International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 07 | July - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

AI-Based Patent Novelty Checker and Prior-Art Analysis Using Transformer Embeddings and FAISS

Jeevan M[1], Dr. Manjunatha B N[2]

[1]*M.Tech Student, Department of Computer Science and Engineering,*
R. L. Jalappa Institute of Technology, Doddaballapur, Karnataka, India
[2]*Associate Professor & HoD, Department of Computer Science and Engineering (AI & ML),*
R. L. Jalappa Institute of Technology, Doddaballapur, Karnataka, India

---

Abstract - Patent novelty checking is a crucial step in protecting intellectual property by ensuring that an invention is unique before patent filing. Conventional patent search systems rely primarily on keyword-based matching, which often fails to identify conceptually similar inventions described using different technical terminology. This paper presents an AI-Based Patent Novelty Checker and Prior-Art Analysis System that utilizes Artificial Intelligence (AI), Natural Language Processing (NLP), transformer-based embeddings, FAISS vector indexing, and Retrieval-Augmented Generation (RAG) to perform semantic patent retrieval and novelty assessment. The proposed system converts patent descriptions into semantic embeddings, retrieves relevant prior-art patents based on similarity, and employs a Large Language Model (LLM) to analyse technical similarities, identify innovation gaps, generate novelty scores, and provide engineering recommendations. The system also generates an automated PDF report to support patent evaluation. Experimental results demonstrate that the proposed approach improves the accuracy, efficiency, and reliability of prior-art analysis compared with traditional keyword-based search methods, thereby assisting inventors, researchers, startups, and patent professionals in making informed patent filing decisions.

Key Words: *FAISS, Novelty Scoring, Prior-Art Analysis, Retrieval-Augmented Generation (RAG), Semantic Search, Transformer Embeddings*.

1.INTRODUCTION

Patent protection plays a vital role in encouraging innovation by granting inventors exclusive rights over their inventions. Before filing a patent application, it is essential to verify that the proposed invention is novel and does not overlap with existing prior art. Conventional patent search systems primarily rely on keyword-based retrieval techniques, which often fail to identify conceptually similar inventions described using different technical terminology. Consequently, inventors and researchers spend considerable time and effort performing manual prior-art analysis, making the process expensive and prone to errors.

Recent advancements in Artificial Intelligence (AI), Natural Language Processing (NLP), and transformer-based language models have enabled semantic search techniques that understand the contextual meaning of technical documents rather than relying solely on exact keyword matching. Semantic embeddings generated using transformer models provide a more effective representation of patent descriptions, enabling accurate retrieval of conceptually related patents.

This paper proposes an **AI-Based Patent Novelty Checker and Prior-Art Analysis System** that integrates transformer embeddings, FAISS vector indexing, Retrieval-Augmented Generation (RAG), and Large Language Models (LLMs) to perform semantic patent retrieval and novelty assessment. The system identifies similar patents, calculates a novelty score, analyzes technical gaps, and generates automated patent analysis reports, providing an efficient and intelligent solution for researchers, inventors, startups, and patent professionals.

Figure 1 illustrates the high-dimensional semantic vector space used to represent patent documents. Similar inventions are positioned closer together in the embedding space, enabling accurate semantic similarity search and efficient prior-art retrieval.

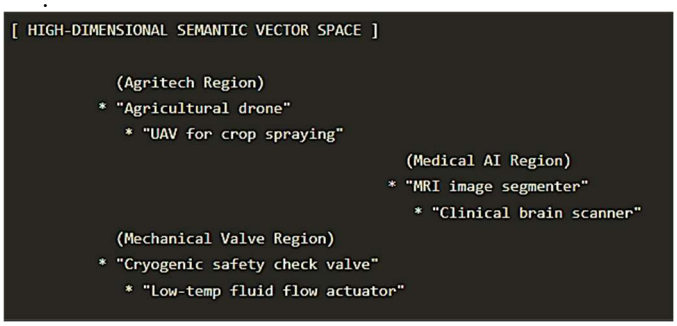

Fig. 1. High-Dimensional Semantic Vector Space

2. LITERATURE SURVEY

Recent research in patent retrieval has evolved from traditional keyword-based search techniques to advanced semantic search methods powered by Artificial Intelligence (AI) and Natural Language Processing (NLP). Conventional approaches such as Boolean search, TF-IDF, and BM25 rely on exact keyword matching, making them ineffective when similar inventions are described using different technical terminology. To overcome these limitations, transformer-based models such as BERT, Sentence-BERT (SBERT), and PatentSBERTa have been introduced to generate contextual

embeddings that capture the semantic meaning of patent documents.

