



Title

Prepared By

Name: Renad D. Alasmari

ID No.:261000861

Name:Ghada S. Alenazi

ID No.: 261002011

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

Supervisor Name: Dr.Sherif Mostafa

Collage of Informatics, Midocean University, Comoros

Email:.....

Acknowledgements

Special thanks to Dr. .

Abstract

List of Publications

Table of Contents

List of Abbreviations.....	vii
List of Figures.....	viii
List of Tables	ix

Chapter 1: Introduction and

Overview.....	1
1.2 Problem Definition	2
1.3 Thesis Objectives	3
1.4 Thesis Contribution	4
1.5 Thesis Organization.....	5

Chapter 2: Background and Related Work

.....	31
-------	----

Chapter 3: Proposed Approach

4.1 Introduction	78
------------------------	----

Chapter 4: Experimental Results and Discussion

Chapter 5: Conclusion and Future Work

7.1 Important Features	127
------------------------------	-----

7.3 Future Work	129
-----------------------	-----

References	130
------------------	-----

List of Abbreviations

- **Acc.:** Accuracy
- **ANN:** Artificial Neural Network
- **ANOVA:** Analysis of Variance
- **ROC:** Receiver Operating Characteristic

List of Figures

2.1 Educational Data Mining Process	7
2.2 Applications of Educational Data Mining	8
2.3 Distribution of Papers that used different attributes for PSP problem	19
2.4 Student Performance Data Set	21
2.5 Educational Process Mining (EPM): A Learning Analytics Data Set ...	23

List of Tables

2.1	Student Performance Data Set Description	21
2.2	Students' Academic Performance Dataset Description	23

Chapter 1 Introduction

1.1 Overview

1.2 Problem Definition

1.3 Thesis Objectives

1.4 Thesis Contribution

1.5 Thesis Organization

Chapter 2 Background and Related Work

Related Work

The introduction of Retrieval-Augmented Generation (RAG) represented a significant advancement in knowledge-intensive natural language processing tasks. Lewis et al. (2020) proposed a framework that combines document retrieval with language generation, enabling language models to access external knowledge sources during answer generation. This approach demonstrated considerable improvements in question-answering tasks and established the foundation for subsequent RAG research. However, the framework primarily focused on retrieval and generation performance without addressing transparency or explainability concerns.

To improve retrieval effectiveness, CRAG introduced a corrective retrieval mechanism capable of assessing and refining retrieved documents before they are processed by the language model. The study demonstrated that retrieval quality directly impacts answer accuracy and that correcting retrieval errors can significantly enhance system performance. Despite these improvements, the framework offered limited insight into how retrieved evidence influenced the generated responses.

Further developments were introduced through Retrieval-And-Structuring (RAS), which aimed to enhance reasoning capabilities by organizing retrieved information into structured representations before generation. The results showed that integrating structured knowledge can improve reasoning performance and answer quality. Nevertheless, the complexity of the architecture may create challenges when deploying the system in practical enterprise environments.

Researchers have also investigated multilingual retrieval-augmented generation to support information retrieval and question answering across different languages. These studies demonstrated the effectiveness of retrieval-augmented architectures in multilingual settings but primarily focused on open-domain scenarios rather than enterprise knowledge management environments.

As enterprise adoption of large language models increased, the need for realistic evaluation environments became evident. EnterpriseRAG-Bench was introduced to address this challenge by providing a large-scale benchmark consisting of more than 500,000 enterprise documents and 500 evaluation questions collected from multiple enterprise platforms. The benchmark highlighted the difficulties associated with retrieving and reasoning over information distributed across heterogeneous organizational systems. While the benchmark provides a valuable evaluation environment, it does not directly address explainability, confidence estimation, or source attribution.

To enhance trustworthiness in enterprise retrieval systems, ADORE proposed a multi-agent architecture that distributes retrieval, verification, and generation tasks across specialized agents. Experimental results indicated improvements in reliability and response quality compared to conventional RAG systems. However, the approach introduces additional implementation complexity and computational overhead.

Explainability has emerged as another important research direction within RAG systems. Recent studies investigated methods for understanding how retrieved passages influence generated answers and for identifying the contribution of individual documents to the final response. These approaches provide valuable transparency by allowing users and developers to inspect retrieval and generation processes. However, most explainability-focused solutions do not incorporate confidence estimation mechanisms or enterprise-specific evaluation scenarios.

RAGXplain extended this direction by introducing explainable evaluation and diagnostic techniques for RAG pipelines. Instead of focusing solely on model performance, the framework helps identify weaknesses in retrieval and generation components, providing

actionable guidance for improving system behavior. Although effective for analysis purposes, the framework does not constitute a complete enterprise retrieval solution.

Additional explainability research has been conducted through open-source reproduction and explainability analysis of CRAG. This work demonstrated that advanced retrieval correction techniques can be reproduced using open-source tools while simultaneously providing greater visibility into model decisions. Nevertheless, explainability remains limited to retrieval analysis rather than comprehensive response justification.

The evaluation of explainability itself has also attracted research attention. RAGBench introduced benchmark datasets and metrics specifically designed to assess explainability in retrieval-augmented generation systems. Such benchmarks contribute to the standardization of explainability evaluation but do not directly propose retrieval architectures capable of generating trustworthy responses.

