

ODA3 INSTITUTE

SECTOR IMPLEMENTATION GUIDANCE

GAISSF™ Insurance Sector Implementation Guide

Operational guidance for secure, resilient, fair and evidence-based use of AI across insurance

GAISSF-SEC-039 | Public | Draft for Publication — External Review Open

Document Control

Document ID	GAISST-SEC-039
Title	GAISST TM Insurance Sector Implementation Guide
Doc Type	Sector implementation guidance — informative
Framework	GAISST TM v1.0 — Global AI Safety and Security Framework (59 controls across nine domains)
Authoritative sources	GAISST-NOR-001 (Framework Standard); GAISST-NOR-004 (Control Catalogue)
Status	Draft for Publication — Public — External Review Open. Open review gates: qualified insurance/actuarial review and jurisdiction-specific legal/regulatory review remain open; named approvals pending.
Dates	Draft dated 30 June 2026; updated 05 October 2026. No final publication date is asserted.
Owner	ODA3 Institute (ODA3 Pvt Ltd)
Review cycle	At least annually and on material GAISST, regulatory or sector change
Contact	contact@oda3.org
Classification	Public (on release)

GAISSTTM Insurance Sector Implementation Guide

Operational guidance for secure, resilient, fair and evidence-based use of AI across insurance

Normative status: This guide is informative. It does not amend GAISSTTM, create certification requirements, or replace applicable law, regulation, actuarial judgement, contractual duties or policyholder protections. Profile membership is defined in GAISST-NOR-001. GAISST certification is not currently offered. © 2026 ODA3 Pvt Ltd. GAISST Ecosystem developed by ODA3 Institute, licensed under GEL v1.0.

Review gates

Role	Required reviewer	Status	Date	Notes
Document owner	ODA3 Institute	Approved for publication-candidate release	30 June 2026	Substantive and control-mapping review complete
Insurance practitioner	Qualified underwriting/claims leader	Open		Cross-line practitioner review required
Actuarial	Qualified actuary	Open		Validate actuarial terminology and examples
Legal/regulatory	Qualified counsel	Open		Jurisdiction-specific statements and notices

Role	Required reviewer	Status	Date	Notes
Cybersecurity / AI security	Qualified reviewer	Open		Threat and control interpretation
Standards editor	Qualified reviewer	Open		Normative language and cross-references

How to Use This Guide

Reader	Priority sections	Key appendices / artifacts
Board, CFO, CRO and Appointed Actuary	Executive summary; prudential and actuarial impact vectors; governance; Notably Absent	Open issues; workbook dashboard and risk register
CISO and security architecture	Risk landscape; security architecture; agentic AI; control mapping	Control Mapping and Incident Register
Claims and underwriting leaders	Function interpretations; human oversight; worked scenarios	Human Oversight and Scenarios sheets
Compliance, conduct and privacy	Transparency and contestability; data governance; incident management	Source Register and Evidence Register
Procurement and vendor risk	Third-party management and concentration	Vendor Assessment and AI/SBOM evidence

Document Hierarchy and Precedence

- Applicable law and binding regulatory or supervisory direction.
- GAISST-NOR-001 Framework Standard and other applicable GAISST normative documents.
- Approved GAISST interpretations and assessment rules.
- This informative sector guide.
- Organisational policies and procedures.

Where requirements conflict, the organisation should record the conflict, obtain qualified advice, escalate through governance, and preserve the rationale. Unresolved interpretation must not be silently converted into a control conclusion.

Normative language

SHALL and SHALL NOT appear only when quoting or accurately restating verified GAISST normative requirements. SHOULD indicates recommended sector practice; MAY indicates permission; CAN indicates capability or possibility. Conformity is determined against governing normative requirements, not this guide alone.

Executive Summary

Insurance AI influences eligibility, price, coverage, claims, fraud investigation, reserving, catastrophe exposure, customer communication and operational resilience. The same model failure can therefore create policyholder harm, balance-sheet effects, conduct exposure and cyber loss. SEC-039 provides the operational layer between GAISST controls and insurance practice.

- Classify systems by decision authority, policyholder impact, financial consequence, reversibility, data sensitivity, scale and dependency concentration.
- Treat model risk, cybersecurity, actuarial validation, conduct risk, privacy and operational resilience as one evidence system.
- Retain accountability when models, data, platforms or decisions are outsourced.
- Make human oversight timely, competent, informed and authorised; ceremonial review is not effective oversight.
- Monitor outcomes, overrides, appeals, complaints, incidents, drift and vendor changes together.
- Preserve decision traceability sufficient to reconstruct material outcomes and support challenge.

1. Introduction

This guide supports insurers, reinsurers, brokers, insurtechs, third-party administrators and material service providers applying GAISST to insurance operations. It translates control objectives into sector interpretations, evidence and implementation patterns without creating new normative obligations.

1.1 How to use this guide

- Start with the system inventory and classification.
- Map applicable GAISST controls using the verified 59-control register.
- Select function-specific interpretations and evidence.
- Record legal, actuarial and jurisdiction-specific obligations separately.
- Use the workbook to assign owners, collect evidence and track remediation.
- Retain open issues and Notably Absent statements with the assessment record.

2. Scope

Included organisations: life, health, property and casualty, general, commercial and specialty insurers; reinsurers; brokers and distributors; insurtech platforms; TPAs; claims, underwriting, data and analytics providers. Included systems span predictive, generative, computer-vision, graph, optimisation, rules-plus-model and agentic systems across acquisition, development, procurement, deployment, operation, change and retirement.

Excluded: legal or actuarial opinions; product approval; rate filing decisions; insurance coverage interpretation; certification conclusions; universal numerical thresholds; and jurisdiction-specific applicability determinations.

3. Insurance operating context

Actor	Primary accountability	Typical AI dependency	Key evidence
Insurer	Policyholder outcomes, risk acceptance, claims, governance and compliance	Models, SaaS, data, decision engines	Inventory, approvals, validation, monitoring, incidents
Reinsurer	Portfolio exposure, treaty analytics and risk transfer	Catastrophe and accumulation models	Model/version records, assumptions, stress tests
Broker / agent	Advice, suitability, disclosure and submission quality	Recommendation and submission tools	Training, approved content, conversation records
TPA / claims provider	Operational claims handling within delegated authority	Triage, document, fraud and settlement tools	Authority matrix, QA, logs, escalation
Insurtech / vendor	Service security, model operation and contractual evidence	Hosted model/API/platform	Assurance, change notices, incident and subprocessor records
Data provider	Data provenance, quality, rights and update integrity	Enrichment, telematics, health, property and cyber data	Lineage, contracts, quality and correction records

Insurance decisions may have long-tail consequences. Errors can remain latent until claims emerge, portfolios mature or catastrophe events expose correlated assumptions. Legacy systems and rules engines can also materially alter model outputs; assessment scope should include the complete decision chain.

