GAISSF

Healthcare Sector Implementation Guide

SEC-037 | Version 1.0

DRAFT FOR TECHNICAL, CLINICAL, LEGAL, PRIVACY AND STANDARDS REVIEW

ODA3 Institute

Document ID: SEC-037 **Version:** 1.0 **Category:** Sector Guide **Priority / Launch Phase:** High / Phase 2 **Publication status:** DRAFT FOR TECHNICAL, CLINICAL, LEGAL, PRIVACY AND STANDARDS REVIEW **Publication date:** 1 May 2026 **Publisher:** ODA3 Institute **Legal entity:** ODA3 Pvt Ltd (legal notices only) **Classification:** Public guidance

> ***Normative status.** This document is sector-specific implementation guidance. It does not amend, replace, extend, or override the normative requirements of GAISSF. Conformance is determined only against the applicable GAISSF normative publication and approved normative interpretations.*

Disclaimer

This guide provides general cybersecurity, AI security, governance, assurance, and implementation information. It is not legal advice, medical advice, clinical guidance, a regulatory determination, or a substitute for professional judgement. Applicable duties vary by jurisdiction, organisational role, system classification, intended use, patient population, and deployment context. Regulatory, legal, clinical, privacy, and standards references must be independently verified before publication and use.

Document control

Field	Value
Document owner	ODA3 Institute
Business owner	ODA3 Institute
Technical owner	[AI SECURITY LEAD TO BE CONFIRMED]
Clinical reviewer	[CLINICAL SAFETY REVIEWER TO BE CONFIRMED]
Privacy reviewer	[PRIVACY REVIEWER TO BE CONFIRMED]
Legal reviewer	[LEGAL REVIEWER TO BE CONFIRMED]
Standards editor	[STANDARDS EDITOR TO BE CONFIRMED]
Review cycle	At least annually and upon a material GAISSF, regulatory, threat, or technology change
Supersedes	New document

Revision history

Version	Date	Status	Summary	Approved by
0.1	2026-05-01	Working draft	Initial complete sector-guide package	Not approved
1.0	2026-05-01	Publication candidate	Final reviewed publication	[TBC]

Approval table

Review	Required approver	Decision	Date	Conditions
Standards integrity	[TBC]	Pending	—	All GAISSF references verified
Clinical safety	[TBC]	Pending	—	Clinical claims and scenarios reviewed
Cybersecurity and AI security	[TBC]	Pending	—	Threat model and controls validated
Privacy	[TBC]	Pending	—	Data-protection interpretations validated
Legal and regulatory	[TBC]	Pending	—	No legal-equivalence claims
Publication	[TBC]	Pending	—	Formatting, licence and marks confirmed

Claim-tier legend

Tier	Meaning	Use in this guide
Tier 1	Normative GAISSF requirement	Used only with a verified GAISSF citation

Tier	Meaning	Use in this guide
Tier 2	Verified external obligation	Used only after authoritative legal or regulatory verification
Tier 3	Sector interpretation	Reasoned healthcare application of GAISSF
Tier 4	Recommended practice	Non-binding practitioner guidance
Tier 5	Illustrative example	Fictional or hypothetical material

1. Executive Summary

Healthcare AI systems operate inside clinical, administrative, research, public-health, medical-device, and patient-facing workflows. Their security properties can affect confidentiality, integrity, availability, clinical judgement, continuity of care, and—in some contexts—patient safety. The relevant risk is not “AI risk” in the abstract. It is the combination of a defined use, a particular patient or operational context, technical dependencies, human decision pathways, and the consequences of failure or manipulation.

This guide translates GAISSF into healthcare implementation language. It helps organisations connect AI security governance with clinical safety, privacy, quality management, cybersecurity, enterprise risk, procurement, medical-device governance, and incident management. It does not create a healthcare edition of GAISSF and does not establish independent certification criteria.

Leadership should require five things before material healthcare AI use: a documented intended use and prohibited-use boundary; accountable ownership spanning clinical, security, privacy, and operational functions; proportionate validation and adversarial testing; monitored deployment with effective human escalation and fallback; and evidence that the controls operate in practice. High-impact systems require stronger independence, traceability, resilience, supplier evidence, and post-deployment surveillance than low-impact administrative tools.

Principal risk domains include data poisoning and quality degradation, manipulation at inference time, prompt and retrieval attacks, health-data leakage, unsafe integration, model or configuration tampering, excessive automation, drift, uneven performance, supplier opacity, availability failure, and inability to reconstruct historical decisions. Notably absent from this guide is any claim that every healthcare AI system is high risk, that GAISSF certification proves clinical effectiveness, or that listed threats have occurred in a particular organisation.

2. Introduction

The purpose of SEC-037 is to provide practical healthcare-sector interpretation of GAISSF. The guide is intended for healthcare CISOs, security architects, clinical safety officers, privacy and compliance leaders, medical-device security teams, AI governance functions, developers, procurement teams, assessors, auditors, and accountable executives.

Implementers should read this guide together with the current GAISSF normative standard, approved interpretations, assessment rules, and applicable legal and regulatory obligations. Assessors may use the guide to understand plausible healthcare evidence and context, but shall not treat a recommendation in this guide as a mandatory criterion unless that criterion is supported by verified GAISSF normative text.

The guide is deliberately jurisdiction-neutral. It identifies categories of obligation and recognised external standards without making legal determinations. Exact titles, editions, statuses, and regulatory applicability remain publication and deployment checks.

3. Scope

This guide applies to hospitals, health systems, clinics, laboratories, pharmacies, telehealth services, payers, public-health bodies, clinical research organisations, pharmaceutical and biotechnology organisations, medical-device manufacturers, health-technology vendors, health information exchanges, digital-health platforms, and relevant support providers.

It covers internally developed, procured, embedded, externally hosted, generative, predictive, autonomous, semi-autonomous, medical-device, research, administrative, cybersecurity, infrastructure, and foundation-model-based AI systems. Scope is determined by actual use and dependency, not by product labelling.

Excluded are clinical practice recommendations, therapeutic protocols, legal conclusions, product approvals, and claims of regulatory conformity. A system may be technically in scope while a particular legal classification remains unresolved.

