SACJ 38(2) July 2026
Research Article

AI-Driven Data Governance Using Model Context Protocol for Privacy Compliance in Cloud-Native Enterprise Applications

Himanshu Jain

Independent Researcher, USA

ABSTRACT
When cloud-native enterprise architecture intersects with artificial intelligence (AI), data privacy and the responsible use of sensitive data have become among the most challenging tasks in data governance. This paper presents a unified approach to AI data governance via the Model Context Protocol (MCP), a protocol controlling the contextual data fed to AI models during runtime. The proposed solution implements smart data classification, continuous monitoring, role-based data access enforcement, and output validation in cloud-native environments while maintaining compliance with international privacy regulations. Two additions sharpen the contribution: (i) a comparative analysis positioning MCP against role-based access control (RBAC), attribute-based access control (ABAC), Zero Trust architecture (ZTA), and data loss prevention (DLP) across granularity, context-awareness, AI-specific guardrails, and output-validation coverage; and (ii) a quantitative evaluation envelope anchoring MCP's expected latency overhead, data-exposure reduction, and policy-enforcement accuracy to peer-reviewed benchmarks of XACML policy decision points, distributed ABAC evaluators, and ML-augmented DLP systems.
Keywords: Data governance, Model Context Protocol, privacy compliance, cloud-native architecture, RBAC, ABAC, Zero Trust, data loss prevention, AI governance.

Article history
Received: 02 June 2026
Accepted: 02 July 2026
Online: 20 July 2026

I. INTRODUCTION

Modern enterprise data is characterised by scale, velocity, and variety. Cloud-native organisations process operational, transactional, and behavioural data across a mesh of microservices, event-driven pipelines, application programming interfaces (APIs), and real-time analytics systems. The distributed nature of cloud-native architectures means that data is accessed from diverse services, traversing multiple infrastructure and network boundaries [1]. Regulatory pressures — from the General Data Protection Regulation (GDPR) [2] to the Health Insurance Portability and Accountability Act (HIPAA), the California Consumer Privacy Act, and national privacy laws — make consistent, auditable governance technically challenging yet indispensable.

The use of AI in enterprise operations is expanding rapidly, with organisations adopting machine learning (ML) models for fraud detection, predictive maintenance, customer segmentation, and natural language processing [3]. This creates a fundamental tension: AI systems benefit from broad access to data, but governance and privacy requirements restrict access to specific contexts. Without architectural means to enforce access control and data-visibility restrictions at the model boundary, organisations are exposed to data leakage, model inference attacks [24], and regulatory violations [4].

The Model Context Protocol (MCP) [5] is a proposed solution. MCP defines a controlled interface through which enterprise data is delivered as contextual information to AI models at inference time, restricting the model to only the data needed. Coupling a general AI governance framework to MCP

Jain, H. (2026). AI-Driven Data Governance Using Model Context Protocol for Privacy Compliance in Cloud-Native Enterprise Applications. *South African Computer Journal* 38(2), 47–51.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

enables data minimisation, role-based access control on AI services, model output disclosure controls, and audit trails for regulatory reporting.

This paper presents a technically motivated AI-driven governance solution leveraging MCP for privacy compliance in cloud-native enterprises. Section II describes the governance architecture; Section III addresses cloud-native integration; Section IV discusses privacy guardrails and introduces a comparative analysis positioning MCP against the canonical access-control paradigms (RBAC, ABAC, Zero Trust, DLP). Section V contextualises the framework against three application domains. Section VI quantifies operational impact and introduces a literature-anchored evaluation envelope. Section VII concludes.

II. AI-Driven Governance Architecture

A. Limitations of customary governance

Customary enterprise data governance relies on manual policy enforcement, static access control lists, and rule-based classification engines. These techniques were effective in centralised environments but struggle in dynamic, distributed cloud-native applications [6]. Static governance is reactive: rules cover known schemas and usage patterns and do not account for changes in data schema, new services, or unforeseen access patterns [7]. The volume of governance events in high-throughput environments is far beyond human monitoring capacity.