4. Sector risk landscape

Risk domain	Description	GAISST domains	Indicative evidence
Underwriting and pricing	Unstable or discriminatory risk classification, proxy effects, unapproved variables, model drift, historical-data bias and correlated portfolio errors.	D1, D6, D8	Model inventory; feature approval; actuarial validation; subgroup analysis; drift reports; override logs.
Claims	Incorrect denial, suppression or prioritisation; fraud false positives; synthetic evidence; deepfakes; hallucinated summaries; automation bias.	D1, D2, D5, D6, D7	Decision trace; evidence integrity checks; manual-review criteria; appeal records; outcome monitoring.
Policyholder and consumer harm	Unfair treatment, inadequate notice or explanation, inaccessible appeal, misleading communications and vulnerable-customer impacts.	D5, D6, D8	Impact assessment; notices; explanation tests; accessibility review; complaint and appeal metrics.
Data	Weak provenance, inaccurate broker data, sensitive health or financial data misuse, retention failures, proxy variables and training-data rights.	D1, D4, D5, D6, D8	Data register; lineage; consent/purpose basis; quality controls; access logs; deletion evidence.

Risk domain	Description	GAISST domains	Indicative evidence
Cybersecurity and adversarial	Prompt injection, poisoning, evasion, extraction, API abuse, credential compromise, insider misuse and vendor compromise.	D1, D2, D3, D4	Threat model; red-team results; IAM evidence; runtime telemetry; incident exercises; DLP records.
Model and decision	Poor specification, uncontrolled updates, non-reproducibility, overfitting, weak thresholds, rules/model interaction and unsupported overrides.	D1, D6, D8	Independent validation; model cards; version control; approval record; threshold rationale; rollback test.
Third-party ecosystem	Opaque vendor models, weak audit rights, subprocessor exposure, update risk, concentration and exit limitations.	D4, D6, D8	Due diligence; contract clauses; audit reports; dependency map; exit test; concentration assessment.
Prudential and systemic	Common model/vendor concentration, tail-risk underestimation, catastrophe-model dependency, cyber accumulation and claims surge.	D4, D6, D8	Scenario analysis; accumulation map; stress test; board risk appetite; contingency and reinsurance review.

4.1 Underwriting and pricing

Unstable or discriminatory risk classification, proxy effects, unapproved variables, model drift, historical-data bias and correlated portfolio errors. Affected functions should assess business, policyholder, security, compliance and prudential consequences separately. Typical impacts include coverage denial, unaffordable pricing, adverse selection, underpricing, complaints and remediation.

Control focus: D1, D6, D8. Evidence: Model inventory; feature approval; actuarial validation; subgroup analysis; drift reports; override logs.

4.2 Claims

Incorrect denial, suppression or prioritisation; fraud false positives; synthetic evidence; deepfakes; hallucinated summaries; automation bias. Affected functions should assess business, policyholder, security, compliance and prudential consequences separately. Typical impacts include delayed or denied benefits, leakage, litigation, vulnerable-customer harm and operational backlog.

Control focus: D1, D2, D5, D6, D7. Evidence: Decision trace; evidence integrity checks; manual-review criteria; appeal records; outcome monitoring.

4.3 Policyholder and consumer harm

Unfair treatment, inadequate notice or explanation, inaccessible appeal, misleading communications and vulnerable-customer impacts. Affected functions should assess business, policyholder, security, compliance and prudential consequences separately. Typical impacts include financial exclusion, loss of trust, remediation, supervisory action and reputational damage.

Control focus: D5, D6, D8. Evidence: Impact assessment; notices; explanation tests; accessibility review; complaint and appeal metrics.

4.4 Data

Weak provenance, inaccurate broker data, sensitive health or financial data misuse, retention failures, proxy variables and training-data rights. Affected functions should assess business, policyholder, security, compliance and prudential consequences separately. Typical impacts include privacy harm, unreliable decisions, legal exposure, model contamination and rework.

Control focus: D1, D4, D5, D6, D8. Evidence: Data register; lineage; consent/purpose basis; quality controls; access logs; deletion evidence.

4.5 Cybersecurity and adversarial

Prompt injection, poisoning, evasion, extraction, API abuse, credential compromise, insider misuse and vendor compromise. Affected functions should assess business, policyholder, security, compliance and prudential consequences separately. Typical impacts include fraud losses, data exfiltration, service disruption, manipulated decisions and ip loss.

Control focus: D1, D2, D3, D4. Evidence: Threat model; red-team results; IAM evidence; runtime telemetry; incident exercises; DLP records.

4.6 Model and decision

Poor specification, uncontrolled updates, non-reproducibility, overfitting, weak thresholds, rules/model interaction and unsupported overrides. Affected functions should assess business, policyholder, security, compliance and prudential consequences separately. Typical impacts include systematic pricing or claims errors, reserving distortion and audit failure.

Control focus: D1, D6, D8. Evidence: Independent validation; model cards; version control; approval record; threshold rationale; rollback test.

4.7 Third-party ecosystem

Opaque vendor models, weak audit rights, subprocessor exposure, update risk, concentration and exit limitations. Affected functions should assess business, policyholder, security, compliance and prudential consequences separately. Typical impacts include correlated failures, service outage, lock-in, evidentiary gaps and regulatory access problems.

Control focus: D4, D6, D8. Evidence: Due diligence; contract clauses; audit reports; dependency map; exit test; concentration assessment.

4.8 Prudential and systemic

Common model/vendor concentration, tail-risk underestimation, catastrophe-model dependency, cyber accumulation and claims surge. Affected functions should assess business, policyholder, security, compliance and prudential consequences separately. Typical impacts include capital, reserving, liquidity, reinsurance dispute and sector-wide operational impacts.

Control focus: D4, D6, D8. Evidence: Scenario analysis; accumulation map; stress test; board risk appetite; contingency and reinsurance review.

Limitations: risk likelihood and materiality are entity-, product-, portfolio- and jurisdiction-specific; examples are not exhaustive safe harbours.

5. Insurance AI system classification

Class	Decision / impact profile	Examples	Minimum governance response
Low-impact assistive	No direct adverse decision; reversible; low sensitivity	Drafting internal text or summarising non-sensitive material	Standard access, logging, output review and acceptable-use controls.

Class	Decision / impact profile	Examples	Minimum governance response
Material operational	Supports important workflow or material volumes	Claims document classification; broker submission extraction	Formal owner, validation, monitoring, fallback and change control.
High-impact policyholder decision	Influences eligibility, price, coverage, claim or adverse action	Life underwriting score; claim denial recommendation	Independent validation, fairness testing, meaningful human review, explanation and appeal.
Prudentially significant	Can materially affect reserves, capital, accumulation or reinsurance	Catastrophe aggregation or reserving support	Actuarial and risk validation, stress testing, concentration analysis and board oversight.
Critical claims or underwriting	Failure creates immediate large-scale customer or operational harm	Automated claims triage during catastrophe event	Resilience, manual fallback, surge testing, incident playbook and executive escalation.

Classification is an implementation aid. It does not replace GAISST applicability rules or any legal definition of high-risk AI. Escalate classification where one factor is extreme even if other factors are moderate.

