COMPANION TECHNICAL DOCUMENT | FEBRUARY 2026

ARQE Technical Specification

AI Risk Quantification Engine — Version 1.0

A FAIR-Based Methodology for Quantifying Autonomous AI Risk

Companion to: The Agentic Risk Doctrine

Kieran Upadrasta | info@kieranupadrasta.com | www.kie.ie

> This document publishes the complete ARQE methodology to enable independent replication, peer review, and regulatory scrutiny. All parameters, distribution families, calibration sources, and worked examples are disclosed in full.

Contents

1. Scope and Purpose

This document specifies the AI Risk Quantification Engine (ARQE), a methodology for converting technical agentic AI risk metrics into probabilistic financial exposure estimates suitable for board-level risk reporting. ARQE extends the FAIR (Factor Analysis of Information Risk) framework — the only international standard Value-at-Risk model for information risk (The Open Group, Open FAIR Standard, 2023) — with three agentic AI-specific extensions.

The specification is published to enable: (a) independent replication by organisations with FAIR-compatible tooling; (b) peer review by quantitative risk analysts; (c) regulatory scrutiny by supervisory authorities evaluating AI governance frameworks; (d) academic validation through third-party empirical testing.

1.1 Normative References

The Open Group. Open FAIR Body of Knowledge. 2023.

Freund, J. & Jones, J. Measuring and Managing Information Risk: A FAIR Approach. Butterworth-Heinemann, 2014.

FAIR Institute. Lebo, J. FAIR-AIR: Adapting FAIR for AI Risk. 2024.

OWASP. AI Vulnerability Scoring System (AIVSS) v0.5. November 2024.

NIST. AI Risk Management Framework (AI RMF 1.0). AI 100-1, January 2023.

2. Mathematical Foundation

2.1 Standard FAIR Decomposition

Risk = Loss Event Frequency (LEF) × Loss Magnitude (LM)

Where:

LEF = Threat Event Frequency (TEF) × Vulnerability (V)

LM = Primary Loss (PL) + Secondary Loss (SL)

TEF = Contact Frequency (CF) × Probability of Action (PA)

V = Susceptibility (S) / (S + Control Strength (CS))

PL = f(Asset Value, Threat Capability)

SL = Secondary Loss Event Frequency (SLEF) × Secondary Loss Magnitude (SLM)

SL includes: regulatory penalties, litigation costs, reputational damage, D&O personal exposure, insurance premium impact, and operational disruption costs.

2.2 ARQE Agentic Extensions

ARQE introduces three modifications to standard FAIR to account for characteristics unique to autonomous AI systems:

Extension 1: Autonomy Amplification Factor (AAF)

AAF modifies TEF to reflect autonomous decision velocity:

TEF_agentic = TEF_baseline × AAF

AAF = Agent Decision Rate (ADR) × Drift Probability (DP) × Cascading Failure Coefficient (CFC)

Where:

ADR = autonomous decisions per time unit / human-reviewed decisions per time unit. Typical range: 100–100,000x. For autonomous trading agents: ~4,000x (based on 4,000 trades/second vs. 1 trade/second for manual review).

DP = probability that an autonomous decision deviates from intended parameters. Calibrated against OECD AI Incident Monitor: mean DP = 0.031 (SD=0.018) across n=847 incidents (2020–2025).

CFC = expected number of downstream agents affected by a single agent failure. Estimated from incident post-mortems: median CFC = 3.2 (IQR: 1.4–6.8) for interconnected agent ecosystems.

Extension 2: NHI Privilege Exposure Index (NPEI)

NPEI modifies Vulnerability to account for machine identity over-permissioning:

V_agentic = V_baseline × NPEI

NPEI = (NHI_count × Privilege_Ratio) / Human_Identity_Baseline

Where:

NHI_count = total non-human identities (service accounts, API keys, tokens, agent credentials).

Privilege_Ratio = proportion of NHIs with privileges exceeding operational requirements. Population mean: 0.97 (Entro Labs H1 2025, n=500 organisations). Range: 0.72–0.99.

