

Machine Learning in Document Retrieval

Abstract

Title: Enhancing Document Retrieval Through Agentic AI Systems

Authors: Dr. Jane Smith, Prof. Robert Johnson, Dr. Maria Garcia

Institution: Institute of Advanced Computing

Date: December 2023

ABSTRACT:

This paper presents a novel approach to document retrieval using agentic artificial intelligence systems. We introduce a framework that combines retrieval-augmented generation (RAG) with intelligent agent-based decision making to improve the accuracy and relevance of document search results.

INTRODUCTION:

Traditional document retrieval systems rely on keyword matching and basic semantic search. Our approach enhances this by incorporating an intelligent agent that can decide when to search, what to search for, and how to synthesize information from multiple sources.

METHODOLOGY:

We developed a three-stage pipeline:

1. Document Processing: PDFs are chunked and embedded using transformer models
2. Intelligent Routing: An agent decides whether to search or answer directly
3. Response Generation: Context-aware answer generation using local LLMs

RESULTS:

Our experiments on a dataset of 10,000 technical documents showed:

- 34% improvement in answer relevance
- 28% reduction in response time
- 92% user satisfaction rate
- Significant reduction in hallucinations

EXPERIMENTS:

We tested three configurations:

- Baseline: Traditional keyword search
- RAG-only: Standard RAG without agent
- Agentic RAG: Our proposed system

The agentic RAG system outperformed both baselines across all metrics.

CONCLUSION:

Agentic AI systems represent a significant advancement in document retrieval technology. By incorporating intelligent decision-making, we can provide more accurate, relevant, and efficient search results.

FUTURE WORK:

- Multi-modal document support
- Real-time learning from user feedback
- Distributed processing for large-scale deployments

Keywords: RAG, Agentic AI, Document Retrieval, NLP, Machine Learning