GABA

GOVERNED AI BOUNDARY ATTESTATION STANDARD

Formal attestation of AI inference boundary risk for governed deployments.

MIT LICENSE · v1.0 · GENERAL REASONING, INC. · gabastandard.com

> ***No security architecture eliminates all risk at the AI inference boundary. Provider infrastructure, key management, personnel, and legal compellability remain outside the deploying organization's control regardless of how well the boundary is hardened. GABA defines the formal framework for identifying, documenting, and accepting these residual risks -- explicitly, attributably, and immutably -- so that regulated organizations can demonstrate governed AI deployment rather than accidental AI deployment.***

I. The Problem GABA Solves

Every compliance framework written before 2024 assumed that the systems an organization deploys are under that organization's control. SOC 2, HIPAA, PCI DSS, FedRAMP, ISO 27001 -- all of them define controls for systems the organization owns, operates, and can audit.

AI inference is different. When a regulated organization calls an external AI provider, it is sending data to infrastructure it does not own, operated by personnel it cannot audit, under a legal jurisdiction it may not control. The provider's security team may be excellent. Their key management may be rigorous. Their personnel practices may be exemplary. None of that is verifiable by the deploying organization. And none of the existing compliance frameworks have a control for it.

The typical organizational response is to treat this gap as invisible -- to call the API, log the call, and move on without formally acknowledging what has been accepted. The invisible acceptance is still an acceptance. The risk is still present. The difference is that when something goes wrong, the organization cannot demonstrate that it understood and deliberately accepted the risk. It can only demonstrate that it didn't think about it.

> ***GABA makes the implicit explicit. The residual risk at the AI inference boundary already exists in every organization that calls an external AI API. GABA does not create new risk -- it creates a formal record that the risk was identified, understood, and accepted by an authorized human before the deployment went live.***

II. Scope and Prerequisites

GABA applies to governed server deployments that satisfy the CRC Governed Server Boundary (GSB) prerequisite. See crcstandard.com for the full GSB definition. In summary: GABA applies when the deployment runs on infrastructure under formal organizational control, changes require an authorized audit trail, and no unmanaged AI agents with computer-use capability have access to the governed boundary.

GABA does not apply to consumer environments, developer workstations, or any deployment where the GSB prerequisite is not satisfied. The framework makes no claims about and provides no attestation value for unmanaged environments.

GABA is the CRC Boundary pillar attestation mechanism. An organization that completes GABA attestation for all external AI inference endpoints achieves a Boundary pillar score of 4 -- the maximum -- under the CRC Minimum Surface Standard.

III. The AI Boundary Risk Acceptance Record (ABRAR)

The core artifact of GABA is the AI Boundary Risk Acceptance Record (ABRAR). One ABRAR is required per governed deployment. A single ABRAR covers all authorized endpoints for that deployment. Each endpoint is individually enumerated within the record.

The ABRAR is not a form to be filed. It is a structured attestation that must be signed by an authorized human signatory and anchored to an immutable Chandra Protocol context unit. The Chandra CU is what makes the ABRAR unforgeable -- it cannot be backdated, edited, or erased. The signatory is attributed. The date is fixed. The chain grows.

ABRAR required fields:

Field	Content / Requirement
Organization	Full legal name of the deploying organization.
Deployment identifier	Unique identifier for this governed deployment instance.
Assessment date	Date this record was created. ISO 8601 format.
Authorized signatory	Full name and title of the human accepting residual risk on behalf of the organization.
Endpoint registry	Complete list of all external AI inference endpoints authorized for this deployment. Each entry includes: provider name, endpoint URL, certificate fingerprint, and purpose.
Hardening controls confirmed	Attestation that all applicable Boundary hardening controls are active: certificate pinning, request signing, response validation, rate limiting, anomaly detection, circuit breaker, payload sanitization.
Computer-use agent status	Explicit statement of whether any AI agent with computer-use capability is authorized to access the governed boundary. If yes: each agent is individually named, scoped, and attested separately.