Human_Identity_Baseline = total human identities. The NHI:Human ratio (population mean: 144:1, Entro Labs H1 2025) serves as a scaling factor.

Practical effect: An organisation with 1,000 humans and 144,000 NHIs where 97% are over-privileged has NPEI = (144,000 × 0.97) / 1,000 = 139.7. This amplifies the baseline vulnerability by two orders of magnitude.

Extension 3: Regulatory Penalty Matrix (RPM)

RPM modifies Secondary Loss to incorporate multi-jurisdictional penalty exposure:

SL_agentic = SL_baseline + ∑(RPM_j) for j in jurisdictions

RPM_j = Penalty_Max_j × Non_Compliance_Probability_j × Enforcement_Activity_j

Jurisdiction	Regulation	Penalty_Max	Enforcement_Activity (2025–26)
EU	AI Act (high-risk)	€35M or 7% turnover	Pending (effective Aug 2026)
EU Financial	DORA	2% turnover + €5M personal	Active; automated supervision underway
EU	NIS2	€10M or 2% turnover	Active; management bans available
UK	CS&R Bill	£17M or 4% turnover	Committee stage Feb 2026
US	SEC Disclosure	Enforcement actions; settlements	41 Form 8-K filings; 4 enforcement actions

Non_Compliance_Probability is assessed per jurisdiction using a 5-point scale mapped to percentage ranges: (1) Full compliance = 0–5%; (2) Minor gaps = 5–20%; (3) Material gaps = 20–50%; (4) Significant non-compliance = 50–80%; (5) No framework = 80–100%.

3. Monte Carlo Simulation Specification

3.1 Simulation Parameters

Each ARQE assessment runs n=10,000 Monte Carlo iterations. This provides: P50 estimate precision of ±2.5% (relative standard error); P95 estimate precision of ±5%; P99 estimate precision of ±10%. These precision levels are consistent with industry practice for operational risk modelling in financial services (Basel II/III AMA).

Input Variable	Distribution Family	Parameters	Calibration Source	Adjustment Method
Agent Error Rate	Log-normal	μ=0.03, σ=0.8	CyberSecEval 1–4 (Meta)	Client agent testing results override
Transaction Volume	Normal	Client-specific μ, σ	12-month trailing client data	Direct measurement
Regulatory Multiplier	Triangular	min, mode, max per jurisdiction	Statutory penalty schedules	Legal counsel input
Cascading Failure Prob.	Beta	α=2, β=5	OECD AI Incident Monitor n=847	Adjusted for agent interconnection density
Detection Latency	Log-normal	μ=5.3, σ=0.6	IBM X-Force 2025 n=553	Adjusted for client MTTD history
NHI Privilege Ratio	Empirical	Client assessment	Entro Labs methodology	Direct assessment during Phase 1
Reputational Impact	Pert	min=0.5x, mode=1x, max=3x base	Brand equity models	Industry-adjusted by sector

3.2 Distribution Selection Rationale

Log-normal: Used for right-skewed loss data where extreme events are possible but rare. Consistent with FAIR methodology literature (Freund & Jones, 2014, Ch.8) and empirical loss distributions in cyber risk (Eling & Loperfido, 2017, "Data Breaches and Information Security Risk," Journal of Risk and Insurance).

Beta: Used for probability parameters bounded on [0,1]. Shape parameters (α=2, β=5) produce a right-skewed distribution with mode at 0.2, reflecting the observation that cascading failures occur but are not the dominant failure mode.

Triangular/PERT: Used for expert-estimated parameters where only minimum, most likely, and maximum values are available. PERT distribution preferred over triangular for smoother tails.

Normal: Used only for transaction volume where sufficient historical data supports Gaussian assumption (Central Limit Theorem applies to aggregate transaction counts).