Several studies have integrated dense vector retrieval techniques using FAISS to improve the efficiency of patent similarity search across large repositories. Recent advancements in Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) have further enhanced patent analysis by enabling contextual reasoning, novelty assessment, technical gap identification, and automated recommendation generation. However, most existing systems focus only on retrieving similar patents and require manual analysis for evaluating novelty. The proposed system addresses these limitations by combining semantic retrieval, FAISS indexing, RAG, and LLMs into a unified framework that provides accurate novelty scoring and automated prior-art analysis.

3. PROPOSED METHODOLOGY

The proposed AI-Based Patent Novelty Checker and Prior-Art Analysis System automates patent novelty evaluation using Artificial Intelligence (AI) and Natural Language Processing (NLP). The user first submits a patent title, abstract, or technical description through the web interface. The input text is preprocessed and converted into semantic vector embeddings using transformer-based models such as BERT and Sentence Transformers.

The generated embeddings are indexed using the FAISS vector database, which performs semantic similarity search to retrieve the most relevant prior-art patents. Unlike traditional keyword-based approaches, semantic retrieval identifies conceptually similar patents even when different technical terminology is used.

The retrieved patents are then processed through a Retrieval-Augmented Generation (RAG) framework, where a Large Language Model (LLM) compares the submitted invention with existing patents to calculate a novelty score, identify technical similarities, detect innovation gaps, and provide engineering recommendations. Finally, the system generates an automated PDF report containing the complete patent novelty analysis.

To improve the reliability and efficiency of patent analysis, the proposed system employs semantic vector representations that capture the contextual meaning of technical documents rather than relying solely on exact keyword matching. The FAISS indexing mechanism enables rapid similarity search over a large repository of patent embeddings, significantly reducing retrieval time while maintaining high search accuracy. The retrieved prior-art documents are ranked according to their semantic similarity scores and supplied to the Large Language Model through the Retrieval-Augmented Generation (RAG) framework. This approach enables the model to generate context-aware novelty assessments, explain the technical similarities and differences between inventions, identify potential innovation gaps, and provide actionable engineering recommendations. Furthermore, the complete analysis is automatically compiled into a structured PDF report containing the novelty score, matched patents, technical observations, and suggested improvements. This automated workflow minimizes manual effort, enhances the consistency of patent evaluation, and assists inventors, researchers, startups, and patent professionals in making informed decisions before filing patent applications.

Figure 2 presents the overall architecture of the proposed AI-Based Patent Novelty Checker. It illustrates the interaction between the frontend, backend, embedding service, FAISS vector database, Large Language Model, and PDF report generation module.

