ARTIFICIAL INTELLIGENCE

AI-Driven Cloud Computing and Intelligent Enterprise Information Management for Secure, Scalable Digital Transformation in Saudi Arabia under Vision 2030

Wael Zouheir Jouni

Published November 2026

Abstract

Cloud platforms are fast becoming the foundation for new digital services in Saudi Arabia, yet moving workloads and enterprise information to the cloud raises three linked problems: how to run elastic infrastructure reliably, how to secure it against a growing volume of attacks, and how to keep data within the legal and sovereign boundaries that national regulation sets. Artificial intelligence (AI) is now applied to each problem, through anomaly detection and workload forecasting in cloud operations, learning-based intrusion detection and identity controls in security, and metadata-driven governance in cloud data platforms. This review asks how these capabilities can be combined into a coherent model for secure, scalable and compliant cloud-based information management in Saudi organisations. A systematic configurative review was conducted of journal articles published between 2020 and 2025, seeded by a Saudi study of deep-learning intrusion detection in cloud networks and a Saudi survey of barriers to cloud adoption, and closed at 30 sources. The synthesis shows that Saudi adoption is shaped above all by perceived security, regulatory compliance, technical capability and connectivity; that AI methods for forecasting, autoscaling and anomaly detection deliver measurable gains in controlled evaluations but remain dependent on metric selection, data preparation and evaluation settings that rarely resemble production; and that sovereignty must be engineered into data placement, identity and metadata rather than assumed from a provider's location. The paper proposes a sovereign intelligent cloud model built from four nested layers, a workload placement matrix that links data classification to deployment choices and controls, and an agenda for research and practice.

Keywords: Cloud computing; artificial intelligence; cloud security; data sovereignty; enterprise information management; Saudi Arabia; Vision 2030

1 Introduction

Digital transformation in large organisations increasingly runs on cloud infrastructure. Applications are rebuilt as containerised services, analytical data are consolidated in cloud data lakes and lakehouses, and records, documents and content services increasingly sit on shared platforms rather than on departmental servers. The appeal is familiar: capacity can be provisioned on demand, costs follow use, and new services can be released quickly. The difficulties are equally familiar. Migrations overrun when legacy applications are moved without adequate upfront planning (Ramchand et al., 2021), the proliferation of interacting services makes failures harder to detect and diagnose (Soldani & Brogi, 2023), and information placed with a provider raises questions about who can access it and under whose law it is held (Galij et al., 2024).

Saudi Arabia makes these questions especially pressing. Vision 2030 has placed digital government and a data-driven economy at the centre of national reform, and national policy has encouraged

public bodies to consider cloud services first when procuring technology. Regulatory instruments have followed. The Communications, Space and Technology Commission has issued a cloud computing regulatory framework, the National Cybersecurity Authority (NCA) has published Essential Cybersecurity Controls and specific controls for cloud services, and the Personal Data Protection Law governs how personal data are processed and transferred. Global and domestic providers have responded by opening, or announcing, cloud regions inside the Kingdom. These developments shift the central question from whether Saudi organisations should use the cloud to how they can do so securely, at scale and within national rules.

Artificial intelligence is relevant on three fronts. In operations, machine learning forecasts workload, scales resources and detects anomalies in metrics and logs, a field often labelled AIOps. In security, deep learning detects intrusions in cloud network traffic, and new identity and access schemes control who can reach which resources. Within information management, metadata systems and governance frameworks determine whether a cloud data lake becomes a trusted asset or an unmanageable store. The literature treats these fronts separately. Studies of Saudi cloud adoption seldom examine operational or security techniques (Alamri & Alzahrani, 2024; Alarefi, 2023); technical studies of autoscaling or intrusion detection rarely consider regulation or data sovereignty (Aljuaid & Alshamrani, 2024; Pintye et al., 2024); and work on data governance and data lakes is largely silent on the jurisdictional questions raised by cloud hosting (Bližnák et al., 2024; Sawadogo & Darmont, 2021).

This review connects the three fronts within the Saudi context. It argues that a secure and scalable cloud is not a single technology choice but a set of nested capabilities: information governed at the core, intelligent operations around it, a zero-trust security fabric around both, and a regulatory and sovereignty perimeter that encloses the whole. Where those layers are designed together, AI strengthens each; where they are designed apart, gains in one layer can create risk in another, as when automated scaling replicates personal data into a region where it should not reside.

2 Aim, objectives and research questions

The aim of this review is to explain how AI-enabled cloud operations, AI-supported cloud security and cloud-native enterprise information management can be combined to deliver secure, scalable and sovereign digital transformation in Saudi organisations, and to translate that explanation into practical guidance on architecture and workload placement. Four objectives support this aim. The first objective is to identify the factors that drive and constrain cloud adoption and migration in Saudi organisations, using evidence from comparable contexts where Saudi evidence is thin. A second objective is to evaluate what AI techniques for anomaly detection, workload forecasting and autoscaling, and for intrusion detection, identity management and privacy-preserving learning, have been shown to achieve and under what conditions. The third objective analyses how data sovereignty, data residency and national regulation constrain cloud architecture and information governance. Finally, the fourth objective synthesises these findings into a layered model, a workload placement matrix and an agenda for research and practice.