3.3 Correlation Structure

Input variables are treated as independent in the base specification. This is a simplifying assumption. In practice, agent error rates and cascading failure probabilities are positively correlated (errors trigger cascades). A Gaussian copula with correlation matrix can be applied in advanced implementations. Correlation estimates are not published in v1.0 due to insufficient paired observations in the calibration dataset.

This is a known limitation. Ignoring positive correlation between error rates and cascading failures will underestimate tail risk (P95/P99). Users are advised to treat P99 estimates as conservative.

4. Worked Examples

4.1 Example A: European Bank with Autonomous Trading Agents

Parameter	Input Value	Source
Autonomous trading agents	800	Client inventory
Average transactions/agent/day	12,000	Client data (12-month)
Agent error rate (observed)	0.024	Client testing (CyberSecEval methodology)
Average transaction value	€42,000	Client financial data
NHI:Human ratio	186:1	Phase 1 assessment
NHI privilege ratio	0.94	Phase 1 assessment
Jurisdictions	EU (DORA, AI Act), UK, Singapore	Client operations
Current MTTD for AI anomalies	312 days	Client incident data

ARQE Output (n=10,000 iterations):

P50 ALE: €18.4M | P75: €27.1M | P95: €41.3M | P99: €68.7M

95% Confidence Interval for P50: [€15.2M, €21.6M]

Dominant risk driver: Agent error rate (tornado sensitivity: ±€9.8M)

Regulatory penalty component: €4.2M (P50) across 3 jurisdictions

4.2 Example B: Healthcare System with Diagnostic AI Agents

Parameter	Input Value	Source
Diagnostic AI agents	45	Client inventory
Diagnoses/agent/day	340	Client operational data
Agent error rate (observed)	0.008	Clinical validation testing
Patient safety incident cost (avg)	£180,000	NHS Litigation Authority data
NHI:Human ratio	62:1	Phase 1 assessment
Jurisdictions	UK (CS&R Bill), EU AI Act (high-risk)	Client operations

ARQE Output (n=10,000 iterations):

P50 ALE: £4.8M | P75: £8.2M | P95: £14.6M | P99: £23.1M

Dominant risk driver: Patient safety litigation (tornado sensitivity: ±£6.4M)

Note: Healthcare ALE is lower in absolute terms but regulatory multiplier is higher due to EU AI Act high-risk classification of medical AI (Annex III).

5. Validation Protocol

5.1 Back-Testing Methodology

ARQE outputs are validated through annual back-testing against realised losses. For each client with 12+ months of post-deployment data:

1. Compare P50 ALE estimate against actual annualised AI-related losses.

2. Verify that realised losses fall within the stated confidence interval in ≥95% of cases.

3. Track calibration drift: if >20% of realised losses fall outside P75, recalibrate input distributions.

Current back-testing results (n=23 clients with 12+ months data): 21 of 23 (91%) fell within P50–P95 range. Two outliers were both upside surprises (lower-than-predicted losses). No downside calibration failures.

5.2 Sensitivity Analysis Protocol

For each ARQE assessment, tornado diagrams are produced showing the sensitivity of ALE to each input variable. Variables are perturbed ±1 standard deviation from their mean. Variables contributing >15% of total ALE variance are flagged for enhanced data collection in subsequent assessment cycles.

5.3 Independent Replication

The following steps enable independent replication:

1. Obtain FAIR-compatible quantitative risk tool (RiskLens, Safe Security, or equivalent).

2. Map ARQE input variables to tool's input taxonomy using the parameter table in Section 3.1.

3. Apply agentic extensions (AAF, NPEI, RPM) as multiplicative modifiers to standard FAIR inputs.

4. Run n=10,000 Monte Carlo iterations with specified distribution families.

5. Compare P50, P95, P99 outputs against ARQE reference ranges published in Section 4.

Deviations >20% from reference ranges should prompt investigation of calibration differences. The author welcomes correspondence reporting replication results.

6. Agentic Governance Index: Full Scoring Rubric

The AGI provides a standardised, reproducible maturity assessment across seven dimensions. Each dimension is scored 1–5 using explicit evidence requirements designed for independent verification through document review, technical testing, and interview. This section publishes the complete rubric.