Field	Content / Requirement
Residual risks accepted	Explicit enumeration of residual risks accepted for each endpoint: provider infrastructure, provider key management, provider personnel, provider government relationships and legal compellability. Each risk is named, not implied.
Chandra CU reference	The context unit ID of the Chandra Protocol record that makes this attestation immutable. This field is mandatory. A GABA record without a Chandra CU reference is not a valid GABA attestation.
Review date	Date by which this attestation must be reviewed and re-signed. Recommended: 12 months or upon any change to the endpoint registry.

> ***The Chandra CU reference field is mandatory and non-negotiable. An ABRAR without a Chandra CU reference is a document. An ABRAR with a Chandra CU reference is an attestation. Only the latter satisfies GABA.***

IV. Residual Risk Classes

The following residual risk classes must be explicitly named and accepted in every ABRAR. These are not optional -- partial acceptance does not constitute GABA compliance. Each risk class must be addressed by name in the record, with the authorized signatory's acknowledgment that they understand and accept it on behalf of the organization.

Residual Risk Class	Description
Provider infrastructure	The physical and virtual infrastructure operated by the AI provider is outside the deploying organization's control. Vulnerabilities in provider infrastructure cannot be eliminated by the deploying organization.
Provider key management	Cryptographic keys used by the provider to sign and encrypt inference traffic are managed by the provider. The deploying organization cannot verify or rotate these keys.
Provider personnel	The provider's employees, contractors, and agents have access to provider systems. The deploying organization cannot audit or control provider personnel practices.
Provider government relationships	The provider may be subject to legal compulsion by government authorities -- including national security orders that may prohibit disclosure -- in any jurisdiction where the provider operates. The deploying organization cannot anticipate or prevent such compulsion.
Computer-use agent scope creep	Any AI agent authorized for computer-use within the governed boundary may, through prompt injection or adversarial input, be induced to take actions outside its intended scope. Scope is enforced by policy and monitoring, not by technical constraint.
Inference response integrity	The content of AI inference responses cannot be cryptographically verified as originating from an unmodified model. Response validation reduces but does not eliminate the risk of manipulated outputs.

Organizations may add deployment-specific residual risks beyond this base set. The base set is the minimum required for GABA compliance. Omitting any base class invalidates the attestation.

V. AI Agent Classification and Attestation Requirements

Not all AI endpoints present the same risk profile. An inference-only API that returns text is materially different from a computer-use agent that can operate a browser and filesystem. GABA distinguishes four agent types, each with distinct attestation requirements.

Agent Type	Capability Profile	GABA Requirement
Inference-only API	Model receives text/structured input, returns text/structured output. No system access.	Standard GABA attestation. One record per endpoint.
Agentic API with tool use	Model can invoke defined tools (search, database query, API calls) within explicit scope.	GABA attestation per endpoint plus tool registry. Each tool explicitly scoped and logged.
Computer-use agent	Agent can operate keyboard, mouse, browser, filesystem, or any graphical interface.	Separate GABA attestation per agent instance. Scope formally bounded. All actions logged as Chandra CUs. Prompt injection risk formally documented.
Autonomous agent with persistence	Agent retains state across sessions, can initiate actions without per-action human authorization.	GABA attestation plus human authorization checkpoint policy. Mandatory Chandra CU on every autonomous action. Human review cadence formally specified.

Computer-use agents and autonomous agents with persistence are highlighted because they represent a qualitatively different threat profile. An inference-only API call can be validated by schema checking the response. A computer-use agent can take actions that are not captured in any response payload. The GABA attestation for computer-use agents must explicitly address prompt injection risk, scope enforcement mechanism, and the logging strategy for actions taken.

