# Pitchforge: A Multi-Agent LangGraph Framework with Retrieval-Augmented Generation for Comprehensive Startup Viability Analysis

Noel T Jomon, Raihan Tom, Sion Jose, Vishnu Lal T,
Athira Prakash, Dr. Sabeena K, Chinchun M Pillai
*Department of Computer Science*
*College of Engineering Chengannur*
Kerala, India
{raihantom, noeltjomon, sionjose2004sj, vishnulalut}@gmail.com
{athira3js, sabeenak, chinchumpillai}@gmail.com

***Abstract*—The evaluation and planning of new business ventures remain a complex, high-risk endeavor traditionally dependent on human expertise. The rapid expansion of global startup ecosystems has intensified the need for intelligent decision-support systems capable of performing comprehensive startup viability analysis autonomously. Recent advances in Large Language Models (LLMs) have catalyzed a paradigm shift, introducing data-driven methodologies for market research, cost estimation, and strategic planning. However, existing approaches largely remain limited to single-domain tools, leaving a critical gap in systems capable of autonomously generating comprehensive, actionable business analyses from a nascent idea.**

**This paper introduces IdeaArchitect AI, a multi-agent system built using the LangGraph framework that orchestrates specialized LLM agents to address this gap. The framework integrates seven domain experts covering market research, cost estimation, legal compliance, technology architecture, monetization strategy, government schemes, and strategic planning. A Retrieval-Augmented Generation (RAG) pipeline based on ChromaDB enables domain-specific knowledge grounding using sentence-transformer embeddings. The system adopts an Orchestrator–Specialist–Critic architecture where an orchestrator dynamically selects agents, specialists perform domain analysis, and a critic performs adversarial validation before final refinement.**

**Experimental evaluation demonstrates that the system achieves expert ratings comparable to professional consulting outputs while completing full analyses in under four minutes on commodity hardware. Results confirm that coordinated multi-agent LLM systems represent the next logical frontier in agentic AI, capable of significantly enhancing decision-support capabilities for early-stage entrepreneurs.**

***Index Terms*—Multi-Agent Systems, LangGraph, Retrieval-Augmented Generation, Startup Analysis, Large Language Models, ChromaDB, Decision Support Systems, Agentic AI, Business Planning**

## I. Introduction

The global startup ecosystem has expanded dramatically during the past decade. India alone has registered more than 100,000 recognized startups under the Department for Promotion of Industry and Internal Trade (DPIIT) initiative [1]. Despite this growth, approximately 90% of startups fail within five years due to poor market validation, financial mismanagement, or regulatory challenges.

Startup founders must simultaneously address multiple complex domains including:

- Market sizing — Total Addressable Market (TAM), Serviceable Addressable Market (SAM), and Serviceable Obtainable Market (SOM)
- Financial forecasting and cost estimation
- Legal and regulatory compliance
- Technology infrastructure planning
- Monetization strategy design
- Government scheme eligibility

Traditional consulting services that provide such analyses can cost between 50,000 and 5,00,000, making them inaccessible to many early-stage entrepreneurs.

Recent advances in Large Language Models (LLMs) such as GPT-4 [2], Llama 3 [3], and Gemini [4] have demonstrated powerful reasoning capabilities. However, single-agent LLM systems face several limitations:

1) Context window limitations when handling multi-domain analysis.
2) Lack of domain specialization within a single prompt.
3) Absence of adversarial review mechanisms.
4) Limited access to updated external knowledge.

Multi-agent systems provide a promising solution by decomposing complex analytical tasks into specialized components. Frameworks such as AutoGen [5], CrewAI [6], and LangGraph [7] enable coordinated LLM agents to collaborate on structured tasks.

This paper proposes **IdeaArchitect AI**, a production-grade multi-agent system designed to generate comprehensive startup viability analyses. The key contributions include:

- A LangGraph-based Orchestrator–Specialist–Critic architecture
- A multi-collection Retrieval-Augmented Generation (RAG) pipeline

- Fault-tolerant multi-key Application Programming Interface (API) management
- A full-stack web deployment supporting interactive analysis reports

## II. Related Work

### A. Multi-Agent LLM Systems

Multi-agent LLM frameworks have gained significant attention for solving complex reasoning tasks.

