Governance & Policy Appendix

Semantic Security Gateway Firewall (SSGF)

A. Purpose

This appendix defines the **governance model, policy authority, and accountability boundaries** of SSGF. Its purpose is to ensure that security decisions are **transparent, configurable, auditable, and aligned with organizational control structures**.

SSGF does not introduce hidden enforcement or autonomous authority. All decisions are governed by explicit policy.

B. Governance Principles

SSGF governance is based on five core principles:

1. **Determinism Over Interpretation**
   Security decisions are enforced through rules and policy, not model opinion.

2. **Explicit Authority**
   Every decision path has a clearly defined owner.

3. **Configurability Without Drift**
   Policies are configurable but constrained to prevent silent behavior changes.

4. **Auditability by Design**
   All decisions are logged and reproducible.

5. **Separation of Concerns**
   Security enforcement is decoupled from reasoning and content generation.

C. Authority and Decision Ownership

SSGF enforces a strict hierarchy of authority:

1. **Hard Security Rules**
   Non-negotiable. Cannot be overridden by models or downstream systems.

2. **Policy Configuration**
   Organization-defined rules governing thresholds, escalation, and actions.

3. **Semantic Inspection (DEEP)**
   Provides structured signals and evidence, not final authority.

4. **Downstream AI Systems**
   Responsible for content generation only after security clearance.

This hierarchy ensures **no model can grant itself permission**.

D. Policy Definition and Management

Policy Scope

Policies define:

- Allowed, warned, and blocked intent categories
- Escalation thresholds
- Logging requirements
- Deployment mode (on-prem, hybrid, cloud)

Policies do **not** define:

- Truth or correctness of content
- Ideological or moral positions
- User profiling rules

Policy Change Control

Policy changes:

- Are versioned
- Require explicit authorization
- Can be rolled back
- Are logged with timestamps and operator identity

This prevents silent or accidental policy drift.

E. Audit and Compliance

Each SSGF decision produces a structured record including:

- Input hash (no raw content required)
- Triggered rules and flags
- Entropy score
- Escalation path (FAST / DEEP)
- Final action (ALLOW / WARN / BLOCK)

These records support:

- Internal audits
- Incident response
- Regulatory reporting
- Forensic analysis

SSGF supports **data minimization** by design.

F. Deployment Governance

SSGF can be deployed in:

- On-premise environments
- Sovereign clouds
- Edge or hybrid configurations

Deployment decisions remain under **organizational control**, not vendor dependency.

SSGF does not require:

- Model retraining
- External telemetry sharing
- Persistent user identification

G. Responsibility Boundaries

SSGF is responsible for:

- Semantic security enforcement
- Policy application
- Decision logging

SSGF is not responsible for:

- Factual correctness of AI outputs
- Business logic decisions
- End-user behavior interpretation

Final responsibility for AI output remains with the system operator.

H. Ethical and Legal Alignment

SSGF avoids:

- Implicit censorship
- Behavioral surveillance
- Hidden bias enforcement
- Unexplainable moderation

All enforcement criteria are:

- Documented
- Inspectable
- Adjustable

This enables alignment with regional legal frameworks and organizational ethics policies.

I. Governance Summary

Dimension	Governance Approach
Authority	Rule- and policy-driven
Transparency	Full audit logs
Change Control	Versioned and reversible
Model Control	Non-authoritative
Compliance	Built-in, not layered

SSGF governance ensures that **control remains human-defined**, while enforcement remains **machine-consistent**.