> ***A Mac running an AI coding agent that can reach the governed server boundary is not an inference-only endpoint. It is a computer-use agent operating from an unmanaged device -- a condition that invalidates the CRC Governed Server Boundary prerequisite and renders GABA attestation inapplicable until the access is formally governed or removed.***

VI. The Chandra Protocol Integration

Chandra Protocol (chandraprotocol.com) is the attestation mechanism for GABA. Every ABRAR must be anchored to a Chandra context unit at the time of signing. The CU captures: the signatory identity, the deployment identifier, the timestamp, and a hash of the ABRAR content. The chain is immutable. The attestation cannot be altered after the fact.

Beyond the initial ABRAR, every AI inference call in a GABA-compliant deployment is logged as a Chandra CU: endpoint called, timestamp, request identifier, response schema validation result. This continuous logging is what constitutes CRC Boundary pillar score 4. The attestation is not a one-time event -- it is an ongoing chain of governed inference activity.

When the ABRAR is reviewed and re-signed -- annually or upon any change to the endpoint registry -- the renewal produces a new Chandra CU that supersedes the previous attestation. The prior attestation remains in the chain. The history is preserved. The auditor can see every version of the attestation and when it changed.

VII. GABA and the CRC Framework

GABA is the attestation standard for the CRC Boundary pillar. The relationship is explicit: an organization that completes GABA for all authorized endpoints achieves Boundary pillar score 4. An organization that has hardening controls in place but no formal GABA attestation achieves Boundary pillar score 3 at most.

CRC Boundary Score	GABA Status
Boundary 0	No endpoint registry. No logging. No validation. GABA not applicable.
Boundary 1	Endpoints inventoried. No formal controls. GABA not initiated.
Boundary 2	Rate limited, logged, response validated. GABA not yet complete.
Boundary 3	All hardening controls active. GABA attestation in progress.
Boundary 4	All controls active. GABA attestation complete. Chandra CU on record. Residual risks formally signed.

Combined with the other three CRC pillars -- Consolidate, Reduce, and Close -- a Boundary score of 4 contributes to a total MSS of 16: full CRC compliance at the Minimum Surface Standard. The General Reasoning reference architecture (DXMachine on AegisGenera, chronicled by Chandra) is designed to achieve MSS 16.

VIII. Governance and Review

A GABA attestation is not permanent. The AI inference landscape changes: providers update their infrastructure, new endpoints are added, agent capabilities expand. The ABRAR must be reviewed and re-signed under the following conditions:

Trigger	Requirement
Annual review	Every ABRAR expires after 12 months and must be re-signed by an authorized signatory. The review confirms that the endpoint registry is current, all hardening controls remain active, and the residual risk acceptance remains deliberate.
Endpoint change	Any addition, removal, or modification of an authorized endpoint requires immediate ABRAR update and re-signing. Adding an endpoint without updating the ABRAR invalidates the attestation.
Agent type change	Upgrading an inference-only endpoint to an agentic or computer-use capability requires a new attestation with the appropriate agent-type requirements applied.
Provider material change	A material change at the provider -- acquisition, jurisdiction change, significant security incident -- triggers a mandatory review. The organization must affirmatively re-accept the residual risk under the new conditions.
Control failure	Any confirmed failure of a hardening control (certificate mismatch, circuit breaker trigger, schema validation failure in production) requires review and re-attestation.

IX. Publication and Licensing

The GABA Standard is published by General Reasoning, Inc. under the MIT License. It is an open standard. Any organization may implement GABA independently of General Reasoning's products. The Chandra Protocol integration is strongly recommended but the ABRAR structure and residual risk framework are usable with any immutable audit mechanism.

General Reasoning publishes GABA as a contribution to the field of governed AI deployment practice, in the same spirit as the CRC Minimum Surface Standard. The goal is not proprietary advantage -- it is a common, auditor-recognizable framework that regulated organizations can use to demonstrate deliberate AI governance rather than accidental AI deployment.

> ***The gap that GABA fills has existed since the first regulated organization called an external AI API. Every organization that has done so without a formal residual risk acceptance record has accepted that risk implicitly. GABA makes it explicit. That is the entire contribution.***