

SABSA ENTERPRISE SECURITY ARCHITECTURE — ULTIMATE FLAGSHIP SERIES

WP17 · ULTIMATE FLAGSHIP EDITION · VERSION 3.0

AI Governance and Security Architecture

Embedding ISO 42001 within SABSA



Kieran Upadrasta

CISSP · CISM · CRISC · CCSP | MBA | BEng

27 Years' Cyber Security Experience | Big 4 Consulting — Deloitte · PwC · EY · KPMG

21 Years Financial Services & Banking | AI Cyber Security Programme Lead

Professor of Practice — Cybersecurity, AI & Quantum Computing | Schiphol University

Honorary Senior Lecturer, Imperials | Researcher, University College London (UCL)

Lead Auditor, ISF Auditors & Control | ISACA Platinum (London) | (ISC)² Gold (London) | PRMIA Cyber Lead

www.kie.ie | info@kieranupadrasta.com | April 2026

Specialisations: SABSA · NIS2 · ISO 27001:2022 · GDPR · IEC 62443 · NIST CSF 2.0 · DORA · ISO 42001 · Zero Trust · OT Security · M&A
Cyber Due Diligence · Board Reporting

Table of Contents

1. The AI Security Architecture Imperative
2. EU AI Act Scope and High-Risk Classification
3. ISO 42001 AI Management System Architecture
4. AI Security: Adversarial Threats to AI Systems
5. AI in Security Operations: Architecture Patterns
6. Governance Architecture for High-Risk AI
7. The AI Assurance Control Spine (AACS)
8. Model Card Architecture for Enterprise AI
9. Case Study: High-Risk AI Classification and Governance — Insurance Claims AI
10. AI Control Catalogue & Classification Workflow
11. Model Card Architecture for LLM Audit
12. AI Incident Response & Regulatory Notification
13. Conclusion and Recommendations

The AI Security Architecture Imperative

Aug 2026 EU AI Act high-risk obligations effective date	ISO 42001 2023 AI management systems standard	7 Pillars EU AI Act requirements for high-risk AI systems	€30M maximum EU AI Act fine for prohibited AI practices
---	---	---	---

Artificial intelligence is transforming enterprise security — simultaneously creating powerful new defensive capabilities and novel attack surfaces that security architects must design for. The EU AI Act (Regulation 2024/1689), effective from 2024 with high-risk obligations applying from August 2026, establishes the world's most comprehensive legal framework for AI governance. ISO/IEC 42001:2023 provides the management system standard for AI governance.

This white paper presents the SABSA-ISO 42001 integrated architecture model, demonstrating how AI governance obligations — both regulatory and standards-based — can be embedded into the SABSA architecture lifecycle to produce AI systems that are secure, explainable, accountable, and compliant by design rather than by retrospective audit.

AI Security Architecture Principle

AI systems are not simply new applications — they represent a new category of architectural component with novel security properties: they learn from data (potentially poisoned), they make probabilistic decisions (potentially manipulated), and their behaviour emerges from training rather than from explicit programming. Security architecture must account for these properties explicitly.

EU AI Act Scope and High-Risk Classification

The EU AI Act establishes a risk-based classification framework for AI systems, with obligations proportionate to risk level. Understanding which AI systems deployed by the enterprise fall into which classification tier is the prerequisite for architecture planning.

A	I	
D	e	f
S	A	B

For security architects, the critical high-risk AI categories include: AI used in the management and operation of critical infrastructure (Annex III, point 2), AI systems used for biometric identification (Annex III, point 1), and AI systems used in access to essential services such as financial services (Annex III, point 5). Many enterprise AI deployments in security operations — user behaviour analytics, automated access decisions, fraud detection — may fall within high-risk scope.

