PRODUCT REQUIREMENTS DOCUMENT

Enterprise AI Governance Studio

AI Use Case Intake · Deterministic Risk Scoring · Policy-as-Code Generation · MCP Runtime Gate · HITL Audit

Author	Version	Status	Date
Minsun Kim	v1.6 — Adds Technical Evidence Intake (F-10) + Risk Engine Rationale	Reference implementation on GitHub	July 2026

Executive Summary

> *"Flagging risk is easy. Enabling responsible deployment is the product."*

The Enterprise AI Governance Studio enables governance decision intelligence: it turns complex, evolving regulatory requirements into explainable, actionable product decisions — evaluating, scoring, and governing AI use cases before deployment, and, as of v1.1, enforcing the resulting policy at the point an agent actually tries to act, not only on paper.

v1.0 of this PRD scoped three workflows: AI Use Case Intake, structured Risk Assessment, and LLM-powered Governance Recommendations. v1.1 added a working reference implementation and a machine-enforceable MCP tool-calling policy manifest. v1.2 added Section 09, scoping how this PRD is packaged for external review. v1.3 documents F-07 (Playbook Engine) and four illustrative demo scenarios. v1.4 adds market and business framing and the Risk Matrix concept in F-02. v1.5 adds the product's lifecycle position, design principles, business-outcome targets, a roadmap-only Policy-as-Code export, and the live demo's wireframed information architecture.

v1.6 does two things: it documents the rationale behind F-02's risk-engine design — specifically, why 8 binary factors, not a different number or a probabilistic model — which existed as a design choice before it existed as a written explanation; and it adds F-10 (Technical Evidence Intake & Review), a staged capability for attaching and reviewing the actual artifact behind a use case (a repository, a prompt, a model card) rather than only a self-reported checklist. F-10 is deliberately staged across three tiers of increasing capability and decreasing certainty, and Tier 3 is named as an explicit non-goal rather than a future roadmap item — see F-10 below for why.

Full source, tests, and setup instructions: github.com/[username]/ai-governance-studio (update link before publishing). All customer personas, metrics targets, and company references remain illustrative — this is a portfolio project, stated plainly rather than implied otherwise.

Problem Statement & Background

Background

Enterprise organizations are deploying AI at an unprecedented pace. According to McKinsey (2024), over 65% of enterprises have integrated generative AI into at least one business function — up from 33% just twelve months prior. Governance processes have not kept pace with adoption speed.

Through direct experience contributing to enterprise AI transformation initiatives, a consistent pattern emerged: business teams struggled to assess AI risks before deployment, governance decisions were fragmented across departments, and risk evaluations relied on ad hoc manual review with no structured framework.

Core Problems

Enterprise organizations increasingly recognize AI as a strategic capability; however, the governance operating model has not evolved at the same pace. AI initiatives are expanding rapidly across business functions, while governance processes remain fragmented, manual, and difficult to operationalize. This creates significant friction between business velocity and responsible AI adoption.

- **1. Governance Cannot Keep Pace with AI Innovation.** Enterprise AI capabilities evolve faster than internal governance processes, regulatory guidance, and organizational policies. The result is a growing gap between innovation and governance, forcing organizations to choose between deployment speed and risk management rather than achieving both simultaneously.
- **2. AI Use Cases Lack a Standardized Intake & Evaluation Framework.** AI initiatives are often initiated independently across business units without a common governance framework, making governance decisions difficult to compare across the enterprise.
- **3. Risk Assessment is Manual, Subjective, and Difficult to Scale.** Risk assessments are frequently performed through spreadsheets and manual interpretation. Identical AI use cases may receive different decisions depending on the reviewer; governance specialists become operational bottlenecks rather than strategic partners.
- **4. Governance Knowledge is Difficult to Operationalize.** Product teams often understand that governance requirements exist but lack clarity on how to apply them during product development. This dependency on manual consultation increases cognitive load and slows product delivery.
- **5. Governance Identifies Risks but Rarely Enables Action.** Most governance processes end once risks have been identified. Product teams are often left without clear guidance on which controls to implement, who owns each action, or how recommendations translate into engineering tasks.