Scope question	Include when	Boundary evidence
Does the system influence care?	It informs, prioritises, automates, delays, or constrains care	Workflow map, intended-use statement
Does it process health data?	It receives, derives, retrieves, logs, or exposes sensitive health information	Data-flow diagram, vendor data terms
Is AI embedded or indirect?	AI functionality is supplied through a device, API, SaaS feature, or subcontractor	Architecture and supplier inventory
Is it only experimental?	Include where real data, real users, or production-adjacent infrastructure create risk	Research protocol, environment controls

4. Conformance and Normative Relationship

Conformance is assessed against GAISSF, not against SEC-037. Where this guide conflicts with the verified normative standard, the normative standard prevails. Where applicable law imposes a stricter or different obligation, the organisation must address that obligation independently.

Interpretation hierarchy: (1) applicable law and binding regulatory requirements; (2) verified GAISSF normative requirements; (3) approved GAISSF interpretations; (4) this sector guide; and (5) organisation-specific procedures. The hierarchy does not imply that GAISSF overrides law.

Recommended practices may be adapted based on system impact, organisational scale, technical feasibility, and risk acceptance. Adaptation must not be used to disregard an applicable normative requirement. Compensating controls should be documented with rationale, ownership, evidence, limitations, and residual risk.

Notably absent: this section does not determine the applicable law, certify a system, or authorise an assessor to create new mandatory evidence requirements.

5. Healthcare Operating Context

Healthcare environments combine time-critical decisions, safety-sensitive workflows, sensitive data, legacy assets, medical devices, high availability needs, and multi-party accountability. A technically small AI failure may therefore have disproportionate operational consequences when it changes patient prioritisation, suppresses an alert, corrupts a record, or removes a service on which staff have become dependent.

Risk is also shaped by heterogeneous populations, differences in local clinical practice, limited downtime windows, complex procurement chains, and transition from research into operational use. Security controls must fit clinical workflow: controls that create unsafe delay, encourage workarounds, or block emergency access can introduce their own hazards.

Healthcare organisations should analyse the full pathway from technical event to operational and clinical consequence. They should not assume that privacy, security, safety, quality, ethics, and compliance are interchangeable disciplines; each asks different questions and requires different evidence.

6. Healthcare AI System Taxonomy

The following taxonomy supports inventory, ownership, validation planning, and risk classification. A system may occupy more than one category.

Category	Typical users and uses	Principal failure modes	Oversight and evidence emphasis
Clinical diagnostic AI	Radiology, pathology, dermatology, laboratory interpretation	Missed detection, false finding, image manipulation, distribution shift	Clinical validation, subgroup analysis, image provenance, override review
Clinical decision support	Treatment, deterioration, medication, risk scoring	Incorrect recommendation, automation bias, stale knowledge	Intended-use control, clinician review, performance surveillance
Patient monitoring	Remote monitoring, ICU alerts, wearable analytics	Missed alert, alert flooding, identity mismatch, connectivity failure	Availability, alarm governance, patient matching, fallback
Medical-device AI	Embedded intelligence, SaMD, robotics	Unsafe actuation, update failure, cyber-physical manipulation	Device safety process, secure update, manufacturer coordination
Administrative AI	Scheduling, coding, billing, workforce planning	Discrimination, denial or delay, integrity errors, fraud	Data quality, appeal, auditability, role separation
Patient-facing generative AI	Navigation, education, symptom guidance	Hallucination, unsafe advice, impersonation, privacy leakage	Clear boundaries, escalation, source grounding, content review
Clinical documentation AI	Ambient notes, transcription, summarisation	Fabricated or omitted facts, patient mismatch, hidden edits	Clinician attestation, provenance, change history, sampling
Research and life-sciences AI	Discovery, trials, genomics, synthetic data	Biased selection, leakage, invalid inference, dual use	Protocol governance, reproducibility, data rights, separation from care
Cybersecurity and infrastructure AI	Detection, access analytics, automation	False positives, unsafe automated response, blind spots	Human approval, containment safeguards, model monitoring

For each inventoried system, record intended and prohibited use, users, patient population, data, integrations, autonomy, human review, deployment location, supplier chain, validation status, monitoring owner, and retirement plan.

7. Healthcare AI Risk Classification

Risk classification should be based on consequence pathways and dependency, not on model type alone. The method below is Tier 4 recommended practice unless incorporated into GAISSF through a verified normative reference.

Score and document: severity and reversibility of possible harm; immediacy; clinical autonomy; effectiveness of human review; scale; data sensitivity; cyber-physical effect; availability dependency; adaptability; external exposure; propagation; vulnerable populations; supplier reliance; and detectability of silent failure.

Illustrative tier	Typical characteristics	Minimum governance response
Minimal operational impact	No patient data or care influence; easy reversal	Named owner, inventory, access control, basic testing
Material administrative impact	Affects scheduling, payment, workforce, or access	Business impact analysis, appeal/correction, monitoring
Significant care-delivery impact	Influences workflow or prioritisation; clinician dependency	Clinical owner, validation, fallback, incident integration
High patient-safety significance	Material potential for delayed, incorrect, or omitted care	Independent assurance, safety-security analysis, rigorous surveillance
Critical dependency	Life-sustaining or time-critical dependency; limited fallback	Executive acceptance, resilience testing, emergency procedures, continuous monitoring

Decision sequence: identify actual use; map affected decisions; identify failure and attack pathways; assess harm and reversibility; test human-review effectiveness; evaluate dependency and fallback; assign provisional tier; obtain multidisciplinary approval; revisit after material change.

Notably absent: no universal numerical threshold is established, and tier assignment does not determine legal classification.

8. Healthcare Threat and Failure Model

The threat model covers malicious action, accidental error, design limitation, environmental change, and organisational misuse. Each threat must be tied to an asset, precondition, consequence, control, evidence, and residual risk.