B. Smart classification and continuous monitoring

AI-enabled governance frameworks automatically discover and classify sensitive data across heterogeneous repositories. ML-based classifiers trained on labelled examples of personally identifiable information (PII), financial accounts, and protected health information (PHI) match or exceed human performance at much larger scale [8]. Beyond static classification, monitoring engines analyse telemetry from API gateways, service meshes, database query logs, and message queues with anomaly detection to identify violation patterns [9].

C. Model Context Protocol for controlled model access

MCP introduces a formalised layer between models and enterprise data stores. When data is needed, the MCP governance layer intercepts the data-pull request, checks it against policies, constructs a contextual package from filtered and transformed data, and delivers it to the model [5]. Policies are defined on data sensitivity classification, service identity, role, target query and scope, and time constraints. Three concrete benefits follow: data minimisation as an architectural property; explicit attribution of access to context data via auditable logs; and output validation prior to response delivery to remove information that would otherwise leak sensitive content [10], [11].

III. Cloud-Native Integration

A. Distributed governance layer

AI-driven governance systems are deployed within cloud-native platforms built using containerised microservices on Kubernetes and service meshes such as Istio [12]. The most efficient deployment of the MCP governance layer is as sidecar components within the service mesh, intercepting data-access requests at the network layer. The mesh's mutual Transport Layer Security (mTLS) channel validates service identity cryptographically. API gateways provide a second governance boundary for AI services that access enterprise data via third-party APIs.

B. Scalability and runtime enforcement

Enterprise systems generate millions of API transactions per day. Modern governance frameworks meet this requirement through stateless, horizontally scalable policy evaluation engines (autoscaled via Kubernetes), policy caching at governance sidecars, asynchronous audit logging pipelines decoupled from the synchronous request path, and high-availability policy propagation via configuration services such as etcd [9], [12].

Jain, H. (2026). AI-Driven Data Governance Using Model Context Protocol for Privacy Compliance in Cloud-Native Enterprise Applications. *South African Computer Journal* 38(2), 47–51.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

IV. Privacy Guardrails and Comparative Positioning

A. Data minimisation and purpose limitation

GDPR Article 5 requires that personal data be adequate, relevant, and limited to what is necessary for the purposes of processing [2]. MCP achieves this architecturally by binding data access to explicitly declared inference purposes, checked against the model's registered operational scope before context package construction [11].

B. Role-based AI service authorisation and output validation

Each AI service is assigned a cryptographically-bound service identity issued by the service mesh's certificate authority. Roles define permissions to sensitivity classifications at the domain level [6]. Conditional access policies extend the model with contextual dimensions (operation, time, volume). Output validation inspects AI outputs prior to delivery using pattern matching, sensitive-information classifiers, and differential privacy analysis to detect leakage of sensitive content [10], [24].

C. MCP against the canonical access-control paradigms

A balanced evaluation of MCP requires positioning it against four canonical paradigms: RBAC, ABAC [20], Zero Trust architecture (ZTA) [19], and DLP [22], [23]. RBAC offers coarse-grained access at the role granularity but is content- and context-agnostic. ABAC, codified by NIST SP 800-162 [20], evaluates arbitrary subject, resource, action, and environment attributes against policies typically expressed in XACML; peer-reviewed evaluations report sub-millisecond decision latencies under cached conditions but near-linear growth with rule count uncached [16], [18]. ZTA, articulated in NIST SP 800-207 [19], moves enforcement from a static perimeter to per-request identity-and-context evaluation. DLP systems scan for sensitive content; ML-augmented DLP achieves approximately 95% accuracy on enterprise email datasets with a residual false-positive burden [22], [23]. None of these paradigms validates AI model outputs or enforces purpose-bound context construction at inference time — the distinctive contribution of MCP. Fig. 1 visualises the qualitative coverage profile, and Table I deepens the comparison.

