have defined responsibilities in the AI-enabled workflow. This role clarity is important because public-sector releases often involve statutory obligations, vendor dependencies, security controls, and citizen-facing services. A risk score is useful only when the organization knows who must interpret it, who can override it, and how the decision is documented.

A third implication concerns procurement and vendor management. Many agencies adopt monitoring, cloud, ITSM, and AIOps tools through vendor platforms. Procurement language should require data exportability, audit logs, model documentation, access controls, and explainability features. These requirements help prevent dependency on opaque vendor models and support later integration with agency-specific ITIL workflows, risk registers, and compliance evidence.

Table 8 summarizes a phased implementation roadmap for AI-enabled change and release management in public-sector IT. The roadmap highlights how agencies can move from foundational data readiness and observability toward predictive decision support, governance integration, and continuous improvement. This staged approach ensures that AI adoption remains aligned with ITIL workflows, human oversight, compliance expectations, and operational reliability.

Table 8: Implementation roadmap for AI-enabled change and release management

Phase	Key actions	Expected outcome
Phase 1: Data readiness	Standardize change tickets, incident taxonomies, configuration items, and service criticality.	Cleaner operational evidence and more reliable baseline metrics.
Phase 2: Observability	Connect logs, metrics, traces, alerts, test outcomes, and deployment records.	Earlier detection of abnormal release behavior and stronger post-release validation.
Phase 3: Predictive support	Pilot change-risk scoring using historical releases and incident outcomes.	Evidence-based CAB discussion without removing human approval.
Phase 4: Governance integration	Map AI outputs to ITIL controls, approval policies, model cards, dataset records, and audit trails.	Traceable decisions and stronger compliance readiness.
Phase 5: Continuous improvement	Monitor model drift, review false positives/negatives, update training data, and evaluate agency outcomes.	Improved reliability, reduced release risk, and sustainable AI operations.

The roadmap also defines measurable outcomes for pilot work. Agencies should evaluate the reduction in post-release incidents, mean time to detect abnormal behavior, false-positive and false-negative risk alerts, CAB cycle time, rollback preparedness, and audit evidence completeness. These metrics are more useful than broad claims about automation because they connect AI adoption to public-sector service reliability and governance outcomes.

7. Limitations

This review has three limitations. First, the available literature contains more studies on monitoring, anomaly detection, and incident response than on AI-enabled change approval or release governance. As a result, some findings are transferred from adjacent areas such as MLOps, DevOps, cloud observability, and public-sector AI governance. Second, the review emphasizes peer-reviewed and high-quality governance sources, but implementation practices in agencies may also be documented in internal playbooks, procurement records, and vendor reports that are not publicly available. Third, the proposed framework is conceptual and should be validated through agency-level case studies, interviews, and pilot deployments.

These limitations do not weaken the main finding; rather, they clarify where empirical work is needed. The literature is sufficiently mature to identify the technical and governance components required for AI-assisted change and release management, but it is not yet mature enough to claim that a single reference architecture will fit all public-sector agencies. Future validation should therefore compare different agency sizes, risk profiles, service portfolios, and data maturities.

8. Conclusion

This systematic literature review examined the transition from ITIL-based service governance toward AI-enabled operations in public-sector IT. The review shows that AIOps is well developed for monitoring, anomaly detection, incident response, and failure prediction. MLOps provides complementary controls for model deployment, validation, monitoring, and retraining. Public-sector AI literature emphasizes transparency, accountability, governance, and organizational readiness.

The central finding is that AI has not yet been sufficiently integrated into formal change and release management. Intelligent monitoring provides useful technical foundations, but public-sector agencies need validated frameworks that connect operational analytics with ITIL workflows, human approval, compliance requirements, and explainable recommendations. The proposed framework positions AI as a decision-support layer that can improve risk assessment, release planning, and post-release validation while preserving governance accountability.

For practice, the review suggests that public-sector IT leaders should begin with high-value, low-risk AI use cases such as anomaly detection, post-release monitoring, and evidence-based change-risk assessment. For research, the next step is empirical validation through public-sector case studies, pilot implementations, and standard evaluation metrics that determine how AI can improve service continuity, operational efficiency, and trustworthy governance in public-sector IT environments.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

The author thanks the reviewers for their constructive feedback, which helped improve the methodological transparency, figure explanations, and practical framing of this review.

References

[1] AXELOS, ITIL Foundation: ITIL 4 Edition. London, U.K.: The Stationery Office, 2019.