6.1 Dimension 1: AI Asset Inventory

Level	Score	Description	Evidence Requirement
Initial	1	No comprehensive inventory of AI agents or autonomous systems	Absence of inventory documentation
Developing	2	Partial inventory maintained manually; covers <50% of known agents	Spreadsheet or wiki-based register; no automated discovery
Defined	3	Comprehensive register of all known agents with decision authority classification	Documented register covering >90% of agents; reviewed quarterly
Managed	4	Automated inventory with <24hr update cycle; includes NHI mapping	Automated discovery tool outputs; NHI:human ratio calculated; deviation alerts
Optimising	5	Auto-discovering inventory with real-time emergent capability tracking	Continuous monitoring; new agent types detected within 1 hour; board-reported

6.2 Dimension 2: Decision Authority Framework

Level	Score	Description	Evidence Requirement
Initial	1	No boundaries defined for autonomous decision-making	No delegation matrix or authority limits documented
Developing	2	Informal guidelines; some agents have soft limits	Draft policy; no formal approval or enforcement mechanism
Defined	3	Formal delegation framework approved by management	Board-approved policy; RACI matrix for agent decision categories
Managed	4	Dynamic delegation with automated enforcement and override logging	Technical enforcement (policy-as-code); override audit trail; quarterly board review
Optimising	5	Adaptive authority boundaries adjusting in real-time to risk context	Context-aware policies; automated escalation; human-in-the-loop for novel decisions

6.3 Dimension 3: Kill Switch Readiness

Level	Score	Description	Evidence Requirement
Initial	1	No containment capability for autonomous agents	No kill switch architecture; no shutdown procedures documented
Developing	2	Manual shutdown procedures documented but untested	Written procedures; no test records; estimated shutdown time >30 minutes

Defined	3	Three-tier kill switch architecture designed and partially implemented	Architecture document; soft contain tested; hard/emergency untested
Managed	4	Full three-tier hierarchy tested monthly; results logged and board-reported	Monthly test records; emergency shutdown <60 seconds demonstrated; board reports
Optimising	5	Automated kill switch with self-healing capabilities and rehearsed recovery	Automated triggering; <30 second containment; tested quarterly with tabletop; zero downtime recovery demonstrated

6.4 Dimension 4: Board Reporting

Level	Score	Description	Evidence Requirement
Initial	1	No AI-specific reporting to board or risk committee	No board papers reference AI agents or autonomous system risk
Developing	2	Ad-hoc reporting when incidents occur	Incident-driven board papers; no regular cadence
Defined	3	Quarterly AI governance report to board risk committee	Standing agenda item; templated report with defined KRIs
Managed	4	Automated KRI dashboard with defined escalation thresholds	Live dashboard access; escalation protocols documented and tested; board training completed
Optimising	5	Real-time board visibility with predictive analytics and peer benchmarking	Predictive KRIs; automated regulatory change alerts; sector benchmark comparison

6.5 Dimension 5: Regulatory Mapping

Level	Score	Description	Evidence Requirement
Initial	1	Unaware of AI-specific regulatory obligations	No regulatory gap assessment conducted
Developing	2	Awareness of key regulations; no systematic mapping	Informal awareness; no documented compliance matrix
Defined	3	Formal gap assessment completed against primary frameworks	Gap analysis document; remediation plan with timelines
Managed	4	Cross-jurisdictional compliance matrix maintained and regularly updated	Multi-framework mapping (EU AI Act, DORA, NIS2, UK CS&R); quarterly review cycle
Optimising	5	Automated regulatory tracking with proactive compliance adaptation	Regulatory intelligence feed; automated impact assessment for new requirements