Threat or failure	Assets / preconditions	Plausible consequence	Prevent / detect / respond	Illustrative evidence
Training-data poisoning	Weak provenance or uncontrolled contribution	Biased or attacker-influenced behaviour	Curated sources; anomaly checks; dataset versioning; retraining rollback	Data lineage, acceptance tests
Inference manipulation / adversarial input	Exposed input channel	Incorrect classification or prioritisation	Input validation; robustness tests; rate controls; human review	Adversarial test report
Prompt or indirect prompt injection	LLM with external content or tools	Instruction override, leakage, unsafe action	Content isolation; tool allowlists; privilege separation; output checks	Attack test cases, tool policy
Model extraction or membership inference	Query access and weak controls	IP loss or disclosure of training participation	Access limits; monitoring; privacy testing	Query logs, privacy assessment
Health-data leakage	Sensitive prompts, logs, caches, telemetry	Confidentiality breach	Data minimisation; redaction; retention controls; DLP	Data-flow and retention evidence
Model/configuration tampering	Weak release or admin controls	Unapproved behaviour	Signing, approvals, privileged access, integrity monitoring	Release manifest, access review
Supply-chain compromise	Third-party model, package, API, update	Backdoor, outage, hidden behaviour change	Supplier due diligence; SBOM where relevant; update validation	Supplier file, dependency register
Hallucination or unsupported content	Generative use without grounding	False clinical or patient information	Retrieval controls; citations; mandatory review; use restrictions	Evaluation corpus, review records
Automation bias / alert fatigue	High trust or excessive alerts	Missed contradiction or ignored genuine alert	Training; interface design; override analysis; threshold review	Usability tests, override trends
Patient matching failure	Weak identity linkage	Wrong-patient recommendation or documentation	Strong identifiers; reconciliation; exception handling	Matching test, incident log
Availability failure	Cloud, network, device, or ransomware dependency	Delayed care or loss of workflow	Redundancy; manual fallback; recovery tests	BCP test, recovery records
Drift or uneven performance	Population or workflow change	Silent degradation for groups or sites	Monitoring; subgroup analysis; revalidation triggers	Drift dashboard, review minutes
Shadow AI / unapproved public tools	Easy user access and unmet workflow need	Data disclosure, unreviewed advice	Acceptable use; approved alternatives; discovery; training	CASB/DLP alerts, attestation

The organisation should explicitly record what was not tested, what populations were outside the validation set, and what incidents were not observed. Listing a scenario does not establish occurrence.

9. Healthcare Assets and Trust Boundaries

Protected assets include patient records, images, laboratory and genomic data, prescriptions, device telemetry, clinician notes, prompts, system instructions, retrieval corpora, model weights, datasets, evaluation results, configurations, audit trails, identity and consent data, provenance, safety cases, submissions, incident records, and supplier documentation.

Trust boundaries commonly exist between patients, clinicians, administrators, researchers, developers, devices, cloud providers, AI vendors, data processors, model providers, laboratories, payers, regulators, emergency services, and downstream care organisations.

Reference architecture specification: depict user channels; identity provider; clinical application; orchestration layer; model endpoint; retrieval store; tool/API layer; EHR or clinical systems; logging and monitoring; supplier services; and security controls. Mark data classes, authentication boundaries, administrative paths, model-update path, external content path, and manual fallback. Each arrow should identify protocol, data, owner, trust level, logging, and failure response.

10. Healthcare Governance Model

AI security governance should be integrated into existing clinical, cybersecurity, privacy, quality, device, procurement, enterprise-risk, and incident structures. A committee without system ownership is insufficient; each system needs an accountable executive, operational owner, technical owner, data owner, and risk owners.

Activity	Board / executive	Clinical owner	CISO / security	Privacy	AI governance	Supplier / procurement	Independent assurance
Approve risk appetite	A	C	C	C	R	I	I
Approve intended use	I	A/R	C	C	R	C	I
Security architecture	I	C	A/R	C	C	C	I
Clinical validation	I	A/R	C	I	C	C	C
Supplier acceptance	I	C	C	C	C	A/R	C
Deployment authorisation	A	R	R	R	R	C	C
Monitoring and incident triage	I	R	A/R	R	R	C	I
Independent assessment	I	C	C	C	C	I	A/R

Role names vary. The required outcome is clear authority, segregation of duties, competent review, escalation, and documented acceptance of residual risk.

11. Healthcare Lifecycle Implementation Guidance

The lifecycle controls below are Tier 3 interpretations and Tier 4 recommended practices unless tied to a verified GAISSF reference.

11.1 Use-case initiation

Define the problem, intended and prohibited use, patient population, setting, benefit and risk hypotheses, stakeholders, data, legal and regulatory screening, provisional risk tier, and go/no-go criteria. Reject use cases whose benefit cannot be articulated or whose risks cannot be controlled or observed.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

11.2 Design and architecture

Apply secure-, privacy-, and safety-by-design; least privilege; environment separation; data minimisation; strong identity; logging; model isolation; resilience; fail-safe behaviour; workflow fit; effective human oversight; provenance; and rollback.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

11.3 Data acquisition and preparation

Document authority, provenance, representativeness, quality, patient matching, de-identification limits, labelling, poisoning controls, access, retention, deletion, secondary use, transfers, and vulnerable-population implications.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

11.4 Development and configuration

Use secure development, controlled model selection, reproducible builds, dependency and secrets management, prompt/system-instruction control, code review, segregated testing, documentation, and change approval.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

11.5 Verification and validation

Separate technical performance, clinical validity, clinical utility, security assurance, privacy, operational safety, and regulatory approval. Test robustness, false outcomes, subgroups, adversarial behaviour, leakage, failover, usability, and acceptance criteria.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

11.6 Deployment

Require deployment authorisation, baseline configuration, access controls, training, escalation, phased rollout where appropriate, monitoring readiness, rollback, continuity procedures, and multidisciplinary sign-off.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

11.7 Operation and monitoring

Monitor performance, drift, security, anomalous outputs, feedback, safety events, supplier status, prompt attacks, leakage, access, subgroups, unapproved use, and workflow changes. Define thresholds and owners locally.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

11.8 Change and update management

Classify the significance of model, prompt, knowledge-base, data-source, vendor, and infrastructure changes. Set revalidation triggers, emergency-change controls, rollback, release notes, and user communication.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

11.9 Incident response and recovery

Integrate clinical, cybersecurity, privacy, supplier, and regulatory escalation. Support suspension, isolation, manual fallback, evidence preservation, notification decisions, root-cause analysis, corrective action, and revalidation.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

11.10 Retirement and decommissioning

Approve retirement; preserve required records; remove interfaces and credentials; manage dependencies; protect continuity of care; obtain supplier exit support; verify destruction or archival; and close ownership.

Illustrative evidence: approved lifecycle record, named owners, risk decision, test results, change log, monitoring record, or retirement certificate as applicable. **Notably absent:** completion of a document alone does not prove effective operation.

12. Interpretation of GAISSF Domains for Healthcare

The matrix below must be reconciled to the final GAISSF taxonomy before publication. Placeholder references are intentional.