The review addresses three research questions. RQ1 asks which technological, organisational and environmental factors most strongly shape cloud adoption and migration in Saudi organisations. RQ2 asks how far AI improves the reliability, efficiency and security of cloud operations, and what limits the transfer of reported gains to production environments. RQ3 asks how Saudi organisations should

place workloads and govern enterprise information in the cloud so that elasticity and analytical value are achieved without breaching sovereignty, residency and cybersecurity obligations.

Some terms are used in a defined sense. Cloud computing refers to on-demand access to shared, configurable computing resources delivered as infrastructure, platform or software services. Hybrid cloud combines on-premises or private infrastructure with public cloud services, and multi-cloud uses more than one public provider. Sovereign cloud denotes cloud services operated so that data, operations and legal control remain within a defined national jurisdiction. Enterprise information management (EIM) is used here for the governance of data and content held on cloud platforms, including data lakes and lakehouses, metadata catalogues, content services and the policies that determine ownership, quality, access and retention.

3 Review methodology

A systematic configurative review was adopted. Unlike aggregative reviews, which pool comparable effect sizes, a configurative review arranges heterogeneous evidence to build an explanatory model. This fits the corpus, which combines adoption surveys, algorithmic evaluations, legal and policy analyses, and design studies of data platforms. The steps of the review were specified in advance: research questions, sources, search terms, eligibility rules, appraisal criteria and the coding frame. Screening decisions and reasons for exclusion were logged so that the corpus can be audited.

Two seed studies provided the starting point. Aljuaid and Alshamrani (2024), working at a Saudi university, developed and evaluated a convolutional neural network for intrusion detection in cloud environments, and Alaboud (2025) surveyed construction professionals to identify the barriers to cloud adoption in the Kingdom. The first anchors the security strand of the review in Saudi research; the second anchors the adoption strand. Both reference lists were examined, but most of the material they cite either predates 2020 or appeared in conference proceedings, so their main role was to define search vocabulary and identify research streams.

The principal searches were run in a multidisciplinary scholarly metadata service covering journal articles published between January 2020 and December 2025. Three groups of terms were combined. The cloud group included "cloud computing", "cloud adoption", "cloud migration", "hybrid cloud", "multi-cloud", "data sovereignty", "data residency" and "data localization". The intelligence and security group included "anomaly detection", AIOps, autoscaling, "workload prediction", "intrusion detection", "zero trust", "identity and access management" and "federated learning". The information group included "data lake", lakehouse, metadata, "data governance" and "big data analytics". Each group was searched with and without context terms such as "Saudi Arabia", "Vision 2030", Gulf, government and "public sector", and the Saudi-specific searches were repeated within the outputs of individual publishers whose journals carry much of the regional literature.

Studies were eligible if they were peer-reviewed journal articles, if both their online and their final issue dates fell between 2020 and 2025, if they carried a resolvable digital object identifier with complete volume and page or article-number metadata, and if they contributed to at least one objective. Conference papers, preprints, theses, editorials and trade articles were excluded, as were studies whose cloud content was incidental, such as medical or financial prediction models hosted in the cloud, and articles subject to retraction. Where several studies addressed the same technique, preference was given to those reporting evaluation against named baselines or real-world data.

Each retained study was appraised against four questions: whether it addressed the research questions directly, whether its method was described clearly enough to judge its claims, whether its evidence was proportionate to its conclusions, and whether its setting allowed transfer to Saudi organisations. Algorithmic studies were read with particular attention to datasets, baselines and evaluation environments, because many report results from public benchmarks or research clouds rather than production systems. Survey studies were read for sample size, sampling method and the constructs measured.

The final corpus contains 30 articles. Nine are surveys or systematic reviews; eleven are technical studies that propose and evaluate a method, model or platform; seven are empirical studies of organisations, based on surveys, mixed methods or case analysis; and three are legal or policy analyses. Seven articles concern Saudi Arabia directly and one concerns the public sector of the United Arab Emirates. Coding recorded, for each study, the layer of the cloud stack addressed (information, operations, security or regulation), the capability examined, the evidence type, the reported outcome and any stated limitation. The synthesis then proceeded in three moves: a within-layer summary, a cross-layer analysis of dependencies and tensions, and the construction of the nested model and placement matrix presented in the following sections.

4 Cloud adoption and migration in the Saudi context

Saudi evidence on cloud adoption converges on a handful of factors. In the seed study by Alaboud (2025), 72 construction professionals rated security and connectivity as the most significant obstacles, and factor analysis grouped the barriers into data security and regulatory compliance, which explained the largest share of variance, followed by technical and organisational infrastructure and by cost and expertise limitations. A survey of 211 information technology professionals in Saudi government organisations found that perceived usefulness, ease of use, external influence and security significantly affected cloud usage, whereas privacy did not, and that technological readiness moderated several of these effects (Alarefi, 2023). Among 285 Saudi medium-sized enterprises using cloud-based technologies, technological capability, attention to regulation and compliance, and internal controls all improved internal audit efficiency, with technological capability the strongest determinant (Hakami, 2025). Alamri and Alzahrani (2024), drawing on government agencies, service providers and small firms in the e-commerce sector, used the technology–organisation–environment (TOE) framework to map the current state of adoption and its challenges.

Two features of this evidence deserve emphasis. First, security and compliance dominate as both barriers and enablers. Organisations hesitate because they doubt that cloud services meet security and regulatory expectations, yet when regulatory and control factors are in place they improve outcomes such as audit efficiency. This dual role suggests that clear national rules can accelerate adoption rather than impede it, provided that organisations can translate them into practical controls. Second, the finding that privacy was not a significant predictor of usage in Saudi government organisations, while security was (Alarefi, 2023), suggests that practitioners may treat privacy as a by-product of security, an assumption that the data protection law, with its distinct obligations, makes harder to sustain.

