



AI-Enabled Cloud Operations for Intelligent IT Service Management in Saudi Enterprises under Vision 2030

Yahya Khalil^{1*} 

¹Independent Researcher

*Corresponding Author

Yahya Khalil

Independent Researcher

Article History

Received: 19.08.2026

Accepted: 12.10.2026

Published: 14.10.2026

Abstract: The move towards a cloud-based digital transformation has taken enterprise reliability from being merely a concern of the back office to making it a strategic capability. In Saudi Arabia this transition is supported by Vision 2030, the quick adoption of cloud technology, national investment in artificial intelligence, and the increasing reliance by both public and private organisations on digital services that are always available. Although traditional IT service management is still useful for controlling processes, its reliance on manual triage, monitoring based on thresholds, fragmented telemetry, and human-led diagnosis is becoming more and more strained as a result of distributed cloud architectures. This study brings together research that was published between 2020 and 2025 on the use of artificial intelligence in IT operations, cloud failure prediction, log and trace analytics, incident intelligence, root-cause analysis, automated remediation, and intelligent service desks. The objective is to see how these capabilities can be combined into a governable operating model for businesses in Saudi Arabia rather than being implemented as separate algorithms. The evidence shows that AI can enhance early detection of failures, the classification of incidents, prioritisation, causal diagnosis, knowledge retrieval, and certain remediation tasks, but its performance is very much dependent on the quality of the telemetry, the service topology, controls against model drift, explainability, and the limits set on autonomous action. The paper puts forward a framework for AI-enabled Cloud Operations and Intelligent IT Service Management based on five layers: observability, context, intelligence, orchestration, and governance. It also defines a closed-loop assurance lifecycle in which machine-generated recommendations are continuously assessed in relation to service-level, risk, security, and business outcomes. The review concludes that the most significant contribution of Vision 2030 will be achieved through the controlled enhancement of human operations, not through uncontrolled automation, and it identifies key areas such as Arabic-aware service intelligence, cross-enterprise benchmarking, sovereign data controls, and measurable human-AI accountability.

Keywords: Artificial Intelligence; Cloud Operations; Intelligent IT Service Management (ITSM); AIOps; Cloud Reliability; Predictive Failure Detection; Root-Cause Analysis.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

1. INTRODUCTION

The structure of enterprise IT operations has changed as applications have shifted from relatively

stable on-premise systems to hybrid clouds, software-as-a-service platforms, containers, microservices, APIs, and continuously delivered

Citation: Khalil, Y. (2026). AI-Enabled Cloud Operations for Intelligent IT Service Management in Saudi Enterprises under Vision 2030; *Glob Acad J Econ Buss*, 8(5), 1863-1874. <https://doi.org/10.36348/gajeb.2026.v08i05.071>

digital products. The result of this change is not just an enlargement of the infrastructure scale but an increase in dependency density: a service that is visible to users may depend on dozens or hundreds of components, the metrics of which, along with the logs, traces, configuration changes, access events, tickets, and business transactions of these components, develop at different speeds. Failure management has thus become a problem involving the drawing of inferences from diverse types of evidence rather than simply being about reading alerts. Research into AIOps has developed in response to this issue, making use of machine learning, natural-language processing, graph reasoning, causal discovery, and increasingly large language models to assist with failure perception, diagnosis, and remediation [1-3].

This development has a direct connection with IT service management. ITSM offers the set of organisational procedures by means of which incidents, requests, problems, changes, service levels, configuration items, and knowledge are managed. Yet conventional workflows generally suppose that people are able to manually interpret the descriptions in tickets, identify the appropriate support group, check the monitoring dashboards, correlate events, and arrive at a reasonable corrective action within the required service timeframes. Research based on enterprise incident data indicates that machine-learning and transformer models are capable of detecting major incidents earlier than purely reactive methods and can automate some of the tasks involved in assignment and resolution [4, 5]. Other work concerning service-desk categorisation shows that text classification can cut down on routing effort where historical labels are reliable and the categories are well represented [6, 7]. Duration prediction introduces an additional stage of decision-making by providing an estimate of how long an incident might stay unresolved, which in turn allows for more effective escalation and allocation of workload [8].

The opportunity in this area is greater than that offered by ticket automation. It is possible for the architectures used for data-centre monitoring to integrate telemetry engineering, analytics, and machine-learning services into an AIOps pipeline [9]. Studies of cloud failures have demonstrated that the characteristics of resources, tasks, and jobs can be used to predict failures before they fully occur [10, 11]. In microservice environments, distributed tracing and the use of log representations allow anomaly detection to be carried out at a level closer to the actual service path [12, 13]. Further improvements in detection can be achieved through multi-modal observability since metrics on their own may not show how a fault spreads across a service

graph [14]. These capabilities form the foundation of an intelligent operations loop in which weak signals are identified early, probable causes are prioritised, incidents are automatically enriched, responders are given contextual recommendations, and low-risk recovery actions can be carried out in accordance with policy.

The issue is of strategic importance for Saudi companies. As part of Vision 2030, digital capability, economic diversification, cloud infrastructure and artificial intelligence have all been identified as key factors in boosting productivity and modernising services. Academic research has connected the Kingdom's strategy for diversification with investment in AI [27], and more recent studies from Saudi Arabia have found that the decision to adopt cloud services is affected by organisational readiness, the availability of skills, security expectations, policy conditions, and the extent to which cloud platforms are believed to enhance efficiency [28, 29]. Research into the use of generative AI in Saudi software engineering also highlights the potential for increased productivity together with the need to address issues of transparency, privacy, accountability and workforce development [30]. Given these considerations, an enterprise AIOps programme cannot be judged solely on the basis of model accuracy; it also has to be assessed in terms of whether it improves continuity, auditability, cybersecurity, service quality, cost discipline and human capability.

