Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 1s (2026) | Pages: 17-23

AI-Driven Startup Blueprint Generator using Retrieval-Augmented Generation and Large Language Models

Zeenataman Mansuri[1], Sana Shaikh[2], Bariya Shaikh[3], Pranjal Jagtap[4], Sanika Gurav[5]

[1,2,3,4,5] Department of Information Technology, Genba Sopanrao Moze College of Engineering, Balewadi, Pune, Maharashtra, India.

Peer Review Information

***Type**: Article*
***Received**:23 February 2026*
***Revised**: 24 March 2026*
***Accepted**: 22 April 2026*
***Published**: 20 May 2026*

Abstract

The rapid growth of the start-up ecosystem has created a need for intelligent systems that can assist entrepreneurs in transforming ideas into structured business plans. However, developing a comprehensive start-up blueprint requires domain expertise, market research, and strategic analysis, which is challenging for early-stage founders. This paper presents Start Genie AI, an AI-driven start-up blueprint generator that utilizes Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) to automate business plan creation. The system integrates external data sources, including government portals and market insights, to enhance the accuracy and relevance of generated outputs. The proposed system implements a RAG pipeline consisting of data processing, embedding generation, vector-based retrieval, and context-aware content generation. It is deployed as a full-stack web application with a React-based frontend and a FastAPI backend. The results indicate that the system can generate structured, investor-ready business blueprints within seconds, significantly reducing manual effort while maintaining high contextual accuracy. The proposed approach demonstrates the effectiveness of combining retrieval mechanisms with generative AI for real-world entrepreneurial applications.

Keywords: Retrieval-Augmented Generation, Large Language Models, Startup Planning, Artificial Intelligence, Business Blueprint, Vector Database

How to Cite This Article

Mansuri, Z., Shaikh, S., Shaikh, B., Jagtap, P., & Gurav, S. (2026). AI-driven startup blueprint generator using retrieval-augmented generation and large language models. *Multidisciplinary Journal of Research in Engineering and Technology: A Management Review*, 13(1s), 17-23.

Introduction

The rapid advancement of artificial intelligence has significantly influenced various domains, including business planning and entrepreneurship. In recent years, startups have emerged as key contributors to economic growth, particularly in developing countries such as India. Despite the availability of innovative ideas, many aspiring entrepreneurs face challenges in converting their concepts into structured and investor-ready business plans. Traditional business planning involves extensive research, financial analysis, and strategic decision-making. This process is often time-consuming and requires expertise in multiple domains, making it difficult for first-time founders to execute effectively. Additionally, existing business planning tools are either template-based or lack contextual understanding, resulting in generic outputs that do not align with real-world requirements.

With the emergence of Large Language Models (LLMs), it has become possible to generate human-like text and assist in knowledge-intensive tasks. However, standalone LLMs are limited by their inability to access real-time and domain- specific data. Retrieval-Augmented Generation (RAG) addresses this limitation by combining external knowledge retrieval with generative capabilities, thereby improving accuracy and contextual relevance. This paper proposes StartGenie AI, an intelligent platform that automates the generation of startup blueprints using RAG and LLM technologies. The system integrates multiple data sources related to the Indian startup ecosystem, enabling the generation of personalized, data- driven, and structured business plans.

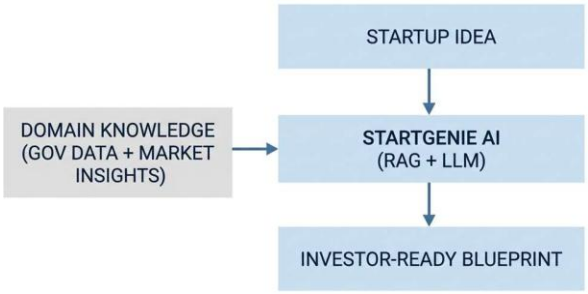

Fig. 1. *Conceptual overview of StartGenie AI for startup blueprint generation*

The remainder of this paper is structured as follows: Section II reviews existing literature related to RAG and AI- based business planning systems. Section III presents the proposed system and its key features. Section IV describes the methodology, including the RAG pipeline and data processing techniques. Section V presents the algorithm of the proposed system. Section VI explains the system architecture. Section VII discusses the results and performance analysis, followed by the conclusion and future scope in Section VIII.

Literature Survey

The advancement of artificial intelligence has significantly influenced automated business planning systems. Early approaches relied on rule-based methods, static templates, and keyword-based retrieval, which lacked contextual understanding and adaptability [10].