6. Control interpretation by insurance function

Function	Key risks	GAISST domain	Implementation	Human oversight
Underwriting	Proxy discrimination, drift, manipulation, weak explainability	D1, D6, D8	Feature governance, actuarial validation, adverse-decision review	Mandatory for adverse or borderline outcomes
Pricing	Unapproved variables, unfair differentiation, optimisation instability	D1, D6, D8	Rate governance, subgroup and stability tests, filing alignment	Required for material exceptions and adverse outcomes
Claims	Deepfakes, false negatives, hallucinations, automation bias	D1, D2, D5, D7	Evidence integrity, confidence thresholds, escalation and appeals	Mandatory for denial, vulnerability and high value
Fraud detection	False positives, evasion, bias, feedback loops	D1, D2, D7	Adaptive testing, investigator review, outcome feedback controls	Investigator confirmation before adverse action
Reserving	Tail error, regime shift, opaque assumptions	D1, D6, D8	Actuarial ownership, scenario testing, reconciliation	Actuarial sign-off
Catastrophe modelling	Vendor concentration, climate shift, tail underestimation	D1, D4, D6, D8	Model plurality, version governance, stress and dependency analysis	Expert review for material capital/reinsurance use
Policy administration	Record corruption, unauthorised changes, privacy leakage	D2, D3, D6	Transaction controls, segregation, reconciliation, rollback	Approval for policy-changing actions
Distribution and intermediary support	Mis-selling, unsuitable recommendations, prompt leakage	D3, D5, D6	Approved knowledge, suitability checks, constrained tools	Human agent accountability retained

Function	Key risks	GAISST domain	Implementation	Human oversight
Customer service	Hallucinated coverage, misleading advice, accessibility failure	D2, D5, D6	Grounding, disclaimers, escalation, transcript retention	Human handoff for disputes and material decisions
Reinsurance	Data mismatch, concentration, reporting error	D1, D4, D6, D8	Data reconciliation, dependency mapping, expert validation	Specialist review for treaty/capital decisions
Cyber-insurance	Gaming, stale signals, correlated vendor risk	D1, D4, D6, D8	Signal provenance, adversarial testing, accumulation scenarios	Underwriter review
Health and life insurance	Sensitive data, proxy discrimination, irreversible harm	D1, D5, D6, D8	Strict data governance, clinical/actuarial input, contestability	Mandatory qualified human review
Telematics and IoT	Sensor spoofing, surveillance, physical safety	D1, D2, D5, D9	Device security, consent, calibration, fallback	Review where data materially changes price/coverage
Generative AI	Prompt injection, hallucination, leakage, unsafe tools	D2, D3, D5	RAG controls, tool allowlists, output validation, DLP	Required for external or decision-relevant output
Internal corporate use	Confidentiality leakage, shadow AI, inaccurate work product	D2, D4, D6	Approved tools, data classification, usage logging, training	Reviewer accountable for work product

7. Governance and accountability

Accountability remains with the insurance entity and accountable business owner. Outsourcing changes evidence sources, not accountability. Governance should integrate board risk appetite, executive ownership, actuarial and model validation, CISO controls, compliance/conduct review, privacy, procurement, resilience and internal audit.

Decision	Accountable	Consulted / responsible	Escalation
System approval	Business owner	AI governance, CISO, model risk, compliance	Board/committee for material systems
Risk classification	AI governance	Business, model risk, legal, actuarial	CRO
Data approval	Data owner	Privacy, security, actuarial/model risk	AI governance
Model validation	Independent validation / actuarial	CISO, business owner	CRO / validation committee
Security testing	CISO	Engineering, vendor, red team	AI governance
Deployment approval	Business owner	CISO, model risk, compliance, operations	Executive committee for critical systems
Monitoring	System owner	Model risk, claims/underwriting, security	CRO

Decision	Accountable	Consulted / responsible	Escalation
Incident escalation	Incident commander	Legal, compliance, business, privacy	Executive / board by severity
Customer notification	Compliance / legal	Claims or product owner	Accountable executive
Model retirement	System owner	Records, vendor, security, business continuity	AI governance

8. Human oversight

Meaningful oversight exists only where the reviewer acts before harm becomes irreversible, has relevant competence and contextual information, can disagree without penalty, can access source evidence, and has authority and time to intervene. A nominal approval click, post-event sampling or blind reliance on a confidence score is ceremonial oversight.

Trigger	Oversight model	Reviewer	Required record
Coverage denial or material exclusion	Human-in-command	Qualified underwriter / claims decision-maker	Sources, model output, rationale, alternative considered, approval
High-value or vulnerable claimant	Mandatory manual review	Senior claims professional	Vulnerability and escalation record
Fraud flag causing adverse action	Human-in-the-loop	Trained investigator	Evidence reviewed, disposition, feedback label
Medical/health-related decision	Qualified human review	Appropriate clinical/actuarial/claims professional	Competence, sources, rationale, appeal route
Agentic action affecting policy/claim	Dual approval or constrained authority	Business owner and control function	Tool calls, authority, approvals, rollback

9. Data governance

The data register should identify source, owner, legal basis or permitted purpose, provenance, sensitivity, quality controls, known limitations, protected-characteristic or proxy risk, update frequency, retention and downstream use. Broker-submitted and third-party enrichment data require correction and challenge mechanisms.

Data ID	Source	Use	Sensitivity	Proxy risk	Owner
DATA-001	Policy administration system	Underwriting history	Confidential personal	Geography and occupation	Data owner
DATA-002	Telematics provider	Behaviour-based pricing	Location / behavioural	Socioeconomic and disability	Product owner

10. Security architecture

Reference architecture: identity-aware channels feed an API gateway and policy enforcement layer; input validation, malware/media integrity and prompt-injection controls precede model or decision services; retrieval is isolated by tenant, purpose and access; tools are allowlisted with least privilege and transaction limits; outputs pass confidence, policy, safety and data-loss checks; material decisions are recorded with model, prompt, data, rule, reviewer and outcome identifiers; telemetry feeds security, model-risk and business-outcome monitoring; tested manual fallback supports critical operations.

- Separate development, validation and production environments.
- Protect model registries, prompts, retrieval corpora, rules, thresholds and secrets as controlled assets.
- Use strong identity, privileged-access management, segmentation, encryption, key management, API security and immutable logging.
- Test backup, rollback, provider outage, manual fallback and catastrophe surge procedures.
- Detect unauthorised AI use and prevent sensitive data exfiltration through DLP and approved-tool controls.

11. Agentic AI in Insurance Workflows

Multi-agent insurance workflows can connect broker, insurer, reinsurer, claims, data and payment services. The decision boundary must include the full delegation chain rather than assessing each agent in isolation.