There is still a gap between studies based on algorithms and the design of how enterprises operate. A large part of the existing literature assesses specific tasks—such as anomaly detection or root-cause localisation—using benchmark data. Research into IT service management usually concentrates on the text of the tickets, while that concerned with cloud reliability looks at telemetry. Governance research, on the other hand, tends to deal with principles at a level that is disconnected from the actual incident workflows. Saudi organisations need an integrated perspective on several points: how telemetry should be structured, how AI outputs should be incorporated into ITSM processes, which actions can be automated, what evidence should be kept for audit purposes, and how performance should be measured in terms of both technical and business outcomes. The present review tackles this issue of integration by developing a governance-focused framework for AI-enabled cloud operations and intelligent IT service management.

2. AIM, OBJECTIVES AND REVIEW QUESTIONS

This study aims at creating an evidence-based governance framework for the adoption of AI-enabled cloud operations within enterprise IT service

management in Saudi Arabia as part of Vision 2030. The review does not regard AIOps as a single product category; instead, it considers it to be a socio-technical operating capability which incorporates cloud telemetry, service context, analytical models, automation, ITSM workflows, human expertise, and governance controls.

The study is guided by four objectives. The first of these is to identify the operational functions for which AI has shown practical value, such as failure prediction, anomaly detection, ticket intelligence, root-cause analysis, knowledge support, and remediation. The second objective is to look at the data and architectural conditions necessary for those functions to operate reliably in today's cloud environments. Thirdly, the review examines the risks which can hinder adoption, namely data fragmentation, class imbalance, concept drift, opaque recommendations, false-positive automation, security exposure, and weak accountability. The fourth objective is to convert the evidence into an implementation model that is appropriate for Saudi enterprises who are looking for measurable outcomes as a result of their digital transformation rather than just isolated proofs of concept.

The work is structured around four research questions. The first of these, RQ1, investigates which AI capabilities generates the greatest operational value throughout the incident lifecycle. RQ2 looks into the factors relating to telemetry, context, and model characteristics that affect reliability and transferability. RQ3 examines how enterprises can manage both human and automated decisions without losing response speed. RQ4 explores how an AIOps-enabled ITSM model can contribute to the Vision 2030 priorities such as resilient digital services, cloud adoption, workforce capability, productivity, and trustworthy AI. These questions intentionally link technical evidence with the design of the operating model since sustainable adoption requires both.

3. REVIEW METHODOLOGY

To carry out the review a structured integrative approach was adopted, bringing together evidence from the fields of software engineering, cloud computing, IT service management, machine learning, and research into digital transformation in Saudi Arabia. The time period covered by the review was limited to 2020–2025 in order to take into account the present state of cloud-native operations and the fast development of AI techniques. The protocol was set out in terms of clear objectives, research questions, search terms, criteria for eligibility, assessment of quality, extraction of data, thematic synthesis, identification of limitations, and the development of practice-oriented

recommendations. This kind of structure was chosen in order to ensure methodological transparency while allowing for the integration of diverse evidence that cannot be combined through statistical methods. By separating the selection process from the synthesis phase, it is possible to reduce the risk of confusing the performance of technical models with evidence relating to organisational adoption, a distinction that is especially significant when combining studies on cloud reliability with research set in the context of Saudi enterprises and which in turn enhances consistency throughout the later analytical sections.

The search concepts were divided into four categories. The first of these was concerned with technology and included the following terms: "AIOps", "artificial intelligence for IT operations", "machine learning", "deep learning", "large language model", and "intelligent automation". The second category related to operational activities and comprised the terms: "incident management", "service desk", "failure prediction", "anomaly detection", "root cause analysis", "event correlation", "self-healing", and "automated remediation". The third category dealt with architecture and evidence sources and included "cloud computing", "microservices", "distributed systems", "logs", "metrics", "traces", "observability", and "service topology". The fourth category was about context and consisted of "Saudi Arabia", "Vision 2030", "cloud adoption", "digital transformation", and "software engineering". Combinations of these terms were employed in order to find journal and conference literature that had been peer-reviewed and was relevant to the review questions.

To be included, a study had to have been published between 2020 and 2025, include a permanent DOI, be substantially relevant to at least one of the research questions, and provide sufficient methodological details so that it could be determined what data had been used and what operational outcome had been assessed. Studies were kept in the set if they offered evidence regarding detection, prediction, diagnosis, ITSM automation, recovery, cloud adoption, or the Saudi digital environment. Those that were excluded were duplicates, non-academic commentaries, promotional material for products, articles lacking a verifiable DOI, and works in which the only link to AI or cloud operations was incidental. The final body of evidence consists of 30 references, all of which were checked against publisher, indexing, institutional, or DOI-resolution records before being included.

The assessment was based on five aspects: the clarity of purpose, the transparency of the data and methods, the appropriateness of the evaluation

metrics, the relevance to actual operational conditions, and the recognition of limitations. A numerical quality score was not employed since the body of evidence includes different types of study designs which cannot be meaningfully compared using a single overall figure. Rather, the conclusions were evaluated qualitatively. Evidence from large industrial datasets and actual production deployments was considered to be stronger support for operational feasibility than that from small benchmark experiments; surveys were used to describe the situation in the field rather than to prove performance; and the Saudi context studies were used to understand the conditions under which adoption would occur rather than to deduce technical accuracy.

The data extraction included the publication year, the operational problem, the source of the evidence, the AI technique used, the evaluation setting, the main performance outcome, the stated limitations, and the governance implications. The synthesis applied thematic coding within six operational themes: observability and the perception of failure; ticket intelligence; predictive incident management; causal diagnosis; remediation and self-

healing; and governance together with human oversight. A cross-theme analysis looked at the recurring dependencies, for instance by comparing anomaly-detection performance with the availability of traces and topology; by examining ticket models in relation to label quality and imbalance; by assessing remediation studies in terms of the amount of human or policy control; and by interpreting the studies carried out in Saudi Arabia in relation to skills, security, regulation, and organisational readiness.

