

Azure KI-Betriebsplattform

Architekturvorschlag für regulatorisch konforme KI-Systeme

EU AI Act | DSGVO | DORA | NIS 2 | ISO 27001

Version 1.0 – Dezember 2024

1. Executive Summary

Dieses Dokument beschreibt eine Azure-basierte Referenzarchitektur für den Betrieb von KI-Systemen unter Berücksichtigung der wichtigsten europäischen und internationalen Regularien. Die Architektur adressiert sowohl technische Anforderungen an moderne KI-Workloads (Agentic AI, LLM-Integration, Vektor-Datenbanken) als auch umfassende Compliance-Anforderungen.

Die vorgeschlagene Lösung basiert auf dem Azure Cloud Adoption Framework und implementiert eine Hub-Spoke-Netzwerktopologie mit durchgängiger Verschlüsselung, Zero-Trust-Prinzipien und umfassendem Audit-Logging.

2. Regulatorischer Rahmen

2.1 EU AI Act

Der EU AI Act klassifiziert KI-Systeme nach Risikoklassen und fordert für Hochrisiko-Systeme umfassende Dokumentation, Risikomanagement, Datenqualität und menschliche Aufsicht. Die Architektur berücksichtigt: Transparenzanforderungen durch Logging und Audit-Trails, technische Dokumentation aller Modelle und Datenflüsse, robuste Überwachung und Qualitätssicherung sowie Human-in-the-Loop-Mechanismen.

2.2 DSGVO

Die Datenschutz-Grundverordnung erfordert: Rechtmäßigkeit der Verarbeitung, Datenminimierung, Speicherbegrenzung, Integrität und Vertraulichkeit, Betroffenenrechte (Auskunft, Löschung, Berichtigung), Privacy by Design und Default sowie Auftragsverarbeitungsverträge mit Cloud-Providern.

2.3 DORA

Der Digital Operational Resilience Act für den Finanzsektor fordert: ICT-Risikomanagement-Framework, Incident Management und Reporting, regelmäßige Resilienztests, Third-Party-Risikomanagement sowie Informationsaustausch über Cyberbedrohungen.

2.4 NIS 2

Die NIS-2-Richtlinie erweitert die Cybersicherheitsanforderungen: Risikomanagementmaßnahmen, Incident Handling, Business Continuity, Supply Chain Security, Netzwerk- und Informationssicherheit sowie Verschlüsselung und Multi-Faktor-Authentifizierung.

2.5 ISO 27001

Der internationale Standard für Informationssicherheits-Managementsysteme (ISMS) bietet den Rahmen für: systematisches Risikomanagement, Sicherheitsrichtlinien und -kontrollen, kontinuierliche Verbesserung sowie regelmäßige Audits und Zertifizierung.

3. Architektur-Übersicht

3.1 Netzwerktopologie

Die Architektur basiert auf einer Hub-Spoke-Topologie mit Azure Virtual WAN oder einem zentralen Hub-VNet. Der Hub enthält zentrale Netzwerkdienste (Firewall, Bastion, VPN Gateway), während die Spokes nach Funktionsbereichen segmentiert sind: Compute-Spoke für Kubernetes und Container-Workloads, Data-Spoke für Datenbanken und Storage, Management-Spoke für DevOps und Monitoring sowie einen DMZ-Spoke für externe Zugänge.

3.2 Sicherheitszonen

Die Architektur definiert drei Sicherheitszonen:

- Public Zone: WAF, Application Gateway, externe Load Balancer
- Private Zone: Kubernetes-Cluster, Datenbanken, interne Services
- Management Zone: Bastion Host, DevOps-Tools, Monitoring

4. Detaillierte Komponentenzuordnung

4.1 Compute & Container

Komponente	Azure Service	Compliance-Relevanz
Kubernetes Cluster	Azure Kubernetes Service (AKS)	ISO 27001: A.12 Operations Security NIS 2: Netzwerksicherheit DORA: ICT-Risikomanagement
Container Registry	Azure Container Registry (ACR)	ISO 27001: A.14 System Acquisition EU AI Act: Dokumentation NIS 2: Supply Chain Security
Cluster Management	Azure Arc + GitOps (Flux)	ISO 27001: A.12.1 Operational Procedures DORA: ICT Asset Management
Frontend Container	AKS + Azure Front Door	DSGVO: Verschlüsselung NIS 2: Zugangskontrolle

