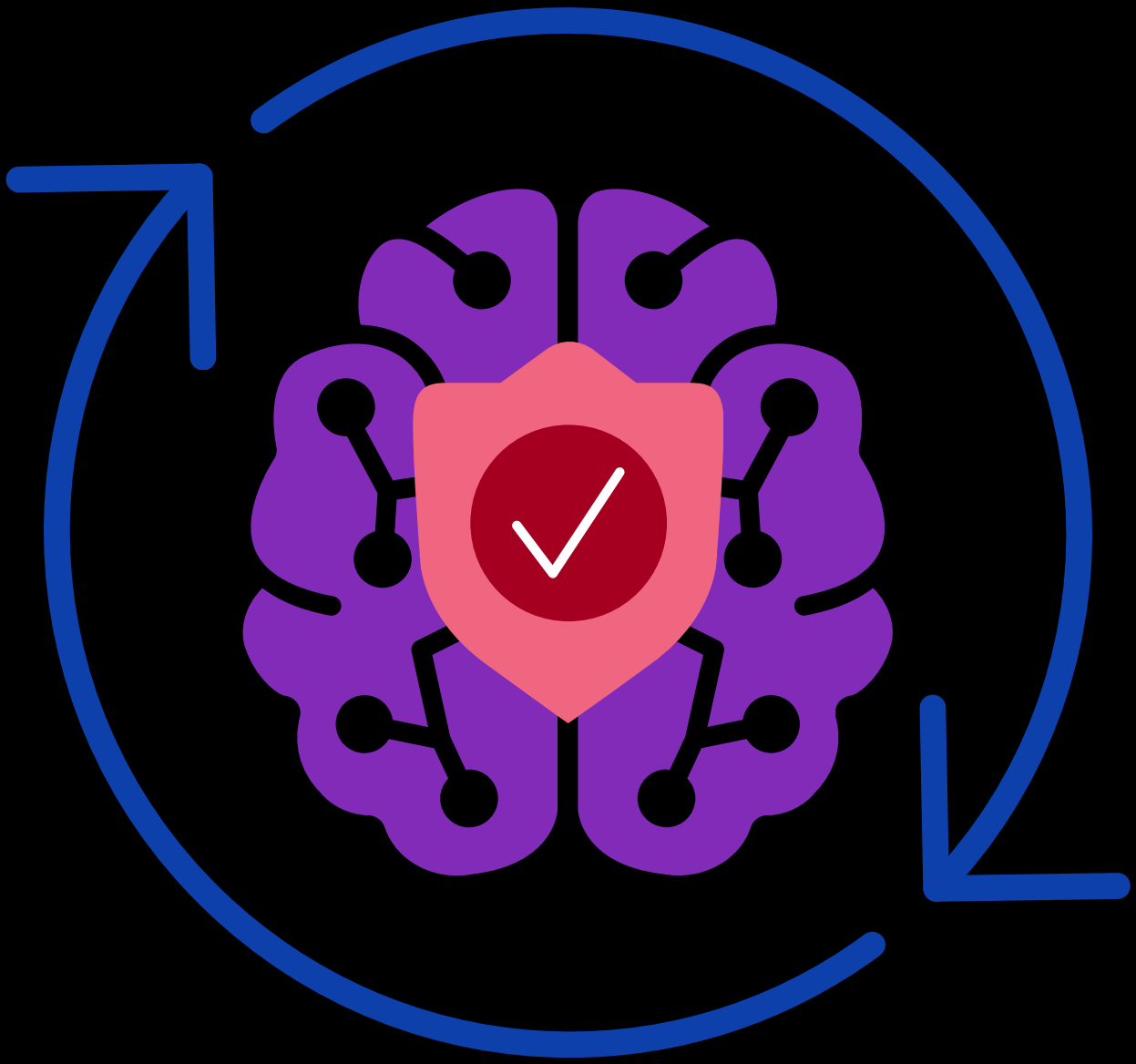
HANDBOOK

Generative AI policy handbook

Turn complexity, risk, and rapid change into a framework for safe, scalable, enterprise-ready AI.



Introduction: The need for a GenAI policy

Generative AI has shifted from experimental deployments to mission-critical infrastructure.

The race is on to enable a stronger competitive edge through AI, but many enterprises still lack the centralized controls, visibility, and policy enforcement required for secure success.

At the same time, the complexity is compounding: new open-source models drop weekly, SaaS tools ship with AI capabilities baked in, and cross-functional teams—from developers to marketers—are experimenting independently. As model capabilities evolve, so do the attack surfaces.