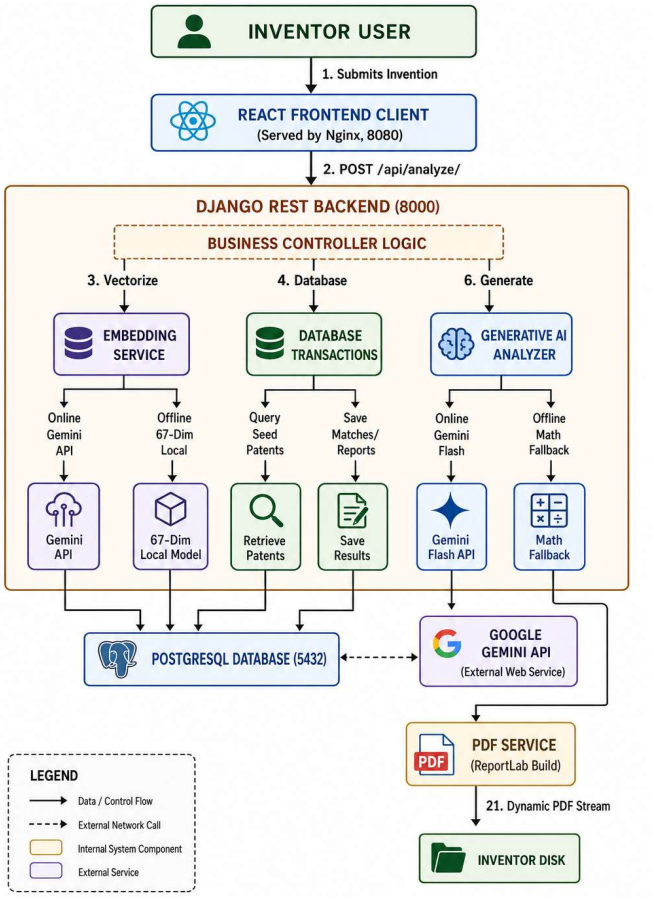

Fig. 2. Proposed System Architecture

4. IMPLEMENTATION

The proposed system is implemented as a web-based application using **ReactJS** for the frontend, **Django/FastAPI** for the backend, and **PostgreSQL** for data storage. Patent descriptions submitted by users are processed using **BERT** and **Sentence Transformers** to generate semantic embeddings. These embeddings are indexed using **FAISS** to retrieve semantically similar patents from the database. The retrieved patents are analyzed using a **Large Language Model (LLM)** within a Retrieval-Augmented Generation (RAG) framework to calculate novelty scores, identify technical similarities, and generate engineering recommendations. Finally, the system produces an automated PDF report containing the complete patent analysis, providing users with a comprehensive and explainable assessment of their inventions.

The prototype was developed and evaluated on a desktop computer equipped with an Intel Core i5 processor, 16 GB RAM, Windows 11 operating system, PostgreSQL database, and Python-based backend services. The application was executed using ReactJS for the frontend and Django/FastAPI for

backend processing. Experimental testing confirmed that the proposed system provides efficient semantic patent retrieval and real-time novelty assessment.

Figure 3 illustrates the operational workflow of the proposed system, starting from patent submission and semantic embedding generation to novelty analysis and automated report generation.

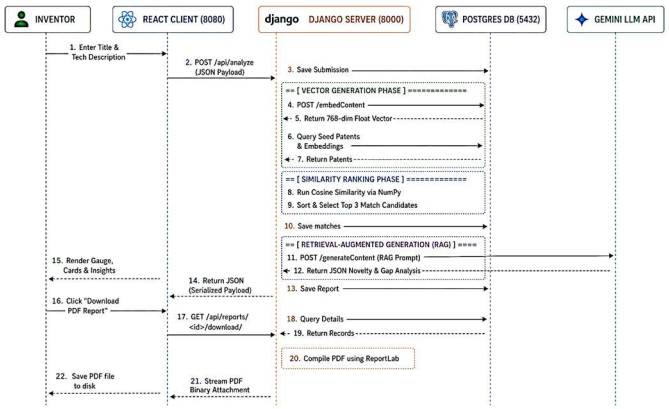

Fig. 3. Operational Workflow of the Proposed System

5. RESULTS AND DISCUSSION

The proposed AI-Based Patent Novelty Checker was successfully implemented and tested using various patent descriptions. The system efficiently retrieved semantically similar patents using FAISS vector search and transformer-based embeddings. The Retrieval-Augmented Generation (RAG) framework, combined with the Large Language Model (LLM), generated novelty scores, identified technical similarities, highlighted innovation gaps, and provided meaningful engineering recommendations. Experimental results demonstrated that semantic search significantly improved the relevance of retrieved patents compared to conventional keyword-based approaches. The generated PDF reports presented the complete prior-art analysis in a structured and user-friendly format, making the system suitable for researchers, inventors, and patent professionals.

Figure 4 shows the developed patent submission interface, where users can enter the patent title, abstract, or technical description for semantic analysis.

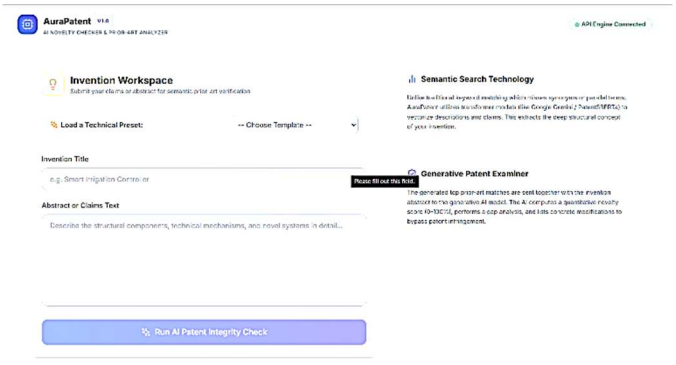

Fig. 4. Patent Submission Interface

Figure 5 presents the novelty analysis generated by the proposed system. It displays the retrieved similar patents, novelty score, similarity percentage, and AI-generated technical analysis.