[2] J. Iden and T. R. Eikebrokk, "Implementing IT Service Management: A systematic literature review," International Journal of Information Management, vol. 33, no. 3, pp. 512-523, 2013, doi:10.1016/j.ijinfomgt.2013.01.004.

[3] J. Serrano, J. Faustino, D. Adriano, R. Pereira, and M. M. da Silva, "An IT Service Management Literature Review: Challenges, Benefits, Opportunities and Implementation Practices," Information, vol. 12, no. 3, article 111, 2021, doi:10.3390/info12030111.

[4] P. Notaro, J. Cardoso, and M. Gerndt, "A Survey of AIOps Methods for Failure Management," ACM Transactions on Intelligent Systems and Technology, vol. 12, no. 6, article 81, pp. 1-45, 2021, doi:10.1145/3483424.

[5] Y. Dang, Q. Lin, and P. Huang, "AIOps: Real-World Challenges and Research Innovations," in 2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), 2019, pp. 4-5, doi:10.1109/ICSE-Companion.2019.00023.

[6] Q. Cheng et al., "AI for IT Operations (AIOps) on Cloud Platforms: Reviews, Opportunities and Challenges," arXiv:2304.04661, 2023, doi:10.48550/arXiv.2304.04661.

[7] M. E. Kanakis, R. Khalili, and L. Wang, "Machine Learning for Computer Systems and Networking: A Survey," ACM Computing Surveys, vol. 55, no. 4, article 71, pp. 1-36, 2022, doi:10.1145/3523057.

[8] D. Kreuzberger, N. Kuehl, and S. Hirschl, "Machine Learning Operations (MLOps): Overview, Definition, and Architecture," IEEE Access, vol. 11, pp. 31866-31879, 2023, doi:10.1109/ACCESS.2023.3262138.

[9] A. Zuiderwijk, Y.-C. Chen, and F. Salem, "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, article 101577, 2021, doi:10.1016/j.giq.2021.101577.

[10] B. W. Wirtz, J. C. Weyerer, and C. Geyer, "Artificial Intelligence and the Public Sector: Applications and Challenges," International Journal of Public Administration, vol. 42, no. 7, pp. 596-615, 2019, doi:10.1080/01900692.2018.1498103.

[11] I. Mainardi, "Change management: artificial intelligence (AI) at the service of public administrations," AI & Society, vol. 40, no. 5, pp. 3953-3981, 2025, doi:10.1007/s00146-024-02136-2.

[12] J. Schwaeke, C. Gerlich, H. L. Nguyen, D. K. Kanbach, and J. Gast, "Artificial intelligence (AI) for good? Enabling organizational change towards sustainability," Review of Managerial Science, vol. 19, pp. 3013-3038, 2025, doi:10.1007/s11846-025-00840-x.

[13] R. D. Zota, C. Barbulescu, and R. Constantinescu, "A Practical Approach to Defining a Framework for Developing an Agentic AIOps System," Electronics, vol. 14, no. 9, article 1775, 2025, doi:10.3390/electronics14091775.

[14] D. Sculley et al., "Hidden Technical Debt in Machine Learning Systems," in Advances in Neural Information Processing Systems 28, 2015, pp. 2503-2511.

[15] E. Breck, S. Cai, E. Nielsen, M. Salib, and D. Sculley, "The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction," in 2017 IEEE International Conference on Big Data, 2017, pp. 1123-1132, doi:10.1109/BigData.2017.8258038.

[16] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," IEEE Access, vol. 6, pp. 52138-52160, 2018, doi:10.1109/ACCESS.2018.2870052.

[17] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," BMJ, vol. 372, article n71, 2021, doi:10.1136/bmj.n71.

[18] C. Okoli, "A Guide to Conducting a Standalone Systematic Literature Review," Communications of the Association for Information Systems, vol. 37, article 43, 2015, doi:10.17705/1CAIS.03743.

[19] V. Braun and V. Clarke, "Using thematic analysis in psychology," Qualitative Research in Psychology, vol. 3, no. 2, pp. 77-101, 2006, doi:10.1191/1478088706qp063oa.

[20] S. Amershi et al., "Software Engineering for Machine Learning: A Case Study," in 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), 2019, pp. 291-300, doi:10.1109/ICSE-SEIP.2019.00042.