> *The Systemic Gap: collectively, these represent a gap between the pace of enterprise AI innovation and the organization's ability to govern AI consistently at scale. Organizations need an operational platform that standardizes evaluation, translates requirements into actionable plans, and embeds responsible AI throughout the product lifecycle.*

Product Rationale: Why AI Governance Studio?

Enterprise AI adoption is accelerating faster than organizations can operationalize governance. AI Governance Studio was created to bridge this gap by transforming governance from a fragmented review process into a standardized product capability — enabling teams to understand what actions must be taken, who is responsible, and what evidence is required to deploy AI responsibly at enterprise scale.

Market Context & The Governance Gap

Market Context: The rapid adoption of generative AI has fundamentally changed how organizations build and deploy AI-powered products. While some enterprises develop AI solutions internally, a significant portion collaborates with system integrators, consulting firms, or technology partners to accelerate implementation.

Accountability: Regardless of the development sourcing strategy (build vs. buy/partner), the legal and reputational accountability for AI deployment ultimately resides with the enterprise.

Operational Bottleneck: This is the same fragmentation described in Core Problems #1 and #3 above, viewed from a delivery-accountability angle. Governance can no longer function as a manual, end-of-cycle bottleneck; it must be embedded as a continuous capability throughout the AI product lifecycle.

The Solution: Enterprise Governance Control Plane

AI Governance Studio is designed to bridge the gap between rapid AI innovation and enterprise risk management. Rather than replacing existing AI development tools or LLM orchestration platforms, it serves as an overarching Enterprise Governance Layer.

- **Core Capabilities:** Standardizes AI use case intake, automates risk assessment, maps regulatory obligations, recommends actionable governance controls, and orchestrates workflows before deployment.
- **Value Proposition:** Transforms governance from a reactive compliance hurdle into a proactive, execution-enabling product capability.
- **Architectural Principle:** AI Governance Studio is strictly platform-agnostic, enabling organizations to govern AI consistently regardless of whether solutions are built internally or delivered through external vendors.

Target Market & Customer Profile

Primary Target Customers (User Personas)

- **Enterprise Product Management:** PMs introducing AI-powered products and copilots who need clear, actionable guardrails to ship faster.
- **Responsible AI & Governance Teams:** Centers of Excellence (CoE) establishing and scaling centralized AI governance standards.
- **Risk, Security, Legal & Compliance:** Teams overseeing AI adoption, requiring auditability and runtime trust boundaries.
- **Digital Transformation & Enterprise Architecture:** Leaders managing cross-functional, multi-cloud AI initiatives.
- **Consulting & System Integration (SI) Partners:** External vendors implementing AI solutions for enterprise clients, requiring a standardized framework to prove compliance.

Ideal Customer Profile (ICP)

- **Company Size:** Mid-market to global enterprise. Early-stage startups typically don't yet have the dedicated governance, security, or compliance roles this platform is designed to support.
- **Adoption Scale:** Managing multiple AI initiatives distributed across various teams or business units.
- **Industry Context:** Industry-Agnostic, with highly regulated sectors as natural early adopters.
- **Technical Stack:** A hybrid mix of commercial LLMs and/or internally developed AI solutions.
- **Operational Need:** Consistent, vendor-neutral governance across distributed and complex AI portfolios.

Product Vision & Goals

> *"To make AI governance as easy as filing an expense report. Flagging a risk is the easy part; telling the business exactly what to do about it, and enforcing that once approved, is the product."*

Goal	Description	Success signal
Standardize intake	Replace ad hoc submission with a structured intake process	100% of evaluated use cases go through the intake form
Quantify risk	Deterministic scoring engine producing reproducible scores	Risk scores identical across repeated runs for the same inputs (verified by test suite)
Translate risk into action	Deterministically derive an owner, review cadence, and control checklist from the risk score — not just a score	100% of assessed use cases receive a Playbook with zero LLM involvement (see F-07)
Translate policy into	Convert LLM-generated guidance into a machine-	100% of generated

enforcement	readable MCP policy manifest	manifests pass schema validation before reaching the audit store
Enforce at runtime	Gate simulated agent tool calls against the generated manifest	Policy-denial outcomes match expected outcomes in the automated test suite

The platform is designed to transform governance from a reactive approval process into a proactive product capability — embedded throughout the AI product lifecycle rather than functioning as a final gate.