- Agent-to-agent authentication failure between broker, insurer, reinsurer and service-provider agents
- Compounding error in autonomous underwriting, claims adjustment or reinsurance placement
- Orphaned execution loops that repeatedly request evidence, change reserves, issue communications or invoke payment tools
- Delegation-chain ambiguity, excessive tool authority and inability to reconstruct which agent made or executed a decision

Minimum implementation controls:

- Unique workload identity and mutual authentication for each agent
- Explicit delegation scope, transaction limits, segregation of duties and expiry
- Loop, cost, time and action-count circuit breakers
- Human approval before binding coverage, denying claims, changing reserves, placing reinsurance or initiating payment
- Complete agent-to-agent message, tool-call, model-version and approval trace

12. Model validation and testing

Test family	Insurance focus	Evidence
Independent validation	Fitness for use, assumptions, data, reproducibility and limitations	Validation report and issue closure

Test family	Insurance focus	Evidence
Actuarial validation	Rate, reserve, mortality, morbidity, severity or frequency relevance	Actuarial review and sign-off
Security / adversarial	Poisoning, evasion, extraction, prompt injection, abuse and tool misuse	Threat model, test cases and remediation
Fairness / outcome	Subgroups, proxies, adverse outcomes and vulnerable customers	Defined metrics, rationale, limitations and actions
Stress / scenario	Regime shift, catastrophe surge, vendor outage and tail events	Scenario design, results and decision record
Operational acceptance	Latency, capacity, integration, rollback and manual fallback	UAT, resilience and fallback test

Universal numerical thresholds are intentionally absent. Thresholds should be justified against use, harm severity, legal requirements, portfolio characteristics, baseline performance and escalation capacity.

13. Monitoring and change management

Metric / change	Frequency	Owner	Escalation approach
Performance and calibration	Risk-based; at least per approved monitoring plan	Model owner	Deviation from approved tolerance or unexplained trend
Fairness and adverse outcomes	Aligned to decision volume and harm	Compliance / model risk	Material subgroup disparity or deterioration
Overrides, appeals and complaints	Monthly or more frequently for high-impact uses	Business owner	Concentration, reversal trend or vulnerable-customer signal
Security events and abuse	Continuous where technically feasible	CISO	Confirmed compromise, injection, exfiltration or privilege misuse
Vendor/model update	Before acceptance and after material change	Vendor owner	Unassessed change, regression or evidence gap
Data/feature change	Before production	Data/model owner	Lineage, rights, quality or proxy-risk change

14. Incident management

Severity	Policyholder / financial impact	Examples	Escalation	Evidence preservation
S1 Critical	Widespread or severe harm; material financial/prudential impact	Systemic denial/pricing error, major breach, critical claims outage	Immediate executive, legal, regulator-assessment and board escalation	Freeze versions, logs, data, rules, prompts, communications and decisions

Severity	Policyholder / financial impact	Examples	Escalation	Evidence preservation
S2 High	Material cohort or product impact	Discriminatory outcome, vendor compromise, material claim error	Incident command and senior control functions	Preserve decision and vendor evidence
S3 Moderate	Limited reversible impact	Local drift, contained data exposure, repeat hallucination	Owner and control-function remediation	Retain samples, root cause and closure
S4 Low	No material external impact	Near miss or minor control deviation	Normal issue management	Document lesson and control update

Notification timeframes are jurisdiction-specific and intentionally not stated as universal rules. The legal/regulatory assessment should identify applicable clocks and triggering facts.

15. Third-party and supply-chain management

- Assess provider security, model documentation, data rights, validation, subcontractors, incident history, resilience, geographic processing and regulatory access before contract.
- Contract for change notice, audit/access rights, incident notification, evidence retention, secure deletion, continuity, portability, exit support and material subprocessor control.
- Map fourth parties and concentration across cloud, foundation models, catastrophe models, data and claims services.
- Test exit and fallback; a contractual right without operational feasibility is weak evidence.

16. Policyholder transparency, explanation and contestability

Notices and explanations should match the decision, audience and harm. They should identify the material basis of the outcome, meaningful factors, applicable limitations, correction and appeal routes, and whether qualified human review is available. Trade-secret or security constraints may limit technical detail but do not remove the need for a meaningful, lawful explanation.

17. Assurance and evidence catalogue

Evidence	Category	Owner	Reliability
Approved AI policy, risk appetite and committee decisions	Governance	Board / CRO	Approver authority, approval date, challenge record and exception completeness.
System classification and impact assessment	Risk	AI governance	Prefer T1 operating evidence; corroborate with T2 independent review and T3 simulation. T4 evidence is supplementary only.
Lineage, rights, quality and proxy analysis	Data	Data owner	Source quality, rights, lineage completeness, validation status and operating period.
Model card, validation, limitations and version history	Model	Model risk	Validation independence, test coverage, version linkage and documented limitations.

Evidence	Category	Owner	Reliability
Threat model, tests, IAM and runtime telemetry	Security	CISO	Testing currency, scope, independence and linkage to the production configuration.
Approval, release, configuration and rollback evidence	Deployment	System owner	Prefer T1 operating evidence; corroborate with T2 independent review and T3 simulation. T4 evidence is supplementary only.
Performance, outcomes, drift, complaints and overrides	Monitoring	Business / model owner	Prefer T1 operating evidence; corroborate with T2 independent review and T3 simulation. T4 evidence is supplementary only.
Timeline, containment, impact, notification and lessons	Incident	Incident commander	Completeness, timeliness, preserved evidence, root cause and remediation closure.
Due diligence, contract, assurance and change records	Vendor	Third-party risk	Prefer T1 operating evidence; corroborate with T2 independent review and T3 simulation. T4 evidence is supplementary only.
Reviewer competence, intervention and override records	Oversight	Business owner	Reviewer competence, timing, authority and record completeness.
Notices, explanations, appeals and accessibility tests	Policyholder	Compliance	Prefer T1 operating evidence; corroborate with T2 independent review and T3 simulation. T4 evidence is supplementary only.
Testing, findings, remediation and limitations	Audit	Internal audit	Prefer T1 operating evidence; corroborate with T2 independent review and T3 simulation. T4 evidence is supplementary only.

17.1 Evidence quality tiers

- T1: Primary verified operating evidence: automated logs, immutable decision/claims trails, access records, telemetry and approved source records.
- T2: Independent or second-line verified evidence: MRM validation, Appointed Actuary sign-off, internal audit, privacy, legal or security review.
- T3: Controlled simulation evidence: adversarial testing, deepfake-claim injection, stress testing, catastrophe surge, failover and out-of-distribution exercises.
- T4: Anecdotal or unverified context: vendor benchmarks, surveys and generic white papers; insufficient alone for high-impact underwriting, pricing or claims conclusions.