Comparative studies sharpen these points. Ukeje et al. (2024), screening 758 articles and retaining 33, found that information security and privacy together accounted for 70% of the significant gaps in research on government cloud adoption, and that some authors called for frameworks to address these challenges without developing them. In small firms operating in disruptive contexts, technolo-

gical features were associated with adoption while organisational and environmental features were not, and interviews identified economic barriers, emigration of skilled staff, weak infrastructure, survival-mode operation and culture as obstacles (Baalbaki et al., 2025). The contrast with the Saudi findings is instructive. In disruptive settings, economic survival and basic infrastructure dominate; in the Saudi studies, conducted where national policy actively promotes digital transformation, security, regulatory compliance and organisational capability carry more weight, although connectivity still matters in sectors such as construction (Alaboud, 2025). In the public sector of the United Arab Emirates, a case study of cloud-based enterprise resource planning found that alignment between institutional work and the logics embedded in the cloud system, shaped by organisational culture, allowed successful assimilation (Alsharari, 2022), a reminder that adoption is an institutional as well as a technical process.

Migration is where adoption decisions become operational. Ramchand et al. (2021) attribute many failed migrations to insufficient planning about which legacy applications suit public or private cloud, and they propose profiling the application portfolio, preferably through automated discovery, and assessing each application quantitatively to estimate run costs before migration. For Saudi organisations with large estates of bespoke government or industrial systems, portfolio assessment of this kind is the natural point at which data classification and residency requirements should be recorded. An application holding personal or critical national data needs a different destination from one serving open data, and that decision is far cheaper to make before migration than after.

Human factors complete the adoption picture. Asfahani (2024) found that Saudi employees' perceptions of their organisation's responsibility for cybersecurity strongly influenced their own security behaviour, with attitudes mediating the effect, while personal experience of incidents did not moderate it. Cloud environments, in which users and administrators configure services themselves, place considerable weight on such behaviour. Mosli (2024) shows how organisations can map the tasks of the Saudi Cybersecurity Workforce Framework onto the NCA Essential Cybersecurity Controls and use optimisation to identify the job roles needed for full compliance, a method that could be extended to cloud-specific controls. Taken together, the Saudi studies suggest that adoption succeeds when technical capability, regulatory translation and workforce design advance together.

5 AI for cloud operations and security

Operating an elastic cloud estate is a prediction problem. Capacity must be provisioned before demand arrives, and faults must be detected before users notice them. Workload forecasting is the first application of AI to this problem. Bi et al. (2021) combined bidirectional and grid long short-term memory networks with smoothing and filtering of the input series, and reported higher accuracy than widely used alternatives on a public trace from a large production cluster. Forecasts become useful when they drive scaling. Jeong et al. (2023) used a convolutional forecasting model to set container resource requests and replica counts, improving average resource utilisation by between roughly 4% and 34% relative to conventional autoscaling, with fewer overloads. Pintye et al. (2024) showed that the choice of metrics matters as much as the model: selecting the most relevant metrics for each application by statistical analysis reduced quality-of-service violations by up to 80% and virtual-machine usage by 3% to 50% on a research cloud supporting more than 300 projects.

Anomaly detection and root-cause analysis address the second half of the problem. Soldani and Brogi (2023) survey techniques for detecting failures in microservice and cloud-native applications and for tracing anomalies to their possible root causes, observing that the proliferation of services and

service interactions makes both tasks harder. Landauer et al. (2023) review deep-learning approaches to anomaly detection in log data and report that they have outperformed conventional machine learning while coping better with unstable log formats; they also stress that encoding raw, unstructured logs for neural networks is non-trivial, and their review deliberately stops short of a quantitative comparison. Sakong et al. (2024) illustrate a promising direction. Their ensemble, built on an attention-based anomaly transformer, fuses parsed and embedded log data with resource metrics to overcome the limits of single-source models, and it outperformed traditional sequence, autoregressive and log-only baselines on metric and log data collected from real cloud environments.

Security is the third operational domain, and here AI is used both to detect attacks and to shape access. In the seed study, Aljuaid and Alshamrani (2024) trained a three-block convolutional network on a public intrusion dataset generated in a cloud environment, using class balancing and correlation-based feature selection, and reported an overall accuracy of 98.67%. Their per-scenario results are as informative as the headline figure: detection was almost perfect for denial-of-service and botnet traffic but fell to about 92% for infiltration attacks, the category in which detection proved hardest, and the authors themselves identify testing in a live cloud and resilience to adversarial manipulation as future work. Albshaier et al. (2025), reviewing 30 studies on federated learning for cloud and edge security, report gains in privacy and threat detection but also persistent communication overhead and resource constraints, with half of the reviewed studies noting latency problems. Federated approaches are attractive in the Saudi context because they allow models to learn from several organisations' data without moving that data across organisational or jurisdictional boundaries.

Zero trust reframes the security architecture within which these detectors operate. Syed et al. (2022) set out the tenets of this paradigm, which withholds default trust from users, devices and network locations, and review the techniques for authentication, access control, encryption, micro-segmentation and security automation that implement it, while identifying unresolved challenges in trust and risk computation. Identity is the hinge of the model. Fugkeaw (2023) proposes a blockchain-based identity and access management scheme for single sign-on across applications outsourced to the cloud, in which smart contracts enforce fine-grained policies and retain access transactions for accountability, and reports that it was roughly four times as efficient as three existing schemes in average processing time. For Saudi organisations, the significance of such work lies less in any particular mechanism than in the principle that access decisions should be continuous, policy-driven and auditable.