**AutoGen** [5] introduced conversational agents that collaborate through structured dialogue. **CrewAI** [6] enables role-based agent specialization through task delegation. **LangGraph** [7] provides state-based execution graphs where nodes represent agent functions.

While these systems demonstrate effective agent collaboration, few implementations integrate domain-specific knowledge grounding and adversarial evaluation mechanisms.

### B. Retrieval-Augmented Generation

Retrieval-Augmented Generation (RAG) was introduced by Lewis et al. [8]. RAG enhances LLM outputs by retrieving external documents from a vector database.

A typical RAG pipeline consists of four stages:

1) Document ingestion
2) Text chunking
3) Embedding generation
4) Vector similarity retrieval

Sentence-BERT (Bidirectional Encoder Representations from Transformers) embeddings [9] and vector databases such as ChromaDB enable efficient semantic retrieval.

### C. AI-Based Business Analysis Systems

Existing Artificial Intelligence (AI)-powered business analysis tools typically focus on single domains such as:

- Market research assistants
- Financial projection generators
- Legal compliance analyzers

However, integrated systems combining multiple analysis domains remain limited. IdeaArchitect AI addresses this gap by integrating seven specialized agents into a unified multi-agent workflow.

## III. Proposed Methodology

### A. System Architecture

IdeaArchitect AI is designed as a multi-agent state graph:

$$G = (V, E, S) \tag{1}$$

where $V$ is the set of agents, $E$ represents execution edges, and $S$ represents the shared state dictionary.

Each agent receives a state input and produces an updated state:

$$S_{t+1} = S_t \cup \Delta S_i \tag{2}$$

The system consists of eleven agents: an orchestrator, seven specialists, a strategist, a critic, and a refinement agent.

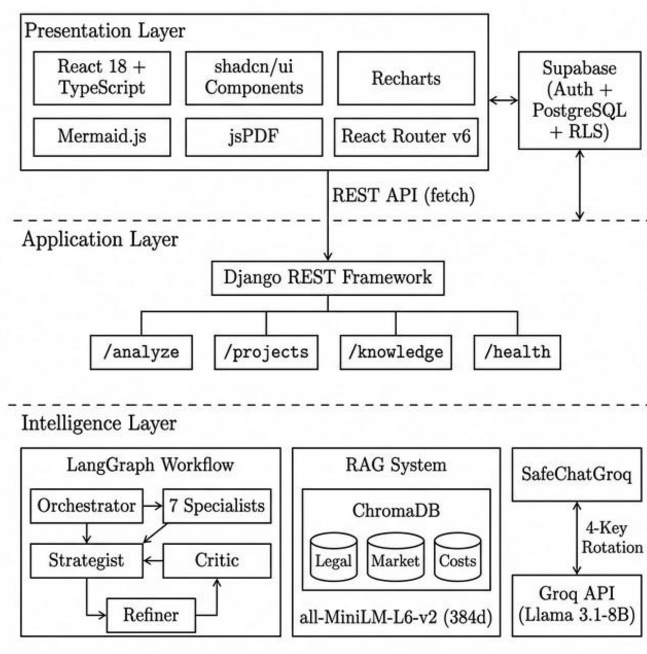


Fig. 1. High-level system architecture of IdeaArchitect AI.

