

SAP AI Core Enterprise Reference Architecture

A Comprehensive Enterprise Architecture for Governed AI Model Operations, GenAI, Integration, Security, Observability, and Lifecycle Management on SAP BTP

Sanjeeve Kumar Gajadi*
*(HCL Technologies, Bengaluru, Karnataka, India)
Email: sanjeevekumar.g@hcltech.com

Abstract:

SAP AI Core provides enterprise capabilities for executing, managing, and operationalizing Artificial Intelligence workloads within SAP Business Technology Platform landscapes. Enterprise adoption requires a reference architecture that connects model runtime, data, applications, identity, security, integration, observability, Responsible AI, and lifecycle governance. This paper proposes a SAP AI Core Enterprise Reference Architecture spanning user and application channels, AI service orchestration, model runtime, data and feature access, SAP HANA Cloud, SAP Business Data Cloud, SAP Datasphere, SAP Integration Suite, model repositories, security controls, monitoring, and governance. The architecture supports predictive AI, custom machine learning, Generative AI, embeddings, and agentic services while maintaining clean-core integration, least privilege, auditability, resilience, and business accountability. It also defines implementation patterns, deployment controls, model lifecycle processes, and operational KPIs/KRIs. The proposed architecture enables organizations to scale AI on SAP BTP in a secure, reusable, and governed manner.

Keywords — SAP AI Core, SAP BTP, Enterprise AI Architecture, MLOps, Generative AI, SAP HANA Cloud, SAP Datasphere, SAP Integration Suite, Responsible AI, AI Governance

I. INTRODUCTION

Enterprise AI must operate as a managed platform capability rather than a collection of disconnected models. SAP AI Core provides runtime and lifecycle services that can be integrated into broader SAP BTP architectures.

A reference architecture is necessary to standardize model access, deployment, security, data integration, observability, and governance across business use cases.

II. ARCHITECTURE DRIVERS

Key drivers include scalable model execution, reuse, integration with SAP business processes, secure access to enterprise data, cost control, and regulatory readiness.

The architecture must support both predictive workloads and newer Generative AI patterns without weakening clean-core or enterprise security principles.

III. REFERENCE ARCHITECTURE PRINCIPLES

The architecture follows API-first, least privilege, separation of concerns, immutable deployment, observable execution, data minimization, versioned artifacts, and lifecycle governance.

Critical business rules and authorization decisions should remain in deterministic services rather than being delegated solely to probabilistic models.

IV. CONSUMPTION AND EXPERIENCE LAYER

AI services may be consumed by SAP applications, SAP Fiori, custom BTP applications, APIs, workflow services, Joule experiences, or background processes.

The consumption layer should capture user identity, business context, transaction scope, and consent before requests reach AI services.

V. AI ORCHESTRATION AND SERVICE LAYER

Reusable services can manage prompt construction, model routing, embeddings, RAG, evaluation, policy checks, and response handling.

Standardizing these services reduces duplication and enables consistent governance across teams.

VI. SAP AI CORE RUNTIME LAYER

SAP AI Core should host approved AI workloads with controlled configurations, versioned deployments, resource limits, and environment separation.

Runtime governance should link each deployment to a model version, use case, owner, risk tier, data source, and approval record.

VII. DATA AND FEATURE ACCESS

AI workloads may consume curated data from SAP Business Data Cloud, SAP Datasphere, SAP HANA Cloud, object stores, or enterprise APIs.

Data access should preserve source authorization, lineage, quality, region, retention, and purpose limitation. Feature and embedding stores require the same governance discipline as operational data.

VIII. INTEGRATION ARCHITECTURE

SAP Integration Suite, API Management, and Event Mesh can expose governed business capabilities and decouple AI services from core applications.

Integration contracts should include schemas, authentication, throttling, error handling, observability, and version management.

IX. IDENTITY AND SECURITY

Human users, applications, workloads, and automation identities require distinct authentication and authorization. Secrets should be centrally managed and rotated.

Network segmentation, private connectivity, secure endpoints, input validation, and threat monitoring should protect model and data access.