Three cautions apply across this evidence. First, almost all reported gains come from benchmarks, public traces or research clouds; none of the operational studies in the corpus evaluates a production estate in the Kingdom. Second, model performance depends on data preparation choices, including metric selection, log parsing and class balancing, that are rarely transferable without adjustment. Third, automated operations and detection create new dependencies: an autoscaler that misreads demand can incur cost or outage, and a detector with high average accuracy may still miss the rare, targeted attacks that matter most. Table 1 summarises the evidence by function, recording the technique, the reported outcome and the principal limitation.

Table 1. Evidence on AI for cloud operations and security.

Function	Technique in the reviewed studies	Reported outcome	Principal limitation	Sources
Workload forecasting	Bidirectional and grid long short-term memory networks with smoothing and filtering	Higher accuracy than widely used baselines on a public cluster trace	Evaluated on a public trace rather than an operational estate	Bi et al. (2021)
Autoscaling	Convolutional forecasting driving container requests and replicas; statistical selection of application-specific metrics	Utilisation gains of about 4%–34%; up to 80% fewer quality-of-service violations and 3%–50% fewer virtual machines	Gains depend on workload pattern and metric choice; research-cloud settings	Jeong et al. (2023); Pintye et al. (2024)
Anomaly detection and root-cause analysis	Deep learning on logs; ensembles fusing logs with resource metrics	Ensemble outperformed traditional baselines; deep learning outperformed conventional machine learning on logs	Log encoding is non-trivial; evaluations are hard to compare	Landauer et al. (2023); Sakong et al. (2024); Soldani and Brogi (2023)
Intrusion detection	Multi-block convolutional network with class balancing and feature selection	Overall accuracy of 98.67%; weaker on infiltration attacks	Benchmark dataset; live and adversarial testing outstanding	Aljuaid and Alshamrani (2024)
Privacy-preserving detection	Federated learning across cloud and edge nodes	Improved privacy and threat detection reported across reviewed studies	Communication overhead, resource limits and latency	Albshaier et al. (2025)
Identity and access	Zero-trust controls; blockchain-based single sign-on with policy enforcement by smart contracts	Continuous, auditable access decisions; faster processing than three earlier schemes	Trust and risk computation remain open problems	Fugkeaw (2023); Syed et al. (2022)

6 Sovereignty, regulation and the cloud information layer

Data sovereignty is the point at which cloud architecture meets national law. Galij et al. (2024) review reference architectures for enforcing sovereignty in public clouds, note the computational and economic costs that some frameworks and cryptographic measures impose, propose an abstract sovereignty model for system designers, and find in experiments on two hyperscale platforms that the geographic location of a sovereignty connector strongly affected its performance while the choice of platform had little effect. Ryan et al. (2024), analysing the Common European Data Space, show that individuals, organisations and states can all claim sovereignty over the same data, and they ask whether current European Union regulation reconciles these competing claims or leaves shortcomings still to be addressed. Islam and Khan (2024) examine the rationales for data localisation policies, which are typically framed around national sovereignty, security and data protection, and question whether restricting cross-border flows achieves these aims on its own. These studies imply that sovereignty is a property of the whole system, including who operates it, where keys are held and how access is logged, rather than a matter of data-centre location alone.

Residency requirements must therefore be engineered into placement decisions. Gharote et al. (2022) present a model-driven decision-support framework that helps enterprises select cloud data centres and place customer data so that residency rules, which differ by country and by data type, are respected. In the Saudi setting, the relevant rules derive from the data protection law, sectoral requirements and national cybersecurity controls. J. Sarabdeen and Mohamed Ishak (2025) observe that Saudi health data were covered by general laws pending the implementation of the newly passed data protection law in 2024, and they judge the 2022 law capable of being considered adequate under the European Union's data protection regime. Their analysis is a reminder that the regulatory baseline has changed recently and that cloud estates designed under earlier assumptions may need review. In practice, every dataset migrated to the cloud should carry a classification that determines the regions, service models and providers in which it may reside, and automated operations should be constrained by those classifications so that scaling, backup and disaster recovery never move data outside permitted boundaries.

The information layer is where these constraints are applied to enterprise content. Sawadogo and Darmont (2021) describe the data lake as a large repository of raw data in any format and argue that architecture and metadata management are the keys to success; without metadata, Boukraa et al. (2024) warn, a lake degenerates into a data swamp. The same authors map the facets of metadata management in data lakes, from modelling and capture to use, and identify research gaps and open problems. Lakehouse designs extend the lake with open table formats that bring warehouse-style management to raw storage. AbouZaid et al. (2025) built a portable lakehouse platform from open, cloud-native components orchestrated on containers and, in an industry-standard benchmark, reported a 33% median improvement in query duration when caching was used. For Saudi organisations, portability of this kind matters because it keeps open the option of moving a data platform between in-Kingdom providers or between public and sovereign environments. Berisha et al. (2022) describe the shift from extract-transform-load to extract-load-transform processing as analytics moved to the cloud, a change that places raw data in the cloud earlier and therefore makes classification at ingestion more important.