4.2 KI-spezifische Komponenten

Komponente	Azure Service	Compliance-Relevanz
Agentic AI Platform	Azure AI Studio + Semantic Kernel	EU AI Act: Art. 9 Risikomanagement EU AI Act: Art. 13 Transparenz ISO 27001: A.8 Asset Management
LLM Adapter / Router	Azure API Management + AI Gateway	EU AI Act: Art. 14 Human Oversight DSGVO: Art. 22 Automatisierte Entscheidungen DORA: Third-Party-Risiko
Prompt Firewall	Azure AI Content Safety + Custom Rules	EU AI Act: Art. 15 Accuracy/Robustness ISO 27001: A.13 Communications Security NIS 2: Incident Prevention
AI Model Scanning	Microsoft Defender for AI + Counterfit	EU AI Act: Art. 9 Risk Management ISO 27001: A.12.6 Vulnerability Mgmt DORA: ICT Testing
AI Interface Scanning	Azure AI Studio Evaluation + Garak	EU AI Act: Art. 15 Robustness Testing ISO 27001: A.14.2 Security in Dev

4.3 Netzwerk & Security

Komponente	Azure Service	Compliance-Relevanz
Web Application Firewall	Azure WAF (Application Gateway / Front Door)	ISO 27001: A.13.1 Network Security NIS 2: Art. 21 Cybersecurity Measures DORA: Art. 9 Protection
Load Balancer + TLS	Azure Application Gateway v2	DSGVO: Art. 32 Verschlüsselung ISO 27001: A.10 Cryptography NIS 2: Verschlüsselungspflicht
API Gateway	Azure API Management (Premium)	ISO 27001: A.9 Access Control DSGVO: Zugriffsprotokolle EU AI Act: API-Dokumentation
Network Firewall	Azure Firewall Premium	ISO 27001: A.13.1 Network Controls NIS 2: Segmentierung DORA: Network Security
DDoS Protection	Azure DDoS Protection Standard	NIS 2: Art. 21 Verfügbarkeit DORA: Art. 11 Business Continuity ISO 27001: A.17 BC Management
Private Connectivity	Azure Private Link / Private Endpoints	DSGVO: Datenminimierung ISO 27001: A.13.2 Information Transfer NIS 2: Secure Communications

4.4 Identity & Access Management

Komponente	Azure Service	Compliance-Relevanz
IAM (RBAC)	Microsoft Entra ID + Azure RBAC	ISO 27001: A.9.1 Access Control Policy DSGVO: Art. 32 Zugriffskontrolle NIS 2: Identity Management
LDAP Integration	Microsoft Entra Domain Services	ISO 27001: A.9.2 User Access Mgmt DORA: Access Management
Privileged Access (PAM)	Microsoft Entra PIM + Bastion	ISO 27001: A.9.2.3 Privileged Access NIS 2: Art. 21 Admin Access DORA: Access Rights Management
Multi-Factor Auth	Microsoft Entra ID MFA + FIDO2	NIS 2: Art. 21 Multi-Factor Auth ISO 27001: A.9.4 System Access DSGVO: Technische Maßnahmen
Conditional Access	Microsoft Entra Conditional Access	ISO 27001: A.9.1.2 Network Access NIS 2: Risiko-basierter Zugang DORA: Logical Access Security