GAISSF domain / reference	Healthcare interpretation	Recommended practices	Illustrative evidence	Common error / notably absent
Governance [REFERENCE TO BE CONFIRMED]	Link AI accountability to clinical and enterprise governance	Named accountable owners; multidisciplinary approval; risk acceptance	Charters, RACI, minutes, approvals	Committee exists but no system owner
Risk management [REFERENCE TO BE CONFIRMED]	Analyse technical-to-clinical consequence pathways	System-specific scenarios; residual-risk ownership	Risk register, safety-security analysis	Generic AI risk list without workflow context
Inventory [REFERENCE TO BE CONFIRMED]	Include embedded, supplier, research, and shadow AI	Discovery, ownership, lifecycle state, dependencies	AI inventory, architecture register	Only internally developed models listed
Data security and privacy [REFERENCE TO BE CONFIRMED]	Protect health data across prompts, logs, retrieval and training	Minimise data; control retention; test leakage	Data flows, privacy assessment, logs	Contract claim accepted without technical verification
Identity and access [REFERENCE TO BE CONFIRMED]	Address clinicians, patients, devices, services, emergency access	Strong identity, least privilege, break-glass governance	Access matrix, recertification, emergency logs	Shared accounts or unreviewed service credentials
Secure development and model security [REFERENCE TO BE CONFIRMED]	Control code, model, prompts, data and release artefacts	Reproducibility, integrity, adversarial testing	Build manifest, test reports, signatures	Only conventional application testing performed
Supply chain [REFERENCE TO BE CONFIRMED]	Treat model, cloud, device and data suppliers as a chain	Due diligence, change notice, incident terms, exit	Supplier assessment, contracts, dependency map	Vendor questionnaire without corroboration
Validation and assurance [REFERENCE TO BE CONFIRMED]	Separate security, clinical, technical, privacy and operational claims	Independent review proportional to impact	Validation plans and reports	Regulatory status treated as complete security evidence
Monitoring [REFERENCE TO BE CONFIRMED]	Observe drift, attacks, outputs, overrides, safety and availability	Defined metrics, thresholds, ownership, revalidation	Dashboards, review minutes, alerts	Monitoring limited to uptime
Incident response and resilience [REFERENCE TO BE CONFIRMED]	Coordinate patient safety and cyber response	Integrated triage, fallback, recovery testing	Playbooks, exercises, incident files	Security team handles incident without clinical escalation
Documentation and traceability [REFERENCE TO BE CONFIRMED]	Reconstruct system, version, data, decision, and change context	Provenance, version records, retention	System card, audit trail, release history	Current state documented but historical state unrecoverable
Human oversight [REFERENCE TO BE CONFIRMED]	Make review effective in real workflow	Authority, time, competence, interface, escalation	Usability tests, training, override analysis	“Human in the loop” asserted but not tested
Competence and improvement	Train role-specific users and learn from operation	Competency criteria, lessons learned, periodic	Training and improvement records	Awareness training substituted for role

GAISSF domain / reference	Healthcare interpretation	Recommended practices	Illustrative evidence	Common error / notably absent
[REFERENCE TO BE CONFIRMED]		review		competence

13. Patient Safety and Clinical Risk Integration

A security event becomes a patient-safety issue when it changes the availability, integrity, timing, presentation, or interpretation of information or system behaviour used in care. Organisations should connect threat models to clinical hazard analysis rather than maintaining disconnected security and safety registers.

Element	Required analysis
Threat or failure	What malicious, accidental, or environmental event occurs?
System function	Which input, model, interface, tool, alert, or integration changes?
Workflow consequence	How does staff or patient action change?
Potential harm	What delay, omission, incorrect action, or exposure could result?
Existing controls	Which clinical and technical controls already exist?
Detection / response	How is the event identified, escalated, contained, and recovered?
Residual risk / owner	What remains and who accepts it?

Illustrative worked example: An emergency triage model receives degraded data after an interface change. Risk scores are systematically lower for a subset of records. Technical monitoring detects a shift in missing-value frequency; clinical monitoring detects reduced escalation rates. Controls suspend model recommendations, restore manual triage, preserve affected records, review potentially impacted patients, and require revalidation before return. This example does not establish a universal implementation requirement. Notably absent: no claim is made that this event occurred or that the stated controls guarantee prevention of harm.

14. Human Oversight and Clinical Accountability

Human oversight is effective only when the reviewer has authority, competence, sufficient information, adequate time, usable presentation, and a practical means to disagree, override, escalate, and document the decision. Merely placing a person after an automated output is not a control if workload, interface design, organisational pressure, or hidden automation makes independent judgement unrealistic.

Define which outputs require review; whether review is prospective or retrospective; what information is shown; time limits; override rights; escalation; disagreement recording; and responsibility where vendors and providers share control. Measure override patterns, missed escalations, alert burden, review latency, and user confidence. Patient-facing systems should make boundaries clear and route urgent or uncertain cases to appropriate human services.

Notably absent: this guide does not transfer professional accountability to a model, prescribe clinical judgement, or treat explainability as a substitute for validation.

15. Health Data Protection and Confidentiality

Health data protection must cover collection, prompts, retrieval, training, fine-tuning, outputs, caches, logs, telemetry, backups, support access, and supplier processing. Data minimisation should be enforced technically and contractually. De-identification and pseudonymisation reduce some risks but do not automatically remove re-identification, linkage, or jurisdictional obligations.

Implementation should document purpose, authority, access, patient rights where applicable, minimum-necessary use, retention, deletion, secondary use, transfer, model-provider training terms, support access, and incident handling. Genomic, voice, image, child, and rare-condition data may require heightened analysis because uniqueness or context can increase identifiability.

External legal requirements are jurisdiction-specific. [EXTERNAL REFERENCE REQUIRES VERIFICATION BEFORE PUBLICATION]

16. Medical-Device and Connected-System Considerations

AI embedded in a medical device or supplied as medical-device software requires coordination among security, clinical engineering, quality, safety, regulatory, and manufacturer functions. Inventory device identity, software and model version, network location, remote-support path, maintenance owner, patch constraints, end-of-support date, and fallback.

Validate updates before deployment where operationally feasible; assess the safety effect of security controls; segment networks; control remote access; monitor vulnerabilities; preserve manufacturer communications; and maintain emergency procedures. Where relevant, use software component transparency and secure-development evidence, but do not assume any single artefact proves device security.

Notably absent: GAISSF assessment does not replace medical-device approval, quality-system obligations, post-market surveillance, or clinical acceptance.

17. Generative AI in Healthcare

Generative AI requires explicit task boundaries. Uses such as summarisation, documentation, education, navigation, coding support, and research assistance present different risks from diagnosis or autonomous action. Tool use, retrieval, and agentic permissions materially change the risk profile.

