

# Enhancing Large Language Models with Agentic Retrieval-Augmented Generation

Noe Javet

May 31, 2025

## Abstract

This paper presents an investigation into improving the performance of large language models (LLMs) through agentic retrieval-augmented generation (RAG). We compare classical RAG, a baseline system, and an enhanced approach using contextual embeddings and tool-using agents. Our experiments show [insert summary of findings].

## 1 Methodology

To evaluate the capabilities of various language models (LLMs) in processing medical discharge reports, a multi-stage evaluation pipeline was implemented. The process consisted of document preprocessing, prompt engineering, model inference, and result evaluation using defined metrics.

Initially, raw discharge summaries in DOCX format were parsed using the `python-docx` library. This stage focused on extracting plain text content from the original documents while preserving section headers and structural elements. The extracted text was then sent to nine different LLMs using a uniform system prompt designed to assess completeness and relevance of medical sections.

Subsequently, to reduce variability introduced by inconsistent formatting, the documents were manually reformatted to adhere to a markdown-like structure. These manually structured texts were then re-evaluated using the same baseline system prompt to measure improvements attributable solely to input clarity.

In the final stage, the same markdown-formatted texts were reprocessed using a refined system prompt with more explicit instructions aimed at improving model focus and output consistency. This allowed for a comparative assessment of prompt sensitivity across models.

The evaluation metrics comprised both structural and content-based indicators: true positives, true negatives, false positives, and false negatives regarding section detection; as well as binary assessments of section relevance and conciseness (0/1 scale). Sections marked as missing or irrelevant were explicitly recorded to aid in qualitative analysis.