Generative AI isn't just software. It's a constantly-evolving, unpredictable system that requires new modes of control.

A single prompt can lead to brand-damaging outputs, sensitive data exposure, or security failures that cascade through applications.

Understanding the risk landscape

Today's generative AI risks are often misunderstood, underestimated, or completely invisible until damage is done. Unlike traditional attack surfaces, generative AI threats exploit the model's behavior, context, and interactivity—meaning even well-intentioned deployments can spiral into unintended outcomes.

The shifting threat surface

Dynamic, compound risk vectors include:

- Malicious actors using new attacks techniques such as prompt injection, jail breaking, and data poisoning
- Data leakage due to users disclosing sensitive, protected, or proprietary information
- Harmful outputs or compliance violations from the models themselves

Unlike traditional software, generative AI systems interact continuously with users and content. That interaction surface is both your greatest innovation lever—and your biggest liability.

Agentic AI and the rise of autonomous risk

As enterprises deploy agentic systems (AI that uses tools to complete tasks), risk becomes multi-dimensional:

- **Unbounded behavior:** Autonomous agents making financial, legal, or HR decisions
- **Chained logic errors:** AI acting on outputs from other AIs
- **Model Context Protocols (MCPs):** New integration pathways for attack entry

GenAI controls and security layers

Why controls must center on runtime

The apex of AI innovation—foundation model training—is controlled by a few hyperscalers who wield immense computational resources and can navigate the complexities of massive datasets.

The vast majority of enterprises operate at the runtime layer, where real work (and real risk) happens. Runtime is where:

- Users interact with models
- Data is exposed or ingested
- Outputs affect business outcomes

This approach minimizes the attack surface and allows for robust security measures to be implemented around the model and application, regardless of the model's internal complexities.

Securing AI at the runtime layer is critical and requires a holistic approach that encompasses three core areas:

1. **Defense:** Establish resilient defenses by applying input sanitization, output filtering, and runtime monitoring tailored to each use case. Controls must account for both direct threats and cross-pipeline vulnerabilities, adapting in real-time to evolving risk.
2. **Offense:** Use proactive red teaming to uncover weaknesses in both models and their applications. Through adversarial testing and agentic attacks, organizations can identify, score, and prepare for emerging attack paths before they're exploited.
3. **Governance, regulation, and compliance:** Map evolving regulatory frameworks to practical controls. Rather than relying solely on "paper policies," translate compliance requirements into enforceable guardrails—ensuring policies are lived, not just written.

Core controls to prioritize

To achieve a holistic strategy, layered protections, dynamic enforcement, and continuous adaptation is required. The following categories represent foundational building blocks for securing generative AI at the point of interaction—the runtime layer—where users, data, and models converge. These controls should be customized based on organizational risk appetite, use case sensitivity, and model complexity.

These controls form the basis of a modern AI policy framework—one that focuses not just on documentation, but on living, adaptable, and enforceable practices:

Input controls

- Prompt sanitization
- Forbidden term filters (e.g., internal project names, employee data)

Output controls

- Toxicity filtering
- PII redaction
- Response scoring (alignment confidence)

User role management

- Role-based access to models
- Tiered use-case approval (e.g., low-risk marketing chatbot vs. finance agent)

Runtime observability

- Logging every interaction
- Anomaly detection (e.g., excessive prompt length, unusual access patterns)

Red-teaming for AI

- Signature attack databases (e.g., jailbreak prompts)
- Agentic resistance (privilege escalation, denial-of-wallet)
- Risk scoring for models and applications
- AI agent risk management (agent specific runtime security controls, identity governance for nonhuman identities, and policy enforcement frameworks that extend to multiagent systems).

Shadow AI detection

- Discovery of unauthorized AI tool usage
- Monitoring for data egress via generative AI interfaces

Customizable control layers

- Unified security platforms that support customizable controls with the ability to adjust policy enforcement to specific needs (i.e., user role, model type, use case, and sensitivity).
- Support both conservative and innovation-oriented teams within the same organization.