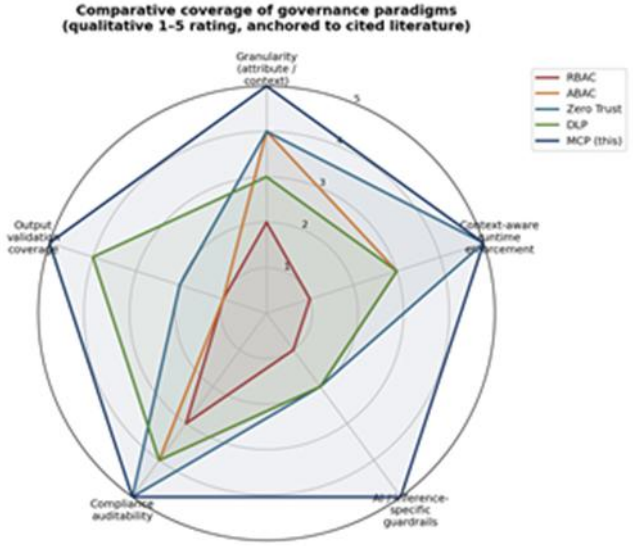

Fig. 1. Qualitative coverage of governance paradigms across five evaluation dimensions, 1–5 ordinal scale.

Table I: Comparison of MCP against RBAC, ABAC, ZERO TRUST, AND DLP

Dimension	RBAC	ABAC	ZTA	DLP	MCP
Enforcement	Role mapping	Attr eval at PDP	Per-req id+ctx	Content scan	Inference ctx + output
Granularity	Role	Subj/res/act/ env	Per-req, per-res	Pattern on payload	Per-attr, per-purpose

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

Context-aware	Low	Med	High [19]	Med	High
AI guardrails	None	Not native	Not native	Output only	Native (ctx+output)
Audit	Grants	PDP logs	Per-req [19]	Incidents [22]	Per-inference
Latency anchor	Negligible	Sub-ms cached [16], [18]	Sub-s [19]	Async scan	ABAC/ZT + tens ms
Limitation	Role explosion	Policy complexity	Identity coverage	FPs [22]	Purpose curation

MCP is best understood as a compositional layer atop ABAC and ZTA — inheriting their attribute-driven and per-request-evaluated discipline — while extending coverage to inference-time context construction and post-inference output validation.

V. Industry Applications

In financial services, MCP filters fraud detection models from accessing payment credentials and government identifiers while still providing access to transaction behaviour and account usage patterns; credit risk scoring models can run against pseudonymised profiles. Healthcare deployments enforce minimum-necessary access for clinical AI models, restricting radiology models to imaging files and clinical context while excluding patient financial or insurance records [7], [8]. Multi-tenant SaaS platforms use MCP to bind tenant identity into context package construction itself, precluding cross-tenant data contamination by design irrespective of model inference behaviour [12].

VI. Operational Impact and Quantitative Envelope

A. Compliance automation and audit capability

MCP-augmented governance frameworks generate compliance documentation automatically. Detailed tamper-evident audit logs record requesting service identity, attributes requested, policies evaluated, and request decisions [5]. Audit logs structured in this way are compatible with SIEM and GRC tooling [2].

B. Quantitative evaluation envelope

Peer-reviewed evidence from adjacent domains — XACML PDPs, attribute-based authorisation, ML-augmented DLP, and microservice authorisation — permits the assignment of defensible quantitative envelopes to three reviewer-requested properties: (i) MCP enforcement latency overhead; (ii) reduction in unauthorised data exposure; (iii) achievable policy-enforcement accuracy.