Business Outcomes (Target — Not Yet Measured)

> **These are business-level outcomes this product is designed to produce at enterprise scale. Unlike Section 06A's metrics, none of these are currently measured — stated as design targets, not verified results.**

Metric	Target	Why it matters
Governance review cycle time	↓ 70% (illustrative target)	Faster deployment of AI initiatives
Early governance adoption	> 80% of AI projects assessed during design phase	Shifts governance left, before build
Manual compliance effort	↓ 50% (illustrative target)	Reduces operational overhead on legal/security
Policy consistency across business units	> 95% inter-reviewer agreement	Standardized enterprise decisions, not reviewer-dependent
Playbook adoption rate	> 75% of recommendations implemented	Measures whether generated guidance is actually used, not just generated

Product Position in the AI Lifecycle

AI Governance Studio embeds governance throughout the AI product lifecycle rather than treating compliance as a final approval checkpoint.

↓ Business Idea / AI Use Case Registration
↓ Risk Assessment (Risk Matrix + Risk Engine — F-02, Built)
↓ Framework Mapping (F-03, Built)
↓ Governance Playbook (F-07, Built)
↓ Approval Workflow (F-04 HITL decision — partially built; multi-step dual sign-off is Roadmap)
↓ Development & Deployment (Not built in this MVP)
↓ Continuous Monitoring (Not built — Roadmap)
↓ Policy Improvement Loop (Not built — Roadmap)

> *This MVP covers the first five stages end-to-end. The last three are a long-term vision, not a current capability.*

Product Design Principles

- **1. Explainable Before Intelligent.** Core governance decisions must be deterministic, transparent, and fully auditable before any AI-generated content is introduced — the same principle documented as "Deterministic Before Intelligent" under F-02 and Design Decision T-06, restated here as a company-wide principle.

- **2. Governance by Design.** Governance should be embedded from product discovery through deployment — not performed as a final approval gate.
- **3. Human Accountability Over Automation.** AI assists decision-making, but governance ownership always remains with accountable human reviewers. See F-04's HITL model.
- **4. Enterprise First.** Every recommendation must align with enterprise governance processes, regulatory obligations, and operational realities.
- **5. Actionability Over Assessment.** Governance is valuable only when risk assessments translate into concrete implementation guidance. This is why F-07 (Playbook) exists.

Functional Requirements — MVP Scope (v1.6)

F-01 | Standardized AI Use Case Intake — P0 — Built

- Structured form capturing use case name, business unit, description, persona, output type, and problem statement (min. 50 characters).
- **Validation:** Pydantic-validated schema shared across every downstream module — the same typed contract that risk scoring, policy generation, and the MCP gate all consume.

F-02 | Deterministic Risk Scoring Engine — P0 — Built

Dynamic Scoring: 8 binary risk factors with fixed point weights, summed into a reproducible 0–130 score (Low 0–29 / Medium 30–59 / High 60–130).

The Governance Risk Engine abandons the probabilistic, black-box approach to risk assessment. Instead, it evaluates every AI use case through a structured, three-stage deterministic process based on AIGP (AI Governance Professional) principles.

Stage 1 — Context Classification (The Risk Matrix)

Before calculating a numerical score, the system plots the use case on the Enterprise AI Risk Matrix to determine its baseline governance tier and required approval routing.

[X-Axis] Level	Enterprise Impact	Typical AI Use Cases
Low	Low business impact, limited to operational efficiency.	Internal productivity assistants, code generation (internal only).
Medium	Direct customer interaction but limited to advisory or support roles.	Customer support chatbots, knowledge base search.
High	Direct impact on revenue, legal exposure, or human well-being.	HR hiring decisions, financial approvals, healthcare diagnostics.

[Y-Axis] Level	Regulatory & Security Impact	Characteristics
Low	Publicly available data, zero regulatory friction.	Processing public web data, general marketing copy.
Medium	Internal business data, limited exposure.	Internal financial summaries, proprietary IP processing.
High	Regulated data, safety implications, high autonomy.	Processing PII/PHI, autonomous DB execution, regulated industries.