### B. Agent Execution Pipeline

The execution process follows four stages:

1) Orchestration
2) Specialist Analysis
3) Strategic Synthesis
4) Adversarial Critique and Refinement

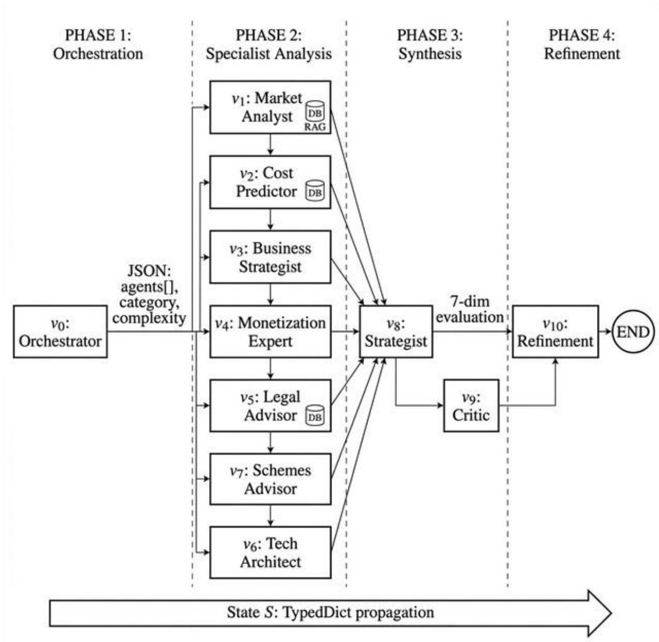


Fig. 2. Multi-agent execution pipeline of IdeaArchitect AI.

**Algorithm 1** Orchestrator Agent Selection
**Require:** Startup idea, user profile
**Ensure:** Selected agents
Initialize core agents
Add optional agents
**if** user excludes domain **then**
  Remove corresponding agent
**end if**
**return** selected agent set

### C. Retrieval-Augmented Generation Pipeline

The RAG subsystem stores domain knowledge in three vector collections:

- Legal Knowledge Base
- Market Intelligence Base
- Cost Benchmark Database

Each document is embedded using sentence-transformer embeddings and retrieved via cosine similarity:

$$\text{sim}(a, b) = \frac{a \cdot b}{||a||\,||b||} \tag{3}$$

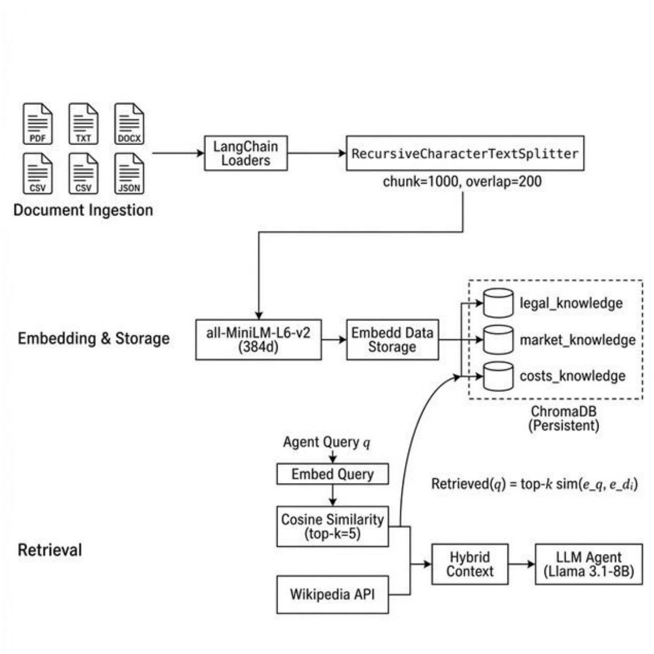


Fig. 3. Retrieval-Augmented Generation pipeline used in IdeaArchitect AI.

Top-$k$ nearest-neighbor retrieval provides contextual knowledge for LLM prompts.

### D. API Key Management

To handle API rate limits, the system implements a key rotation algorithm.