18. Insurance implementation profiles

Profile	Operating model	Priority controls	Constraints
Large composite insurer	Multiple lines, jurisdictions and legacy platforms	Portfolio inventory, federated governance, concentration and shared-service controls	Fragmented ownership and evidence

Profile	Operating model	Priority controls	Constraints
Life insurer	Long-duration products and sensitive mortality/health data	Underwriting fairness, actuarial validation, long-tail drift and contestability	Sparse outcomes and legacy assumptions
Property and casualty insurer	High-volume claims, images, catastrophe and telematics	Claims integrity, surge resilience, computer vision and accumulation	Catastrophe regime shift and vendor concentration
Health insurer	Sensitive data and high-impact benefit decisions	Purpose limitation, qualified review, fairness, explanation and appeal	Clinical context and jurisdictional complexity
Reinsurer	Portfolio aggregation, catastrophe and treaty analytics	Model plurality, concentration, assumptions and stress testing	Opaque cedant data and common-model dependency
Insurtech / digital insurer	Cloud-native, vendor-heavy and rapid change	Secure SDLC, third-party controls, monitoring and evidence automation	Resource constraints and dependency lock-in

19. Practical implementation roadmap

Phase	Indicative timeframe	Objective	Activities	Exit evidence
Phase 0	0-30 days	Mobilisation and scope	Sponsor, owner, boundaries, sources, approval gates	Charter and scoped plan
Phase 1	30-60 days	Inventory and classification	Systems, models, data, vendors, decision authority and impact	Approved inventory and classes
Phase 2	60-100 days	Risk and control mapping	Map 59 controls, obligations, evidence and gaps	Applicable-Control Baseline and gap register
Phase 3	100-180 days	Remediation and governance	Implement priority technical, process and contractual controls	Closed critical gaps and operating evidence
Phase 4	180-240 days	Validation and assurance	Independent validation, red team, fairness, resilience and audit	Assurance report with limitations
Phase 5	Ongoing	Operational monitoring	Outcomes, drift, incidents, complaints, vendor and change monitoring	Dashboards and issue records
Phase 6	6-12 months	Continuous improvement	Thematic review, concentration, lessons and maturity	Improvement plan and readiness decision

20. Worked scenarios

20.1 AI-assisted life insurance underwriting

Element	Implementation interpretation
Context	Applicant data, medical and financial indicators support risk classification.
Threats and failure modes	Proxy discrimination; sensitive-data misuse; drift; opaque adverse decisions.
Control interpretation	Independent actuarial/model validation; restricted features; subgroup testing; meaningful underwriter review; explanation and appeal evidence.
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.
Notably absent	No assumption that model approval establishes lawfulness or actuarial adequacy in every jurisdiction.

20.2 Automated motor claims triage

Element	Implementation interpretation
Context	Images and claim data prioritise cases and estimate complexity.
Threats and failure modes	Adversarial images; deepfakes; misclassification; catastrophe surge failure.
Control interpretation	Media integrity checks; confidence thresholds; surge testing; high-value/vulnerable escalation; manual fallback.
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.
Notably absent	No autonomous denial based solely on triage output.

20.3 Fraud detection using network analytics

Element	Implementation interpretation
Context	Graph analytics identify suspicious claimant, provider or broker relationships.
Threats and failure modes	False positives; feedback loops; evasion; group bias.
Control interpretation	Investigator confirmation; outcome feedback; precision/recall and subgroup monitoring; feature provenance.

Element	Implementation interpretation
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.
Notably absent	A fraud flag is not proof of fraud.

20.4 Property damage assessment using computer vision

Element	Implementation interpretation
Context	Images estimate damage severity or repair scope.
Threats and failure modes	Image manipulation; poor generalisation; geographic bias; hidden damage.
Control interpretation	Capture controls; tamper detection; uncertainty reporting; adjuster review for material settlements.
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.
Notably absent	No claim that image-only assessment is universally sufficient.

20.5 Health insurance decision support

Element	Implementation interpretation
Context	AI supports benefit, risk or utilisation decisions.
Threats and failure modes	Sensitive health data; discrimination; clinical context loss; irreversible harm.
Control interpretation	Strict purpose limitation; qualified human review; explanation; appeal; clinical and actuarial validation.
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.
Notably absent	No medical advice or clinical approval is provided by this guide.

20.6 Dynamic or behaviour-based pricing

Element	Implementation interpretation
Context	Telematics or behavioural signals influence premium.
Threats and failure modes	Surveillance, consent limits, sensor spoofing, unfair differentiation.
Control interpretation	Signal provenance; device security; proxy analysis; customer notice; correction process.
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.
Notably absent	No endorsement of legality or fairness of any variable.

20.7 Generative AI claims assistant

Element	Implementation interpretation
Context	LLM summarises evidence and drafts communications.
Threats and failure modes	Hallucination; prompt injection; confidential-data leakage; misleading coverage statements.
Control interpretation	Grounded retrieval; citation to source records; prohibited-action controls; human approval; transcript logging.
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.
Notably absent	No direct claim denial or binding coverage interpretation without qualified review.

20.8 Cyber-insurance risk scoring

Element	Implementation interpretation
Context	External security telemetry and questionnaires support underwriting.
Threats and failure modes	Gaming; stale data; concentration; weak causal assumptions.
Control interpretation	Source assurance; freshness controls; adversarial testing; underwriter override; portfolio accumulation analysis.

Element	Implementation interpretation
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.
Notably absent	No representation that a score predicts breach with certainty.

20.9 Catastrophe modelling dependency

Element	Implementation interpretation
Context	Vendor models inform aggregation, reinsurance and capital decisions.
Threats and failure modes	Vendor concentration; model change; tail underestimation; correlated market use.
Control interpretation	Multi-model challenge; change impact; stress/scenario testing; exit and continuity plan.
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.
Notably absent	No assumption that vendor validation transfers accountability.

20.10 Third-party foundation model in customer service

Element	Implementation interpretation
Context	External model answers policyholder questions using approved knowledge.
Threats and failure modes	Provider update; retention; prompt injection; hallucinated coverage; outage.
Control interpretation	Contract controls; version monitoring; RAG isolation; evaluation; escalation; fallback.
Required evidence	System classification; data lineage; threat model; validation; approval; monitoring; oversight and incident records proportionate to risk.
Monitoring	Performance, outcomes, overrides, complaints, incidents, vendor/model changes and drift.
Incident triggers	Material decision error, security compromise, unexplained disparity, loss of traceability, vendor change or failure of manual fallback.
Residual risks	Data limitations, regime shift, human error, unknown attacks and jurisdiction-specific obligations remain.

Element	Implementation interpretation
Notably absent	No provider assurance treated as substitute for insurer testing.

21. GAISST Control Mapping

Control IDs and titles are those in GAISST-NOR-004. Insurance interpretations are informative and do not replace the control requirement; owners are indicative.

Control	Title	Insurance interpretation	Indicative owner
D1-CTL-01	Dataset Provenance & Poisoning Prevention	Verify provenance, rights and poisoning resistance of underwriting, claims, broker and third-party enrichment data before training or recalibration.	Underwriting; pricing; claims; fraud; reserving; catastrophe modelling
D1-CTL-02	Model Extraction Resistance	Rate-limit and monitor scoring APIs so brokers, claimants or competitors cannot reconstruct pricing, fraud or underwriting models.	Underwriting; pricing; claims; fraud; reserving; catastrophe modelling
D1-CTL-03	Behavioral Drift Detection	Monitor pricing, claims and fraud models for behavioural drift against approved tolerances, including after catastrophe events and economic shifts.	Underwriting; pricing; claims; fraud; reserving; catastrophe modelling
D1-CTL-04	Federated Learning Poisoning Prevention	Where models learn across insurers, partners or devices, protect federated updates against poisoning by a single contributor.	Underwriting; pricing; claims; fraud; reserving; catastrophe modelling
D1-CTL-05	Embedding Space Robustness	Test robustness of embeddings used for claims similarity, document retrieval and fraud network analysis against crafted inputs.	Underwriting; pricing; claims; fraud; reserving; catastrophe modelling
D1-CTL-06	Post-Quantum Model Signing & Crypto Hardening	Sign approved pricing, reserving and claims models so production versions are verifiable over long policy and claims lifetimes.	Underwriting; pricing; claims; fraud; reserving; catastrophe modelling
D1-CTL-07	LoRA/Adapter Integrity Verification	Verify integrity of fine-tuning adapters applied to claims, underwriting or customer-service models before deployment.	Underwriting; pricing; claims; fraud; reserving; catastrophe modelling
D1-CTL-08	Model Merge Attack Detection	Detect hidden behaviour introduced when vendor or open models are merged into insurer models.	Underwriting; pricing; claims; fraud; reserving; catastrophe modelling

