

Architekturvorschlag für den konformen KI-Betrieb auf Microsoft Azure unter Einhaltung von EU AI Act, DSGVO, DORA, NIS 2 und ISO 27001

- Hochskalierbare, sichere und konforme KI-Architektur auf Azure mit klarer Trennung der Layer: Data, AI/ML, Application, Security, Operations, CI/CD.
- Integration spezifischer Azure-Dienste wie AKS, Azure Machine Learning, Azure API Management, Azure Sentinel und Azure Defender für Cloud.
- Umfassendes Compliance-Mapping der Azure-Komponenten zu EU AI Act (Hochrisiko-KI), DSGVO (Datenschutz, Datenminimierung), DORA (Resilienz im Finanzsektor), NIS 2 (Cybersicherheit) und ISO 27001 (ISMS).
- Implementierung von Zero Trust, Datenhoheit (EU-Residenz), Auditierbarkeit (Immutable Logs) und Zugriffskontrollen (RBAC, PAM).
- DevSecOps- und Security-by-Design-Ansatz mit automatisierten CI/CD-Pipelines, Vulnerability Scanning und Incident Response.

1. Übersicht der Architektur und Compliance-Mapping

Architekturdiagramm

Die Architektur ist als mehrschichtiges System konzipiert, das die folgenden Layer umfasst:

- **Data Layer:** Datenhaltung und -verarbeitung mit Azure SQL DB, Azure Cosmos DB, Azure Blob Storage, Azure Vector Database, Azure Redis Cache und Azure Graph Database.
- **AI/ML Layer:** KI-Modellentwicklung, Training und Inference mit Azure Machine Learning (AML), Azure Cognitive Services, Azure OpenAI Service und Azure Functions für Agentic AI.
- **Application Layer:** Frontend und API-Gateway mit Azure App Service, Azure Load Balancer, Azure API Management und Azure Application Gateway.
- **Security Layer:** Schutzmaßnahmen mit Azure Web Application Firewall (WAF), Azure Front Door, Azure Sentinel, Azure Defender for Cloud und Azure Key Vault für Zertifikatsmanagement.
- **Operations Layer:** Überwachung, Logging und Incident Response mit Azure Monitor, Azure Log Analytics, Azure Event Hubs und Azure Stream Analytics.
- **CI/CD Layer:** Automatisierte Build-, Test- und Deployment-Pipelines mit Azure DevOps, Azure Pipelines und Azure Repos.



Compliance-Mapping-Tabelle

Regulatorische Anforderung	Azure-Komponente	Erfüllung der Anforderung	Hinweise
EU AI Act (Hochrisiko-KI)	Azure Machine Learning (AML)	Risikoklassifizierung, Dokumentation, Transparenz, Bias-Detection, Explainability	AML Responsible AI Dashboard, Model Catalog, Prompt Flow für Governance und Auditierbarkeit
DSGVO (Datenschutz)	Azure SQL DB, Azure Cosmos DB, Azure Blob Storage	Datenverschlüsselung, Zugriffskontrolle, Datenminimierung, Löschkonzept	Always Encrypted, Immutable Blob Storage, Azure Purview für Datenklassifizierung und -labeling
DORA (Resilienz Finanzsektor)	Azure Event Hubs, Azure Stream Analytics, Azure Monitor	Kontinuierliches Monitoring, Resilienztests, Incident Response, Audit-Trails	Integration mit Azure Sentinel für SIEM und automatisierte Playbooks
NIS 2 (Cybersicherheit)	Azure Sentinel, Azure Defender for Cloud, Azure WAF, Azure Bastion	Risikomanagement, Vorfallsmeldung, Zugriffskontrolle, Netzwerksegmentierung, PAM	Azure Bastion für sichere VM-Zugriffe, Azure WAF für Webanwendungen, Defender für Vulnerability Scanning
ISO 27001 (ISMS)	Azure Policy, Azure RBAC, Azure Key Vault, Azure Monitor	Risikoanalyse, Sicherheitsziele, Dokumentation, interne und externe Audits	Azure Policy für Compliance-Enforcement, Key Vault für Zertifikatsmanagement, Monitor für Logging und Alerting

2. Detaillierte Komponentenauswahl mit Azure-Diensten

a) Kubernetes & Container-Orchestration

Zweck: Skalierbare, isolierte Ausführung von KI-Workloads (Training/Inference), Multi-Tenancy-Fähigkeit, Compliance mit EU AI Act (Transparenz, Dokumentation).