Policies are only useful if they are:

- **Context-aware** (tailored to use case, data sensitivity, model risk)
- **Operationally enforceable** (can be embedded in workflows and tooling)
- **Adaptable** (updated as models or integrations evolve)

Building a generative AI policy framework

Key policy domains to cover

Acceptable use

- Prohibited generative AI uses (e.g., legal advice, financial forecasting without oversight)
- Authorized model providers and plugins

Prompt guidelines

- Redline prompt inputs (e.g., “do not include PII,” “do not recreate copyrighted content”)
- Templates for safe prompting

Data classification and protection

- Data tiering (public, internal, sensitive, regulated)
- Controls for regulated inputs (e.g., healthcare, finance, personal data)

Model access policies

- Permissions by role and department
- Approval processes for new models or integrations

Generative AI governance is about building adaptive, enforceable policies that reflect how it’s being used, who’s using it, and what’s at risk

Testing and evaluation

- Red team thresholds before production use
- Scoring systems for model security and alignment

Logging, retention, and incident response

- Required logging of model interactions
- Escalation paths for prompt or output incidents
- AI-specific incident response mapped to SOC procedures

Third-party and SaaS AI governance

- Procurement policies for AI-enabled software
- Vendor security assessments and risk tiers

Employee awareness and training

- Role-specific AI usage education
- Safe prompting workshops and regular refreshers

Governing AI

Navigating the evolving regulatory landscape is a critical aspect of AI security. Jurisdictions and industry sectors are rolling out regulations that are specific to generative AI—ranging from the EU AI Act to sector-specific guidance in finance, healthcare, and government.

An effective governance, regulation and compliance (GRC) approach must bridge the gap between static documentation and active enforcement. This includes:

- Mapping legal and regulatory obligations to technical controls
- Ensuring auditability of AI use and decisions
- Implementing retention policies that meet sector standards
- Demonstrating policy compliance during internal or external reviews

Crucially, compliance should not exist in isolation. Your GRC posture should be integrated with policy frameworks, runtime controls, and governance structures to ensure defensibility and agility.

Governance as continuous oversight

Traditional governance is retrospective. Generative AI requires continuous oversight across lifecycle stages:

- Model selection and usage
- Prompt construction and ingestion
- Output review and escalation
- Monitoring, logging, and tuning

Governance structures to consider

- AI steering committees: Blend legal, security, innovation, data, and business roles
- Risk assessment panels: Score and review use cases before deployment
- Federated champions: Business-unit leaders who help scale policy adherence

Financial governance

Gartner predicts that by 2027, 60% of enterprises will adopt Financial Operations practices to manage AI cost unpredictability.[1] Therefore, AI-specific FinOps disciplines should be embedded into your governance strategy to track, forecast, and optimize spending across model usage, infrastructure, and orchestration workflows. Failure to implement robust financial controls is now considered a leading risk to generative AI initiatives.

Governance should not stifle innovation –it should provide clear, trusted lanes for secure adoption.

Example RACI chart

Task	Responsible	Accountable	Consulted	Informed
AI use case approval	Head of AI	CDO/CPO	Legal, CISO	Line manager
Prompt security review	Security	CISO	DevOps	All developers
Incident investigation	SOC	CISO	Legal, privacy	Executive team
Policy update and distribution	CDO	CIO	Legal, HR	All staff

Maturity model and rollout roadmap

Where are you now?

Assess your current state across three tiers:

Maturity level	Description	Traits
Reactive	Uncoordinated usage, no visibility	Shadow AI, ad hoc rules, audit risk
Proactive	Policy and controls in place	Runtime enforcment, red-teaming, AI use registry
Optimized	Fully integrated, adaptive governance	Dynamic risk-scoring, runtime-layer observability

12-month roadmap example

Quarter	Focus area	Key activities
Q1	Inventory and baseline	Model audit, stakeholder alignment
Q2	Core policy and controls	Input/output filters, red-team, role-based use
Q3	Governance setup	Launch steering group, set review cadence
Q4	Scaling secure adoption	Expand to new use cases, federated champion model

Conclusion

Top questions to ask now

1. What models are being used across our organization?
2. Who can access generative AI tools and for what use cases?
3. Do we log, audit, and analyze AI interactions?
4. Have our models been red-teamed for security and misuse?
5. Do we have acceptable use policies tied to enforcement?
6. Are we blocking unsafe prompts or responses at runtime?
7. Do we have visibility into shadow AI?
8. Are roles and permissions tailored by risk level?
9. Can we demonstrate responsible usage to regulators?
10. Who owns AI governance—and how often is it updated?

Secure GenAI usage with F5

Whether your AI is deployed as SaaS, edge-hosted, cloud-hosted, or self-hosted, F5 has you covered. Secure your AI models, applications, agents, and data from both internal misuse and external threats by preventing data leakage, monitoring endpoints, and governing AI interactions.

Explore AI runtime security solutions.

For more information, visit fullproxy.com to schedule a demo.

1 Gartner Predicts 2025: AI's Impact on the Future of Enterprise Technology