6.6 Dimension 6: NHI Governance

Level	Score	Description	Evidence Requirement
Initial	1	No visibility over non-human identities	NHI:human ratio unknown; no service account inventory
Developi	2	Partial inventory of service accounts and	Manual inventory; no privilege assessment; no

ng		API keys	rotation policy
Defined	3	Comprehensive NHI inventory with privilege classification	Inventory covering >90% of NHIs; privilege levels documented
Managed	4	Privilege ratio <20%; credential rotation <90 days; audit trail maintained	Automated rotation; privilege reviews quarterly; anomaly detection active
Optimising	5	JIT (just-in-time) privilege automation with zero standing privileges	Zero standing privilege architecture; JIT provisioning; all NHI actions logged

6.7 Dimension 7: Incident Response

Level	Score	Description	Evidence Requirement
Initial	1	No AI-specific incident response playbook	No documented procedures for AI agent failures
Developing	2	General incident response adapted informally for AI events	Existing IR plan with informal AI addendum; untested
Defined	3	Dedicated AI incident response playbook documented	Playbook covering agent containment, rollback, notification; assigned roles
Managed	4	Playbook rehearsed quarterly; automated containment capability tested	Quarterly tabletop records; automated containment demonstrated; regulatory notification workflow tested
Optimising	5	Fully automated response with self-healing and post-incident learning loop	Automated containment <60s; self-healing demonstrated; post-incident ML model retraining triggers

7. Population Baselines and Self-Diagnosis

This section provides reference baselines enabling organisations to benchmark their current posture without external engagement. Baselines are derived from Phase 1 (ASSESS) measurements across n=47 organisations prior to Doctrine implementation.

7.1 Pre-Deployment Baseline Scores by Sector

Dimension	Financial Services (n=18)	Critical Infra (n=9)	Healthcare (n=8)	Defence (n=5)	All Sectors (n=47)
AI Asset Inventory	2.1 (SD=0.8)	1.8 (SD=0.7)	1.6 (SD=0.9)	2.8 (SD=0.6)	2.0 (SD=0.8)
Decision Authority	1.8 (SD=0.9)	1.5 (SD=0.6)	1.3 (SD=0.7)	2.4 (SD=0.8)	1.7 (SD=0.8)
Kill Switch Readiness	1.5 (SD=0.7)	1.2 (SD=0.5)	1.1 (SD=0.4)	2.1 (SD=0.7)	1.4 (SD=0.6)
Board Reporting	2.3 (SD=0.6)	1.9 (SD=0.8)	1.7 (SD=0.6)	2.6 (SD=0.5)	2.1 (SD=0.7)
Regulatory Mapping	2.6 (SD=0.7)	2.2 (SD=0.9)	2.0 (SD=0.8)	2.9 (SD=0.4)	2.4 (SD=0.8)
NHI Governance	1.2 (SD=0.5)	1.0 (SD=0.3)	1.1 (SD=0.4)	1.8 (SD=0.6)	1.2 (SD=0.5)
Incident Response	2.0 (SD=0.8)	1.7 (SD=0.7)	1.5 (SD=0.6)	2.5 (SD=0.7)	1.9 (SD=0.7)
COMPOSITE AGI	1.9	1.6	1.5	2.4	1.8

Interpretation: Financial services and defence show higher baselines than healthcare and infrastructure, consistent with longer histories of regulatory-driven governance investment. NHI Governance is the weakest dimension across all sectors (composite mean: 1.2), confirming this as the highest-leverage investment area.

7.2 Post-Deployment Outcome Ranges

Dimension	Pre Mean	Post Mean	Mean Improvement	Min Improvement	Max Improvement
AI Asset Inventory	2.0	4.1	+2.1	+1.2	+3.0
Decision Authority	1.7	3.8	+2.1	+1.0	+2.8
Kill Switch Readiness	1.4	4.2	+2.8	+1.5	+3.5
Board Reporting	2.1	4.0	+1.9	+0.8	+2.6
Regulatory Mapping	2.4	4.4	+2.0	+1.0	+2.8
NHI Governance	1.2	3.6	+2.4	+1.3	+3.2
Incident	1.9	4.0	+2.1	+1.0	+3.0

Response					
COMPOSITE AGI	1.8	4.0	+2.2	+1.1	+3.0

Note: Improvement ranges reflect heterogeneity in starting position, organisational complexity, and implementation fidelity. Organisations starting at Level 1 show larger absolute gains; organisations starting at Level 2–3 achieve higher final scores.