Decision factor	Lower-risk pattern	Higher-risk pattern
Purpose	Drafting or retrieval support	Diagnosis, treatment, autonomous action
Data	Controlled, minimised, approved	Identifiable data sent to unapproved service
Grounding	Curated sources with provenance	Open-web or unverified retrieval
Review	Competent mandatory review	Direct patient or system action without review
Tools	Read-only, allowlisted	Write, prescribe, schedule, or modify records
Failure handling	Clear escalation and fallback	Output accepted as authoritative

Controls should address hallucination, fabricated citations, prompt injection, retrieval poisoning, sensitive-data disclosure, provider retention, system-prompt exposure, unsafe tool calls, patient impersonation, provenance, source verification, and disclosure to users. Public model use should be governed through acceptable-use rules and approved alternatives.

18. Third-Party and Supply-Chain Risk

Supplier assurance should cover the full chain: application vendor, model provider, cloud provider, data provider, device manufacturer, API, subcontractor, and open-source dependency. Contractual statements are useful but should be corroborated where impact warrants.

Due diligence should ask about architecture, data use and retention, training, security testing, vulnerability handling, incidents, audit evidence, regulatory status, availability, changes, model versions, subcontractors, location, deletion, logging, IP, validation, continuity, and exit. Material suppliers should provide timely notification of security incidents and changes that could invalidate prior assessment.

Contract-control category	Intended outcome
Data use and deletion	Limit purpose, retention, training, support access, and residual copies
Security and assurance	Define controls, testing, evidence, vulnerability and audit rights
Change management	Notify model, infrastructure, data-source, and subcontractor changes
Incident response	Set notification, cooperation, evidence preservation, and remediation duties
Availability and exit	Support continuity, portability, transition, and secure termination

This is implementation guidance, not legal drafting advice.

19. Identity, Access, and Privileged Operations

Identity controls must cover clinicians, patients, researchers, administrators, service accounts, devices, model endpoints, and suppliers. Apply least privilege, role or attribute controls, strong authentication, session protection, segregation of duties, and periodic recertification.

Emergency “break-glass” access should be limited, logged, reviewed, and tested so that security does not obstruct urgent care while emergency privilege does not become a routine bypass. Privileged changes to models, prompts, tools, thresholds, retrieval content, and integrations should require approval and traceability.

20. Resilience and Continuity of Care

Continuity planning must assume loss, degradation, corruption, or withdrawal of AI functionality. Identify clinical and operational dependencies, maximum tolerable disruption, manual fallback, staffing requirements, data reconciliation, recovery sequence, supplier dependencies, and criteria for safe return.

Continuity template: system and owner; dependent workflows; trigger; degraded mode; manual process; safety checks; required data; communications; recovery objective; backup and model artefacts; supplier contact; test frequency; last test; gaps; approval.

Exercises should include cloud outage, network interruption, ransomware, vendor failure, model withdrawal, corrupted update, and loss of supporting data. Notably absent: a documented fallback is not evidence that staff can execute it under realistic workload.

21. Monitoring and Performance Surveillance

Monitoring should combine security, model, clinical, operational, privacy, and supplier signals. Metrics may include abnormal inputs and outputs, drift, subgroup performance, overrides, false outcomes, incidents, availability, latency, prompt attacks, leakage, unauthorised use, version changes, workarounds, complaints, near misses, and safety events.

Each metric needs an owner, data source, review frequency, threshold or decision rule, escalation route, retention, and revalidation trigger. Universal thresholds are not prescribed because acceptable values depend on use and harm pathways.

Signal	Owner	Example decision
Input distribution change	Model owner / clinical owner	Investigate data pipeline; restrict use if material
Prompt-attack indicator	Security operations	Contain session, review tool actions and exposed data
Override-rate change	Clinical governance	Review model performance, workflow and training
Supplier version change	System owner / procurement	Assess significance and revalidate before acceptance
Availability degradation	IT operations / clinical operations	Activate fallback and continuity communications

22. Healthcare AI Incident Management

AI incidents may be cybersecurity, privacy, safety, data-quality, performance, availability, supplier, misuse, regulatory, or near-miss events. Classification should support coordinated action rather than force an incident into one organisational silo.

Triage factor	Questions
Patient impact	Actual or plausible harm, delay, omission, or wrong-patient effect?
Scale and duration	How many people, sites, decisions, and how long?
Data sensitivity	What health or identity data were exposed, altered, or unavailable?
Exploitability	Is there active exploitation or repeatability?

Triage factor	Questions
System criticality	Is there manual fallback and is it effective?
Regulatory relevance	Could notification, preservation, or authority engagement be required?
Evidence	Can affected versions, inputs, outputs, logs, and decisions be reconstructed?

Workflow: detect; protect patients and operations; contain; preserve evidence; classify; notify internal owners; determine external obligations; communicate; recover; assess affected decisions; remediate; revalidate; document lessons. Legal notification decisions require qualified advice.

23. Documentation and Evidence Expectations

Evidence should demonstrate design and operation. Typical evidence includes system inventory, intended and prohibited use, data flows, architecture, system/model cards, clinical and security validation, privacy assessment, risk and safety analysis, supplier assessment, contracts, change records, approvals, training, monitoring, incidents, access reviews, audit trails, performance reports, and retirement records.

Evidence	Purpose	Likely owner	Lifecycle / review
Intended-use statement	Bound authorised use and users	Clinical / system owner	Initiation; material change
Data-flow and architecture	Identify data, interfaces and trust boundaries	Architecture / privacy / security	Design; integration change
Validation package	Support technical, clinical, security and operational claims	Model and clinical owners	Pre-deployment; revalidation trigger
Supplier assurance file	Support reliance on external service	Procurement / TPRM	Onboarding; annual; material change
Monitoring record	Show continuing control and review	Operations / model owner	Continuous / periodic
Incident and CAPA record	Show detection, response and improvement	Incident owner / quality	Per event; trend review
Retirement certificate	Close access, data and dependencies	System owner	Decommissioning

Retention should reflect legal, clinical, security, contractual, and evidentiary needs. No universal retention period is asserted.

24. Assessment Guidance

Assessors should establish scope, sample systems by impact and type, interview clinical and technical stakeholders, inspect evidence, test selected controls, and compare documented use with actual workflow. Supplier evidence may support but not automatically replace the organisation's own risk decision.

Strong evidence is specific, current, approved, traceable, and corroborated by operation. Weak evidence is generic, stale, incomplete, or unowned. Misleading evidence uses regulatory approval, policy existence, or vendor certification to imply broader assurance than it supports. Absent evidence is a gap; evidence requiring corroboration includes self-attestation, screenshots without provenance, and reports whose scope excludes the deployed configuration.