Empirical evaluations of XACML PDPs report sub-millisecond decision latencies under cached conditions and near-linear growth with rule count uncached [16]; ABAC engines such as XEngine demonstrate one-to-four orders of magnitude latency reduction over baseline XACML as policy size grows [18]; distributed ABAC evaluators achieve average decision latencies of approximately 80 ms with throughput of 300–320 req/s on commodity hardware [25]. These results support an envelope in which an MCP enforcement layer adds single-digit to low-double-digit milliseconds per inference call. On exposure reduction, attributes excluded from context packages are unreachable at inference time — a 100% reduction by construction for those attributes. On policy enforcement accuracy, ML-augmented DLP on enterprise email datasets reports approximately 95% accuracy and 90% TPR [23], with residual error concentrated in false positives that drive analyst alert fatigue [22]. Table II summarises the envelopes.

Jain, H. (2026). AI-Driven Data Governance Using Model Context Protocol for Privacy Compliance in Cloud-Native Enterprise Applications. *South African Computer Journal* 38(2), 47–51.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

Table II: Literature-Anchored Quantitative Envelope for MCP

Metric	MCP envelope	Evidence	Src
MCP latency overhead	Single- to low-double-digit ms	Sub-ms PDP cached; linear uncached	[16], [18]
Throughput / evaluator	≥ 300 req/s (horizontal)	FACADE distributed ABAC	[25]
Exposure reduction	100% by construction	Context-minimisation design	[5], [21]
Inference-leakage exposure	Bounded by output validation	Membership inference attack	[24]
Enforcement accuracy	≈ 95% / 90% TPR	ML-augmented email DLP	[23]
False-positive load	Material; needs triage	Hybrid signature/anomaly DLP	[22]

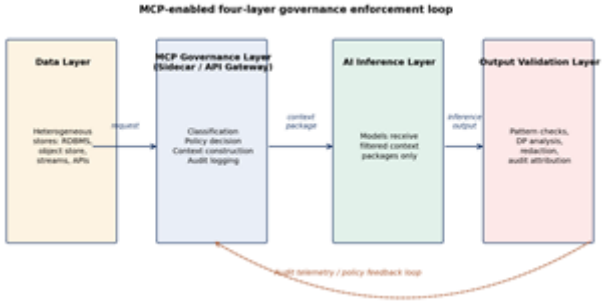

Fig. 2. MCP four-layer enforcement loop: classification and policy at the governance layer, filtered context to the AI layer, output validation, and audit telemetry feeding policy refinement.

VII. Conclusion

AI governance in cloud-native enterprise architectures requires controls at four boundary layers: classification and monitoring at the data layer, contextual access control via MCP at the AI inference layer, response validation at the response layer, and audit logging at the compliance layer. The comparative analysis in Section IV.C shows that MCP is best understood as a compositional layer atop ABAC and Zero Trust — inheriting their attribute-driven discipline while extending coverage to inference-time context construction and output validation that none of RBAC, ABAC, ZTA, or DLP addresses natively. The quantitative envelope in Section VI.B anchors latency overhead, data exposure reduction, and policy enforcement accuracy to peer-reviewed evidence. Future work includes federated MCP across multi-cloud and hybrid systems, standardised MCP policy languages, and governance mechanisms for large language models.

References

1. M. E. Cambronero et al., "Towards a GDPR-compliant cloud architecture with data privacy controlled through sticky policies," PMC, 2024.
2. European Union, "General Data Protection Regulation (GDPR)," Reg. 2016/679, 2018.
3. P. Mehrotra et al., "AI-based fraud detection in financial transactions," Int. J. Embed. Syst. Emerg. Technol., vol. 35, no. 8, p. 101641, 2023.
4. H. Bae et al., "Security and privacy issues in deep learning," IEEE Trans. Artif. Intell. (preprint), arXiv:1807.11655, 2020.
5. R. Liu, W. Ding, Y. Zhang, and Z. Wang, "Context-aware AI-augmented access control for dynamic MFA environments," Research preprint, 2023.