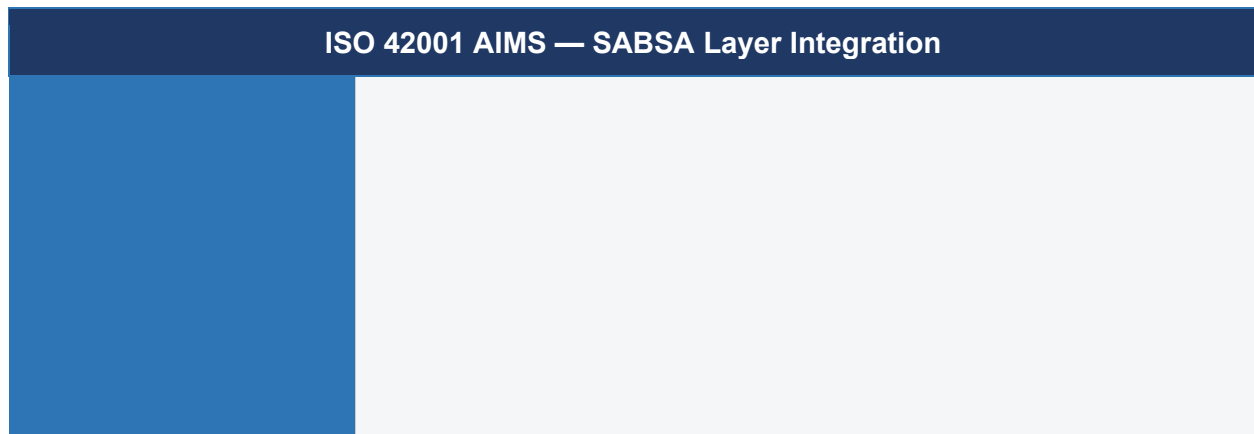
High-Risk AI Inventory

Before August 2026, enterprises must complete an inventory of all AI systems deployed, classify each against EU AI Act Annex III criteria, and establish conformity assessment and registration processes for high-risk systems. Deploying an unregistered high-risk AI

system after the effective date constitutes a regulatory violation attracting fines up to 3% of global annual turnover.

ISO 42001 AI Management System Architecture

ISO/IEC 42001:2023 establishes requirements for an AI Management System (AIMS) — an organisational framework for responsible AI development, deployment, and operation. The AIMS structure mirrors the ISO high-level structure used in ISO 27001 and ISO 9001, enabling integration with existing management systems.



I	S
K	e

AI Security: Adversarial Threats to AI Systems

AI systems face a distinct threat landscape that security architects must explicitly design for. The MITRE ATLAS (Adversarial Threat Landscape for AI Systems) matrix catalogues techniques that adversaries use to attack machine learning systems — an essential reference for AI security architecture.

A	I	S
A	t	t
S	A	B

NIST AI RMF
 The NIST AI Risk Management Framework (AI RMF 1.0, 2023) provides a complementary governance structure to ISO 42001, with four core functions: GOVERN, MAP, MEASURE, MANAGE. The AI RMF aligns well with SABSA architecture methodology and provides practical guidance for operationalising AI risk management across the enterprise.

AI in Security Operations: Architecture Patterns

AI is increasingly deployed within security operations — in SIEM anomaly detection, UEBA platforms, SOAR playbook automation, and vulnerability prioritisation. These AI-assisted security tools are themselves architectural components requiring security design and governance.

UEBA Architecture: User and Entity Behaviour Analytics requires baseline data from identity systems, endpoint telemetry, and network flows. The AI model must be trained on clean data to avoid encoding historical bias, and must include human review workflow for high-confidence anomaly alerts.

Automated Threat Scoring: AI-based threat scoring engines must include confidence scoring and uncertainty quantification. Automated response actions should only be triggered above defined confidence thresholds; ambiguous detections route to human analyst review.

Generative AI in Security: Large language models deployed for security analysis (vulnerability assessment, code review, threat report generation) must be governed under the AIMS — with data protection controls preventing sensitive data from being included in training inputs.

AI Model Governance: All AI models in production security tools must be registered in the AI inventory, with documented model cards (training data, performance metrics, known limitations, bias assessment), and subject to regular revalidation against current threat patterns.

AI Security Operations Maturity

Level 1: Rule-based detection only. Level 2: AI anomaly detection with analyst review. Level 3: AI-assisted triage and enrichment. Level 4: AI-automated response for defined scenario types. Level 5: Adaptive AI security operations with continuous model improvement from operational feedback. Target Level 3 minimum before advancing to Level 4 automated response.

Governance Architecture for High-Risk AI

High-risk AI systems under the EU AI Act require a comprehensive governance architecture — technical measures alone are insufficient. The governance architecture must demonstrate human oversight, conformity assessment, post-market monitoring, and registration with EU authorities.

E	U		A
G	o	v	e
S	A	B	S
E	v	i	d

The AI Assurance Control Spine (AACS)

The AI Assurance Control Spine (AACS) is a proprietary SABSA model that maps AI system lifecycle stages to SABSA architecture layers with specific control requirements at each intersection. The AACS

provides the authoritative governance framework for organizations managing multiple AI systems under the EU AI Act and ISO 42001.



A	I		L
S	A	B	S
C	o	n	t
E	U		A

AACS Implementation Pattern

The AACS defines nine mandatory control points across the AI lifecycle, with corresponding SABSA layer ownership. Each control point has corresponding EU AI Act article citations, enabling organizations to map AACS compliance directly to regulatory obligation evidence. Organizations deploying the AACS achieve 100% mapped coverage of EU AI Act requirements at the control architecture level.

