R&D Work Plan vission

Overall Development Framework

The proposed R&D program is structured as a multi-phase technological development process aimed at building a hybrid AI-driven semantic evaluation and grant intelligence infrastructure. The work plan reflects staged risk reduction, architectural validation, calibration research, and production hardening. The development contains interdependent research components in natural language processing, semantic similarity modeling, distributed systems engineering, cryptographic architecture, and AI governance mechanisms.

The total R&D timeline is projected at 18–24 months and is divided into core engine development and extended intelligent automation modules.

Phase 1: Core Semantic Evaluation Engine (Months 1–6)

The first phase focuses on designing and validating the hybrid probabilistic–deterministic evaluation architecture. During this period, the team will formalize the claim decomposition methodology, define structured JSON schema enforcement mechanisms, and develop the deterministic scoring matrix framework. The primary technological uncertainty at this stage concerns the ability to reliably transform unstructured answers into atomic claims while preserving semantic fidelity.

In parallel, embedding-based semantic similarity infrastructure will be implemented. This includes experimentation with document chunking strategies, similarity threshold calibration, cosine similarity modeling, and validation datasets to measure false-positive and false-negative evidence validation rates. Controlled experiments will be conducted to determine optimal embedding models and segmentation parameters.

By the end of this phase, a functioning prototype capable of evaluating answers against reference materials with structured scoring outputs will be operational in a controlled environment. The milestone for completion will be the successful demonstration of stable, reproducible evaluations across a statistically significant internal validation dataset.

Budget justification for this phase is primarily allocated to senior AI engineers, a backend architect, and research personnel specializing in NLP and embedding calibration. Additional costs include secure cloud infrastructure for model interaction and experimentation environments. The concentration of budget in this phase reflects the core technological risk inherent in semantic validation and deterministic scoring alignment.

Phase 2: Parallel Validation and Observability Infrastructure (Months 5–9)

The second phase overlaps partially with Phase 1 and focuses on scalability and production-grade reliability. The objective is to transform the prototype engine into a resilient system capable of handling concurrent evaluations under rate-limited API environments.

This stage includes the development of asynchronous orchestration logic, controlled concurrency management, retry and backoff strategies, and token usage optimization. A significant technical effort will be dedicated to maintaining deterministic outputs under asynchronous processing conditions.

Simultaneously, observability infrastructure will be implemented, including telemetry collection for token consumption, latency measurement, semantic similarity distribution tracking, scoring variance monitoring, and error rate analysis. This layer is essential for long-term calibration and institutional adoption.

The milestone for this phase will be a validated load-tested system operating under simulated production traffic, demonstrating stable performance metrics and acceptable latency thresholds.

Budget allocation during this stage is directed toward backend engineering resources, DevOps infrastructure design, monitoring tools, and secure logging systems. The investment reflects the complexity of distributed AI orchestration rather than feature development.

Phase 3: Automated Grant Schema Extraction Engine (Months 7–13)

The third phase introduces a major innovation component: automatic extraction and structuring of grant applications from regulatory documents. This phase involves research into document structure recognition, heading hierarchy detection, semantic clustering of field instructions, and the creation of a reusable field archetype ontology.

The R&D uncertainty lies in accurately mapping heterogeneous grant formats into a unified structured schema while minimizing misclassification risks. Iterative experimentation will be required to calibrate semantic similarity thresholds for mapping new grant fields to existing instruction templates.

During this period, the team will construct a domain-specific grant field knowledge graph that captures relationships between evaluation criteria across funding bodies. This knowledge representation layer is expected to evolve over multiple experimental iterations.

The milestone for this phase will be successful automated structuring of at least three heterogeneous grant programs with verified field mapping accuracy above a predefined confidence threshold.

Budget justification reflects the need for NLP research engineers, ontology modeling expertise, and data annotation personnel for validation datasets. Additional infrastructure costs will support document ingestion pipelines and embedding storage.

Phase 4: Feedback-Driven Calibration Engine (Months 10–16)

This phase focuses on creating a closed-loop learning system that incorporates evaluator feedback letters into the scoring calibration process. The central research question is how to reliably extract structured criticism from narrative evaluation reports and map it to specific claim categories and scoring weights.

The team will develop a feedback parsing module capable of identifying recurring evaluation weaknesses, implicit scoring deficiencies, and qualitative critique patterns. Statistical methods will be applied to adjust scoring weight distributions and similarity thresholds without modifying underlying foundation models.

The milestone for this phase will be empirical evidence demonstrating measurable improvement in alignment between platform evaluation scores and external evaluator scores across pilot datasets.

Budget during this period is allocated to data scientists, NLP engineers, and calibration researchers. Experimental cycles will require controlled testing datasets and evaluation benchmarking processes.

Phase 5: Multi-User Collaboration and Cryptographic Isolation (Months 12–18)

The fifth phase addresses distributed collaboration and data security architecture. This includes designing a field-level concurrency system, version control logic, cross-field semantic consistency validation, and encrypted application isolation.

The primary technological risk lies in integrating real-time collaboration with AI validation while maintaining data consistency and preventing semantic contradictions across fields. A field-level optimistic concurrency model will be implemented and tested under simulated multi-user workloads.

In parallel, per-application encryption key infrastructure will be developed. The architecture will likely follow an envelope encryption model, where each grant application is encrypted using a unique symmetric key, and key access is controlled via asymmetric encryption and secure key management services. Research challenges include key revocation mechanisms, multi-user access governance, and secure in-memory decryption practices.

The milestone will be a secure, production-ready environment in which multiple authorized users can collaborate on encrypted grant applications without data leakage or concurrency conflicts.

Budget justification reflects cybersecurity engineering resources, backend distributed systems development, and secure infrastructure services. Given the regulatory sensitivity of grant content, this phase represents a strategic technological investment rather than an operational feature.

Phase 6: Microservice Hardening and Production Deployment (Months 16–24)

The final phase converts the integrated system into a standalone microservice architecture with health monitoring, resilience mechanisms, deployment pipelines, and institutional-grade governance controls.

This stage includes implementing health checks, circuit breakers, failover handling, prompt version control systems, model version tracking, and full evaluation trace logging. A comprehensive audit trail framework will be finalized to ensure regulatory-grade traceability.

The milestone for completion will be a production-deployed microservice validated in pilot customer environments with monitored reliability and compliance-grade logging.

Budget allocation during this phase is focused on DevOps engineering, cloud architecture optimization, security compliance validation, and penetration testing.

Budget Structure Overview

The majority of the R&D budget is allocated to highly specialized personnel in artificial intelligence engineering, backend distributed systems architecture, NLP research, and cybersecurity design. A smaller but essential portion of the budget is dedicated to secure cloud infrastructure, model interaction costs, data storage, encryption key management services, and testing environments.

The cost structure reflects genuine technological research and engineering complexity rather than product interface development. The investment supports the creation of proprietary semantic validation methodologies, calibration datasets, grant ontology assets, and security infrastructure, all of which form long-term defensible intellectual property.

Technological Risk and R&D Justification

The project involves significant technological uncertainty in semantic similarity calibration, structured extraction of regulatory documents, probabilistic-to-deterministic output enforcement, and secure distributed collaboration architecture. These challenges require iterative experimentation and interdisciplinary expertise.

The outcome of this R&D program will not be a simple AI-powered application but a structured, secure, and self-improving institutional-grade evaluation intelligence platform.