4.5 Daten-Layer

Komponente	Azure Service	Compliance-Relevanz
SQL Datenbank	Azure SQL Database (Hyperscale)	DSGVO: Art. 17 Lösrecht ISO 27001: A.8.2 Information Classification DORA: Data Integrity
NoSQL Datenbank	Azure Cosmos DB	DSGVO: Data Residency (EU) ISO 27001: A.18.1 Compliance NIS 2: Data Protection
Object Storage (S3)	Azure Blob Storage (Data Lake Gen2)	DSGVO: Art. 32 Encryption at Rest ISO 27001: A.10.1 Cryptographic Controls EU AI Act: Training Data Storage
Vektor Datenbank	Azure AI Search + Cosmos DB vCore	EU AI Act: Art. 10 Data Governance DSGVO: Embedding-Verarbeitung ISO 27001: A.8 Asset Management
Redis Cache	Azure Cache for Redis (Enterprise)	ISO 27001: A.12.1 Operational Procedures DSGVO: Session Management NIS 2: Data in Transit
Graph Datenbank	Azure Cosmos DB (Gremlin API)	EU AI Act: Knowledge Graph Governance ISO 27001: A.8.1 Asset Inventory DSGVO: Beziehungsdaten
Big Data Platform	Azure Synapse Analytics + Databricks	EU AI Act: Art. 10 Training Data DSGVO: Pseudonymisierung ISO 27001: A.8.2 Classification

4.6 Messaging & Integration

Komponente	Azure Service	Compliance-Relevanz
Apache Kafka	Azure Event Hubs (Kafka API)	ISO 27001: A.12.4 Logging DORA: Event Traceability EU AI Act: Audit Trail
Message Queues	Azure Service Bus (Premium)	ISO 27001: A.13.2 Message Security DORA: Transaction Integrity NIS 2: Secure Messaging
Event Processing	Azure Event Grid	ISO 27001: A.12.1 Event Management DORA: Real-time Processing NIS 2: Incident Triggers

4.7 DevOps & CI/CD

Komponente	Azure Service	Compliance-Relevanz
Git Repositories	Azure Repos / GitHub Enterprise	ISO 27001: A.14.2 Secure Development EU AI Act: Code Documentation DORA: Change Management
CI/CD Pipelines	Azure Pipelines / GitHub Actions	ISO 27001: A.14.2.2 Change Control NIS 2: Secure Deployment DORA: Release Management
Build Runner	Azure DevOps Agents (Self-hosted)	ISO 27001: A.14.2.6 Secure Dev Env NIS 2: Supply Chain Security
IDE Integration	GitHub Codespaces / VS Code	ISO 27001: A.14.2 Development Security DORA: Developer Access Controls
Infrastructure as Code	Terraform + Azure Verified Modules	ISO 27001: A.12.1.2 Change Management NIS 2: Configuration Management DORA: ICT Asset Documentation
AI Agent Build/Test/Deploy	Azure ML Pipelines + AI Studio	EU AI Act: Art. 9 Testing Requirements EU AI Act: Art. 17 Quality Management ISO 27001: A.14.2.8 System Testing

4.8 Observability & Security Operations

Komponente	Azure Service	Compliance-Relevanz
Logging	Azure Monitor Logs + Log Analytics	ISO 27001: A.12.4 Logging & Monitoring EU AI Act: Art. 12 Record-keeping DORA: Art. 12 Logging Requirements
SIEM (Splunk)	Microsoft Sentinel + Splunk Integration	NIS 2: Art. 23 Incident Reporting ISO 27001: A.16 Incident Management DORA: Art. 19 Incident Reporting
Monitoring	Azure Monitor + Application Insights	DORA: Art. 10 Detection ISO 27001: A.12.4.1 Event Logging NIS 2: Continuous Monitoring
Metriken (Grafana)	Azure Managed Grafana + Prometheus	ISO 27001: A.12.1.3 Capacity Management DORA: Performance Monitoring EU AI Act: Model Performance
Real-time Analytics	Azure Stream Analytics + Event Hubs	DORA: Real-time Detection NIS 2: Threat Detection ISO 27001: A.12.4 Monitoring
Vulnerability Management	Microsoft Defender for Cloud	ISO 27001: A.12.6 Vulnerability Mgmt NIS 2: Art. 21 Risk Management DORA: Art. 8 Vulnerability Assessment
Bastion Host	Azure Bastion (Premium)	ISO 27001: A.9.4.2 Secure Log-on NIS 2: Secure Admin Access DORA: Privileged Access