8. ARQE-to-Board Translation Protocol

ARQE outputs are technical. This section specifies the protocol for translating quantitative results into board-consumable reporting.

8.1 Board Report Template

Report Element	Content	Frequency	Audience
AI Risk Exposure Summary	P50 ALE, P95 ALE, trend vs. prior quarter	Quarterly	Full Board
Top 3 Risk Drivers	Tornado sensitivity ranking with mitigation status	Quarterly	Risk Committee
Regulatory Compliance Status	Multi-jurisdictional heat map with deadlines	Quarterly	Audit Committee
Kill Switch Test Results	Pass/fail rate, containment times, trend	Quarterly	Risk Committee
AGI Maturity Scorecard	7-dimension radar chart with sector benchmark	Semi-annual	Full Board
Incident Summary	AI-specific incidents, near-misses, containment outcomes	Quarterly	Risk Committee
ARQE Calibration Update	Any parameter changes, back-testing results	Annual	Risk Committee

8.2 KRI Dashboard Specification

KRI	Metric Definition	Green	Amber	Red	Data Source
Agent Decision Accuracy	% of autonomous decisions within defined parameters	>97%	93–97%	<93%	Agent monitoring platform
Model Drift Index	Statistical distance from baseline distribution (PSI)	<0.10	0.10–0.25	>0.25	Model monitoring tool
NHI Privilege Ratio	% of NHIs with privileges exceeding requirements	<20%	20–50%	>50%	IAM platform; quarterly assessment
Kill Switch Latency	Time from trigger to full containment (seconds)	<60	60–300	>300	Monthly test records
Regulatory Gap Count	Number of unresolved compliance gaps across frameworks	0	1–3	>3	Compliance tracking system
Human Override Frequency	Overrides per day (trend matters more than absolute)	Stable/decreasing	Increasing 10–25%	Increasing >25%	Agent platform logs
AI Incident Rate	Incidents per quarter (normalised by agent count)	<0.5/100 agents	0.5–2.0/100	>2.0/100 agents	Incident management system

Board Confidence Score	Annual board self-assessment of AI risk oversight (1–5)	> 4.0	3.0–4.0	< 3.0	Annual board survey

9. Additional Worked Examples

9.1 Example C: European Energy Utility with Grid AI

Parameter	Input Value	Source
Autonomous grid management agents	12	Client inventory
Decisions/agent/hour	8,400	SCADA system logs
Agent error rate	0.0006	Historical grid event data (36 months)
Customers served	15.2M	Regulatory filing
Average outage cost (per minute)	€340,000	Regulatory compensation framework
NHI:Human ratio	98:1	Phase 1 assessment
Jurisdictions	EU (NIS2, DORA via energy designation)	Client operations

ARQE Output (n=10,000 iterations):

P50 ALE: €8.1M | P75: €14.2M | P95: €28.7M | P99: €52.3M

Dominant risk driver: Cascading failure coefficient (CFC = 8.4 due to grid interconnection density)

Note: Despite low agent error rate (0.06%), the extreme consequence magnitude of grid failures and high CFC produce significant tail risk. This example illustrates why ARQE tail analysis (P95/P99) is critical for CNI.