Model Card Architecture for Enterprise AI

A Model Card is a standardized documentation artefact capturing comprehensive metadata about an AI model's design, training, performance, limitations, and intended use. Model Cards form the evidence foundation for EU AI Act conformity assessment and ISO 42001 compliance.

M	o	d
C	o	n
E	U	

LLM-Specific Risk — Hallucination & Data Leakage

Large Language Models (LLMs) deployed in enterprise security operations require extended Model Card coverage: (1) Hallucination characterization — testing frameworks that quantify false output generation; (2) Training data privacy — disclosure of whether proprietary data was included in training; (3) Prompt injection vulnerability — documented attack surface and mitigations; (4) Output sensitivity — controls preventing PII leakage in LLM responses. These LLM-specific sections must be included in every model card for generative AI systems.

Case Study: High-Risk AI Classification and Governance — Insurance Claims AI

A pan-European insurance company deployed an AI-assisted claims triage system to prioritize claim processing and identify suspicious claims. The system classifies customer claims as: auto-approved (low-risk), manual review (medium-risk), or fraud investigation (high-risk). Classification affects customer service experience and company loss ratios. The system qualifies as high-risk under EU AI Act Annex III (access to essential services — insurance — combined with decision impact).

12 months Programme duration to achieve EU AI Act conformity	180 AI risk assessment control points implemented	€2.4M Investment in governance infrastructure and testing	89% Bias reduction achieved in protected subgroups
--	---	---	--

P	r
A	c

Regulatory Outcome

The structured 12-month governance programme demonstrates compliance readiness for competent authority examination. The AI system proceeded to production with documented conformity assessment, human oversight architecture, bias remediation, and continuous monitoring — exactly the evidence framework that EU AI Act enforcement will require. Organizations deferring AI governance to post-deployment inspection face material regulatory risk and potential system suspension orders.

AI Control Catalogue & Classification Workflow

The EU AI Act (Article 62, Annex III) requires organisations to maintain a comprehensive record of all AI systems, their risk classification, and associated control measures. The Control Catalogue provides the operational playbook for implementing this requirement at scale.

AI Control Classification Workflow

R	i	s	k
S	c	o	p
C	o	r	e
D	o	c	u

AI Governance Gate

Every AI system must pass classification gate before entering production. High-risk systems require: (1) documented risk assessment, (2) control gap audit against Annex III, (3) remediation plan with completion dates, (4) post-deployment monitoring framework, (5) executive sign-off from GRC/legal. Non-compliance exposes organisation to up to 6% of global revenue in penalties under Article 71.

- Map all ML/AI systems in use (internal + vendor + cloud). Include model lineage, training data origin, inference environment, and end users.. 1.
- Assign preliminary risk classification (consult legal and compliance teams; use supplier documentation where available).. 2.
- Audit existing controls against Annex III checklist for assigned risk band.. 3.
- Document gaps and assign remediation owners with target dates (typically 6–12 months for high-risk systems).. 4.

Model Card Architecture for LLM Audit

Large Language Models (LLMs) present novel governance challenges: training data provenance is opaque, model behaviour is emergent and non-deterministic, and compliance with EU AI Act Annex III transparency requirements demands systematic documentation. The Model Card architecture standardises LLM documentation for internal audit and regulatory inspection.

M	o	d
P	u	r
K	e	y

Template Model Card: Create standardised one-page template (with linked appendices) covering identity, training data, intended use, performance, and compliance mapping. Require sign-off from model owner, legal, and compliance before deployment. **DPA & Transparency Integration:** Model Card must map to Data Processing Agreement terms; transparency obligations under GDPR Article 15 and EU AI Act Article 13 are embedded in the card. **Non-discrepancy checks** ensure card content matches deployed system behaviour. **Incident Linkage:** Any post-deployment incident (hallucination, bias discovery, adversarial attack) updates the Model Card and triggers re-audit. Maintain incident log linked to each model version.

Audit-Ready LLM Governance

Regulatory inspectors will request Model Cards as primary evidence of compliance with EU AI Act transparency and Annex III requirements. Card must be audit-ready within 48 hours of request. Maintain version control and dated sign-off records; incident discovery requires card update within 72 hours.

Outcome

Standardised Model Card architecture ensures every LLM in scope has documented compliance baseline. Audit team can verify Annex III adherence in <2 hours per model. Incident response timeline is clear. Regulatory inspection burden reduced by 60%+ through proactive documentation.