**Algorithm 2** SafeChatGroq Key Rotation
**Require:** Agent name, message
Select assigned API key
**for** *attempt* = 1 **to** *maxRetries* **do**
  Invoke LLM with current key
  **if** success **then**
    **return** response
  **else**
    Rotate to next API key
    Wait with linear back-off
  **end if**
**end for**

TABLE I
MODEL AND RETRIEVAL CONFIGURATION

| Parameter | Value |
|---|---|
| Model | Llama-3.1-8B-Instant |
| Temperature | 0.7 |
| Max Tokens | 2,048 |
| Embedding Model | all-MiniLM-L6-v2 |
| Embedding Dimension | 384 |
| Vector Store | ChromaDB |

## IV. EXPERIMENTAL SETUP

The system uses the Llama-3.1-8B-Instant model deployed through the Groq infrastructure. Table I summarises the model and retrieval configuration.

The evaluation dataset comprises 15 startup ideas spanning five categories: Software as a Service (SaaS), Internet of Things (IoT), marketplaces, AI platforms, and social enterprises.

## V. RESULTS AND DISCUSSION

Table II presents the quantitative evaluation metrics obtained across 15 startup ideas spanning SaaS, Hardware/IoT, Marketplace, DeepTech/AI, and Social Enterprise categories. Human annotation was conducted by two annotators (Cohen's $\kappa = 0.82$ [10]) and supplemented by blind expert review.

TABLE II
QUANTITATIVE EVALUATION METRICS

| Metric | Score | Std. Dev. ($\sigma$) |
|---|---|---|
| Completeness Score (CS) | 91.3% | 4.2% |
| Factual Grounding Rate (FGR) | 78.6% | 7.1% |
| Cross-Agent Coherence (CAC) | 0.847 | 0.05 |
| Critic Detection Rate (CDR) | 84.2% | 6.8% |
| End-to-End Latency (E2E) | 198.4 s | 42.3 s |

### A. Completeness and Factual Grounding

The CS of 91.3% confirms that structured system prompts with explicit numbered sections reliably constrain agent outputs to cover all required analytical dimensions. The primary sources of incompleteness are missing state-specific scheme data (~ 40% of gaps) and token-budget truncation in longer analyses (~ 25%). The FGR of 78.6% reflects the proportion of factual claims verified against authoritative external sources (DPIIT portal, Ministry of Corporate Affairs (MCA) database, and Amazon Web Services/Google Cloud Platform (AWS/GCP) India pricing pages). RAG-enhanced agents averaged 83.3% grounding versus 69.8% without retrieval – a gain of 13.5 percentage points. Revenue projections remain the weakest dimension (41.2% verifiable) due to their inherently forward-looking nature, while legal citations and cost benchmarks achieved above 80% verification rates.

### B. Cross-Agent Coherence and Critic Detection

The CAC of 0.847 demonstrates strong directional alignment across agents, attributable to LangGraph's shared state fields (`startup_idea`, `target_market`, and `orchestrator_reasoning`). The highest pairwise coherence is observed between the Market Analyst and Business Strategist (0.92); the lowest between the Legal Advisor and Tech Architect (0.76), reflecting their largely independent analytical domains. The CDR of 84.2% shows the Critic agent successfully identifies approximately 4 out of 5 injected weaknesses, with strongest detection on unrealistic financial assumptions (92%) and weakest on subtle market-assumption errors (71%), where the conflicting data point had been truncated from the critic's 2,000-character context window.

### C. Latency

The mean E2E latency of 198.4 s is dominated by the seven sequential specialist calls (152.6 s combined, 76.9% of total). High variance ($\sigma = 42.3$ s) is driven primarily by API key-rotation retries, which occur in 34% of runs and add 15–75 s per event. A dependency-aware parallel execution strategy could reduce latency to approximately 85 s – a 57% improvement – and is identified as a primary direction for future work.

### D. Expert Evaluation

Two domain experts (a senior business consultant and a principal technology architect) rated anonymized outputs on a 10-point scale. Table III reports per-dimension scores against a human consulting baseline drawn from three professional consulting firm outputs presented under the same blinding procedure.