Assessors should verify effective fallback and human oversight, historical traceability, monitoring response, change significance, and treatment of high-impact systems. They must not create certification criteria from this guide.

25. Healthcare Implementation Scenarios

All scenarios are fictional Tier 5 examples. **This example is illustrative and does not establish a universal implementation requirement.**

25.1 Radiology image-analysis system

Context: large hospital network. **Use:** suggests findings for radiologist review. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-

specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** image integrity attack, drift, wrong-patient association. **Implementation:** PACS integration controls, clinical validation, image provenance, radiologist override monitoring, fallback to standard reading. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

25.2 Emergency-department triage model

Context: regional health system. **Use:** prioritises queue and flags deterioration risk. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** data-interface degradation, subgroup performance, automation bias. **Implementation:** manual triage authority, real-time data-quality monitoring, periodic subgroup review, downtime protocol. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

25.3 Patient-facing generative assistant

Context: digital-health platform. **Use:** provides navigation and education. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** hallucination, privacy leakage, unsafe urgency advice. **Implementation:** curated retrieval, prohibited diagnostic use, urgent-care escalation, session data controls, sampled review. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

25.4 Ambient documentation tool

Context: multi-site clinic. **Use:** drafts notes from clinical conversations. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** omission, fabrication, patient mismatch, audio leakage. **Implementation:** clinician attestation, identity reconciliation, encrypted audio handling, provenance and correction log. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

25.5 Remote patient-monitoring platform

Context: home-care provider. **Use:** detects deteriorating measurements. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** connectivity loss, alert flooding, device spoofing. **Implementation:** device identity, availability monitoring, alarm escalation, patient contact fallback. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

25.6 AI-enabled medical device

Context: device manufacturer and hospital. **Use:** adapts a device function within approved bounds. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** unsafe update, remote-access compromise, model drift. **Implementation:** quality and safety integration, signed updates, network controls, post-market surveillance. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

25.7 Administrative scheduling system

Context: small rural provider. **Use:** optimises appointments and capacity. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** access inequity, incorrect cancellation, outage. **Implementation:** human appeal, audit log, manual scheduling, impact review. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

25.8 Clinical-trial selection model

Context: clinical research organisation. **Use:** matches candidates to protocol criteria. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** biased exclusion, data misuse, non-reproducible ranking. **Implementation:** protocol governance, reproducibility, privacy controls, investigator review. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

25.9 Cybersecurity anomaly-detection platform

Context: hospital security operations. **Use:** prioritises identity and network alerts. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** false positive isolation, blind spot, unsafe automation. **Implementation:** human approval for disruptive action, tuning, recovery and audit. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

25.10 Externally hosted clinical decision support

Context: hospital network and SaaS supplier. **Use:** offers treatment-support information. **Data and architecture:** relevant clinical or operational data passes through controlled interfaces to an identified model or service; exact design is organisation-specific. **Provisional risk:** determined using Section 7 and may range from material operational to critical dependency. **Threats/failures:** supplier outage, hidden version change, unsupported recommendation. **Implementation:** supplier change notification, local validation, version monitoring, manual fallback. **Evidence:** intended-use statement, architecture, validation, access and monitoring records, supplier file where relevant, incident and change records. **Residual risk:** incorrect output, dependency, human factors, and unobserved distribution change

remain possible. **Notably absent:** no real organisation, product, incident, performance rate, or regulatory outcome is represented.

26. Small and Resource-Constrained Healthcare Organisations

Smaller organisations should use proportionate governance without removing accountability. Minimum viable practice includes a complete inventory; named accountable clinical and technical owners; documented intended use; supplier due diligence; basic security and privacy review; validation appropriate to impact; manual fallback; monitoring; and incident escalation.

Shared security services, regional clinical networks, managed providers, and standardised templates may reduce burden. Supplier reliance increases the importance of clear evidence, change notice, incident cooperation, data terms, and exit plans. Where the system can materially affect patient care, specialised clinical safety, privacy, security, or legal expertise should be obtained.

Notably absent: resource constraints do not by themselves justify accepting uncontrolled high-impact risk.

27. Multi-Jurisdictional Considerations

Organisations should maintain a jurisdictional applicability worksheet covering health-data and privacy law, medical-device regulation, AI regulation, cybersecurity duties, breach and safety reporting, localisation, retention, patient rights, research, and cross-border service delivery.

Jurisdiction / role	System and use	Data / patients	Potential obligation category	Authority source	Owner	Verified date	Open decision
[TBC]	[TBC]	[TBC]	Privacy / device / AI / cyber / clinical	[AUTHORITATIVE SOURCE]	[TBC]	[TBC]	[TBC]

This worksheet supports issue spotting only. Binding status and applicability require authoritative verification and qualified advice.

28. Relationship to External Standards and Frameworks

External frameworks may support implementation but should not be represented as equivalent to GAISSF. Titles, editions, and statuses below require verification before publication.

Framework	Relevance / overlap	Distinction and remaining work
ISO/IEC 27001	Information-security management system	Does not alone address all AI model, clinical, or patient-safety concerns
ISO/IEC 27701	Privacy information management	Does not determine healthcare law or AI security conformance
ISO/IEC 42001	AI management system	Broad AI governance; healthcare security and clinical application still require sector controls
ISO/IEC 23894	AI risk-management guidance	Supports risk method; does not create GAISSF equivalence
ISO 14971	Medical-device risk management	Device safety focus; not a complete enterprise AI-security standard
IEC 62304	Medical-device software lifecycle	Software lifecycle focus; model, supplier, and operational security need additional treatment
IEC 81001-5-1	Health software and health IT security lifecycle	Relevant to secure lifecycle; scope and edition must be verified
IEC 80001 series	Risk management for IT networks incorporating medical devices	Relevant to connected clinical environments; not full AI assurance
NIST Cybersecurity Framework	Cybersecurity outcomes and profiles	Requires AI- and healthcare-specific interpretation
NIST AI Risk Management Framework	AI risk functions and profiles	Broad voluntary framework; no automatic

Framework	Relevance / overlap	Distinction and remaining work
		certification equivalence

[EXTERNAL REFERENCE REQUIRES VERIFICATION BEFORE PUBLICATION]