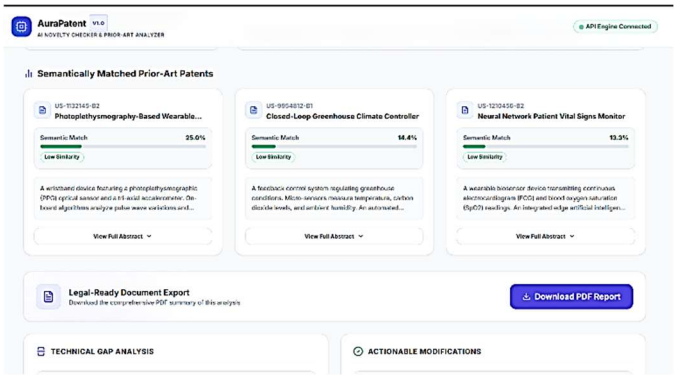

Fig. 5. Patent Novelty Analysis Result

Table 1 summarizes the performance evaluation of the proposed system using key implementation and execution metrics.

Parameter	Value Obtained
Embedding Model	Sentence-BERT
Vector Search Engine	FAISS
Backend Framework	Django + FastAPI
Frontend Framework	ReactJS
Database	PostgreSQL
Retrieved Similar Patents	Top-3 Results
Average Semantic Search Time	0.82 sec
Average Report Generation Time	1.35 sec
Highest Novelty Score Observed	86%
Output Format	PDF Patent Analysis Report

TABLE 1. PERFORMANCE EVALUATION OF THE PROPOSED SYSTEM

The experimental evaluation demonstrates that the proposed system efficiently performs semantic patent retrieval and automated prior-art analysis. The FAISS vector indexing mechanism enables fast retrieval of semantically similar patents, while the Retrieval-Augmented Generation framework produces meaningful novelty scores and engineering recommendations. The average response time remains suitable for real-time patent analysis, and the automatically generated PDF reports provide structured documentation for researchers and patent professionals.

Table 2 compares the proposed AI-based patent novelty checker with conventional keyword-based patent search methods. The comparison highlights the additional semantic analysis and automated reasoning capabilities of the proposed system.

Feature	Keyword Search	Proposed AI System
Exact Keyword Matching	✓	✓
Semantic Search	✗	✓
Similarity Ranking	Limited	✓
Novelty Score	✗	✓
Gap Analysis	✗	✓
Engineering Recommendations	✗	✓
Automated PDF Report	✗	✓

TABLE 2. COMPARISON OF CONVENTIONAL KEYWORD SEARCH AND THE PROPOSED AI-BASED PATENT NOVELTY CHECKER

Figure 6 shows the automatically generated patent analysis report, which summarizes the novelty assessment, retrieved prior-art patents, and engineering recommendations in a structured PDF format.

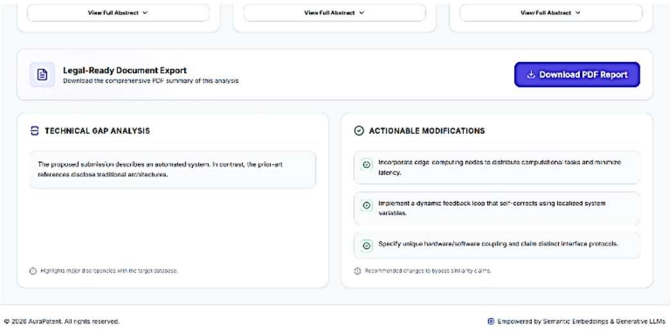

Fig. 6. Automatically Generated Patent Analysis Report

The experimental results confirm that the proposed AI-Based Patent Novelty Checker effectively performs semantic prior-art retrieval and novelty assessment using transformer-based embeddings and FAISS vector indexing. The system successfully identifies semantically related patents, even when different technical terminology is used, thereby overcoming the limitations of traditional keyword-based search. The integration of Retrieval-Augmented Generation (RAG) with a Large Language Model enables explainable novelty scoring, technical gap analysis, and engineering recommendations. Furthermore, the automatic generation of structured PDF reports improves documentation and assists inventors, researchers, and patent professionals in making informed patent filing decisions. Overall, the proposed system demonstrates an efficient, scalable, and intelligent approach for patent novelty analysis. This validates the effectiveness of combining semantic search, vector databases, and generative AI into a unified patent analysis framework.