4.9 Zertifikatsmanagement

Komponente	Azure Service	Compliance-Relevanz
Private Zertifikate	Azure Key Vault (Managed HSM)	ISO 27001: A.10.1 Cryptographic Policy DSGVO: Art. 32 Verschlüsselung NIS 2: Key Management
Öffentliche Zertifikate	Azure App Service Certificates + DigiCert	ISO 27001: A.10.1.2 Key Management NIS 2: PKI Requirements
Zertifikats-Lifecycle	Key Vault Auto-Renewal + Event Grid	ISO 27001: A.10.1 Certificate Lifecycle DORA: Cryptographic Controls
Secrets Management	Azure Key Vault + Managed Identity	ISO 27001: A.9.4.3 Password Management DSGVO: Credential Protection NIS 2: Secrets Handling

5. Compliance-Matrix

Die folgende Matrix zeigt die Zuordnung der Architekturkomponenten zu den regulatorischen Anforderungen:

Architekturbereich	EU AI Act	DSGVO	DORA	NIS 2	ISO 27001
Compute & Container	○	●	●	●	●
KI-Komponenten	●	●	○	○	●
Netzwerk & Security	○	●	●	●	●
Identity & Access	●	●	●	●	●
Daten-Layer	●	●	●	○	●
Messaging & Integration	●	○	●	○	●
DevOps & CI/CD	●	○	●	●	●
Observability & SecOps	●	●	●	●	●
Zertifikatsmanagement	○	●	●	●	●

● = *Primäre Relevanz* | ○ = *Sekundäre Relevanz*

6. Architekturdiagramm-Beschreibung

Die folgende textuelle Beschreibung dient als Grundlage für die Erstellung eines visuellen Architekturdiagramms in Tools wie Visio, Draw.io oder Lucidchart.

6.1 Layer-Struktur (von außen nach innen)

Internet / External Layer

- Azure Front Door mit WAF-Policies (OWASP, Custom Rules)
- Azure DDoS Protection Standard
- Public DNS (Azure DNS)

DMZ / Edge Layer

- Azure Application Gateway v2 (TLS-Terminierung, SSL-Offloading)
- Azure API Management (Premium Tier, VNET-integrated)
- Azure AI Content Safety (Prompt Firewall)

Application Layer

- AKS Cluster (User Node Pools) für Frontend-Container und Backend-Services
- AKS Cluster (GPU Node Pools) für KI-Worker und Model Inference
- Azure AI Studio für Agentic AI Orchestration
- LLM Router (Custom Service auf AKS)

Data Layer

- Azure SQL Database (Hyperscale) - Transaktionsdaten
- Azure Cosmos DB - NoSQL und Graph (Gremlin)
- Azure AI Search - Vektor-Datenbank für RAG
- Azure Blob Storage (Data Lake Gen2) - Object Store
- Azure Cache for Redis - Session und Caching
- Azure Synapse Analytics - Big Data und Analytics

Integration Layer

- Azure Event Hubs (Kafka API) - Event Streaming
- Azure Service Bus - Message Queues
- Azure Event Grid - Event Routing

Management Layer

- Azure Bastion - Sichere Verwaltungszugriffe
- Azure DevOps / GitHub Enterprise - CI/CD und Repos
- Azure Monitor + Log Analytics - Logging und Metriken
- Microsoft Sentinel - SIEM und SOAR
- Azure Managed Grafana - Dashboards und Visualisierung

Security & Identity Layer

- Microsoft Entra ID - Zentrales IAM
- Microsoft Entra PIM - Privileged Identity Management
- Azure Key Vault (Managed HSM) - Secrets und Zertifikate
- Microsoft Defender for Cloud - Security Posture Management
- Microsoft Defender for AI - AI-spezifische Threat Detection

6.2 Netzwerk-Verbindungen

- Hub-VNet: Enthält Azure Firewall, VPN Gateway, Bastion
- Spoke-VNets: Compute, Data, Management - jeweils via VNet Peering verbunden
- Private Endpoints: Alle PaaS-Services über Private Link erreichbar
- NSG/ASG: Network Security Groups pro Subnet, Application Security Groups für Workloads
- UDR: User Defined Routes für Firewall-Routing (0.0.0.0/0 → Azure Firewall)

7. Implementierungsempfehlungen

7.1 Phasenplanung

Phase 1 – Foundation (4-6 Wochen): Aufbau der Netzwerkinfrastruktur (Hub-Spoke), Identity Foundation mit Entra ID, Key Vault und Secrets Management sowie Bastion und grundlegendes Monitoring.