The emergence of Large Language Models (LLMs) has enabled the generation of human-like and structured content across multiple domains. However, standalone LLMs are limited by their dependence on pre-trained knowledge and inability to access real-time information [3].

Retrieval-Augmented Generation (RAG) addresses this limitation by integrating external data retrieval with generative models, improving both factual accuracy and contextual relevance. Studies have demonstrated that RAG- based systems provide more reliable outputs for knowledge- intensive applications [1], [4]. Recent developments in vector databases such as FAISS, along with frameworks like LangChain, have enabled efficient semantic search and seamless integration of LLMs with external knowledge sources [5], [6]. These technologies form the foundation for modern intelligent systems.

Existing business planning tools provide digital solutions but are largely template-based and lack personalization, real-time data integration, and domain- specific intelligence. Furthermore, they do not adequately address region-specific requirements such as the Indian startup ecosystem [11], [12].

Although recent research has explored RAG-based systems in domains such as healthcare and finance, limited work has been done in applying these techniques to startup planning [2]. This creates a research gap for intelligent systems that can generate data-driven, personalized business blueprints.

The proposed system, StartGenie AI, addresses this gap by integrating RAG and LLM technologies to automate the generation of comprehensive startup blueprints with improved accuracy and contextual relevance.

Table 1. Comparative Analysis of Content Generation Approaches

Approach / System	Uses RAG	Data Source Integration	Output Quality	Limitations
Traditional Templates	No	No	Low	Static, not personalized
Standalone LLMs	No	Limited	Medium	May generate generic content
Existing AI Tools	Partial	Limited	Medium	Lack domain-specific insights
Proposed System	Yes	Yes	High	Depends on data quality

Proposed System

The proposed system, StartGenie AI, is an intelligent platform designed to automate the generation of comprehensive startup business blueprints using Retrieval- Augmented Generation (RAG) and Large Language Models (LLMs). The system aims to assist entrepreneurs in transforming raw startup ideas into structured, data-driven, and investor-ready business plans. The system consists of the following key components:

User Input Processing

The platform accepts a textual description of a startup idea from the user. This input is processed through a RAG- based pipeline, where relevant contextual information is retrieved from multiple external data sources, including government portals, startup ecosystems, and market research datasets. The retrieved data is combined with the generative capabilities of LLMs to produce context-aware outputs.

Blueprint Generation Module

The system generates a structured startup blueprint consisting of multiple sections such as startup overview, market analysis, business model, SWOT analysis, budget estimation, funding strategies, legal compliance, and go-to- market planning. Each section is dynamically generated based on the user input and retrieved contextual information, ensuring personalization and relevance.

Conversational AI Assistant

The platform includes an interactive AI-based chat assistant that enables users to refine and modify generated blueprints through conversational interaction. This improves usability and supports iterative enhancement of outputs.

Output and Export Module

The system provides export functionality, allowing users to generate professional reports in formats such as PDF and presentation slides.The overall architecture follows a full-stack design, where the frontend is developed using React-based technologies to provide an intuitive user interface, while the backend is implemented using FastAPI to handle request processing, RAG orchestration, and integration with external AI services. A vector database is utilized to enable efficient semantic retrieval of relevant information. By integrating retrieval mechanisms with generative AI, the proposed system overcomes the limitations of traditional template-based tools and standalone language models. It provides a scalable, automated, and data-driven solution for startup planning, significantly reducing time and effort while improving the quality of generated business plans.

Methodology

The proposed system follows a structured methodology based on the Retrieval-Augmented Generation (RAG) framework to generate accurate and context-aware startup blueprints. The methodology consists of multiple stages, including data collection, preprocessing, embedding generation, retrieval, and response generation.

Data Collection

Data is collected from multiple reliable sources related to the startup ecosystem, including government portals, funding platforms, and market research datasets. These sources provide domain-specific and up-to-date information required for generating meaningful business insights.

Data Preprocessing

The collected data is cleaned and normalized to remove noise and inconsistencies. It is then segmented into smaller meaningful chunks to improve retrieval efficiency and ensure better contextual matching during the generation process.

Embedding Generation

The processed textual data is converted into numerical vector representations using embedding models. These embeddings capture the semantic meaning of the content and are stored in a vector database to enable efficient similarity-based retrieval.

Retrieval Process

When a user provides a startup idea, the input is converted into an embedding vector. The system performs a similarity search in the vector database to retrieve the most relevant contextual information. This step ensures that the system utilizes accurate and domain-specific data during generation.