Data governance supplies the rules by which this layer is managed. Bližnák et al. (2024), reviewing 38 publications from 2017 to 2023, find that data governance has matured as a discipline but that its definition and implementation remain ambiguous, and that new data protection and AI regulation increases the pressure for compliance. For cloud-based EIM, governance must operate at the scale and speed of the platform: ownership, classification, quality rules, retention and access policies should be expressed as metadata in a catalogue and enforced automatically by the platform, rather than documented in policies that the platform cannot read. Metadata thus becomes the mechanism that connects the information layer to the sovereignty perimeter and to the security fabric, because the same classification that determines residency can also determine encryption requirements and access policies.

Figure 1 draws these strands together in the sovereign intelligent cloud model. At its core is the governed information layer, comprising lakehouse storage, a metadata catalogue, content services and data-governance policies. Around it sits the intelligent operations layer, in which workload forecasting, metric-aware autoscaling, anomaly detection and root-cause analysis keep services available and efficient. Around both lies a zero-trust security fabric of continuous identity verification, fine-grained access control, encryption with keys held under local control, and learning-based intrusion detection, including federated models that learn across organisations. The outer perimeter represents sovereignty and regulation: the data protection law, the national cloud

regulatory framework and cybersecurity controls, data residency rules and the choice of in-Kingdom regions. Each layer constrains the one inside it, and each relies on the classification metadata held at the core.

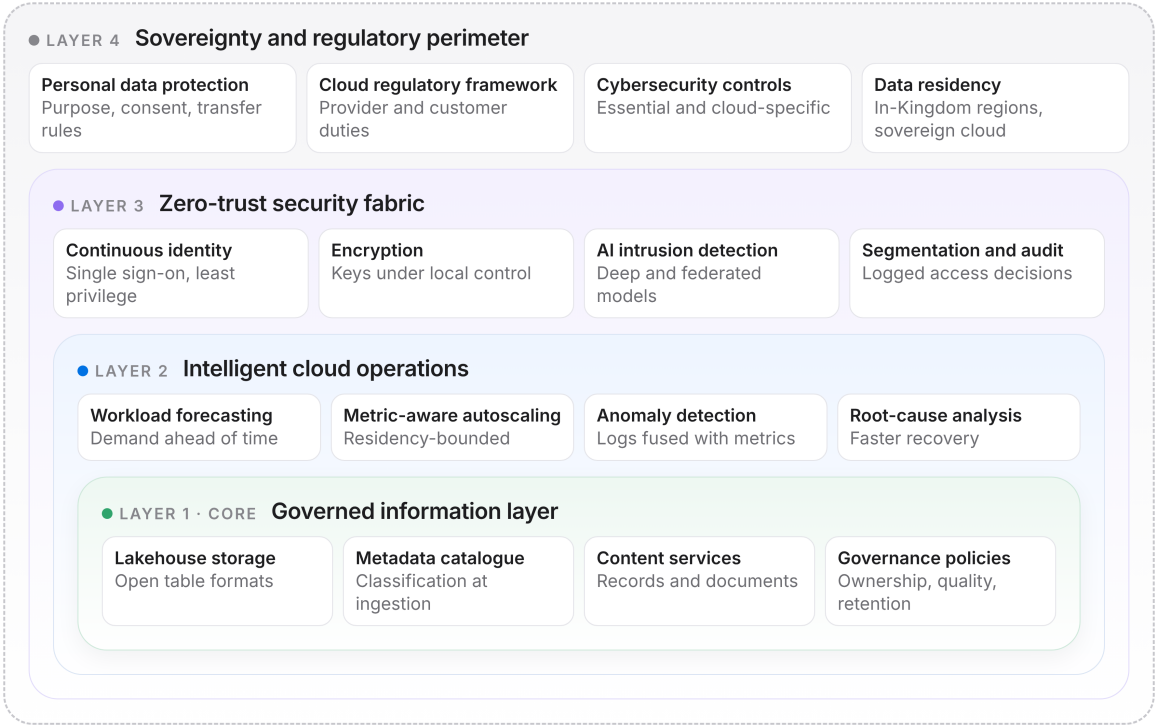


Figure 1. Sovereign intelligent cloud model: a governed information core within intelligent operations, a zero-trust security fabric and a sovereignty perimeter.

7 Workload placement and implementation

The nested model leaves open a practical question that every Saudi organisation faces: where should each workload run? Figure 2 proposes a placement matrix built on two dimensions that the reviewed evidence identifies as decisive. The vertical axis is data sensitivity, ranging from open or public data to restricted data such as critical national information and sensitive personal data; this dimension reflects the dominance of security, compliance and residency concerns in the Saudi adoption studies and in the sovereignty literature. The horizontal axis is the need for elasticity and advanced managed services, ranging from stable, predictable workloads to highly variable, analytics-intensive or AI-intensive ones; this dimension reflects the operational evidence that forecasting and autoscaling deliver most value where demand fluctuates.

Four placements follow. Workloads with low sensitivity and high elasticity needs, such as public websites, open-data services and development environments, suit public cloud regions located in the Kingdom, where managed services and autoscaling can be used freely. Low-sensitivity workloads with stable demand can be consolidated on any compliant platform chosen on cost. Where sensitivity is high but demand is stable, as with core registries and critical records, the appropriate home is a sovereign or government cloud environment, or private infrastructure, with keys and operations under local control. The most demanding quadrant combines high sensitivity with high elasticity, as

in national analytics on personal data or AI training on confidential records. Here a hybrid pattern is appropriate: sensitive data remain in a sovereign environment, while elastic computation is brought to the data or performed on de-identified derivatives, and federated learning is used where models must learn across organisations.