Phase 2 – Core Platform (6-8 Wochen): AKS-Cluster mit GPU-Nodes, Container Registry und GitOps, Basis-Datenbanken (SQL, Cosmos DB, Redis) und API Management.

Phase 3 – AI Components (6-8 Wochen): Azure AI Studio und AI Search, LLM-Integration und Routing, Prompt Firewall und Content Safety sowie Model Scanning und Evaluation.

Phase 4 – Advanced Features (4-6 Wochen): Big Data Platform (Synapse, Databricks), Event Streaming (Event Hubs, Service Bus), Advanced Monitoring (Sentinel, Grafana) und Vulnerability Management.

Phase 5 – Compliance & Hardening (4-6 Wochen): Compliance-Dokumentation, Penetrationstests und Security Audits, ISO 27001 Pre-Assessment sowie Runbook- und Prozessdokumentation.

7.2 Azure Landing Zone Alignment

Die Architektur folgt dem Azure Cloud Adoption Framework mit Enterprise-Scale Landing Zones. Dies umfasst Management Groups für Governance-Hierarchie, Azure Policy für Compliance-Enforcement, Azure Blueprints für standardisierte Deployments und Azure Resource Manager Templates über Terraform.

7.3 Data Residency

Alle personenbezogenen Daten und KI-Modelle werden ausschließlich in EU-Regionen gehostet (empfohlen: West Europe, North Europe, Germany West Central). Azure Policy erzwingt regionale Einschränkungen, und geo-redundante Backups verbleiben innerhalb der EU.

7.4 Kostenoptimierung

Für eine kosteneffiziente Implementierung werden folgende Maßnahmen empfohlen:

- Reserved Instances für vorhersehbare Workloads (1-3 Jahre)
- Spot Instances für KI-Training und Batch-Verarbeitung
- Auto-Scaling für variable Workloads
- Azure Hybrid Benefit für Windows- und SQL-Lizenzen
- Storage Tiering für kosteneffiziente Datenhaltung

8. Anhang: Azure Service Quick Reference

Kategorie	Azure Service	SKU / Tier
Container Orchestration	Azure Kubernetes Service	Standard Tier + GPU Node Pools
Container Registry	Azure Container Registry	Premium (Geo-Replication)
API Management	Azure API Management	Premium (VNET Integration)
Load Balancer	Azure Application Gateway	v2 WAF_v2
CDN / Global LB	Azure Front Door	Premium (WAF)
Firewall	Azure Firewall	Premium (TLS Inspection)
SQL Database	Azure SQL Database	Hyperscale
NoSQL / Graph DB	Azure Cosmos DB	Provisioned Throughput
Vector Database	Azure AI Search	Standard S2+
Object Storage	Azure Blob Storage	Data Lake Gen2
Cache	Azure Cache for Redis	Enterprise E10+
Event Streaming	Azure Event Hubs	Premium (Kafka API)
Message Queue	Azure Service Bus	Premium
Identity	Microsoft Entra ID	P2
PAM	Microsoft Entra PIM	Included in P2
Secrets / Certificates	Azure Key Vault	Premium (HSM)
SIEM	Microsoft Sentinel	Pay-as-you-go
Monitoring	Azure Monitor	Log Analytics Workspace
Dashboards	Azure Managed Grafana	Standard
DevOps	Azure DevOps / GitHub	Enterprise
AI Platform	Azure AI Studio	Enterprise
AI Safety	Azure AI Content Safety	Standard
Security Posture	Microsoft Defender for Cloud	CSPM + CWP

— Ende des Dokuments —