Azure-Komponenten: - **Azure Kubernetes Service (AKS)** mit:

- Confidential Computing-Nodes (für sensible Datenverarbeitung, DSGVO-konform).



- AKS Workload Identity (statt Service Principals für sichere Pod-Identitäten, IAM-Integration).
- Azure Policy + Gatekeeper (Enforcement von Compliance-Richtlinien wie "keine privileged Container").
- AKS Node Pools mit Spot Instances (Kosteneffizienz) + dedizierte Node Pools für Hochrisiko-KI (Isolation nach EU AI Act).
- Azure Container Registry (ACR) mit Content Trust (Notarization) und Vulnerability Scanning (integriert in CI/CD).

Begründung: AKS bietet automatisiertes Patch-Management (NIS 2), Network Policies (Microsegmentierung) und Azure Monitor für AKS (Logging/Metriken).

Compliance-Hinweis: KI-Modell-Artefakte werden in Containern versioniert und unveränderlich gespeichert (ISO 27001:2022 A.12.3.1) durch Integration mit Azure DevOps und ACR.

b) Agentic AI mit LLM-Adapter

Zweck: Orchestrierung autonomer KI-Agenten mit Prompt-Sicherheit, Halluzinationskontrolle und Audit-Trails.

Azure-Komponenten: - **Azure Machine Learning (AML)** für:

- Prompt Flow (Versionierung und Governance von Prompts, EU AI Act-Konformität).
- Model Catalog (Metadaten-Management für LLM-Adaptionen, inkl. Risikoklassifizierung).
- AML Responsible AI Dashboard (Bias-Detection, Explainability für EU AI Act Art. 13).
- **Azure OpenAI Service** (mit private Endpoints und Customer-Managed Keys (CMK) für Datenhoheit).
- **Azure Functions/Durable Functions** (serverless Agent-Orchestration mit Retries & Dead-Letter Queues für Resilienz).

Begründung: AML bietet built-in Compliance-Templates (z. B. für DSGVO-Löschanfragen via Data Subject Requests).

Offene Frage: Agent-Aktionen werden über Azure Monitor und Azure Log Analytics protokolliert, um DORA-Anforderungen an "Complete Activity Logs" zu erfüllen.

c) Security Layer (WAF, Prompt Firewall, AI-Scanning)

Zweck: Schutz vor Prompt Injections, Datenexfiltration und Model Poisoning (NIS 2, ISO 27001 A.14).

Azure-Komponenten: - **Azure Web Application Firewall (WAF)** auf Azure Front Door (OWASP Top 10 Abdeckung, Rate Limiting).

- **Prompt Firewall:**

- Azure API Management (APIM) mit Custom Policies (Regex/ML-basierte Prompt-Validation).
- AML Content Safety (Toxizitätsfilterung + Custom Rules für domänenspezifische Compliance).

- **AI Model Scanning:**

- AML Model Scanning (Schadcode in Modellen, z. B. via Trivy-Integration in Pipelines).



- Azure Defender for Containers (Runtime-Schutz für KI-Worker).
- **AI Interface Scanning:**
- Microsoft Defender for APIs (in APIM integriert, Detects Anomalies in LLM-Aufrufen).

Begründung: Kombiniert Pre-Deployment Scanning (CI/CD) mit Runtime Protection (Defender).

Compliance-Hinweis: Scanning-Ergebnisse werden für Audits (EU AI Act Art. 72) 10 Jahre in Azure Monitor Log Analytics gespeichert.

d) Frontend & API-Gateway

Zweck: Sichere Exposition von KI-Services mit TLS 1.3, mTLS und Rate Limiting.

Azure-Komponenten: - **Frontend:** Azure Static Web Apps (Container-basiert, mit Managed Identity für Backend-Zugriff).

- **Load Balancer:** Azure Load Balancer (Layer 4) + Application Gateway (Layer 7, TLS-Terminierung, WAF-Integration).
- **API Gateway:** Azure API Management (APIM) mit:
 - OAuth2/JWT Validation (IAM-Integration).
 - Policy Expressions für Request/Response Transformation (z. B. PII-Redaktion).
 - Developer Portal (Dokumentation für EU AI Act-Transparenz).
- **LLM Router:** Azure Traffic Manager (Global Routing) + APIM Routing Rules (A/B-Testing, Canary Deployments).