X. MLOPS AND MODEL LIFECYCLE

Models should move through development, validation, approval, deployment, monitoring, recertification, and retirement with auditable evidence.

Changes to model versions, prompts, data, or preprocessing logic should trigger regression testing proportional to risk.

XI. GENERATIVE AI AND RAG PATTERNS

Generative AI services may combine model endpoints with governed retrieval from documents, business data, or Knowledge Graphs.

Prompt injection, sensitive-data leakage, hallucination, and malicious retrieved content should be addressed through layered controls outside the model.

XII. OBSERVABILITY AND FINOPS

Telemetry should capture model calls, latency, token or compute usage, failures, data-source access, policy decisions, and business outcomes.

Cost controls can include quotas, routing, concurrency limits, capacity planning, and chargeback or showback to consuming business units.

XIII. RESPONSIBLE AI AND GOVERNANCE

AI governance should cover use-case approval, risk classification, evaluation, fairness, privacy, explainability, human oversight, and incident response.

High-impact AI services should have stronger independent review and documented limitations.

XIV. RESILIENCE AND OPERATIONS

Production AI services require backup dependencies, retry controls, rollback, fail-safe behavior, and alternative manual processes for critical business functions.

Operational teams should have runbooks for model degradation, data-source failure, security incidents, and provider outages.

XV. IMPLEMENTATION ROADMAP

Enterprises should begin with a bounded use case, build reusable platform patterns, and establish common security, model, data, integration, and observability services.

As maturity increases, teams can add automated governance, model catalogs, evaluation pipelines, reusable RAG services, and agentic capabilities.

XVI. CONCLUSION

SAP AI Core can serve as a central enterprise AI runtime when integrated with SAP BTP architecture, trusted data, governed APIs, security, observability, and lifecycle controls.

The proposed reference architecture provides a scalable path for predictive, generative, and agentic

AI while preserving enterprise control and business accountability.

REFERENCES

- [1] ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system.
- [2] ISO/IEC 27001:2022, Information security, cybersecurity and privacy protection — Information security management systems.
- [3] ISO 31000:2018, Risk management — Guidelines.
- [4] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), 2023.
- [5] National Institute of Standards and Technology, AI RMF Generative Artificial Intelligence Profile, 2024.
- [6] National Institute of Standards and Technology, SP 800-207: Zero Trust Architecture, 2020.
- [7] The Open Group, TOGAF Standard, 10th Edition, 2022.
- [8] ISACA, COBIT 2019 Framework: Governance and Management Objectives, 2019.
- [9] SAP SE, SAP Business Technology Platform Documentation.
- [10] SAP SE, SAP AI Core and SAP AI Launchpad Documentation.
- [11] SAP SE, SAP Integration Suite Documentation.
- [12] SAP SE, SAP HANA Cloud Documentation.
- [13] SAP SE, SAP Cloud Identity Services Documentation.
- [14] SAP SE, SAP Cloud ALM Documentation.
- [15] SAP SE, SAP Business Data Cloud and SAP Datasphere Documentation.
- [16] OWASP Foundation, OWASP Top 10 for Large Language Model Applications.
- [17] MITRE, ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems.
- [18] OECD, Recommendation of the Council on Artificial Intelligence.
- [19] IEEE, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems.
- [20] DAMA International, DAMA-DMBOK2: Data Management Body of Knowledge, 2017.

AUTHOR BIOGRAPHY

Sanjeev Kumar Gajadi is an SAP Enterprise Architect and Lead Consultant with extensive experience in SAP BTP, SAP AI, SAP Datasphere, SAP HANA, integration, cloud architecture,

security, governance, and enterprise transformation. His professional interests include SAP AI Core, enterprise AI platforms, MLOps, Generative AI, Responsible AI, and agentic architectures.

ACKNOWLEDGMENT

The author acknowledges the contributions of enterprise architecture practitioners, international standards bodies, AI governance communities, security researchers, and SAP documentation teams whose published guidance informed the concepts synthesized in this paper.