9.2 Example D: Insurance Group with Underwriting AI

Parameter	Input Value	Source
AI underwriting agents	34	Client inventory
Policies processed/agent/day	420	Underwriting platform data
Agent error rate	0.018	Back-testing against actuarial review
Average policy value	£185,000	Client financial data
NHI:Human ratio	112:1	Phase 1 assessment
NHI privilege ratio	0.89	Phase 1 assessment
Jurisdictions	UK (PRA/FCA), EU (Solvency II, DORA)	Client operations

ARQE Output (n=10,000 iterations):

P50 ALE: £6.4M | P75: £10.8M | P95: £19.2M | P99: £31.6M

Dominant risk driver: Regulatory multiplier (dual PRA/FCA + EU DORA jurisdiction)

Unique finding: Insurance underwriting AI errors compound through the reserve chain — a single systematic pricing bias across 14,000 policies created £2.1M in reserve adequacy risk before detection. This highlights the importance of model drift monitoring (KRI: PSI threshold) for financial services AI.

9.3 Example E: Defence Contractor with Autonomous Supply Chain AI

Parameter	Input Value	Source
Supply chain optimisation agents	8	Client inventory
Procurement decisions/agent/day	2,800	ERP system logs
Agent error rate	0.012	Quality assurance testing
Average procurement value	£420,000	Client financial data
NHI:Human ratio	74:1	Phase 1 assessment (classified environment)
Jurisdictions	UK (MOD standards), NATO (STANAG 4778)	Client operations

ARQE Output (n=10,000 iterations):

P50 ALE: £3.2M | P75: £5.8M | P95: £11.4M | P99: £18.9M

Note: Defence applications show lower ALE due to smaller agent populations and more disciplined baseline controls (mean pre-deployment AGI: 2.4). However, consequence severity for classified data exposure is modelled separately under national security frameworks and is not included in the financial ALE.

10. Known Limitations

10.1 Correlation structure: Input variables are treated as independent (Section 3.3). This underestimates tail risk where positive correlations exist between agent errors and cascading failures.

10.2 Calibration currency: Parameters calibrated against 2020–2025 data. Agentic AI is evolving rapidly; distribution parameters may require annual recalibration as the incident base matures.

10.3 Regulatory enforcement uncertainty: RPM penalty probabilities are forward-looking estimates. Actual enforcement patterns under DORA and EU AI Act are not yet observed (Section "Limitations" in main whitepaper).

10.4 Reputational loss modelling: Brand equity impact is the least well-calibrated component. PERT distributions based on expert estimation introduce subjective variability. Alternative approaches (event study methodology on equity prices) would strengthen this component but require publicly traded entities.

10.5 NHI data source concentration: NPEI relies primarily on Entro Labs data. Independent NHI surveys (e.g., from CyberArk, Silverfort, or Veza) may produce different baseline ratios. Cross-validation against multiple sources is recommended.

10.6 Sample size for back-testing: n=23 is below the minimum (n=30) for robust statistical inference on back-testing accuracy. Confidence in calibration quality will increase as the deployment base matures.

10.7 Geographic limitation: All deployments are EU/UK/ME. ARQE parameters have not been validated in North American, APAC, or emerging market regulatory environments. RPM penalty matrices require jurisdiction-specific adaptation.

10.8 Sector concentration: Financial services (38% of sample) is over-represented. Findings may not generalise equally to manufacturing, retail, or technology sectors where agentic AI deployment patterns differ.

10.9 Temporal scope: The 2020–2025 calibration period captures the emergence phase of agentic AI incidents. As the field matures and detection capabilities improve, incident frequency distributions will likely shift. Annual recalibration is recommended.

11. Version History and Peer Review

Version	Date	Changes	Review Status
v0.1	January 2024	Initial FAIR extension for agentic AI	Internal review
v0.5	September 2025	Added NPEI and RPM extensions	Reviewed by FAIR-certified analyst
v1.0	February 2026	Full specification; calibration published	Independent methodology review completed

Methodology review statement: The ARQE v1.0 specification was reviewed by an independent quantitative risk analyst holding Open FAIR Analyst certification. The reviewer confirmed: (a) mathematical consistency of agentic extensions with FAIR ontology; (b) appropriateness of distribution families for stated use cases; (c) adequate disclosure of calibration sources and limitations. The reviewer has no commercial relationship with the author.

---

info@kieranupadrasta.com | www.kie.ie