Control	Title	Insurance interpretation	Indicative owner
D1-CTL-09	Quantization Backdoor Screening	Screen quantised or compressed models, including on telematics devices, for backdoors introduced during optimisation.	Underwriting; pricing; claims; fraud; reserving; catastrophe modelling
D2-CTL-01	Direct Prompt Injection Prevention	Defend customer, broker and agent chat channels against direct prompt injection aimed at coverage, claims or pricing outcomes.	Claims platforms; customer portals; APIs; GenAI assistants; decision engines
D2-CTL-02	Indirect Prompt Injection Prevention	Treat claims documents, emails, medical reports and uploaded images as untrusted inputs that may carry indirect injection.	Claims platforms; customer portals; APIs; GenAI assistants; decision engines
D2-CTL-03	Jailbreak Resistance Testing	Test generative claims and customer-service assistants for jailbreaks that bypass coverage, disclosure or escalation rules.	Claims platforms; customer portals; APIs; GenAI assistants; decision engines
D2-CTL-04	Multi-Modal Injection Defense	Defend image-based damage assessment and document extraction against multimodal injection and adversarial images.	Claims platforms; customer portals; APIs; GenAI assistants; decision engines
D2-CTL-05	Function Call/Tool Call Injection Prevention	Prevent injected instructions from triggering payment, policy-change, reserve or notification tools.	Claims platforms; customer portals; APIs; GenAI assistants; decision engines
D2-CTL-06	Cross-Context Hijacking Mitigation	Isolate sessions so one claimant's or broker's context cannot influence another policyholder's outcome.	Claims platforms; customer portals; APIs; GenAI assistants; decision engines
D3-CTL-01	Least Agency Enforcement	Grant claims, underwriting and service agents only the authority required; binding cover, denial, reserving and payment stay outside autonomous scope without approval.	Agentic claims; workflow automation; broker support; autonomous triage
D3-CTL-02	Inter-Agent Communication Security	Authenticate and protect messages between broker, insurer, reinsurer, TPA and service-provider agents.	Agentic claims; workflow automation; broker support; autonomous triage
D3-CTL-03	Agentic Prompt Chaining Detection	Detect chained agent actions that cumulatively exceed claims, payment or reserve limits.	Agentic claims; workflow automation; broker support; autonomous triage

Control	Title	Insurance interpretation	Indicative owner
D3-CTL-04	Embodied AI Safety Controls	Where drones, inspection robots or vehicles act on AI output, apply embodied-safety controls alongside D9.	Agentic claims; workflow automation; broker support; autonomous triage
D3-CTL-05	Multi-Agent Trust Chain Attestation	Attest the trust chain across multi-agent placement, claims and recovery workflows so the acting agent is always identifiable.	Agentic claims; workflow automation; broker support; autonomous triage
D3-CTL-06	Persistent Memory Exfiltration Prevention	Prevent agents from exfiltrating claimant, medical or policy data retained in persistent memory.	Agentic claims; workflow automation; broker support; autonomous triage
D3-CTL-07	Secure Memory Lifecycle Management	Set retention and deletion for agent memory consistent with claims records, privacy and legal-hold requirements.	Agentic claims; workflow automation; broker support; autonomous triage
D4-CTL-01	AI Bill Of Materials (AI BOM) Maintenance	Maintain an AI bill of materials covering pricing, claims, catastrophe, fraud and foundation-model components and their versions.	All outsourced models, data, SaaS, cloud, TPA and reinsurer analytics
D4-CTL-02	Model File & Artifact Scanning	Scan vendor and open model files and artefacts before use in insurance systems.	All outsourced models, data, SaaS, cloud, TPA and reinsurer analytics
D4-CTL-03	Model Hub & Registry Vetting	Vet model hubs and registries used by data-science and insurtech teams.	All outsourced models, data, SaaS, cloud, TPA and reinsurer analytics
D4-CTL-04	MCP Server Behavioral Monitoring	Monitor MCP servers connecting assistants to policy, claims and payment systems for unexpected behaviour.	All outsourced models, data, SaaS, cloud, TPA and reinsurer analytics
D4-CTL-05	Third-Party AI Api Security Assessment	Assess security of third-party AI APIs, including catastrophe-model, data-enrichment and foundation-model providers.	All outsourced models, data, SaaS, cloud, TPA and reinsurer analytics

Control	Title	Insurance interpretation	Indicative owner
D4-CTL-06	Shadow AI Discovery & Governance	Discover unsanctioned AI used by underwriters, claims handlers, actuaries and intermediaries, and bring it under governance.	All outsourced models, data, SaaS, cloud, TPA and reinsurer analytics
D4-CTL-07	AI Software Composition Analysis (SCA)	Run software composition analysis on AI libraries in pricing, reserving and claims pipelines.	All outsourced models, data, SaaS, cloud, TPA and reinsurer analytics
D5-CTL-01	Harmful Content Blocking	Block, constrain and escalate hallucinated, misleading, unsafe or unauthorised outputs in policyholder communications, claims summaries, underwriting support and agent-facing content.	Customer communications; claim summaries; policy documents; recommendations
D5-CTL-02	PII Leakage Prevention	Prevent leakage of medical records, government identifiers, financial identifiers, claims data and telemetry through prompts, retrieval, generated outputs, logs and model-provider interfaces; apply redaction and output inspection before disclosure.	Customer communications; claim summaries; policy documents; recommendations
D5-CTL-03	Copyright Detection	Verify training, retrieval and reference-content provenance for generative underwriting and claims assistants; detect unauthorised use of proprietary actuarial models, policy wordings, manuals and third-party copyrighted material.	Customer communications; claim summaries; policy documents; recommendations
D5-CTL-04	AI Watermarking Robustness	Apply provenance marking or watermarking where appropriate to insurer-generated synthetic media and communications, and operate detection and escalation pipelines for inbound deepfake or synthetically altered claims evidence.	Customer communications; claim summaries; policy documents; recommendations
D5-CTL-05	Privacy-By-Design Verification	Embed privacy requirements in insurance AI design, including purpose limitation, data minimisation, role-based access, retention, privacy testing and review of sensitive health, biometric, financial and telematics data.	Customer communications; claim summaries; policy documents; recommendations