Begründung: APIM bietet Quota Management (DDoS-Schutz) und Usage Analytics (DORA-Reporting).

e) IAM & PAM (RBAC, LDAP, Just-in-Time-Zugriff)

Zweck: Least Privilege, Separation of Duties (ISO 27001 A.9.1.2) und Notfallzugriff (DORA).

Azure-Komponenten: - **Azure Active Directory (AAD)** mit:

- Privileged Identity Management (PIM) (Just-in-Time Admin-Zugriff).
- Conditional Access Policies (MFA, Device Compliance).
- Entra Permissions Management (Cloud Infrastructure Entitlement Management, CIEM).
- **LDAP:** Azure AD Domain Services (für Legacy-Systeme, read-only Replicas).
- **Bastion Host:** Azure Bastion (RDP/SSH ohne öffentliche IPs, Session Recording für Audits).

Begründung: AAD Access Reviews erfüllen DSGVO-Recertification (Art. 32).

Offene Frage: Break-Glass-Zugriff (DORA) wird über PIM mit zeitlich begrenzten Rollen und MFA umgesetzt, um Compliance nicht zu brechen.

f) Datenlayer (SQL, NoSQL, Vektordatenbank, Cache, GraphDB, S3)

Zweck: Datenminimierung (DSGVO), Verschlüsselung (NIS 2), Performance für KI.



- Azure-Komponenten:**
- **SQL:** Azure SQL DB mit Always Encrypted (Client-side) + Ledger (unveränderliche Tabellen für Audits).
 - **NoSQL:** Azure Cosmos DB (Global Distribution, Customer-Managed Keys).
 - **Vektordatenbank:** Azure AI Search (mit Vector Index) oder Weaviate/Pinecone auf AKS (falls Open-Source erforderlich).
 - **Object Store:** Azure Blob Storage mit Immutable Blob Storage (DSGVO-Löschsperren).
 - **Redis Cache:** Azure Cache for Redis (Enterprise Tier mit TLS und AAD-Auth).
 - **GraphDB:** Azure Cosmos DB Gremlin API oder Neptune auf AKS (für Knowledge Graphs).

Begründung: Azure Purview klassifiziert und labelt Daten automatisch (DSGVO Art. 30).

Compliance-Hinweis: Datenresidenz (EU-only) für Kafka-Topics und Cosmos DB wird durch Azure Policy und Azure Monitor erzwungen.

g) Big Data & Echtzeit-Analytics

Zweck: Skalierbare Datenpipelines für Model Training/Monitoring (EU AI Act Art. 10).

- Azure-Komponenten:**
- **Apache Kafka:** Azure Event Hubs (Kafka-kompatibel) + MirrorMaker für Multi-Region.
 - **Stream Processing:** Azure Stream Analytics (SQL-basierte Echtzeit-Aggregation).
 - **Batch Processing:** Azure Databricks (mit Delta Lake für ACID-Compliance).
 - **Data Lake:** Azure Data Lake Storage (ADLS) Gen2 mit Hierarchical Namespace.

Begründung: Event Hubs Geo-DR erfüllt NIS 2-Anforderungen an Ausfallsicherheit.

h) Logging, Monitoring & Vulnerability Management

Zweck: Immutable Logs (DORA), Anomalieerkennung (NIS 2), Compliance-Reporting.

Azure-Komponenten:

- **Logging:**

- Azure Monitor (AKS, APIM, SQL Logs) + Log Analytics Workspace (7 Jahre Retention).
- Azure Storage Lifecycle Management (Automatische Archivierung alter Logs).

- **Monitoring:**

- Grafana auf Azure Managed Grafana (mit Azure Monitor Datenquelle).
- Azure Dashboards für SLO/SLI-Tracking (DORA-Metriken).

- **Vulnerability Management:**

- Microsoft Defender for Cloud (CVE-Scanning, Regulatory Compliance Dashboard).
- Azure Security Benchmark (NIS 2-Alignment).
- **Splunk:** Azure Native Splunk Integration (für SIEM, falls erforderlich).