	Low Governance Exposure	High Governance Exposure
High Business Criticality	Enhanced Review (Business VP Approval)	AI Governance Committee (Mandatory HITL + Board-level Review)
Low Business Criticality	Fast Track Approval (Automated Guardrails, No Manual Review)	Standard Review (Legal & Compliance Approval)

Precedence rule: the Risk Matrix sets an initial routing hint before the detailed dimensions are scored in Stage 2. The cumulative point score computed in Stage 3 is the final, authoritative determinant of risk level and approval route — if the two disagree, Stage 3 takes precedence, since it reflects the specific combination of risk factors rather than a coarse 2×2 approximation.

Stage 2 — Risk Dimension Assessment

Once the baseline context is established, the engine evaluates 8 standardized governance dimensions. Each dimension is assessed as either Present (1) or Not Present (0), contributing a fixed point weight to the total score.

Risk Dimension	Weight	Compliance & Enterprise Justification
Personal Data / Privacy	20 pts	GDPR, CCPA, and PII protection mandates.
Sensitive Data / Enterprise IP	20 pts	Protection of trade secrets and internal confidential assets, distinct from personal data — see overlap note below.
Autonomous Decision Making	20 pts	Lack of human oversight increases systemic failure impact.
High-Impact Domain	20 pts	EU AI Act (e.g., Healthcare, Hiring, Financial decisions).
External Users / Reputation	15 pts	Customer-facing outputs carry direct brand and legal exposure.
Model Transparency	15 pts	"Black-box" limitations impact explainability and user awareness.
Security Exposure	10 pts	Direct write access or autonomous execution against internal production systems (narrower than general system integration — see note below).
Third-Party Dependency	10 pts	Reliance on external foundation models (vendor risk management).
Maximum Cumulative Risk	130 pts	Determines the strictness of the generated Playbook

Overlap note: Sensitive Data/Enterprise IP and Security Exposure can both trigger for the same use case (e.g., an agent with production write access to proprietary data) — this is intentional, not double-counting: one flags what's at risk (the data), the other flags how it could be exposed (the access path), and a use case with both deserves the combined weight. The same logic applies to High-Impact Domain and External Users/Reputation overlapping on a customer-facing healthcare or financial product. Stated here explicitly so a reviewer doesn't have to guess whether this is a design flaw.

Why 8 factors, and why binary

Binary (Present/Not Present) rather than a 1-to-5 subjective scale exists to solve Core Problem #3 directly: a 1-to-5 scale is exactly where reviewer-to-reviewer inconsistency creeps in — "is this a 3 or a 4" is a judgment call, while "is PII present" is a fact check most reviewers will answer the same way. Severity is encoded once, in the fixed weight assigned to each factor at design time, not re-judged on every submission.

Eight factors, specifically, reflects three constraints working together, not an arbitrary round number: (1) it's a realistic upper bound for a form a reviewer can complete in the couple of minutes the Product Vision's "as easy as an expense report" goal implies — twenty or thirty factors would be more precise but would break that goal; (2) eight binary factors produce 256 distinct combinations, giving the 0–130 range enough resolution to distinguish "barely High" from "maximally High" rather than collapsing everything into three flat buckets; and (3) each factor is

designed to correspond to a specific regulatory concern surfaced later in F-03's framework mapping (Personal Data → GDPR, Autonomous Decision Making → EU AI Act high-risk provisions, Third-Party Dependency → vendor risk) — so the count of 8 is a product of F-02 and F-03 being designed to interlock, not a coincidence.

Stage 3 — Governance Decision & Playbook Routing

The aggregated score does not just label the risk; it acts as an operational trigger that generates the Governance Playbook and enforces the approval route.

Example workflow for a score of 72:

- **Triggered Playbook (Technical):** PII Redaction Enabled ➔ Encryption Required ➔ Output Disclaimer Forced.
- **Triggered Playbook (Business):** Quarterly Review Scheduled ➔ Human Review (HITL) Required.
- **Enforced Approval Route:** Legal ➔ Security ➔ Responsible AI Committee ➔ Safe Deployment.