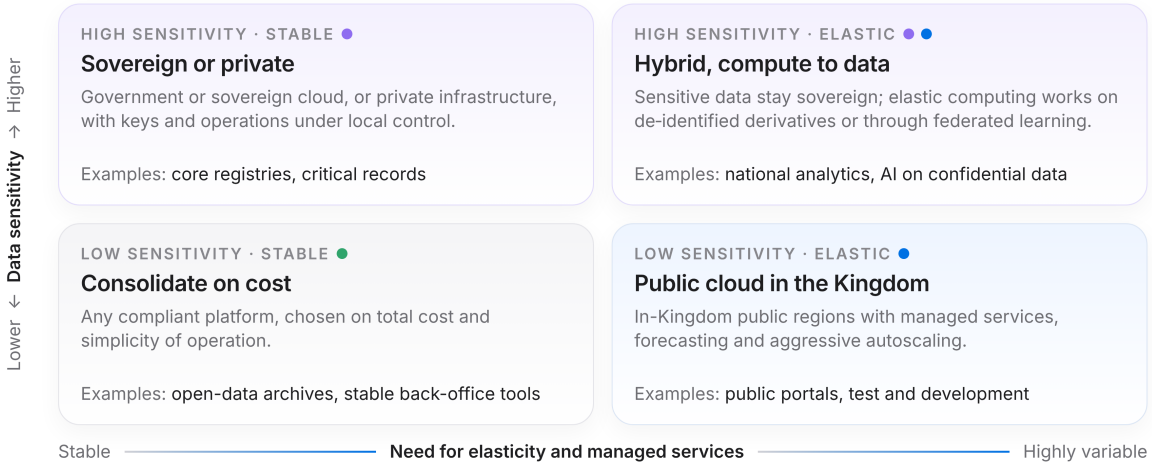


Figure 2. Workload placement matrix linking data sensitivity and elasticity needs to deployment choices.

Table 2 converts the matrix into a placement guide organised by data class. For each class it states the recommended deployment, the residency and regulatory considerations, the minimum security controls and the operational emphasis for AI. The guide is deliberately conservative for restricted data and permissive for open data. This asymmetry follows from the argument of the review: excessive caution with low-risk workloads would forfeit the efficiency gains that motivate cloud adoption, while insufficient caution with high-risk workloads would expose organisations to the security and compliance failures that dominate the concerns of Saudi practitioners.

Table 2. Workload placement guide by data class for Saudi organisations.

Data class	Recommended deployment	Residency and regulatory considerations	Minimum security controls	AI operations emphasis
Open or public	Public cloud region in the Kingdom; multi-cloud where useful	Licensing and integrity of published data	Integrity protection; standard identity controls	Aggressive autoscaling; cost optimisation
Internal	Public or hybrid cloud in the Kingdom	Organisational policy; contractual controls on the provider	Single sign-on with multi-factor authentication; encryption at rest	Forecasting and anomaly detection on shared telemetry
Confidential or personal	Hybrid or sovereign cloud with in-Kingdom residency	Personal data protection law: purpose limitation and transfer rules; sectoral requirements	Encryption with locally controlled keys; fine-grained access; full logging	Residency-aware scaling and backup; federated learning across organisations

Data class	Recommended deployment	Residency and regulatory considerations	Minimum security controls	AI operations emphasis
Restricted or critical	Sovereign or government cloud, or private infrastructure	National cybersecurity controls, including cloud-specific controls; strict residency	Zero-trust segmentation; privileged access management; continuous monitoring	Advisory AI only; automated actions under strict guard-rails

Implementation should proceed through four disciplined steps. Organisations should first build an application and data inventory that records classification, residency constraints and dependencies, following the portfolio-assessment logic of Ramchand et al. (2021). They should then establish the identity, encryption and logging foundations of a zero-trust fabric before migrating sensitive workloads, so that controls precede data rather than follow it. Third, they should introduce AI operations incrementally, starting with forecasting and anomaly detection in advisory mode, validating models on their own telemetry, and only then allowing automated scaling and remediation within guard-rails that respect residency rules. Finally, they should bring the information layer under catalogue-driven governance, so that classification, quality and retention policies are enforced by the platform itself.

Progress should be measured across the layers. Operational indicators include availability, mean time to detect and resolve incidents, resource utilisation and cost per transaction. Security indicators include the share of identities under continuous verification, the proportion of sensitive data encrypted with locally controlled keys, detection rates for red-team exercises and the time taken to revoke access. Sovereignty and governance indicators include the proportion of datasets with complete classification metadata, the number of residency exceptions, the coverage of NCA controls and audit findings. Transformation indicators include the time to deliver new digital services, the reuse of shared data assets across organisations and user satisfaction. Reporting these together prevents the familiar pattern in which efficiency improves while compliance debt quietly accumulates.

8 Discussion, research agenda and limitations

The findings answer the three research questions as follows. For RQ1, Saudi adoption is shaped above all by perceived security and regulatory compliance, technical capability and connectivity, with technological readiness and internal controls moderating or strengthening outcomes; organisational culture and workforce design influence whether adoption deepens into assimilation. For RQ2, AI measurably improves forecasting accuracy, resource utilisation, quality of service and intrusion detection in controlled evaluations, but the gains depend on careful choice of metrics and data preparation, they are seldom demonstrated on production estates, and rare or adversarial events remain difficult to detect. Turning to RQ3, sovereignty and residency must be engineered into placement, identity and metadata; a classification-driven approach that links each dataset to permitted locations, controls and automated behaviours allows organisations to obtain elasticity where risk is low and to retain control where it is high.