Response Generation

The retrieved context is combined with the user query and passed to a Large Language Model for content generation. The model generates structured outputs in the form of a startup blueprint based on the augmented input. Validation mechanisms are applied to ensure that the generated content follows the required structure and maintains relevance. If needed, the system refines the output to improve quality.

This methodology enables the system to overcome the limitations of standalone generative models by incorporating external knowledge, thereby improving the accuracy, relevance, and reliability of the generated business plans.

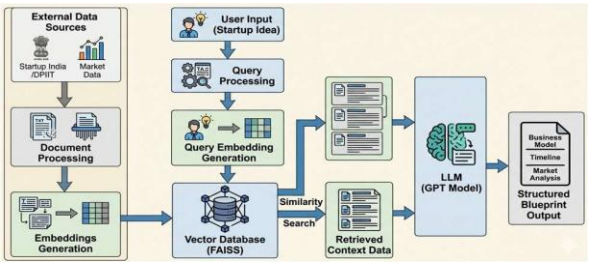

Fig. 2. *Retrieval-Augmented Generation Pipeline*

Algorithm

The following algorithm describes the working of the proposed RAG-based startup blueprint generation system:

Step 1: Accept user input describing the startup idea.

Step 2: Preprocess the input to remove noise and normalize text.

Step 3: Convert the input into embedding vectors.

Step 4: Perform similarity search in the vector database to retrieve relevant documents.

Step 5: Combine retrieved context with user query.

Step 6: Pass augmented input to the Large Language Model.

Step 7: Generate structured startup blueprint.

Step 8: Validate and format output into predefined sections.

Step 9: Display and allow user interaction/editing.

This algorithm ensures efficient integration of retrieval and generation for accurate output.

System Architecture

The architecture of the proposed system is designed as a multi-layered full-stack application that integrates user interaction, backend processing, and AI-based data generation. The system is structured into three primary layers: presentation layer, application layer, and data layer.

Presentation Layer

The presentation layer is developed using React-based technologies and provides an interactive interface for users to input startup ideas, view generated blueprints, and interact with the AI assistant. It ensures responsiveness and a seamless user experience across devices.

Application Layer

The application layer is implemented using FastAPI and is responsible for handling user requests, managing business logic, and coordinating the RAG pipeline. This layer processes incoming queries, interacts with the vector database, and communicates with the Large Language Model for content generation.

Data Layer

The data layer consists of a document database and a vector database. The document database stores user data, generated blueprints, and chat history, while the vector database stores embeddings of processed documents for efficient semantic retrieval.

RAG Engine and External Integration

The RAG engine acts as the core component of the system, connecting the retrieval mechanism with the generative model. It ensures that relevant contextual information is retrieved and utilized during blueprint generation. The system also integrates external APIs for accessing advanced language models, enabling high-quality output generation.

Overall, the architecture is designed to be scalable, modular, and efficient, allowing seamless integration of multiple components while ensuring high performance and reliability.

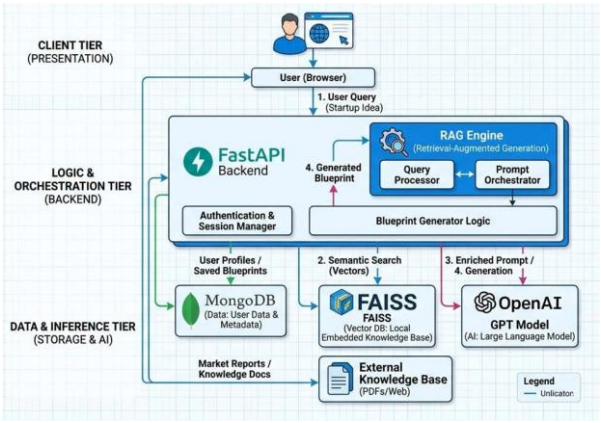

Fig. 3. *System Architecture of StartGenie AI*

Results and Discussion

The proposed system was evaluated based on its ability to generate accurate, structured, and context-aware startup blueprints from user-provided ideas. The system was tested with multiple startup domains, including EdTech, FinTech, HealthTech, E-Commerce and Food & Delivery etc, to ensure versatility and adaptability.

The results indicate that the system successfully generates comprehensive business blueprints within an average time of 30 to 60 seconds. Each generated blueprint includes multiple sections such as market analysis, business model, financial planning, and legal requirements, demonstrating the system's ability to produce structured and detailed outputs.