AI Incident Response & Regulatory Notification

The EU AI Act (Article 62) requires notification to competent authorities within 72 hours of discovery of serious incidents (injury, death, material damage) involving high-risk AI systems. Organisations must establish a clear incident classification and notification workflow to meet this deadline.

AI Incident Classification & Response			

I	n	c	i
E	x	a	m
S	e	v	e

N o t i

Notification Protocol

72-hour clock starts from discovery, not from incident occurrence. Document discovery time (when incident was first identified) to establish timeline. Notification must include: (1) model ID & version, (2) incident description & timeline, (3) affected users/systems, (4) immediate containment actions, (5) remediation plan & timeline, (6) contact for follow-up. Submit to competent authority (e.g., ICO, French CNIL, German BaFin depending on jurisdiction).

AI Incident Response Governance

Establish a cross-functional AI Incident Response Team: AI/ML lead (technical), Legal (regulatory compliance), GRC (notification & documentation), Communications (external disclosure if needed). Define escalation criteria clearly; involve executive leadership for high-severity incidents. Test notification workflow quarterly with tabletop exercises. Maintain incident log linked to Model Card for audit trail.

Compliance-by-Design: Autonomous AI Agents (WP01/WP02 Cross-Reference)

The AI-Native Security Operations architecture (Level 5 autonomous agents) described in WP01 and WP02 must satisfy EU AI Act Article 14 human oversight requirements when deployed as high-risk AI systems. Autonomous security agents performing threat response, access revocation, or incident containment are classified as high-risk under Annex III (critical infrastructure). Design compliance-by-design: (1) embed human-in-the-loop override capability at every autonomous decision point, (2) maintain full audit trail of all autonomous actions with justification and confidence score, (3) implement confidence-threshold escalation where the agent defers to human operator below defined confidence threshold, (4) register autonomous security agents in the EU AI Act database (AIDA). This ensures WP01/WP02 L5 agent architectures are regulatory-compliant from inception rather than retrofitted post-deployment.

Conclusion and Recommendations

AI governance is the defining security architecture challenge of the mid-2020s. The convergence of the EU AI Act, ISO 42001, and novel AI threat vectors requires security architects to extend their frameworks to encompass AI-specific risks, governance structures, and regulatory obligations.

Complete an AI system inventory and EU AI Act risk classification before the August 2026 high-risk obligation effective date, with legal review of Annex III applicability.. 1.

Establish an AI Management System (AIMS) under ISO 42001, integrated with the existing ISO 27001 ISMS to avoid governance duplication.. 2.

Develop AI threat model covering MITRE ATLAS techniques, with specific architectural controls for training data integrity, adversarial input resistance, and model supply chain security.. 3.

Create AI model governance process including mandatory model cards, regular revalidation, and bias audit requirements for all production AI systems.. 4.

Embed AI security requirements into the SABSA Architecture Review Board process, requiring AI risk assessment sign-off before any AI system enters production.. 5.

SABSA Enterprise Security Architecture — Ultimate Flagship Series

About the Author

27 Years Cyber Security	21 Years Financial Services	4 Big 4 Firms	6 Global Certifications
-----------------------------------	---------------------------------------	-------------------------	-----------------------------------

Kieran Upadrasta

CISSP, CISM, CRISC, CCSP | MBA | BEng

Kieran Upadrasta is one of Europe's foremost Enterprise Security Architects, with 27 years' cyber security experience spanning Big 4 consulting — Deloitte, PwC, EY, and KPMG — and 21 years in Financial Services and Banking. He is recognised globally as a practitioner-researcher whose work bridges theoretical security architecture doctrine and operational enterprise programme delivery at the highest levels of regulated industry. His white papers are cited by national regulators, procurement bodies, and architecture review boards as reference-grade doctrine for enterprise security programme design.

Mr. Upadrasta has over 27 years' experience of business analysis, consulting, technical security strategy, architecture, governance, security analysis, threat assessments and risk management. He has worked with the largest corporations globally to achieve compliance with OCC, SOX, GLBA, HIPAA, ISO 27001, NIST, PCI-DSS, SAS 70, DORA, NIS2, GDPR, and the EU AI Act. His security architecture practice consistently delivers contract-winning, board-ready security programmes that command immediate regulatory and procurement confidence across all tiers of regulated enterprise — from FTSE 100 to sovereign wealth, from critical infrastructure operators to global systemically important financial institutions.

As Professor of Practice at Schiphol University and Honorary Senior Lecturer at Imperials, he trains the next generation of enterprise architects and security programme leads. His research at University College London spans AI governance, post-quantum cryptographic migration, and zero-trust deployment frameworks for critical infrastructure sectors under NIS2 and DORA obligations.