Three propositions follow for empirical testing. First, Saudi organisations that record data classification and residency constraints during application portfolio assessment will experience fewer compliance exceptions and lower remediation costs after migration than those that classify

afterwards. Second, AI-driven autoscaling and anomaly detection will yield larger and more stable gains when models are trained and validated on the organisation's own telemetry and when metric selection is tailored to each application, than when generic models are deployed. Third, catalogue-driven governance, in which classification metadata control residency, encryption and access automatically, will mediate the relationship between cloud adoption and digital transformation outcomes, because it allows shared data assets to be reused without renegotiating controls for each use.

The research agenda follows from the gaps the review exposes. Production evaluations of AI operations in Saudi cloud estates, including the frequency and cost of erroneous automated actions, are missing. Intrusion detection research needs datasets that reflect regional traffic and targeted attacks, and evaluation against adversarial manipulation. Federated and privacy-preserving learning across Saudi organisations deserves pilot studies that measure latency, accuracy and governance overhead. Legal research should examine how the data protection law, the cloud regulatory framework and cybersecurity controls interact in multi-cloud and hybrid deployments, including key management and access by foreign authorities. Finally, longitudinal studies should test whether classification-driven governance delivers the reuse and compliance benefits proposed here.

Practitioners can act on four implications now. Chief information officers should require data classification and residency fields in every migration business case. Security leaders should prioritise identity, key management and logging before migrating sensitive data. Operations teams should treat AI models as components that need validation, monitoring and rollback, not as autonomous controllers. Data leaders should invest in metadata catalogues and policy automation, because governance that the platform cannot read will not scale with the platform.

Several limitations apply. The corpus is limited to 30 journal articles, and excluding conference proceedings omits much of the fast-moving technical literature on cloud operations and security. Searches used a single metadata service together with reference-list checks of the two seeds, and Arabic-language publications fell outside the search. One reviewer conducted screening, appraisal and coding. Saudi regulatory instruments are described at the level of their scope rather than their detailed provisions, because peer-reviewed analyses of their cloud-specific requirements were not found within the review window. The nested model and placement matrix rest on synthesis rather than testing and await empirical validation.

9 Conclusion

The review set out to show how intelligent operations, intelligent security and cloud-native information management fit together in support of secure and scalable transformation in the Kingdom. Three conclusions stand out. First, Saudi organisations are ready to adopt cloud services once concerns about security, compliance and capability are addressed. Second, AI can make cloud estates more efficient and better defended, provided that its models are grounded in local data and constrained by policy. Third, sovereignty has to be designed into data placement, identity and metadata; it cannot be inferred from the address of a provider.

The sovereign intelligent cloud model and the workload placement matrix offer a structure for these design choices. Their central claim is that classification metadata, held at the core of the information layer, should drive residency, security and automated operations alike. Organisations that adopt this discipline can use the cloud's elasticity where risk is low, keep firm control where it is high, and build the trusted, shared data foundations on which the digital services envisaged by Vision 2030 depend.

References

[1] AbouZaid A, Barclay PJ, Chrysoulas C, Pitropakis N (2025). Building a modern data platform based on the data lakehouse architecture and cloud-native ecosystem. *Discover Applied Sciences*, 7(3), Article 166. https://doi.org/10.1007/s42452-025-06545-w

[2] Alaboud NS (2025). Identifying challenges to cloud computing adoption in Saudi Arabia's construction industry: An empirical analysis using structural equation modelling and artificial neural networks. *International Journal of Computational Intelligence Systems*, 19(1), Article 107. https://doi.org/10.1007/s44196-025-01101-1

[3] Alamri N, Alzahrani S (2024). Navigating the clouds: An exploratory research of cloud computing adoption in Saudi Arabia's small and medium enterprises. *Engineering, Technology & Applied Science Research*, 14(6), 17859–17869. https://doi.org/10.48084/etasr.8435

[4] Alarefi M (2023). Cloud computing usage by governmental organizations in Saudi Arabia based on Vision 2030. *Uncertain Supply Chain Management*, 11(1), 169–178. https://doi.org/10.5267/j.uscm.2022.10.010

[5] Albshaier L, Almarri S, Albuali A (2025). Federated learning for cloud and edge security: A systematic review of challenges and AI opportunities. *Electronics*, 14(5), Article 1019. https://doi.org/10.3390/electronics14051019

[6] Aljuaid WH, Alshamrani SS (2024). A deep learning approach for intrusion detection systems in cloud computing environments. *Applied Sciences*, 14(13), Article 5381. https://doi.org/10.3390/app14135381

[7] Alsharari NM (2022). Cloud computing and ERP assimilation in the public sector: Institutional perspectives. *Transforming Government: People, Process and Policy*, 16(1), 97–109. https://doi.org/10.1108/tg-04-2021-0069

[8] Asfahani AM (2024). Perceptions of organizational responsibility for cybersecurity in Saudi Arabia: A moderated mediation analysis. *International Journal of Information Security*, 23(4), 2515–2530. https://doi.org/10.1007/s10207-024-00859-3