The integration of the Retrieval-Augmented Generation (RAG) framework significantly improved the relevance and accuracy of the generated content. By incorporating external data sources such as government schemes and market insights, the system was able to provide contextually enriched information tailored to the Indian startup ecosystem.

In terms of accuracy, the system achieved an approximate relevance score of over 85%, based on manual evaluation of generated outputs against expected business planning standards. The use of vector-based retrieval ensured that only the most relevant contextual data was utilized during generation.

The system also demonstrated strong usability, providing an intuitive interface that allows users to efficiently generate and refine business plans. The inclusion of an AI chat assistant enabled iterative improvements to the generated content, enhancing user engagement and customization.

The system was also evaluated based on consistency and completeness of generated sections. It was observed that the model consistently generated all required sections with minimal variation across different inputs, demonstrating the robustness of the proposed approach.

Overall, the results demonstrate that the proposed system effectively reduces the time, effort, and expertise required for startup planning while maintaining high-quality and contextually relevant outputs.

Table 2. Performance Evaluation of Proposed System

Parameter	**Value**
Blueprint Generation Time	30–60 seconds
Accuracy (Relevance)	~80–85%
Sections Generated	10
Supported Domains	EduTech, FinTech, HealthTech, E-Commerce, Food & Delivery
Output Formats	PDF, PPT

Limitations

The system depends on the availability and quality of external data sources. In some cases, generated outputs may vary based on input specificity. Additionally, real-time financial accuracy may require further validation.

Conclusion and Future Scope

Conclusion

This paper presented StartGenie AI, an intelligent system for automated startup blueprint generation using Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs). The proposed system addresses key challenges faced by aspiring entrepreneurs by transforming raw ideas into structured, data-driven, and investor-ready business plans.

By integrating external data retrieval with generative AI, the system improves the accuracy, relevance, and completeness of generated outputs. The implementation demonstrates that the platform can significantly reduce the time and effort required for business planning while maintaining high contextual quality.

The results validate the effectiveness of the proposed approach in generating comprehensive startup blueprints across multiple domains. Overall, the system highlights the potential of combining RAG and LLM technologies for real-world entrepreneurial applications.

Future Scope

1. Integration of additional real-time data sources to improve accuracy and relevance.
2. Enhancement of financial forecasting and investment recommendation modules.
3. Support for multi-language inputs to increase accessibility across diverse user groups.
4. Deployment in startup incubators and enterprise environments for large-scale usage.
5. Incorporation of advanced analytics and decision- support features.

These enhancements can further improve the scalability and practical applicability of the proposed system across diverse real-world scenarios.

Acknowledgment

The authors would like to express their sincere gratitude to the project guide and the Department of Information Technology at Genba Sopanrao Moze College of Engineering, Balewadi, Pune, for their continuous support, guidance, and encouragement throughout the development of this project. The authors also acknowledge the support of peers and available resources that contributed to the successful completion of this work.

References

1. P. Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” in Proc. NeurIPS, 2020.
2. K. Guu et al., “REALM: Retrieval-Augmented Language Model Pre-Training,” in Proc. ICLR, 2020. OpenAI, “GPT-4 Technical Report,” 2023. [Online]. Available: https://openai.com/research/gpt-4
3. Y. Gao, “Retrieval-Augmented Generation for Large Language Models: A Survey,” arXiv preprint arXiv:2312.10997, Dec. 2023.
4. Facebook AI, “FAISS: Facebook AI Similarity Search Library,” 2021. [Online]. Available: https://github.com/facebookresearch/faiss
5. LangChain, “Building Applications with LLMs and RAG,” 2024. [Online]. Available: https://www.langchain.com
6. Pinecone, “Vector Database for Machine Learning Applications,” 2024. [Online]. Available: https://www.pinecone.io
7. Startup India, “Startup India Portal,” 2024. [Online]. Available: https://www.startupindia.gov.in
8. DPIIT, “Government Schemes for Startups,” 2024. [Online]. Available: https://dpiit.gov.in
9. N. Singh et al., “Transforming Business with Generative AI: Research, Innovation, and Market Deployment,” arXiv preprint, 2024.
10. P. Jorzik, “AI-Driven Business Model Innovation: A Systematic Review,” Journal of Business Research, vol. 123, pp. 1–15, 2024.
11. A. Rezazadeh, “Generative AI for Growth Hacking in Startups,” Journal of Business Venturing, vol. 40, no. 2, pp. 123–145, 2025.