The method makes a deliberate effort to avoid including screening counts that are not supported. Instead of citing database totals which were not obtained through a formal bibliometric export, the review corpus was based on 30 studies whose DOIs have been verified and which were chosen for their direct relevance. This strategy enhances transparency since the unit of analysis is the evidence that has actually been synthesised in the paper. Table 1 gives a summary of the review design and the way in which the data were extracted. The thematic synthesis that follows is interpretive and not meta-analytic because the various studies employ different tasks, datasets, metrics, and operational environments.

Table 1: Review design and evidence synthesis protocol

Dimension	Specification	Analytical purpose
Review type	Structured integrative review	Combines heterogeneous evidence across AIOps, ITSM, cloud reliability, and Saudi digital context.
Time window	2020-2025	Reflects cloud-native operating conditions and current AI methods.
Evidence eligibility	Peer-reviewed journal or conference studies with a verifiable DOI and substantive relevance	Supports traceability and a research-quality evidence base.
Search domains	AIOps/AI; ITSM/incident; cloud/microservices/observability; Saudi/Vision 2030	Captures both technical mechanisms and adoption context.
Quality appraisal	Clarity, method transparency, evaluation appropriateness, operational relevance, limitations	Weights evidence without forcing heterogeneous designs into an invalid composite score.
Extraction fields	Problem, data source, AI method, evaluation, outcome, limitations, governance implication	Enables comparable synthesis across technical and organisational studies.
Thematic synthesis	Observability; ticket intelligence; prediction; causal diagnosis; self-healing; governance/human oversight	Builds the evidence base for the proposed governance framework.

4. Evidence Synthesis: From Cloud Telemetry to Intelligent Service Management

4.1 Observability and Early Failure Perception

AIOps begins with the quality of the operational evidence. Today's cloud systems produce time-series metrics, as well as structured and unstructured logs, distributed traces, deployment events, changes to configuration, security signals, and records from the service desk. Surveys repeatedly refer to failure detection as a basic AIOps function [1, 2]. The issue is not a lack of data but rather the fragmentation and the large scale of the data. Since monitoring tools can generate thousands of alerts

without making it clear whether several different symptoms are due to a single dependency failure, an intelligent way of operating therefore requires a common telemetry infrastructure that standardises timestamps, entity identifiers, service ownership, and topology before applying the more advanced models.

Studies of cloud failure prediction show the benefits of drawing on both resource and task behaviour. By using workload characteristics, machine-learning models are able to classify failed jobs or estimate the probability of failure, though

their performance is dependent on the representativeness of the traces and on the stability of the feature relationships [10, 11]. At the infrastructure level, AIOps architectures combine monitoring, data processing, analytical services, and operational applications in such a way that anomaly detection is linked to a practical workflow rather than remaining as an offline model [9]. This aspect is especially important in hybrid environments since cloud resources, network devices, enterprise applications, and data-centre systems have different sampling rates and ownership boundaries.

Logs still play a central role since they record the observations of both developers and the system, observations which may not be reflected in the metrics. Although surveys of deep-learning methods have shown considerable advancement in the area of automated detection of log anomalies, they also point out ongoing difficulties concerning preprocessing, changes to log templates, explainability, the representativeness of benchmarks, and the overall reproducibility [20-22]. It is useful to apply cross-system transfer learning in cases where a newly deployed service has insufficient labelled anomalies, since knowledge gained from more established systems can improve the detection process without needing a long history of local incidents [23]. More recent end-to-end methods decrease the reliance on manually constructed parsing pipelines and demonstrate that the models are able to generalise across a number of log datasets [19]. The practical consequence is that Saudi enterprises should regard log quality, retention, semantic consistency, and privacy classification as AI-readiness controls, rather than just as observability settings.

Distributed traces provide causal context by illustrating the path that requests take as they move between different services. Trace-based anomaly detection is able to spot abnormal spans and assist in identifying performance regressions [12]. ServiceAnomaly proves that when traces are combined with profiling metrics, relationships can be captured that would otherwise be missed by a single telemetry stream [14]. Micro2vec also shows that numerical representations obtained from microservice logs can reveal abnormal relationships between different components [13]. These results support a multi-modal approach: AI should reason about correlated logs, metrics, traces, changes, and tickets rather than trying to optimise each source separately.

4.2 Ticket Intelligence and Predictive Incident Management

ITSM achieves an intelligent level when the operational telemetry is connected to the service record that is used by the support teams. One of the

most compelling examples of this is incident severity prediction. A study carried out on about 500,000 incident records revealed that transformer, deep-learning, and gradient-boosting models were able to predict major incidents with a useful lead time, thereby providing an opportunity to escalate the risk before service degradation becomes catastrophic [4]. Another related knowledge-based system automated the prediction of assignment groups and provided support for incident resolution by making use of large amounts of industrial ticket data, demonstrating how historical service knowledge can be turned into immediate operational assistance [5].

Ticket routing is a more focused yet highly practical feature. When the descriptions of tickets contain enough semantic information, deep neural classifiers can perform better than traditional methods [6]. Furthermore, transformer-based representations enhance the process of categorisation when the service-desk datasets are imbalanced or show linguistic variability [7]. Duration prediction takes ticket intelligence beyond the question of "where should this go?" to include "how difficult is this likely to be?", thereby aiding in workload balancing and enabling proactive escalation [8]. However, these models carry over the historical behaviour that existed; if previous assignment groups illustrated organisational silos, inadequate categorisation, or inconsistent escalation, the AI could instead automate those same weaknesses more quickly. Governance must therefore involve ownership of the taxonomy, regular relabelling, and validation based on outcomes rather than simply relying on accuracy against historical labels.

A practical solution would be to produce an enriched incident object. Whenever an anomaly causes a record to be created, the platform could include the business service that has been affected, the most recent deployments, the relationships between configuration items, similar past incidents, an estimated severity, a predicted duration, the candidate group of people who should resolve it, and any relevant knowledge articles. Instead of receiving a raw alert, the human operator is given a compact context pack. Although large language models are capable of providing conversational summaries and enabling retrieval from knowledge bases, the 2025 AIOps literature cautions that approaches based on LLMs lead to new requirements when it comes to evaluation since fluent explanations may hide uncertainty or unsupported reasoning [3]. Retrieval grounding, source citation, confidence thresholds, and restricted action permissions are therefore necessary.