[21] L. E. Lwakatare, A. Raj, J. Bosch, H. H. Olsson, and I. Crnkovic, "A Taxonomy of Software Engineering Challenges for Machine Learning Systems: An Empirical Investigation," in Agile Processes in Software Engineering and Extreme Programming, Cham, Switzerland: Springer, 2019, pp. 227-243, doi:10.1007/978-3-030-19034-7_14.

[22] A. Paleyes, R.-G. Urma, and N. D. Lawrence, "Challenges in Deploying Machine Learning: A Survey of Case Studies," ACM Computing Surveys, vol. 55, no. 6, article 114, pp. 1-29, 2022, doi:10.1145/3533378.

[23] J. Soldani and A. Brogi, "Anomaly Detection and Failure Root Cause Analysis in (Micro)Service-Based Cloud Applications: A Survey," ACM Computing Surveys, vol. 55, no. 3, article 59, pp. 1-39, 2022, doi:10.1145/3501297.

[24] M. Du, F. Li, G. Zheng, and V. Srikumar, "DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning," in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1285-1298, doi:10.1145/3133956.3134015.

[25] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network," in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2019, pp. 2828-2837, doi:10.1145/3292500.3330672.

[26] J. Su et al., "Large Language Models for Forecasting and Anomaly Detection: A Systematic Literature Review," arXiv:2402.10350, 2024, doi:10.48550/arXiv.2402.10350.

[27] M. Shahin, M. A. Babar, and L. Zhu, "Continuous Integration, Delivery and Deployment: A Systematic Review on Approaches, Tools, Challenges and Practices," IEEE Access, vol. 5, pp. 3909-3943, 2017, doi:10.1109/ACCESS.2017.2685629.

[28] L. Riungu-Kalliosaari, S. Mäkinen, L. E. Lwakatare, J. Tiihonen, and T. Männistö, "DevOps Adoption Benefits and Challenges in Practice: A Case Study," in Product-Focused Software Process Improvement, Cham, Switzerland: Springer, 2016, pp. 590-597, doi:10.1007/978-3-319-49094-6_44.

[29] R. Madan and M. Ashok, "AI adoption and diffusion in public administration: A systematic literature review and future research agenda," Government Information Quarterly, vol. 40, no. 1, article 101774, 2023, doi:10.1016/j.giq.2022.101774.

[30] P. Mikalef et al., "Enabling AI capabilities in government agencies: A study of determinants for European municipalities," Government Information Quarterly, vol. 39, no. 4, article 101596, 2022, doi:10.1016/j.giq.2021.101596.

[31] B. W. Wirtz and W. M. Müller, "An integrated artificial intelligence framework for public management," Public Management Review, vol. 21, no. 7, pp. 1076-1100, 2019, doi:10.1080/14719037.2018.1549268.

[32] J. Berryhill, K. K. Heang, R. Clogher, and K. McBride, Hello, World: Artificial Intelligence and Its Use in the Public Sector. Paris, France: OECD Publishing, 2019, doi:10.1787/726fd39d-en.

[33] OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, Paris, France: OECD, 2019.

[34] M. Mitchell et al., "Model Cards for Model Reporting," in Proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, pp. 220-229, doi:10.1145/3287560.3287596.

[35] T. Gebru et al., "Datasheets for Datasets," Communications of the ACM, vol. 64, no. 12, pp. 86-92, 2021, doi:10.1145/3458723.

[36] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A Survey on Bias and Fairness in Machine Learning," ACM Computing Surveys, vol. 54, no. 6, article 115, pp. 1-35, 2021, doi:10.1145/3457607.

[37] R. K. E. Bellamy et al., "AI Fairness 360: An Extensible Toolkit for Detecting and Mitigating Algorithmic Bias," IBM Journal of Research and Development, vol. 63, no. 4/5, pp. 1-15, 2019, doi:10.1147/JRD.2019.2942287.

[38] E. Tabassi, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1. Gaithersburg, MD, USA: National Institute of Standards and Technology, 2023, doi:10.6028/NIST.AI.100-1.

[39] ISO/IEC, ISO/IEC 42001:2023, Information Technology - Artificial Intelligence - Management System. Geneva, Switzerland: International Organization for Standardization, 2023.

CATHERINE GANDURI is a Systems Analyst at the Texas Commission on Environmental Quality (TCEQ), USA. She holds a Master's degree in Business Analytics from The University of Texas at Dallas. Her research interests include Artificial Intelligence in IT Operations (AIOps), IT Service Management (ITIL), change and release management, and digital transformation in public sector systems. She has experience working with data analytics, enterprise systems, and process optimization to improve operational efficiency.