Design Decision: Deterministic Before Intelligent. While many AI governance platforms rely on LLM-generated recommendations, this platform intentionally prioritizes deterministic governance logic for core risk evaluation. LLMs are used to augment governance by generating contextual summaries and framework citations after the deterministic assessment — not to replace the governance decision itself, and not to author the enforcement manifest (see Design Decision T-06: the MCP tool-calling policy manifest is generated by deterministic code, not the LLM).

Design Decision: Why Binary and Why 130 Points? Not machine learning. Not probabilistic. 100% deterministic. Identical inputs always produce identical outcomes. The maximum score (130) is not a percentage — it represents cumulative risk severity, letting individual dimensions be recalibrated independently as external laws evolve.

> **Consistency note: because this engine is explicitly non-probabilistic by design, it cannot legitimately produce a "confidence" percentage — there is no probability distribution to be confident about. A wireframed "Confidence 94%" UI element (documented in Section 09-B) contradicts this design decision and should be removed or re-scoped before implementation.**

Reproducibility by design: pure function with zero LLM involvement. Visualization: real-time score breakdown showing each active factor's point contribution, verified by an automated test suite.

F-07 | Playbook Engine (Risk-to-Action Plan) — P0 — Built

This platform does not stop at flagging risk. Following each assessment, this engine deterministically derives what to actually do about it — the same non-LLM principle used for the MCP policy manifest in F-03 (see Design Decision T-06): "what should we do about this risk" is exactly the kind of question a model sounds fluent answering and is hardest to hold accountable for.

- **Recommended Actions:** A technical/business control checklist, where every recommended control is traceable to the specific risk factor or risk-level rule that triggered it.
- **Ownership & cadence:** A suggested owner (mapped from risk level to role) and a review cadence (30 / 90 / 180 days by risk tier).
- **Estimated residual risk:** An estimated residual risk assuming recommended controls are applied — explicitly labeled as an illustrative, one-tier-down heuristic, not a measured outcome.

> *This is the functional difference between a checklist tool and an operational one: a reviewer doesn't just see "High risk" — they see who owns it, what to enable this week, and when it gets re-reviewed.*

F-03 | Actionable Guardrail & Policy-as-Code Generation Engine — P0 — Built

Dynamic Framework Mapping Pipeline. Maps the computed risk profile to the most relevant governance, regulatory, and security frameworks through a structured LLM classification pipeline. The system selectively maps frameworks based on the deterministic risk context (e.g., surfacing HIPAA only if the use case involves healthcare data).

- **Core AI Governance Frameworks:** EU AI Act, NIST AI RMF, ISO/IEC 42001, ISO/IEC 23894 — mapped based on general AI risk tier.
- **Responsible AI Principles:** OECD AI Principles — baseline alignment on fairness, transparency, accountability.
- **Privacy & Security:** GDPR (if PII is flagged), SOC 2 (external data sharing / cloud dependencies).
- **Industry-Specific:** ISO 27001, HIPAA, PCI DSS, CCPA/PIPEDA — conditional, industry-specific triggers.

PM Rationale: by structuring the LLM prompt to map against these predefined categories, product teams receive a targeted "Compliance Bill of Materials" reflecting their specific implementation, reducing cognitive load on legal and security reviewers.

- **Developer-ready output:** Generates a paste-able system-prompt constraint targeting the use case's top active risk factor.
- **MCP Tool Calling Policy Manifest:** Generates an MCP (Model Context Protocol) tool-calling policy manifest defining which tool categories an agent may invoke, and whether an HITL approval token is required — derived by deterministic code, not by the LLM. See Design Decision T-06.
- **LLM prompt discipline:** Structured system prompt and Pydantic schema validation enforce deterministic, syntax-error-free JSON output.

F-10 | Technical Evidence Intake & Review — P1 — Partially Scoped (Staged Rollout)

Most AI governance tools stop at a self-reported questionnaire. This platform is designed to let a reviewer look at the actual artifact — the repository, the prompt, the model card — rather than only what the intake form's checkboxes claim. "Look at the artifact" spans a wide range of actual capability, from "store the link" to "rewrite the code," so this is staged across three tiers rather than claimed as one feature.