4.3 Root-Cause Analysis as Causal Reasoning

Detection informs operators that there is a problem; root-cause analysis decides where action should start. In microservice systems this is hard since failures spread. Correlation can detect components that change together, but it is not able to reliably tell the origin of a failure from downstream symptoms. CausalRCA employs causal structure learning based on monitoring metrics in order to identify fine-grained causes [15]. Neural Granger methods explicitly model temporal relationships so that an anomaly in a resource that occurs earlier can be differentiated from later service latency [16]. Other work in causal discovery reduces the search space in a hierarchical manner to make diagnosis scalable across large service graphs [17]. Nevertheless, comparative studies show that causal methods are still sensitive to assumptions, parameter selections, data quality, and the difference between synthetic benchmarks and real systems [18].

In the context of enterprise governance, root-cause outputs should be regarded as ranked hypotheses rather than as absolute truths. An intelligent ITSM process should make use of the ranking while still maintaining evidential traceability. Each recommendation must make clear the telemetry that has been used, the affected dependency path, the confidence level, the relevant recent changes, and the reason why a particular component has been given priority. Explainability is operationally useful not since every model needs to be mathematically transparent to every user, but because the people taking action have to be able to decide whether the evidence makes it worthwhile to carry out a disruptive action such as a rollback, a restart, a shift in traffic, or a change to an access policy.

4.4 Automated Remediation and Self-Healing

The last stage of the AIOps loop is action. Research carried out in production environments on automated intelligent healing shows that machine learning is capable of selecting repair actions on a cloud scale and of reducing the time during which the system is unavailable, provided that the failure patterns and the recovery actions are well structured [25]. Surveys of proactive self-healing classify the

recovery mechanisms according to the stages of detection, diagnosis, planning, and execution, yet they also indicate that reliability is dependent on fault coverage, policy design, validation, and on the ability to prevent an automated action from making the incident worse [26]. Research in the area of edge-oriented AIOps includes one additional factor: since some analytics can operate close to the distributed infrastructure, latency is reduced but there is an increased requirement for consistent model and policy management across different locations [24].

Graduated autonomy is thus the most effective enterprise pattern. Automated actions can be taken in the case of low-risk, reversible, and well-observed situations provided that the predetermined controls are met; examples of this are restarting a stateless service instance, scaling a resource within approved limits, clearing a known transient queue, or carrying out a validated runbook. Medium-risk actions should require human approval together with AI evidence. High-risk actions which have an impact on data integrity, security boundaries, financial processing, core ERP configuration, or regulated services should be handled by humans unless a particular control framework permits them to be automated. The level of automation should be achieved through actual observed performance rather than being assumed simply because a model is technically able to carry out an action.

5. PROPOSED GOVERNANCE FRAMEWORK FOR SAUDI ENTERPRISES

The evidence points to a five-layer model for AI-enabled Cloud Operations and Intelligent IT Service Management, with Figure 1 showing the architecture. The first layer is the observability plane, in which metrics, logs, traces, events, ticket text, configuration data, identity signals, and business-service indicators are gathered using a common time and entity semantics. The second layer is the context plane, and it links the telemetry to the service topology, configuration items, ownership, criticality, data classification, recent changes, and service-level commitments; without this layer, the models may detect anomalies but will not be able to explain their business significance.

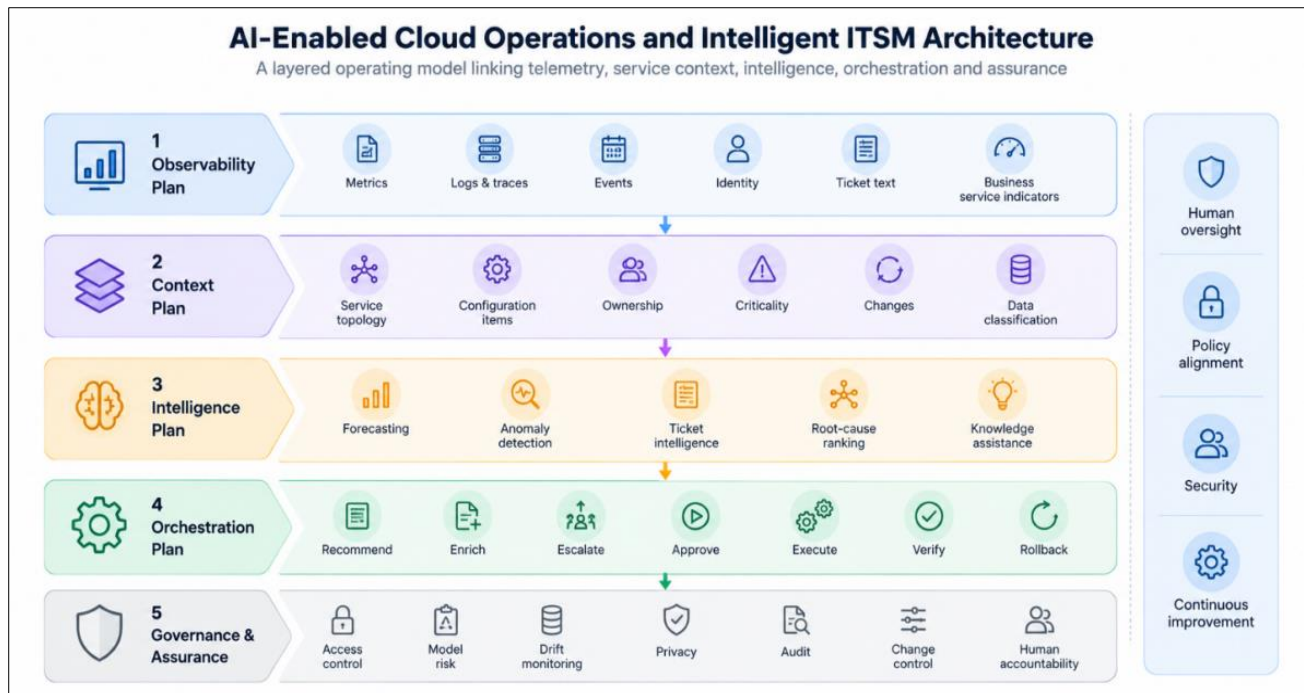


Figure 1: Five-layer architecture for AI-enabled cloud operations and intelligent IT service management