[9] Baalbaki C, Shamieh S, Bodolica V (2025). Cloud computing adoption in disruptive contexts: A TOE framework approach. *VINE Journal of Information and Knowledge Management Systems*, 55(6), 1696–1721. https://doi.org/10.1108/vjikms-03-2024-0111

[10] Berisha B, Mëziu E, Shabani I (2022). Big data analytics in cloud computing: An overview. *Journal of Cloud Computing*, 11(1), Article 24. https://doi.org/10.1186/s13677-022-00301-w

[11] Bi J, Li S, Yuan H, Zhou M (2021). Integrated deep learning method for workload and resource prediction in cloud systems. *Neurocomputing*, 424, 35–48. https://doi.org/10.1016/j.neucom.2020.11.011

[12] Bližnák K, Munk M, Pilková A (2024). A systematic review of recent literature on data governance (2017–2023). *IEEE Access*, 12, 149875–149888. https://doi.org/10.1109/access.2024.3476373

[13] Boukraa D, Bala M, Rizzi S (2024). Metadata management in data lake environments: A survey. *Journal of Library Metadata*, 24(4), 215–274. https://doi.org/10.1080/19386389.2024.2359310

[14] Fugkeaw S (2023). Achieving decentralized and dynamic SSO-identity access management system for multi-application outsourced in cloud. *IEEE Access*, 11, 25480–25491. https://doi.org/10.1109/access.2023.3255885

[15] Galij S, Pawlak G, Grzyb S (2024). Modeling data sovereignty in public cloud—A comparison of existing solutions. *Applied Sciences*, 14(23), Article 10803. https://doi.org/10.3390/app142310803

[16] Gharote M, Mondal S, Roy S, Sahu P, Ramamurthy A (2022). Decision support framework for data residency compliance in cloud. *CSI Transactions on ICT*, 10(1), 61–69. https://doi.org/10.1007/s40012-022-00349-8

[17] Hakami TA (2025). Enhancing internal audit efficiency in Saudi Arabian SMEs: The impact of cloud adoption determinants. *SN Business & Economics*, 5(12), Article 230. https://doi.org/10.1007/s43546-025-00992-8

[18] Islam MT, Khan BU (2024). Impacts of data localization policies and lessons for Bangladesh. *International Data Privacy Law*, 14(2), 150–168. https://doi.org/10.1093/idpl/ipae002

[19] Jeong B, Jeon J, Jeong Y-S (2023). Proactive resource autoscaling scheme based on SCINet for high-performance cloud computing. *IEEE Transactions on Cloud Computing*, 11(4), 3497–3509. https://doi.org/10.1109/tcc.2023.3292378

[20] Landauer M, Onder S, Skopik F, Wurzenberger M (2023). Deep learning for anomaly detection in log data: A survey. *Machine Learning with Applications*, 12, Article 100470. https://doi.org/10.1016/j.mlwa.2023.100470

[21] Mosli R (2024). Optimal job role allocation for compliance with NCA-ECC controls using the Saudi Cybersecurity Workforce Framework. *IEEE Access*, 12, 128235–128245. https://doi.org/10.1109/access.2024.3457025

[22] Pintye I, Kovács J, Lovas R (2024). Enhancing machine learning-based autoscaling for cloud resource orchestration. *Journal of Grid Computing*, 22(4), Article 68. https://doi.org/10.1007/s10723-024-09783-1

[23] Ramchand K, Baruwal Chhetri M, Kowalczyk R (2021). Enterprise adoption of cloud computing with application portfolio profiling and application portfolio assessment. *Journal of Cloud Computing*, 10(1), Article 1. https://doi.org/10.1186/s13677-020-00210-w

[24] Ryan M, Gürtler P, Bogucki A (2024). Will the real data sovereign please stand up? An EU policy response to sovereignty in data spaces. *International Journal of Law and Information Technology*, 32, Article eaae006. https://doi.org/10.1093/ijlit/eaae006

[25] Sakong W, Kwon J, Min K, Wang S, Kim W (2024). Anomaly transformer ensemble model for cloud data anomaly detection. *IEEE Transactions on Cloud Computing*, 12(4), 1305–1313. https://doi.org/10.1109/tcc.2024.3466174

[26] Sarabdeen J, Mohamed Ishak MM (2025). A comparative analysis: Health data protection laws in Malaysia, Saudi Arabia and EU General Data Protection Regulation (GDPR). *International Journal of Law and Management*, 67(1), 99–119. https://doi.org/10.1108/ijlma-01-2024-0025

[27] Sawadogo P, Darmont J (2021). On data lake architectures and metadata management. *Journal of Intelligent Information Systems*, 56(1), 97–120. https://doi.org/10.1007/s10844-020-00608-7

[28] Soldani J, Brogi A (2023). Anomaly detection and failure root cause analysis in (micro) service-based cloud applications: A survey. *ACM Computing Surveys*, 55(3), 1–39. https://doi.org/10.1145/3501297

[29] Syed NF, Shah SW, Shaghaghi A, Anwar A, Baig Z, Doss R (2022). Zero trust architecture (ZTA): A comprehensive survey. *IEEE Access*, 10, 57143–57179. https://doi.org/10.1109/access.2022.3174679

[30] Ukeje N, Gutierrez J, Petrova K (2024). Information security and privacy challenges of cloud computing for government adoption: A systematic review. *International Journal of Information Security*, 23(2), 1459–1475. https://doi.org/10.1007/s10207-023-00797-6