Tier 1 | Evidence Attachment — P1 — Roadmap

Technical Evidence becomes part of intake (F-01): a GitHub repository URL, an uploaded prompt/system-prompt file, or a model card document, attached to the use case record and stored alongside the audit trail.

- **Presence as a risk factor:** Whether required evidence (e.g., a model card) is present or missing becomes a new deterministic risk factor in F-02 — no LLM involved, consistent with Deterministic Before Intelligent.
- **Scope limit:** No content analysis at this tier — the evidence is stored and linked, not read.

Tier 2 | LLM-Assisted Evidence Observations — P2 — Roadmap

Once evidence is attached, an LLM call (structurally similar to F-03's narrative generation) reads the attached prompt file or model card text and surfaces observations for a human reviewer — e.g., "no PII-masking instruction found in the system prompt."

- **Observations, not judgments:** Worded and displayed as suggestions for the reviewer to verify, never as a pass/fail determination — the same narrative-only role the LLM plays in F-03.
- **Scope limit:** This tier reads only the documents the user directly uploads, not the codebase behind a linked GitHub URL.

Tier 3 | Repository-Level Code Remediation — Not Scoped (Explicit Non-Goal)

> **Unlike the rest of this roadmap, Tier 3 is not a "not built yet" item — it is a deliberate non-goal, stated with its reasoning rather than left as an open gap.**

Generating specific code-level remediation from a linked repository carries a materially different risk than a framework citation being wrong (see T-06): a confidently-worded but incorrect remediation could be applied as-is and introduce a real security vulnerability. Reliable repository-level remediation is its own product category (static analysis / SAST tooling for LLM applications), not a feature that can responsibly be added to a governance intake form without dedicated evaluation infrastructure this project does not have.

Tier	Capability	LLM role	Status
1	Attach GitHub URL / prompt file / model card; presence/absence as a risk factor	None (deterministic)	Roadmap (P1)
2	Surface observations from attached prompt/model-card text	Narrative only, human-reviewed	Roadmap (P2)
3	Generate specific code-line remediation from a linked repository	Would require judgment-bearing output this project won't claim	Not scoped — explicit non-goal

F-04 | HITL Approval Dashboard & Audit Trail — P0 — Built

- **Reviewer actions:** Dedicated review surface where reviewers issue Approve / Reject / Conditional Approve decisions per use case.
- **Audit logging:** Every generated manifest and every simulated tool-call attempt — allowed or denied — is logged with a timestamp, visible via the live audit feed.
- **Not yet built:** Mandatory dual sign-off (separation of duties) for High-risk use cases, and PDF export, are scoped for Phase 2.

F-09 | Policy-as-Code Export — P2 — Roadmap (Not Built)

> **Included here, explicitly labeled Roadmap, as a natural extension of this project's policy-as-code theme — not because it exists.**

Concept: after governance recommendations are approved, convert them into structured, machine-readable artifacts — JSON guardrail policies, YAML configuration templates, Markdown governance reports, PDF compliance documentation.

Design Decision T-06 — Policy Enforcement: Deterministic Code vs. LLM-Decided

	Chosen	Not chosen
Approach	LangGraph orchestration + a deterministic, code-based MCP policy gate; LLM scoped to narrative generation only	Letting the LLM directly decide and output which tools an agent may call as part of the policy manifest

An LLM-decided allow-list is a plausible-sounding string, not an enforced boundary. Routing the enforcement decision through ordinary, reviewable code means a security reviewer can read the rule set and know exactly what will happen for a given risk profile, independent of any single model generation.

> *Trade-off, stated plainly: the deterministic rule set is hand-written and does not automatically adapt to novel risk combinations the way a fully LLM-driven policy might. The Phase 2 roadmap item is a rules DSL a compliance team can edit without a code change.*

Section 06A — AI System Evaluation Framework

A Product Manager building enforcement-adjacent AI tooling has to own the measurable quality of what the system claims. Two metrics are implemented and verifiable directly from the automated test suite; a third is explicitly listed as not yet measured.