The intelligence plane makes up the third layer and includes specialised models for forecasting, for detecting anomalies, for classifying tickets, for predicting severity and duration, for carrying out similarity retrieval, for ranking possible root causes, and for generating knowledge. As the literature reviewed shows, there is no single universal model. Since different operational tasks call for different data structures and validation methods [1-3], anomaly detectors should be assessed according to the number of false positives and detection latency; ticket models according to routing quality and the effectiveness of downstream resolution; RCA models according to top-k localisation and the amount of time saved; and language models according to their groundedness, factual accuracy, and usefulness to the responders.

The fourth layer is known as the orchestration plane. In this layer, the outputs from the AI are fed into the ITSM workflows, collaboration channels, change controls, runbooks, and automation platforms. It is necessary for this layer to turn the model's outputs into clear operational states such as

recommend, enrich, escalate, request approval, execute, verify, or rollback. The main design principle is to separate inference from authority. Although the model may infer that the likely cause is the database connection pool, the permission to restart a service or modify a policy is determined by the workflow rules, the service's criticality, and human accountability.

The fifth level is that of governance and assurance and covers all of the other levels. It involves model ownership, data access, cybersecurity, privacy, audit logging, model-risk review, drift monitoring, change management, segregation of duties, and post-incident evaluation. Table 2 converts these controls into operational measures. Governance should be built into the platform rather than being set out separately. Each recommendation and any action taken autonomously must be traceable to a particular model version, evidence set, policy decision, actor, timestamp, and outcome. Thus an auditable chain is established that runs from telemetry through to the decision and on to the business effect.

Table 2: Governance controls and outcome measures for AI-enabled IT service management

Capability	AI-enabled role	Primary operational measure	Required governance control	Vision 2030 contribution
Observability	Correlate metrics, logs, traces, events	Detection latency; false-positive burden	Retention, access, data classification, topology quality	Resilient digital services
Ticket intelligence	Classify, route, summarise, enrich	Routing quality; reassignment rate; first-response time	Taxonomy ownership; label audits; confidence thresholds	Workforce productivity

Capability	AI-enabled role	Primary operational measure	Required governance control	Vision 2030 contribution
Predictive incident management	Predict severity, duration, or failure	Lead time; avoided escalations; service-level attainment	Drift monitoring; imbalance controls; calibration	Proactive continuity
Root-cause analysis	Rank probable causes and dependency paths	Top-k localisation; diagnosis time	Evidence traceability; human validation	Service resilience
Knowledge assistance	Retrieve and summarise runbooks and incident history	Grounded answer rate; reuse; operator acceptance	Approved corpora; source traceability; leakage controls	Institutional knowledge
Remediation	Recommend or execute validated runbooks	Recovery success; rollback rate; mean time to restore	Policy gates; reversibility; blast-radius limits	Reliable automation
Assurance	Monitor model and action outcomes	Drift alerts; policy violations; override rate	Model registry; audit logs; segregation of duties	Trustworthy AI
Workforce	Augment human diagnosis and decision-making	Time saved; training attainment; escalation quality	Role definitions; competency development	Digital skills and human capital

Figure 2 shows the closed-loop assurance lifecycle. The various stages of the process start with observation and detection, continue with correlation and diagnosis, and then proceed to decision and action; the loop is only considered complete when the results are checked against service health and are recorded as learning evidence. If a false positive occurs, the thresholds or the training data should be

adjusted; if a runbook is found to be ineffective it should be revised; if a recovery is successful then that should be turned into reusable knowledge; and in the case of a major incident both the technical cause and the model's behaviour should be reviewed. This learning loop is essential since both cloud environments and organisational processes are in a state of constant change.