Begründung: Defender for Cloud bietet automatisierte Response Playbooks (ISO 27001 A.16.1).

i) CI/CD & DevSecOps (Git, Pipelines, IDE, Agent Builds)

Zweck: Compliance-as-Code (ISO 27001 A.14.2.5) und auditierbare Deployments.



Azure-Komponenten: - **Git Repos:** Azure Repos (mit Branch Policies für Code Reviews).

- **Pipelines:** Azure Pipelines mit:

- Static Application Security Testing (SAST) via Checkmarx oder SonarQube.

- Dynamic Scanning in Load Tests (Azure Load Testing).

- Terraform: Azure Bicep (native IaC) + Terraform in Azure DevOps (mit Plan Policy Validation).

- **IDE:** VS Code + Azure DevOps Extension (mit Roslyn Analyzers für Secure Coding).

- **Agent Builds:** Azure VM Scale Sets (ephemere Agenten für Sicherheit).

Begründung: Pipeline Compliance Reports (z. B. "Keine Hardcoded Secrets") für Audits.

j) Queues & Message Broker

Zweck: Entkopplung von KI-Workloads (Resilienz nach DORA).

Azure-Komponenten: - **Azure Service Bus** (FIFO, Dead-Letter Queues) oder Azure Queue Storage (für einfache Szenarien).

- **Event-Driven Architecture:** Azure Event Grid (für KI-Trigger, z. B. neue Daten in Blob Storage).

k) Zertifikatsmanagement (öffentlich/privat)

Zweck: TLS Everywhere (NIS 2) und Code Signing (EU AI Act).

Azure-Komponenten: - **Azure Key Vault** (mit HSM-backed Keys für Code Signing).

- **Azure App Service Certificates** (automatische Rotation).

- **Private CA:** Azure Private Link + DigiCert/GlobalSign Integration.

3. Cross-Cutting Concerns & Risikomitigation

Datenflüsse

- **Datenflussdiagramm (DFD):** Markierung von DSGVO-relevanten Verarbeitungen (z. B. PII in LLM-Prompts).
- **Incident Response:** Azure Sentinel Playbooks für KI-spezifische Vorfälle (z. B. Model Drift, Data Leakage).
- **Cost Optimization:** Azure Cost Management + Billing und Reserved Instances (für AKS/Cosmos DB).
- **Vendor Lock-in:** Bewertung von Open-Source-Alternativen (z. B. Kubernetes statt AKS, OpenSearch statt Azure AI Search).

4. Empfohlene Next Steps & Proof-of-Concept

- **Priorisierte Implementierungsschritte:**
 1. AKS mit Policy Guardrails,
 2. APIM + WAF,
 3. Azure Sentinel Integration,



4. Azure DevOps CI/CD Pipeline mit Security Scanning.

- **POC-Umfang:** Isoliertes AKS-Cluster mit AML und APIM für einen einzelnen KI-Agenten.
- **Stakeholder-Kommunikation:** Compliance-Readiness-Dokument für interne Audits.

Anhang

- **Architekturdiagramm:** Visio/Lucidchart mit Legende für Compliance-Symbole.
- **Tabelle:** Mapping von Azure-Diensten zu Compliance-Anforderungen (DSGVO, NIS 2 etc.).
- **Terraform/Bicep-Snippets:** Kritische Komponenten (z. B. AKS mit Policy, Key Vault Setup).
- **Beispiel-Pipeline:** YAML für sicheres KI-Agent Deployment.

Quellen

- EU AI Act, GDPR, DORA, NIS 2, ISO 27001 Guidelines.
- Azure Kubernetes Service (AKS) Dokumentation.
- Azure Machine Learning (AML) Dokumentation.
- Azure API Management (APIM) Dokumentation.
- Azure Sentinel und Defender for Cloud Dokumentation.
- Azure DevOps und Azure Pipelines Dokumentation.
- Azure Monitor und Log Analytics Dokumentation.
- Azure Key Vault und Certificate Services Dokumentation.
- Azure Event Grid und Service Bus Dokumentation.

Diese Architektur stellt einen umfassenden, regulatorisch konformen und technisch robusten Vorschlag für den KI-Betrieb auf Microsoft Azure dar, der die Anforderungen von EU AI Act, DSGVO, DORA, NIS 2 und ISO 27001 adressiert und gleichzeitig Skalierbarkeit, Observability, Security-by-Design und DevSecOps-Prinzipien integriert.