29. Implementation Roadmap

Stage	Objective and activities	Accountable roles	Deliverables / completion indicators
1. Governance and scope	Establish authority, scope, policy and escalation	Executive, clinical, CISO, privacy, AI lead	Charter, RACI, scope approved
2. Inventory	Discover internal, embedded, supplier and shadow AI	System owners, architecture, procurement	Inventory with ownership and lifecycle state
3. Risk classify	Apply Section 7 and identify obligations	Clinical, security, privacy, legal	Approved provisional tiers and applicability issues
4. Prioritise	Focus first on high-impact and weakly controlled systems	Governance body	Remediation portfolio and risk decisions
5. Baseline assessment	Compare actual controls and evidence to GAISSF	Control owners, assurance	Gap register with evidence
6. Remediate	Implement technical, clinical, supplier and process controls	System and control owners	Closed critical gaps, accepted residual risks
7. Validate	Perform independent, proportionate testing	Clinical, model, security, privacy	Approved validation package
8. Deploy and monitor	Authorise use; operationalise metrics and fallback	Operations and owners	Deployment approval, dashboards, tested fallback
9. Incident integration	Connect AI events to safety, privacy and cyber response	Incident, quality, clinical, legal	Playbook and exercise record
10. Assess and improve	Internal assessment, external readiness, lessons	Assurance and governance	Findings, CAPA, periodic review

Sequencing is indicative; no universal completion timeline is prescribed.

30. Implementation Checklist

The consolidated checklist is supplied as `SEC-037_Healthcare_Implementation_Checklist.xlsx` and CSV/JSON derivatives. Status values should be evidence-based. “Not applicable” requires rationale and approval.

ID	Implementation question	Applicability	Responsible role	Evidence	Status	Gap	Action	Target date
H-GOV-01	Is accountable ownership defined for each AI system?	All	Executive / system owner	Inventory, RACI	Not Started	—	—	—
H-RISK-01	Has the system been classified using actual use and harm pathways?	All	Risk owner	Risk assessment	Not Started	—	—	—
H-VAL-01	Are technical, clinical, security, privacy and operational claims separately validated?	As applicable	Validation owners	Validation package	Not Started	—	—	—
H-MON-01	Are monitoring signals,	Production	Model / clinical / security	Monitoring plan	Not Started	—	—	—

ID	Implementation question	Applicability	Responsible role	Evidence	Status	Gap	Action	Target date
	thresholds, owners and revalidation triggers defined?		owner					
H-IR-01	Is the system integrated into clinical, cyber, privacy and supplier incident response?	All	Incident owner	Playbook / exercise	Not Started	—	—	—

31. Common Failure Patterns

Failure pattern	Why it matters / warning signs	Corrective action
Regulatory approval treated as full security assurance	Approval scope may not cover deployed integration, threats, or operations	Map approval evidence to actual claims; perform local security assessment
Security testing treated as clinical validation	Technical robustness does not establish clinical utility or safety	Separate plans, owners, criteria and approvals
Embedded or supplier AI omitted from inventory	Unknown dependency and change risk	Discover by architecture, procurement, data flow and vendor review
Vendor claims accepted without evidence	Scope, version or exclusions may not match deployment	Obtain corroborating reports, tests, contracts and local validation
Human oversight asserted but ineffective	Reviewer lacks time, authority, context or usable interface	Test workflow and human factors; monitor override and escalation
Downtime procedures untested	Manual fallback may fail under pressure	Exercise realistically and remediate staffing/data dependencies
Generative output over-trusted	Plausible false content may enter records or advice	Bound use, ground sources, require competent review and provenance
Monitoring limited to uptime	Drift, subgroup failure, unsafe output and misuse remain unseen	Build multi-domain surveillance and decision rules
Updates not revalidated	Behaviour can change without local awareness	Require change notice, significance assessment, rollback and revalidation
Historical decisions unrecoverable	Incident and care review become unreliable	Preserve version, input, output, source and configuration context

32. Notably Absent

This guide does not certify clinical effectiveness or medical-device safety; grant regulatory approval; guarantee legal compliance; eliminate patient risk; require a particular vendor; replace clinical judgement or professional standards; create new GAISSF controls; establish universal thresholds; assert that all healthcare AI is high risk; claim every listed threat has occurred; guarantee incident prevention; replace jurisdiction-specific analysis; provide legal or medical advice; create a safe harbour; or establish automatic equivalence with ISO, IEC, NIST, healthcare, privacy, or medical-device frameworks.

The guide also does not establish that a documented control is operating effectively, that a supplier assertion is correct, that a model is fair across all populations, or that absence of observed incidents proves absence of risk. Items not tested, populations outside scope, unavailable evidence, unresolved legal questions, and technical limitations must be stated in system-level records.

33. Limitations

Limitations include rapidly evolving models and agentic capabilities; changing regulation and standards; jurisdictional variation; incomplete incident visibility; limited public sector data; local workflow differences; supplier opacity; research-to-care transitions; medical-device classification differences; and the simplification inherent in fictional scenarios and generic templates.

The guide should be revised when GAISSF changes materially, authoritative healthcare obligations change, credible new attack or failure patterns emerge, incident learning invalidates guidance, or implementation experience reveals ambiguity.

34. Maintenance and Review

ODA3 Institute should assign a document owner and maintain a controlled issue register. Review at least annually and on material triggers. Changes should be classified as editorial, clarifying, substantive non-normative, or dependent on normative GAISSF change. Stakeholder consultation should include clinical, security, privacy, legal, device, quality, supplier, assessor, and where appropriate patient representation.

Published versions should include compatibility with the applicable GAISSF version, change summary, effective date, deprecation status, and archive location. Normative changes must not be introduced through this guide.

35. Conclusion

Healthcare implementation of GAISSF requires integration rather than parallel governance. Cybersecurity and AI security controls must connect to patient safety, privacy, clinical workflow, quality, procurement, resilience, and incident response. Proportionality should be based on real consequence pathways, dependency, reversibility, data sensitivity, autonomy, and the effectiveness of human and operational controls.

Evidence, monitoring, fallback, and change control are central. The objective is not to produce more paperwork; it is to demonstrate that defined controls operate for the deployed system and that limitations, residual risks, and absent evidence are visible to accountable decision-makers.