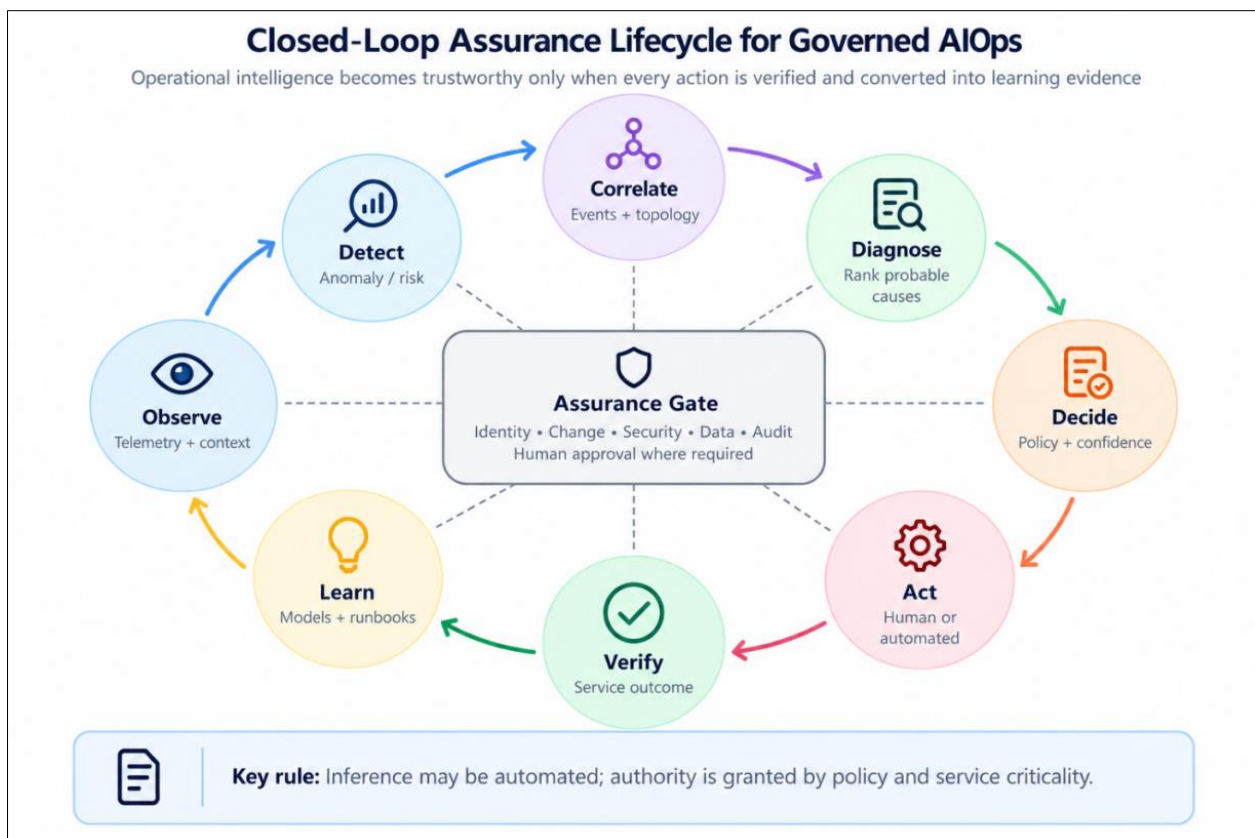


Figure 2: Closed loop assurance lifecycle

Three governance metrics are of particular importance. The first of these is automation precision, which refers to the number of times automated or AI-recommended actions result in the desired recovery without the need for a rollback or causing any secondary effects. The second is operational value, which assesses the reductions in the average time taken to detect, acknowledge, diagnose, and restore a problem, as well as the number of incidents avoided and the improvement in service-level achievement. The third is trust and control, which looks at the proportion of decisions that are supported by traceable evidence, the rate at which humans have to override actions, the number of drift alerts, policy violations, and security exceptions. These metrics stop programmes from considering themselves successful just because the number of automated actions has gone up. They also enable executives to compare the benefits of automation with operational risk, making sure that a faster response to incidents does not compromise accountability, security, service ownership, or regulatory assurance requirements.

6. VISION 2030 ALIGNMENT AND IMPLEMENTATION PRIORITIES

Saudi companies have a favourable strategic position with regard to this approach since cloud adoption and AI capabilities are already linked to the nation's digital objectives. However, the relevant literature shows that uptake relies on skills, organisational readiness, security, clear policy, and confidence in cloud services [27-30]. It follows that an AIOps programme should be seen as an initiative for developing enterprise capabilities, not merely as a purchase of a monitoring tool.

The implementation can be carried out in four stages. The first of these is achieving telemetry readiness by standardising the ownership of services, the relationships between configurations, log retention procedures, tracing methods, identity context handling, and the classification of incidents. The second stage is that of assisted intelligence, which involves deploying low-risk applications such as enriching tickets, retrieving similar incidents, ranking anomalies, and predicting severity, while at the same time monitoring operator acceptance and error patterns. The third stage is governed orchestration, consisting of integrating the validated models with the incident, problem, change, and knowledge workflows and involving the introduction of runbooks based on approval. The fourth stage is bounded autonomy, whereby only actions that have been shown to be precise, which can be reversed, have clear controls on the potential impact, and are continuously verified should be automated.

It is just as important to plan the workforce. Although AI reduces the need for repetitive search and correlation tasks, it increases the importance of knowledge of service architecture, causal reasoning, automation engineering, model validation, and governance. Training should therefore be directed at service owners, operations engineers, security teams, data specialists, and the audit function rather than being limited to data scientists. Saudi organisations have the opportunity to use this change to create more valuable digital roles at the same time as enhancing the resilience of their services. Particularly worthwhile are research and implementation in the areas of Arabic-aware incident summarisation and bilingual knowledge retrieval since enterprise support environments might combine Arabic user reports with English technical telemetry.

Cybersecurity and data governance must be considered as operational design constraints at every stage of this process. AIOps platforms often collect logs, records of identity events, configuration data, the text from tickets, and traces which may include credentials, personal data, commercially sensitive information or security indicators. While centralising these different sources can make diagnosis easier, it also results in the creation of a highly valuable repository for analysis. Saudi companies should therefore implement the principle of least privilege, use encryption, establish retention rules, separate the operational environment from the environment used for developing models, and put in place specific controls regarding the export of telemetry to external model services. The datasets used for training the models should retain information about their origin so that the teams can work out which systems, time periods and types of incidents influenced a given recommendation. Security teams should also check whether manipulated logs, adversarially altered ticket text or corrupted knowledge articles could affect classification or the steps taken to remedy a problem. Human review is especially crucial when AI recommendations have an impact on identity management, network segmentation, financial systems or the controls within enterprise resource planning systems. Furthermore, when making procurement decisions it is necessary to evaluate the portability of both the telemetry and the models in order to prevent becoming operationally dependent on a single cloud or AIOps vendor. Such controls contribute to resilience since the organisation will be able to investigate model behaviour, reproduce decisions and carry on with its operations even if an AI component is not available. They also enhance accountability by enabling the distinction between model error, poor data quality, policy misconfiguration and operator override during a post-incident review. In this way, trustworthy AIOps is not obtained by simply adding an ethics statement

after a system has been deployed; it is instead achieved through architecture, access permissions, the retention of evidence, testing and the continuous monitoring of control effectiveness as part of daily service operations. Regular red-team exercises should be carried out to examine potential failure scenarios before the automated scope is extended across the organisation's services.