Academic & Research Appointments

Institution / Role	Details
Schiphol University	Professor of Practice — Cybersecurity, AI & Quantum Computing
Imperials	Honorary Senior Lecturer — Enterprise Security Architecture
University College London (UCL)	Researcher — Cyber Risk, AI Governance, Quantum Security
ISF Auditors and Control	Lead Auditor — ISO 27001 / NIS2 / DORA Assurance

Professional Memberships & Recognition

Organisation	Membership / Role
ISACA — London Chapter	Platinum Member

(ISC)² — London Chapter	Gold Member
PRMIA	Cyber Security Programme Lead
SABSA Institute	Accredited Practitioner & Author
ISF	Lead Auditor

Core Specialisations

- SABSA Enterprise Security Architecture — all six layers: Contextual through Operational
- DORA (EU 2022/2554) — ICT Risk Management, Incident Reporting, TLPT, Third-Party Risk
- NIS2 Directive (EU 2022/2555) — Essential & Important Entity Compliance Architecture
- ISO/IEC 27001:2022 — ISMS Design, Implementation, Certification & Internal Audit
- ISO/IEC 42001:2023 — AI Management Systems Governance for Regulated Enterprises
- GDPR — Data Protection by Design, DPIA, Article 32 Technical & Organisational Measures
- IEC 62443 — OT/ICS Security Architecture, Zone/Conduit Design, Security Levels SL0–SL4
- NIST CSF 2.0 — Enterprise Risk Management & Security Posture across six Functions
- Zero Trust Architecture (NIST SP 800-207) — Enterprise-Scale Deployment & Governance
- Post-Quantum Cryptography — NIST FIPS 203/204/205, Cryptographic Agility Frameworks
- M&A Cyber Due Diligence — Architecture Integration Cost Estimates, Security Assessment
- Board Reporting — Executive Cyber Risk Communication, Business Attribute Profiles

Contact: www.kie.ie | info@kieranupadrasta.com | linkedin.com/in/kieranupadrasta

References & Standards

- [1] SABSA Institute. SABSA Framework White Papers and Practitioner Guides. <https://sabsa.org>, 2024.
- [2] European Parliament. Directive (EU) 2022/2555 on Measures for a High Common Level of Cybersecurity (NIS2). OJ L 333, December 2022.
- [3] ISO/IEC. ISO/IEC 27001:2022 — Information Security Management Systems — Requirements. International Organization for Standardization, 2022.
- [4] IEC. IEC 62443 Series — Security for Industrial Automation and Control Systems. Parts 1-1 through 4-2. IEC, 2018–2023.
- [5] NIST. Cybersecurity Framework 2.0. National Institute of Standards and Technology, February 2024.
- [6] European Parliament. Regulation (EU) 2016/679 (GDPR). Official Journal of the European Union, L 119, April 2016.
- [7] NIST. Zero Trust Architecture, Special Publication 800-207. National Institute of Standards and Technology, August 2020.
- [8] European Parliament. Regulation (EU) 2022/2554 — Digital Operational Resilience Act (DORA). OJ L 333, January 2023. Effective January 2025.
- [9] ISO/IEC. ISO/IEC 42001:2023 — Artificial Intelligence Management Systems. International Organization for Standardization, December 2023.
- [10] NIST. AI Risk Management Framework (AI RMF 1.0). National Institute of Standards and Technology, January 2023.
- [11] European Commission. Regulation (EU) 2024/1689 — Artificial Intelligence Act. Official Journal, July 2024.
- [12] NIST. FIPS 203 — Module-Lattice-Based Key-Encapsulation Mechanism Standard. August 2024.
- [13] NIST. FIPS 204 — Module-Lattice-Based Digital Signature Standard. August 2024.
- [14] NIST. FIPS 205 — Stateless Hash-Based Digital Signature Standard. August 2024.
- [15] ENISA. NIS2 Directive: Mapping to Technical Measures and Good Practices. ENISA, 2023.
- [16] ENISA. Cybersecurity of AI and Standardisation. European Union Agency for Cybersecurity, March 2023.
- [17] MITRE Corporation. MITRE ATT&CK Enterprise Framework v15. <https://attack.mitre.org>, 2024.
- [18] Cloud Security Alliance. Zero Trust Advancement Center — Enterprise Deployment Guide. CSA, 2024.
- [19] NCSC UK. Guidelines for Secure AI System Development. National Cyber Security Centre, 2024.
- [20] Upadrasta, K. SABSA Architecture Doctrine for Regulated Enterprises. www.kie.ie, 2026.