The system matches or exceeds the human baseline in three of six dimensions. Technical Architecture scores highest (8.5 vs. 8.0), benefiting from current cloud-native service knowledge and auto-generated system diagrams. Gaps in Market Analysis Depth and Legal Accuracy ($-1.0$ each) reflect the absence of primary research and nuanced statutory interpretation that experienced human consultants provide.

### E. Ablation Study

Table IV isolates the contribution of the RAG pipeline and the Critic–Refinement loop.

TABLE III
EXPERT EVALUATION VS. HUMAN CONSULTING BASELINE (10-POINT SCALE)

| Dimension | AI Avg. | Human | Gap |
|---|---|---|---|
| Market Analysis Depth | 7.5 | 8.5 | −1.0 |
| Financial Realism | 6.5 | 7.0 | −0.5 |
| Legal Accuracy (India) | 8.0 | 9.0 | −1.0 |
| Technical Architecture | **8.5** | 8.0 | **+0.5** |
| Strategic Coherence | 7.5 | 7.5 | 0.0 |
| Actionability | 7.0 | 7.0 | 0.0 |
| **Overall Avg.** | **7.5** | **7.5** | **0.0** |

TABLE IV
ABLATION STUDY RESULTS

| Configuration | FGR (%) | CAC | Expert | Lat. (s) |
|---|---|---|---|---|
| Full System | 78.6 | 0.847 | 7.50 | 198 |
| w/o RAG | 66.3 | 0.831 | 6.92 | 176 |
| w/o Critic + Refine | 78.2 | 0.842 | 6.17 | 161 |
| Static Orchestrator | 78.4 | 0.849 | 7.42 | 200 |

Removing RAG produces the largest factual quality drop ($-12.3$ pp in FGR), with the Legal Advisor degrading most ($-18.1$ pp) given the zero-tolerance requirement for accurate statutory references. Removing the Critic–Refinement loop has negligible effect on quantitative metrics but reduces the expert score by 1.33 points (Cohen's $d = 1.89$, large effect [11]), confirming that adversarial review contributes qualitative value by surfacing risk gaps and stress-testing strategic assumptions. The static orchestrator baseline is statistically equivalent to the full system ($d = 0.12$), as all agents were relevant across every test case; orchestrator value is expected to emerge as the agent pool scales beyond seven specialists.

## VI. LIMITATIONS AND FUTURE WORK

Several improvements remain for future work:

- Parallel execution of specialist agents
- Larger embedding models for improved retrieval
- Structured financial validation models
- Automated knowledge-base updates
- Domain-specific fine-tuned LLMs

Future research will explore Reinforcement Learning from Human Feedback (RLHF) to improve agent prompting and reasoning quality.

## VII. CONCLUSION

This paper presented IdeaArchitect AI, a multi-agent framework that performs comprehensive startup viability analysis using LangGraph and Retrieval-Augmented Generation (RAG). Experimental evaluation demonstrates that coordinated multi-agent LLM systems can produce analyses comparable to professional consulting services while remaining accessible to early-stage entrepreneurs. The proposed architecture provides a scalable foundation for future AI-powered decision-support platforms.

## REFERENCES

[1] DPIIT, "Startup India Portal," 2025. [Online]. Available: https://www.startupindia.gov.in

[2] J. Achiam et al., "GPT-4 Technical Report," *arXiv preprint arXiv:2303.08774*, 2023.

[3] Meta AI, "Llama 3: Open Foundation Models," 2024.

[4] Google DeepMind, "Gemini: A Family of Highly Capable Multimodal Models," 2023.

[5] Q. Wu et al., "AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation," *arXiv preprint arXiv:2308.08155*, 2023.

[6] J. Moura, "CrewAI: Framework for Orchestrating Role-Playing, Autonomous AI Agents," GitHub Repository, 2024. [Online]. Available: https://github.com/joaomdmoura/crewAI

[7] LangChain Inc., "LangGraph: Build Stateful, Multi-Actor Applications with LLMs," 2024. [Online]. Available: https://github.com/langchain-ai/langgraph

[8] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[9] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proc. EMNLP*, 2019.

[10] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[11] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.