7. RESEARCH GAPS AND FUTURE DIRECTIONS

The issue of transferability is the first one to be considered. Although many AIOps models work well on a particular dataset, their performance drops when there are changes to the services, the log formats, the workload patterns, or the organisation's taxonomies. Future research should therefore focus on carrying out evaluations across different enterprises, on incorporating drift-aware learning, and on the use of benchmarks that reflect hybrid cloud environments. The second gap lies in causal validation. Although methods for root cause analysis are advancing quickly, businesses still need evidence to show that the causes which are ranked remain reliable during actual change events, in the case of partial loss of telemetry data, and when cascading incidents occur [15-18].

The fourth area involves the evaluation of large language models in operational settings. Although current research points to their potential in the areas of data preprocessing, diagnosis, and assisted remediation, reliable methods for assessing hallucinations, grounding, action safety, and operator over-reliance are still in their early stages [3]. The third gap concerns evidence that is specific to Saudi Arabia. While studies on cloud adoption do offer some useful contextual information [28, 29] and recent work on AI touches upon software engineering [30], there is currently very little longitudinal evidence available regarding enterprise AIOps outcomes in the Kingdom. Future research should look into regulated sectors, Arabic service data, sovereign-cloud restrictions, workforce adaptation, and measurable improvements in MTTR, availability, cost, and audit performance.

There is still one area relating to the design of human-AI teams. The key issue is not whether it is possible for AI to automate an operational task, but rather when doing so leads to better system-level results. Studies should examine various approval thresholds, different formats for explanations, confidence indicators, and escalation policies. This type of research would enable businesses to design autonomy which is fast enough to generate value yet still remain controllable in the event of rare, high-impact situations.

8. LIMITATIONS

The review includes 30 studies which have been verified by DOI but does not assert that it provides a complete bibliometric coverage. Since the evidence comes from various datasets, metrics, and system architectures, a statistical meta-analysis is not possible. Many of the technical studies are based on benchmark environments that might not match the complexity of large Saudi companies, and the body of literature in the Saudi context is wider than that which relates specifically to IT operations. Given the rapid advances in generative AI, it is likely that the capabilities and risks are changing more quickly than the cycles for journal publications. It is therefore appropriate to regard the proposed framework as a governance model informed by evidence and one that needs to be validated through empirical testing in specific sector applications.

9. CONCLUSION

AI-enabled cloud operations have the potential to make IT service management more predictive, contextual, and responsive, but there is no evidence to justify replacing disciplined service governance with unrestricted automation. The best outcomes are achieved when AI is used to tackle clearly defined operational bottlenecks, such as identifying weak failure signals, classifying and prioritising incidents, correlating multi-modal telemetry, ranking the most likely root causes, retrieving relevant knowledge, and carrying out validated low-risk recovery actions. For these capabilities to function, it is necessary to have good observability, service context, trustworthy models, and clearly defined authority boundaries.

For any Saudi companies aiming to achieve their Vision 2030 objectives, the practical chance lies in combining cloud modernisation with an intelligent way of operating one that enhances resilience and fosters digital skills. The five-layer framework suggested connects telemetry, context, intelligence, orchestration and governance, and the closed-loop assurance lifecycle makes sure that each recommendation or action is verified and turned into learning. Success should be assessed in terms of service outcomes, risk reduction, auditability and human effectiveness rather than the amount of automation. A step-by-step approach moving from telemetry readiness through to assisted intelligence, governed orchestration and bounded autonomy offers a justifiable route for implementation. Because of research carried out in Saudi Arabia on Arabic operational data, regulated cloud environments and human-AI accountability, AIOps can turn into a practical foundation for reliable digital growth rather than yet another standalone technology project.