Metric	Target	Method
Schema / manifest validity rate	100%	Every manifest reaching the audit store has passed Pydantic validation; malformed LLM output triggers an automatic retry.
Policy denial accuracy	100%	The MCP gate is deterministic code, so expected deny/allow outcomes are asserted directly in the automated test suite.
Regulatory citation grounding	Not yet measured	Requires a labeled evaluation set reviewed by someone with compliance expertise. Logged as an open gap (target from v1.0: >85%).

MVP Feature Prioritization (v1.6)

Feature	Priority	Status
AI Use Case Intake Form	P0	Built
Deterministic Risk Scoring Engine (incl. Risk Matrix)	P0	Built
Playbook Engine (Risk-to-Action Plan)	P0	Built
Actionable Guardrail & Policy-as-Code Generation	P0	Built
HITL Approval Dashboard + live audit feed	P0	Built
Dual sign-off (separation of duties) for High-risk use cases	P1	Roadmap
Assessment PDF export	P1	Roadmap
Technical Evidence Attachment (F-10 Tier 1)	P1	Roadmap
Regulatory citation grounding eval set	P1	Roadmap
Article-level framework mapping (Article → Control → Evidence)	P2	Roadmap
Technical Evidence LLM Observations (F-10 Tier 2)	P2	Roadmap
Policy-as-Code Export (F-09: JSON/YAML/Markdown/PDF)	P2	Roadmap
Red Team Module (adversarial prompt testing)	P2	Roadmap
Monitoring Dashboard (portfolio-level risk visibility)	P2	Roadmap
Policy Center (custom policy authoring DSL)	P3	Roadmap
Repository-level code remediation (F-10 Tier 3)	—	Explicit non-goal, not roadmap

Long-Term Vision (North Star)

> *The following is a long-term aspiration, not a near-term roadmap commitment.*

Build the operating system for enterprise AI governance — where every AI product is designed, assessed, governed, deployed, and continuously improved through a unified, policy-driven platform. Rather than serving as a compliance checkpoint, AI Governance Studio becomes the governance layer that enables organizations to innovate faster while maintaining trust, accountability, and regulatory readiness.

Risks & Dependencies

Risk	Severity	Mitigation
LLM output quality	High	Structured prompts + Pydantic schema validation + retry loop; enforcement boundary kept out of LLM's hands entirely (see T-06).
Hand-written policy rules don't generalize	Medium	Explicit MVP ceiling; Phase 2 roadmap item is a reviewable rules DSL.
Regulatory change	Medium	Prompt templates versioned; quarterly framework review planned.
Portfolio validation	Low	Structured usability tests with 5–8 target-persona reviewers planned as proxy validation.
Risk-factor overlap (Sensitive Data/IP vs. Security Exposure; High-Impact Domain vs. External Users) reads as double-counting	Low	Explicitly documented as intentional combined-weight scoring in F-02, not a design defect.
Technical Evidence review (F-10) mistaken for full code analysis	Medium	Three-tier staging with Tier 3 named as an explicit non-goal, not a roadmap gap, so the boundary is stated rather than implied.

Section 09 — External Case Study Packaging Requirements

Background

A reference implementation is only useful to a hiring evaluation if it is legible in a bounded review window. Industry data on AI product management hiring loops in 2026 indicates reviewers spend approximately 90 seconds on a resume and approximately 8 minutes on a case study before forming a judgment.

F-05 | Case Study Narrative Structure — P0 — Required for external review

- **Problem framing —** drawing on the Background section above and direct enterprise experience with AI governance gaps.
- **Product decision and trade-off —** drawing directly on Design Decision T-06 — chosen approach, rejected alternative, rationale, and one acknowledged limitation.
- **System workflow —** a single diagram of the LangGraph sequence (Intake → Risk Score → Playbook → Policy Manifest → MCP Gate → Audit Log), annotated to distinguish deterministic steps from LLM-invoked steps.
- **Working demonstration —** using the existing frontend interface, showing a risk score changing in response to input, the automated generation of the Playbook, and the MCP gate denying a disallowed simulated tool call.
- **Evaluation and open gaps —** reusing Section 06A directly, including the explicit statement that regulatory citation grounding is not yet measured.