Annex A — Healthcare Terminology and Definitions

Term	Working definition	Verification note
Clinical AI	AI used in or materially influencing clinical care or clinical workflow	Organisational and legal scope may differ
Clinical decision support	Information or recommendations intended to assist a healthcare decision	Regulatory classification varies
Clinical validation	Evidence that performance is valid for the intended clinical use, population and setting	Requires clinical review
Technical validation	Evidence that technical requirements and performance criteria are met	Not equivalent to clinical utility
Clinical utility	Evidence that use produces meaningful benefit in practice	Not established by model accuracy alone
Medical device / SaMD	Product categories defined by applicable medical-device law	Jurisdictional definition must be verified
Patient safety / clinical safety	Prevention or control of avoidable harm arising in care or health technology use	Local safety terminology may differ
Health data	Data about health, care, treatment, biometrics, genetics or related identity/context	Legal definitions vary
Model drift / data drift	Material change in model behaviour or input distribution over time	Thresholds are use-specific
Human oversight	Effective human authority, review, intervention, escalation and accountability	Must be tested in workflow

Term	Working definition	Verification note
Automation bias	Excessive reliance on automated output despite contrary information	Human-factors concept
Adverse event / near miss	Harmful event, or event that could have caused harm but did not	Reporting definitions vary
Intended use / foreseeable misuse	Authorised purpose and reasonably predictable use outside that purpose	Regulatory phrasing may vary
Ambient clinical documentation	Capture and AI-assisted drafting of clinical documentation from encounters	Privacy and recording rules vary
Safety case	Structured argument supported by evidence that risk is acceptably controlled	Form and legal status vary

Annex B — Healthcare AI System Inventory Template

Fields: system ID; name; owner; vendor; model provider; purpose; intended use; prohibited use; clinical/non-clinical; users; patient population; environment; data classes; integrations; autonomy; human oversight; device status; regulatory status; risk tier; lifecycle state; monitoring owner; validation dates; dependencies; incident history; evidence location; retirement date. A JSON schema and spreadsheet sheet are supplied in the package.

Annex C — Healthcare AI Risk-Assessment Template

Record system context; affected stakeholders; intended and prohibited use; patient-safety, privacy, security, availability and operational impact; threat and failure scenarios; preconditions; existing controls; gaps; likelihood and impact rationale; detectability; reversibility; residual risk; risk owner; approval; monitoring; and review trigger.

Annex D — Supplier Assessment Questionnaire

Minimum question groups: organisation and service scope; architecture and hosting; model and version management; data collection, use, retention, training and deletion; identity and administrative access; secure development; testing and red teaming; vulnerability disclosure; incident history and notification; availability and recovery; subcontractors and locations; change notice; logging and customer evidence; clinical validation; regulatory status; IP and licensing; support; portability and exit. Require evidence and note scope exclusions.

Annex E — Clinical and Security Validation Checklist

Separate workstreams for clinical validity and utility; technical performance; cybersecurity and adversarial robustness; privacy and leakage; operational workflow and usability; human factors; resilience and fallback. Each workstream must define scope, environment, population, version, criteria, results, limitations, unresolved findings, approval, and revalidation triggers.

Annex F — Incident Triage and Escalation Matrix

Use the incident matrix in Section 22 and identify escalation to clinical safety, security operations, privacy, quality, regulatory, legal, supplier, executive, communications, and patient-support functions. The matrix must include out-of-hours contacts, authority to suspend use, evidence preservation, affected-decision review, and external-notification decision ownership.

Annex G — Evidence Catalogue

The package includes a structured evidence catalogue in JSON and spreadsheet form. Evidence should be mapped to lifecycle stage, control objective, owner, system version, period, assessor use, retention, and limitations.

Annex H — Healthcare Implementation Maturity Model

Level	Governance	Technical / clinical practice	Monitoring and evidence
1 Initial	Ad hoc ownership	Inconsistent review	Limited evidence; reactive incidents
2 Defined	Policies, roles and inventory established	Standard methods documented	Basic records and planned metrics
3 Implemented	Accountable governance operates	Controls and validation applied by risk	Regular monitoring and incident integration
4 Measured	Decisions use performance and risk data	Independent assurance and trend analysis	Reliable metrics, traceability and CAPA
5 Continuously improved	Governance adapts to learning	Controls are tested and improved systematically	Cross-system learning and predictive risk indicators

Maturity is not equivalent to GAISSF conformance.

Annex I — Example Responsibility Matrix

Use the RACI in Section 10 as a baseline. Tailor role titles, ensure one accountable role per decision, and document shared responsibilities with suppliers. Patient or public representation may be appropriate for material patient-facing uses.

Annex J — Healthcare Implementation Checklist

The complete checklist is provided in XLSX, CSV, JSON, and Markdown. Categories include governance, inventory, risk, clinical safety, cybersecurity, privacy, data, development, suppliers, validation, deployment, oversight, monitoring, incidents, resilience, change, documentation, assessment and retirement.

Annex K — Cross-Reference Register

The package includes a register with fields: GAISSF clause; clause title; SEC-037 section; interpretation summary; related evidence; external framework; verification status; reviewer; notes. All GAISSF identifiers remain placeholders until reconciled to the approved normative standard.

Annex L — Document Verification Checklist

Verify: GAISSF references; external-standard titles and editions; regulations; terminology; absence of invented identifiers; normative/guidance separation; internal cross-references; table and figure numbering; defined terms; copyright, marks and licence; version/date; links; accessibility; spelling and grammar; unsupported compliance claims; certification disclaimers; Notably Absent coverage; and final legal, clinical, privacy, cybersecurity and standards approvals.

Publication Readiness

Unresolved placeholders

1. Applicable GAISSF version, domain names, clauses and control identifiers.
2. Publication date, approval identities, document owner and reviewers.
3. Publication licence and trademark-status wording.
4. Verified external standard titles, editions and statuses.
5. Jurisdiction-specific regulatory references, if any are added.
6. Final GitHub paths, website URL and document compatibility statement.

External references requiring verification

ISO/IEC 27001; ISO/IEC 27701; ISO/IEC 42001; ISO/IEC 23894; ISO 14971; IEC 62304; IEC 81001-5-1; IEC 80001 series; NIST Cybersecurity Framework; NIST AI Risk Management Framework; and any healthcare, medical-device, privacy, cybersecurity or AI regulation cited in the publication version.

Decisions required from the GAISSF document owner

Confirm the authoritative GAISSF release; approve sector interpretation boundaries; determine whether any healthcare interpretations require formal normative interpretation; approve risk-tier terminology; approve assessment-use wording; select licence; confirm marks; and accept the maintenance process.

Recommended reviewers

Clinical safety and practising clinician; healthcare CISO / security architect; AI and model security specialist; privacy and health-data counsel; medical-device security and regulatory specialist; healthcare quality and patient-safety lead; standards editor; assessor/auditor; supplier-risk practitioner; accessibility and publication reviewer.

Publication-readiness verdict

Not ready for publication. The substantive implementation package is complete as a controlled draft, but material GAISSF cross-references, external references, legal/trademark/licensing statements, and multidisciplinary approvals remain unresolved.