REFERENCES

1. Notaro P, Cardoso J, Gerndt M. A Survey of AIOps Methods for Failure Management. *ACM Transactions on Intelligent Systems and Technology*. 2021;12(6):81. <https://doi.org/10.1145/3483424>
2. Diaz-de-Arcaya J, Torre-Bastida AI, Zarate G, Minon R, Almeida A. A Joint Study of the Challenges, Opportunities, and Roadmap of MLOps and AIOps: A Systematic Survey. *ACM Computing Surveys*. 2024;56(4):84. <https://doi.org/10.1145/3625289>
3. Zhang L, Jia T, Jia M, Wu Y, et al. A Survey of AIOps in the Era of Large Language Models. *ACM Computing Surveys*. 2025; advance online publication. <https://doi.org/10.1145/3746635>
4. Ahmed S, Singh M, Doherty B, Ramlan E, Harkin K, Bucholc M, Coyle D. An Empirical Analysis of State-of-Art Classification Models in an IT Incident Severity Prediction Framework. *Applied Sciences*. 2023;13(6):3843. <https://doi.org/10.3390/app13063843>
5. Ahmed S, Singh M, Doherty B, Ramlan EI, Harkin K, Bucholc M, Coyle D. Knowledge-based Intelligent System for IT Incident DevOps. 2023 IEEE/ACM International Workshop on Cloud Intelligence and AIOps. 2023:1-7. <https://doi.org/10.1109/AIOps59134.2023.00005>
6. Paramesh SP, Shreedhara KS. A Deep Learning Based IT Service Desk Ticket Classifier Using CNN. *ICTACT Journal on Soft Computing*. 2022;13(1):2805-2812. <https://doi.org/10.21917/ijsc.2022.0399>
7. Koukal F, Darena F, Jezdik R, Prichystal J. Improving Automated Categorization of Customer Requests with Recent Advances in Natural Language Processing. *European Journal of Business Science and Technology*. 2024;10(2):173-184. <https://doi.org/10.11118/ejobsat.2024.010>
8. Caturkusuma RM, Alzami F, Nurhindarto A, Sulistiyono MT, Irawan C, Kusumawati Y. Predicting IT Incident Duration using Machine Learning: A Case Study in IT Service Management. *Sinkron*. 2025;9(1):8-19. <https://doi.org/10.33395/sinkron.v9i1.14310>
9. Dong W. AIOps Architecture in Data Center Site Infrastructure Monitoring. *Computational Intelligence and Neuroscience*. 2022;2022:1988990. <https://doi.org/10.1155/2022/1988990>
10. Tengku Asmawi TN, Ismail A, Shen J. Cloud failure prediction based on traditional machine learning and deep learning. *Journal of Cloud Computing*. 2022;11:47. <https://doi.org/10.1186/s13677-022-00327-0>
11. Jassas MS, Mahmoud QH. Analysis of Job Failure and Prediction Model for Cloud Computing Using Machine Learning. *Sensors*. 2022;22(5):2035. <https://doi.org/10.3390/s22052035>
12. Kohyarnejadfar I, Aloise D, Azhari SV, Dagenais MR. Anomaly detection in microservice environments using distributed tracing data analysis and NLP. *Journal of Cloud Computing*. 2022;11:25. <https://doi.org/10.1186/s13677-022-00296-4>
13. Cinque M, Della Corte R, Pecchia A. Micro2vec: Anomaly detection in microservices systems by mining numeric representations of computer logs. *Journal of Network and Computer Applications*. 2022;208:103515. <https://doi.org/10.1016/j.jnca.2022.103515>
14. Panahandeh M, Hamou-Lhadj A, Hamdaqa M, Miller J. ServiceAnomaly: An anomaly detection approach in microservices using distributed traces and profiling metrics. *Journal of Systems and Software*. 2024;209:111917. <https://doi.org/10.1016/j.jss.2023.111917>
15. Xin R, Chen P, Zhao Z. CausalRCA: Causal inference based precise fine-grained root cause localization for microservice applications. *Journal of Systems and Software*. 2023;203:111724. <https://doi.org/10.1016/j.jss.2023.111724>
16. Lin CM, Chang C, Wang WY, Wang KD, Peng WC. Root Cause Analysis in Microservice Using Neural Granger Causal Discovery. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2024;38(1):206-213. <https://doi.org/10.1609/aaai.v38i1.27772>
17. Ikram A, Chakraborty S, Mitra S, Saini SK, Bagchi S, Kocaoglu M. Root Cause Analysis of Failures in Microservices through Causal Discovery. *Advances in Neural Information Processing Systems*. 2022. <https://doi.org/10.52202/068431-2259>
18. Pham L, Ha H, Zhang H. Root Cause Analysis for Microservice System based on Causal Inference: How Far Are We? *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering*. 2024:706-715. <https://doi.org/10.1145/3691620.3695065>
19. Hashemi S, Mantyla M. OneLog: towards end-to-end software log anomaly detection. *Automated Software Engineering*. 2024;31:37. <https://doi.org/10.1007/s10515-024-00428-x>
20. Wei X, Wang J, Sun C, Towey D, Zhang S, Zuo W, et al. Log-based anomaly detection for distributed systems: State of the art, industry experience, and open issues. *Journal of Software: Evolution and Process*. 2024;36(8):e2650. <https://doi.org/10.1002/smr.2650>
21. Han P, Li H, Xue G, Zhang C. Distributed system anomaly detection using deep learning-based log analysis. *Computational Intelligence*. 2023. <https://doi.org/10.1111/coin.12573>

22. Landauer M, Onder S, Skopik F, Wurzenberger M. Deep learning for anomaly detection in log data: A survey. *Machine Learning with Applications*. 2023;12:100470.
<https://doi.org/10.1016/j.mlwa.2023.100470>
23. Chen R, Zhang S, Li D, Zhang Y, Guo F, Meng W, Pei D, Zhang Y, Chen X, Liu Y. LogTransfer: Cross-System Log Anomaly Detection for Software Systems with Transfer Learning. 2020 IEEE 31st International Symposium on Software Reliability Engineering. 2020:37-47.
<https://doi.org/10.1109/ISSRE5003.2020.00013>
24. Becker S, Schmidt F, Gulenko A, Acker A, Kao O. Towards AIOps in Edge Computing Environments. 2020 IEEE International Conference on Big Data. 2020.
<https://doi.org/10.1109/BigData50022.2020.9378038>
25. Li R, Cheng Z, Lee PPC, Wang P, Qiang Y, Lan L, He C, Lu J, Wang M, Ding X. Automated Intelligent Healing in Cloud-Scale Data Centers. 2021 40th International Symposium on Reliable Distributed Systems. 2021:244-253.
<https://doi.org/10.1109/SRDS53918.2021.00032>
26. Rouholamini SR, Mirabi M, Farazkish R, Sahafi A. Proactive self-healing techniques for cloud computing: A systematic review. *Concurrency and Computation: Practice and Experience*. 2024;36(24):e8246.
<https://doi.org/10.1002/cpe.8246>
27. Hassan O. Artificial Intelligence, Neom and Saudi Arabia's Economic Diversification from Oil and Gas. *Political Quarterly*. 2020;91(1):222-227.
<https://doi.org/10.1111/1467-923X.12794>
28. Alamri N, Alzahrani S. Navigating the Clouds: An Exploratory Research of Cloud Computing Adoption in Saudi Arabia's Small and Medium Enterprises. *Engineering, Technology & Applied Science Research*. 2024;14(6):17859-17869.
<https://doi.org/10.48084/etasr.8435>
29. Alaboud NS, Elmansoury A, Jia P. Implementation of Cloud Computing in Saudi Construction: An Exploration of Political, Economic, Social, Technological, Environmental, and Legal Factors Using Fuzzy DEMATEL. *Advances in Civil Engineering*. 2025;2025:1825108.
<https://doi.org/10.1155/adce/1825108>
30. Albaroudi E, Mansouri T, Hatamleh M, Alameer A. Saudi Arabia's Vision 2030: Leveraging Generative Artificial Intelligence to Enhance Software Engineering. 2025 Eighth International Women in Data Science Conference at Prince Sultan University. 2025.
<https://doi.org/10.1109/WiDS-PSU64963.2025.00024>