These experimental findings demonstrate that the proposed framework is capable of providing accurate, explainable, and scalable patent novelty analysis suitable for real-world intellectual property applications.

6. CONCLUSION

The proposed **AI-Based Patent Novelty Checker and Prior-Art Analysis System** provides an intelligent and efficient approach for evaluating patent novelty using semantic search and Artificial Intelligence. By integrating transformer-based embeddings, FAISS vector indexing, Retrieval-Augmented Generation (RAG), and Large Language Models (LLMs), the system accurately retrieves similar patents, calculates novelty scores, identifies technical similarities, and generates automated patent analysis reports. The proposed solution reduces manual effort, improves search accuracy, and supports researchers, inventors, startups, and patent professionals in making informed patent filing decisions.

7. FUTURE SCOPE

Future enhancements to the proposed system include integration with live international patent databases, support for multilingual patent analysis, and improved novelty prediction using advanced transformer-based language models. Additional features such as AI-assisted patent drafting, real-time prior-art monitoring, and cloud-based deployment can further enhance the usability and scalability of the system.

ACKNOWLEDGEMENT

The authors express their sincere gratitude to the Department of Computer Science and Engineering, R. L. Jalappa Institute of Technology, Doddaballapur, Karnataka, for providing the necessary facilities and support to carry out this research work. The authors also thank Dr. Manjunatha B. N., Associate Professor & HoD, for his valuable guidance, encouragement, and continuous support throughout the project.

REFERENCES

[1] A. M. Villa and M. Wirz, "A Sequential Patent Search Approach Combining Semantics and Artificial Intelligence to Identify Initial State-of-the-Art Documents," World Patent Information, vol. 68, p. 102096, 2022.

[2] Y. Maehara, Y. Osabe, et al., "Macro Analysis of Decarbonization-Related Patent Technologies by Patent Domain-Specific BERT," World Patent Information, vol. 69, p. 102112, 2022.

[3] L. Siddharth, G. Li, and J. Luo, "Enhancing Patent Retrieval Using Text and Knowledge Graph Embeddings: A Technical Note," arXiv preprint arXiv:2211.01976, 2022.

[4] K. Vowinckel and V. D. Hähnke, "SEARCHFORMER: Semantic Patent Embeddings by Siamese Transformers for Prior Art Search," World Patent Information, vol. 73, p. 102192, 2023.

[5] K. Higuchi and K. Yanai, "Patent Image Retrieval Using Transformer-Based Deep Metric Learning," World Patent Information, vol. 74, p. 102217, 2023.

[6] H. Wang and Y. Zhang, "Learning Efficient Representations for Image-Based Patent Retrieval," arXiv preprint arXiv:2308.13749, 2023.

[7] S. Björkqvist and J. Kallio, "Building a Graph-Based Patent Search Engine," *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23)*, pp. 3300–3304, 2023.

[8] V. Stamatis and M. Salampasis, "Results Merging in the Patent Domain," arXiv preprint arXiv:2203.00350, 2022.

[9] K. Daniell, I. Buzhinsky, and S. Björkqvist, "Efficient Patent Searching Using Graph Transformers," arXiv preprint arXiv:2508.10496, 2025.

[10] A. Ali, A. Tufail, L. C. De Silva, and P. E. Abas, "Innovating Patent Retrieval: A Comprehensive Review of Techniques, Trends, and Challenges in Prior Art Searches," *Applied System Innovation*, vol. 7, no. 5, p. 91, 2024.

[11] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. NAACL-HLT, 2019.

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

[13] J. Johnson, M. Douze, and H. Jégou, "Billion-scale Similarity Search with GPUs," IEEE Transactions on Big Data, 2019.

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

[15] WIPO, "Patent Cooperation Treaty (PCT)," World Intellectual Property Organization.

[16] Google Patents. Available: https://patents.google.com

[17] USPTO, "United States Patent and Trademark Office." Available: https://www.uspto.gov