> *Acceptance criterion: every claim in the packaged case study must trace to a section already defined in this PRD (F-01–F-04, F-07, F-09, F-10, T-06, 06A) or to the Risks & Dependencies table. No new capability claim may be introduced at the packaging stage that does not exist in the implementation.*

Section 09-A | Illustrative Demonstration Scenarios

Use case	Risk level	Why

Marketing Content Assistant	Low	Baseline case — third-party LLM only, no customer-facing or PII exposure.
HR Resume Screener	Medium	Contains PII and is bias/fairness-sensitive.
Customer Support Bot	Medium	Customer-facing, contains PII, uses a third-party LLM.
Finance Report Generator	High	Autonomous, regulated industry, financial decision, contains PII.

> *These are illustrative scenarios constructed to exercise the risk engine's full range, not real customer or company data.*

Section 09-B | Live Demo Information Architecture (Wireframed)

> **Everything in this subsection describes a wireframe design, not shipped software. The reference implementation currently ships a six-item navigation (per F-05 above); the structure below is the target design from the latest wireframe pass and has not yet been rebuilt into the codebase.**

- **Overview —** Organization-wide summary (Risk Distribution, Recent Activity, Top AI Systems by Risk) plus a detail panel for the selected use case.
- **New Intake —** Structured form including F-10 Tier 1 evidence fields. Submitting reveals an AI Analysis Summary panel.
- **Risk Assessment —** Select any previously submitted use case and view the same analysis panel shown after intake.
- **Playbook —** Unchanged from the current build (F-07).
- **Audit Trail —** Unchanged from the current build (F-04).

> *Two consistency gaps this wireframe pass surfaced, carried forward rather than silently resolved: (1) a "Confidence 94%" figure contradicts F-02's non-probabilistic design (see F-02 note above). (2) Several new intake fields (Email, Foundation Model, Applicable Enterprise Policies) are not yet part of the F-01 schema.*

F-06 | Scope Exclusions at Packaging Stage — P0 — Required for external review

The following are explicitly excluded from the case study's built-and-demonstrated claims: AI use-case registry, human feedback loop, prompt versioning, dual sign-off workflow, PDF export, red team module, monitoring dashboard, policy authoring DSL, Policy-as-Code export (F-09), article-level framework mapping, F-10 Tier 2 (LLM evidence observations), and F-10 Tier 3 (repository-level remediation — a non-goal, not a roadmap gap).

Appendix — Governance Frameworks Referenced

NIST AI RMF

A voluntary framework for managing AI risk across four functions: Govern, Map, Measure, Manage.

EU AI Act

Classifies AI systems into risk tiers; high-risk systems face conformity assessment, transparency, and human-oversight requirements.

ISO 42001

An internationally recognized AI management system standard for organizations seeking formal governance certification.

ISO/IEC 23894

AI risk management guidance, an AI-specific extension to ISO 31000.

OECD AI Principles

International principles for trustworthy AI: inclusive growth, human-centered values, transparency, robustness, and accountability.

GDPR

The EU's data protection regulation; triggered when a use case involves PII, profiling, or automated decisions affecting EU individuals.

SOC 2

An AICPA audit standard covering security, availability, and confidentiality of a service organization's systems.

ISO/IEC 27001

The international standard for information security management systems.

HIPAA

US law governing the privacy and security of Protected Health Information; triggered exclusively for healthcare-context use cases.

PCI DSS

Security requirements for organizations handling cardholder data.

CCPA / PIPEDA

Regional consumer data privacy regulations (California / Canada), analogous to GDPR for non-EU jurisdictions.

Author's Note

This PRD was authored by Minsun Kim as a portfolio project demonstrating AI Product Manager capabilities. v1.1 built and tested the thing being specified; v1.2 packaged it for external review; v1.3 corrected a documentation gap; v1.4 added market framing; v1.5 added lifecycle and design-principle framing; v1.6 documents the reasoning behind the risk engine's core design choice and adds a staged, explicitly-bounded approach to reviewing technical evidence — including naming what this system will not attempt, and why.

Minsun Kim · minsun.kim0819@gmail.com · linkedin.com/in/minsun-kim0926