Control	Title	Insurance interpretation	Indicative owner
D5-CTL-06	Privacy-Preserving ML Validation	Validate privacy-preserving techniques used in distributed or sensitive insurance analytics, including re-identification, inference and utility testing; do not assume anonymisation, federation or synthetic data eliminates privacy risk.	Customer communications; claim summaries; policy documents; recommendations
D6-CTL-01	Human-In-The-Loop For High-Risk Actions	Require competent human approval before denying claims, declining cover, changing reserves or initiating payment.	Enterprise governance; model risk; actuarial; security; audit
D6-CTL-02	Audit Trail Completeness	Keep audit trails sufficient to reconstruct underwriting, pricing and claims decisions for appeals, examinations and litigation.	Enterprise governance; model risk; actuarial; security; audit
D6-CTL-03	AI Model Card Completeness	Maintain model cards stating purpose, data, limitations, validation and approved use for each material insurance model.	Enterprise governance; model risk; actuarial; security; audit
D6-CTL-04	AI Incident Response Readiness	Prepare AI incident playbooks covering mispricing, wrongful denial, data exposure and catastrophe-surge failure.	Enterprise governance; model risk; actuarial; security; audit
D6-CTL-05	Model Deprecation & Decommissioning	Retire models with records preserved for the policy and claims lifetime and with run-off decisions documented.	Enterprise governance; model risk; actuarial; security; audit
D6-CTL-06	Third-Party AI Vendor Governance	Govern insurtech, TPA, catastrophe-model and foundation-model vendors through contracts, assurance and exit plans.	Enterprise governance; model risk; actuarial; security; audit
D6-CTL-07	AI Resilience & Business Continuity	Maintain manual fallback and continuity for claims and underwriting, tested for catastrophe surge.	Enterprise governance; model risk; actuarial; security; audit
D7-CTL-H01	Ai-Generated Phishing Simulation	Run AI-generated phishing simulations for claims, finance and broker-facing staff.	Security Awareness / Fraud / CISO
D7-CTL-H02	Deepfake Detection Training	Train claims handlers and investigators to recognise deepfake images, voices and documents submitted as evidence.	Security Awareness / Fraud / CISO
D7-CTL-H03	Out-Of-Band Authentication	Use out-of-band verification before changing payee details, releasing large settlements or acting on voice instructions.	Security Awareness / Fraud / CISO

Control	Title	Insurance interpretation	Indicative owner
D7-CTL-H04	AI Social Engineering IR	Include AI-enabled social engineering, such as impersonated policyholders or brokers, in incident response.	Security Awareness / Fraud / CISO
D7-CTL-H05	AI-Enhanced External Attack Defense	Defend against AI-enhanced external attacks on portals, call centres and broker channels.	Security Awareness / Fraud / CISO
D8-CTL-01	EU AI Act Risk Tier Mapping	Map insurance AI uses to EU AI Act risk tiers where applicable — for example, life and health risk assessment and pricing.	Legal / Compliance
D8-CTL-02	ISO 42001 Gap Analysis	Perform an ISO/IEC 42001 gap analysis where the insurer operates or plans an AI management system.	Legal / Compliance
D8-CTL-03	GPAI Technical Documentation Verification	Verify technical documentation for general-purpose AI models used in insurance workflows.	Legal / Compliance
D8-CTL-04	DORA ICT Incident Reporting (Financial Sector)	Integrate AI incidents into DORA ICT incident classification and reporting for in-scope EU insurers.	Legal / Compliance
D8-CTL-05	NIST SP 800-218A Compliance Check	Check secure-development practice for insurer-built AI against NIST SP 800-218A where adopted.	Legal / Compliance
D9-CTL-01	Physical Harm Boundary Enforcement	Where insurers deploy drones, inspection robots or loss-prevention devices that act physically, enforce harm boundaries.	Engineering / Safety / Operations
D9-CTL-02	Safe State And Graceful Degradation	Define safe states for insurer-operated devices, such as telematics-triggered interventions, when AI fails.	Engineering / Safety / Operations
D9-CTL-03	Human Override And Emergency Stop	Provide human override and emergency stop for insurer-operated physical AI.	Engineering / Safety / Operations
D9-CTL-04	Cyber-Physical Attack Detection	Detect cyber-physical attacks on insurer-operated sensors and devices, including telematics spoofing.	Engineering / Safety / Operations
D9-CTL-05	Physical Environment Integrity Monitoring	Monitor integrity of the physical environment and sensor inputs used in inspection and loss prevention.	Engineering / Safety / Operations
D9-CTL-06	Actuator Command Verification	Verify actuator commands issued by insurer-operated physical AI.	Engineering / Safety / Operations
D9-CTL-07	Physical Incident Evidence Preservation	Preserve physical-incident evidence from insurer-operated devices for claims, liability and regulatory purposes.	Engineering / Safety / Operations

22. Regulatory and Standards Landscape

External sources provide context, not automatic obligations or equivalence. Implementing organisations should verify current status, jurisdiction, adoption, effective dates, amendments and supervisory interpretation before reliance.

ID	Instrument	Issuer / date	Status / limitation
SRC-01	Application Paper on the Supervision of Artificial Intelligence	IAIS, July 2025	Non-binding application paper; jurisdictional implementation varies
SRC-02	Opinion on Artificial Intelligence Governance and Risk Management (EIOPA-BoS-25-360)	EIOPA, 6 August 2025	Addressed to national supervisors; excludes AI systems that are high-risk or prohibited under the EU AI Act; applicability must be assessed
SRC-03	Model Bulletin: Use of Artificial Intelligence Systems by Insurers	NAIC, 4 December 2023	Model bulletin; adoption and wording vary by state
SRC-04	AI Risk Management Framework (AI RMF 1.0), NIST AI 100-1	NIST, January 2023	Voluntary cross-sector framework; NIST states it is being revised
SRC-05	AI RMF Generative AI Profile, NIST AI 600-1	NIST, July 2024	Voluntary profile; requires tailoring to insurance use
SRC-06	Artificial Intelligence Governance Principles: Towards Ethical and Trustworthy AI in the European Insurance Sector	EIOPA Consultative Expert Group, June 2021	Expert-group report; not binding law
SRC-07	Solvency II framework (Directive 2009/138/EC and delegated acts)	European Union / EIOPA	No direct GAISST equivalence; legal review required
SRC-08	IFRS 17 Insurance Contracts	IASB, effective 1 January 2023 subject to adoption	Accounting linkage to AI failures requires qualified review

23. Financial impact vectors

Vector	Impact pathway
Claims leakage	Incorrect estimation, weak fraud detection or uncontrolled settlement recommendations.
Incorrect claims denial	False negatives or automation bias create remediation and litigation exposure.
Underpricing / overpricing	Model or data errors distort premium adequacy or customer outcomes.
Adverse selection	Poor segmentation or drift changes portfolio risk mix.
Reserving error	Model assumptions or regime shifts distort liabilities and management information.
Capital and reinsurance impacts	Accumulation or catastrophe-model error affects risk transfer and solvency decisions.
Regulatory and legal cost	Examination, remediation, penalties, disputes and record reconstruction.
Operational disruption	Vendor outage, model compromise or claims surge impairs critical services.
Fraud losses	Adversarial evasion, synthetic evidence or compromised scoring increases leakage.
Reputation and attrition	Unfair, opaque or inaccurate outcomes damage trust and distribution relationships.