Jain, H. (2026). AI-Driven Data Governance Using Model Context Protocol for Privacy Compliance in Cloud-Native Enterprise Applications. *South African Computer Journal* 38(2), 47–51.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)

6. E. Bertino et al., "Privacy-preserving digital identity management for cloud computing," 2009.
7. K. Datta et al., "Data anonymization and privacy preservation in healthcare systems," EPJ Web Conf., vol. 328, art. 01069, 2025.
8. V. Chamola et al., "A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G," IEEE Access, vol. 8, pp. 90225–90265, 2020.
9. C. Nwachukwu et al., "Anomaly detection in cloud computing environments," Int. J. Sci. Res. Arch., vol. 13, no. 2, pp. 692–710, 2024.
10. F. Mireshghallah et al., "Quantifying privacy risks of masked language models using membership inference attacks," in Proc. EMNLP, 2022, pp. 8332–8347.
11. R. Shokri et al., "On the privacy risks of model explanations," arXiv:1907.00164, 2021.
12. U. Y. Khan and T. R. Soomro, "Design of a federated learning-based resource management algorithm in fog computing," Int. J. Adv. Appl. Sci., vol. 11, no. 2, pp. 195–205, 2024.
13. G. Danezis et al., "Privacy and data protection by design," ENISA, 2015.
14. ITU-T Tech. Watch, "Privacy in cloud computing," ITU-T Tech. Watch Rep., Mar. 2012.
15. M. Park et al., "A trustworthy AI and data governance architecture for LLM deployments," ResearchGate preprint, 2024.
16. F. Turkmen and B. Crispo, "Performance evaluation of XACML PDP implementations," in Proc. ACM SWS, 2008, pp. 37–44, doi: 10.1145/1456492.1456499.
17. R. Marx et al., "A performance analysis of the XACML decision process and the impact of caching," in Proc. IEEE SITIS, 2015, pp. 657–664.
18. A. X. Liu, F. Chen, J. Hwang, and T. Xie, "XEngine: A fast and scalable XACML policy evaluation engine," in Proc. ACM SIGMETRICS, 2008, pp. 265–276, doi: 10.1145/1375457.1375488.
19. S. Rose, O. Borchert, S. Mitchell, and S. Connelly, "Zero trust architecture," NIST SP 800-207, Aug. 2020, doi: 10.6028/NIST.SP.800-207.
20. V. C. Hu et al., "Guide to attribute based access control (ABAC)," NIST SP 800-162, Jan. 2014 (upd. 2019), doi: 10.6028/NIST.SP.800-162.
21. A. Pereira-Vale, G. Marquez, H. Astudillo, and E. B. Fernandez, "Security mechanisms used in microservices-based systems: A systematic mapping," in Proc. IEEE CLEI, 2019, doi: 10.1109/CLEI47609.2019.235060.
22. E. Costante, D. Fauri, S. Etalle, J. den Hartog, and N. Zannone, "A hybrid framework for data loss prevention and detection," in Proc. IEEE SPW, 2016, pp. 324–333, doi: 10.1109/SPW.2016.24.
23. M. F. Faiz, J. Arshad, M. Alazab, and A. Shalaginov, "Predicting likelihood of legitimate data loss in email DLP," Future Gener. Comput. Syst., vol. 110, pp. 744–757, Sep. 2020, doi: 10.1016/j.future.2019.11.004.
24. R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in Proc. IEEE SP, 2017, pp. 3–18, doi: 10.1109/SP.2017.41.
25. [T. Bui and S. D. Stoller, "Fast distributed evaluation of stateful attribute-based access control policies," in Proc. DBSec 2017, LNCS 10359, Springer, 2017, pp. 101–119, doi: 10.1007/978-3-319-61176-1_6

Jain, H. (2026). AI-Driven Data Governance Using Model Context Protocol for Privacy Compliance in Cloud-Native Enterprise Applications. *South African Computer Journal* 38(2), 47–51.

ISSN 1015-7999 (print) ISSN 2313-7835 (online)