These are pathways, not guaranteed outcomes. Quantification should use entity-specific exposure, decision volume, remediation cost, claims severity, capital and reinsurance assumptions.

23.1 Prudential and actuarial impact vectors

The following table translates security and AI-control failures into potential balance-sheet, actuarial and prudential pathways. It does not assert that a control failure automatically creates an accounting, capital or regulatory breach.

GAISST domain	Insurance failure pathway	Potential prudential / actuarial consequence
D1 — Model Integrity & Adversarial Robustness	Pricing or underwriting instability, poisoning or drift	Adverse selection, margin erosion, reserve assumption distortion and rate-filing challenge.
D4 — Supply Chain & Third-Party AI Security	Common catastrophe, cloud, data or foundation-model dependency failure	Correlated portfolio error, accumulation underestimation, service interruption and reinsurance treaty dispute.
D5 — Content Safety & Output Integrity	Misleading claims, policy or financial-reporting content	Claims leakage, customer remediation, reporting control failure and audit qualification risk.
D6 — Governance, Accountability & Human Oversight	Weak accountability, oversight or continuity	Operational resilience failure, delayed catastrophe response and governance or supervisory findings.
D8 — Regulatory Alignment & Compliance	Reporting, filing or supervisory obligations for reserving, capital or catastrophe models not met	Potential IFRS 17 reporting misstatement, Solvency II capital or governance findings and board risk-appetite breach; applicability requires qualified review.

24. Metrics and reporting

Metric	Definition	Owner	Frequency	Interpretation / gaming caveat
Inventory coverage	Material systems recorded / systems identified	AI governance	Monthly	Unknown shadow AI or stale records can distort metric
High-impact validation currency	High-impact systems with current validation	Model risk	Quarterly	A current report may still contain unresolved limitations
Adverse outcome rate	Adverse decisions by product and subgroup	Compliance / business	Monthly	Small samples and confounding require careful interpretation
Appeal reversal rate	Decisions reversed after challenge	Claims / underwriting	Monthly	Low appeal access can artificially lower the rate

Metric	Definition	Owner	Frequency	Interpretation / gaming caveat
Override rate and reason	Human overrides by system and outcome	Business owner	Monthly	Targeting low override rates can suppress appropriate challenge
Drift exceptions	Open performance/data drift breaches	Model owner	Risk-based	Threshold choice can conceal slow deterioration
Security abuse events	Injection, extraction, evasion or unauthorised-use events	CISO	Continuous / monthly report	Detection coverage varies
Vendor concentration	Critical services sharing provider/model/cloud	Third-party risk	Quarterly	Contract-level count may miss common fourth parties
Manual fallback readiness	Critical systems with tested fallback	Operations	Quarterly	Documented procedure is not proof of usable capacity

25. Notably Absent

- No legal, regulatory, actuarial, financial, insurance or investment advice.
- No regulator, actuarial or certification approval.
- No guaranteed compliance, fairness, accuracy, security, resilience or insurability.
- No universal risk, performance, fairness, confidence or notification threshold.
- No substitute for independent validation, policyholder-impact assessment or jurisdiction-specific review.
- No complete list of insurance threats, regulations, standards, vendors or use cases.
- No endorsement of any model, vendor, technology, data source or methodology.
- No determination that a particular insurance decision, variable, rate, claim outcome or coverage interpretation is lawful.
- No universal definition of high-risk AI.
- No presumption that human review, explainability or provider assurance is effective without operating evidence.
- No representation that the source register is complete for any jurisdiction; India (IRDAI) and other markets require separate review.

26. Maintenance

Review this guide at least annually and after a material GAISSTTM update, insurance supervisory development, significant sector incident, material technology change or substantiated practitioner feedback. Updates are issued as dated revisions; the framework remains GAISSTTM v1.0.

Appendix A — Terms and definitions

Term	Working definition
Adverse consumer outcome	A decision or process outcome that may unlawfully or unfairly disadvantage a policyholder, applicant, claimant or beneficiary.
Decision authority	The degree to which an AI system recommends, determines, executes or constrains an insurance action.
Meaningful human oversight	Timely review by a competent and authorised person with sufficient information and ability to intervene.
Insurance AI system	An AI-enabled component or decision chain used in insurance operations, including models, rules, prompts, retrieval, tools and integrations.
Policyholder impact	Actual or reasonably foreseeable effect on eligibility, price, coverage, claim, service, privacy, dignity or ability to challenge.

Appendix B — Insurance AI System Inventory

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

System ID	Name	Business owner	Function	Decision authority	Classification	Data sensitivity	Vendor	GAISST scope	Status
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix C — Insurance AI Risk Assessment

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Risk ID	System ID	Scenario	Threat/failure	Policyholder impact	Financial impact	Likelihood/rationality	Controls	Residual risk	Owner	Action
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix D — Control Interpretation Matrix

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Control ID	Control title	Applicable?	Insurance interpretation	Implementation	Evidence	Owner	Test	Gap	Status
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix E — Evidence Catalogue

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Evidence ID	Category	Title	Control links	Owner	Source	Period	Retention	Reliability	Location
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix F — Human Oversight Matrix

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Decision	Trigger	Oversight model	Reviewer competence	Information required	Authority	Time limit	Override record	Escalation
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix G — Vendor Due Diligence Questionnaire

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Question ID	Domain	Question	Evidence requested	Response	Risk rating	Reviewer	Action
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix H — Incident Classification Matrix

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Incident ID	System	Description	Severity	Policyholder impact	Financial impact	Security impact	Regulatory assessment	Owner	Status
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix I — Implementation Roadmap Tracker

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Task ID	Phase	Task	Owner	Start	Due	Dependency	Evidence	Status	Completion %
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix J — Worked Scenario Record

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Scenario ID	Context	Data	Decision authority	Threats	Controls	Evidence	Oversight	Monitoring	Residual risk
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix K — Source Register

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

See Section 21 (sources).

Appendix L — Notably Absent Register

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Item ID	Excluded claim/assumption	Reason	Affected section	Owner	Review date
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix M — Document Review Checklist

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Check ID	Category	Check	Owner	Status	Evidence / note
EXAMPLE - illustrative	To be completed	To be completed	To be completed	To be completed	To be completed

Appendix N — Open Issues and Deferred Decisions

Template. Complete material fields, retain source evidence and record assumptions; example rows are not thresholds.

Current open issues:

ID	Issue	Owner	Impact	Status
OI-001	Jurisdiction-specific legal and regulatory review	Qualified counsel	High	Open - legal review required
OI-002	Actuarial terminology and examples	Qualified actuary	High	Open - actuarial review required
OI-003	Formal approval identities	Document owner	Medium	Open
OI-004	Final website and repository URLs	Publication owner	Medium	Open
OI-005	Trademark-status wording	Legal owner	Medium	Open - trademark confirmation required