Source attribution and citation generation have become increasingly important in enterprise environments where users require evidence supporting generated responses. Citation-aware retrieval approaches attempt to associate answers with supporting documents, thereby improving transparency and reducing unsupported claims. Although these methods strengthen traceability, they often stop at presenting citations without explaining why specific sources were selected or how reliable the generated answer is.

Another growing area of research focuses on confidence-aware retrieval and uncertainty estimation. Privacy-Aware Distributed RAG with Confidence-Aware Retrieval Routing proposed mechanisms that adapt retrieval behavior according to confidence levels. The study demonstrated improved reliability by reducing overconfident responses when evidence is insufficient. However, confidence estimation was addressed independently from explainability and citation attribution.

Trustworthy explanation generation has also been explored in domains outside enterprise knowledge retrieval. Research on autonomous systems and robotics showed that context-aware explanations can improve transparency, user trust, and decision interpretability. Although these studies operate in different application domains, their findings provide valuable insights into the design of trustworthy explanation mechanisms for future enterprise RAG systems.

Overall, the existing literature demonstrates substantial progress in retrieval quality, explainability, confidence estimation, and trustworthiness. Nevertheless, these capabilities are generally investigated as separate components. Current enterprise RAG systems still lack a unified framework that simultaneously integrates hybrid retrieval, explainability, citation attribution, confidence scoring, and conflict-aware reasoning within a realistic enterprise benchmark environment. This gap motivates the development of the proposed Trustworthy Explainable Enterprise RAG framework evaluated using EnterpriseRAG-Bench.

Paper Title	Author s	Year	Dataset	Methodolo gy	Evaluation Metrics	Results	Limitation	Research Gap	Relevance to My Research
Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks	Lewis et al.	2020	Wikipedi a, QA Datasets	Dense Retriever + Generator	EM, F1	SOTA perfor mance	No explainability	Lack of transparency	Baseline RAG
Corrective Retrieval-Augmented Generation (CRAG)	Yan et al.	2024	Multiple QA Datasets	Retrieval Evaluator + Corrective Retrieval	Accuracy, Recall	Improv ed retriev al quality	Limited explainability	No user-facing transparency	Retrieval improvement
EnterpriseRAG-Bench	Sun et al.	2026	Enterpris eRAG-Bench	Enterprise Benchmark	Recall, Answer Quality	Large-scale enterpr ise evaluat ion	Benchmark only	No retrieval framework	Main Dataset
Open-Source Reproduction and Explainability Analysis of CRAG	Yalavar thi et al.	2026	CRAG Datasets	Open-source CRAG + Explainabili ty	Accuracy	Improv ed interpr etabilit y	Limited enterprise focus	No enterprise validation	Explainability

RAS: Retrieval-And-Structuring for Knowledge-Intensive LLM Generation	Jiang et al.	2025	QA Benchmarks	Retrieval + Knowledge Structuring	F1, Accuracy	Better reasoning	High complexity	Difficult deployment	Knowledge structuring
Multilingual Retrieval-Augmented Generation for Open-Domain QA	Ranaldi et al.	2025	Multilingual QA	Multilingual RAG	Accuracy	Better multilingual retrieval	Not enterprise-focused	Limited enterprise evaluation	Background study
ADORE: Orchestrating Specialized Agents for Trustworthy Enterprise RAG	You et al.	2026	Enterprise Benchmarks	Multi-Agent RAG	Faithfulness, Accuracy	Improved trustworthiness	Complex architecture	Deployment complexity	Trustworthy RAG
Towards Explainable RAG	Various Authors	2025	RAG Benchmarks	Explainable Retrieval Analysis	Explainability Metrics	Improved transparency	No confidence scoring	Missing reliability assessment	Explainability Layer
RAGXplain	Various Authors	2025	RAG Pipelines	Explainable Evaluation Framework	Explainability Score	Better pipeline understanding	Evaluation-focused	Not a full solution	Explainability Evaluation
RAGBench	Various Authors	2025	RAGBench	Explainability Benchmark	Explanation Quality	Standardized evaluation	Benchmark only	No retrieval enhancement	Explainability Metrics
Privacy-Aware Distributed RAG with Confidence-Aware Retrieval Routing	Various Authors	2025	Enterprise Retrieval	Confidence-Aware Routing	Reliability Metrics	Improved reliability	Limited explainability	No source attribution	Confidence Scoring
Generating Trustworthy and Context-Aware Explanations for Autonomous Robots Using an LLM Agent-Based RAG Architecture	Various Authors	2025	Robotics Knowledge Base	Agent-Based RAG	Trust Metrics	Better trustworthy explanations	Non-enterprise domain	Domain transfer needed	Trustworthy Explanations
Source Attribution in Retrieval-Augmented Generation	Zhang et al.	2025	RAG Benchmarks	Evidence Attribution Framework	Citation Accuracy	Improved traceability	No confidence estimation	Attribution only	Citation Layer
Citation-Aware Retrieval-Augmented Generation	Wang et al.	2025	QA Datasets	Citation-Aware Generation	Citation Accuracy, Relevance	Better source-grounded responses	Limited explainability	No reasoning explanation	Citation Attribution
Confidence-Aware Retrieval-Augmented Generation	Various Authors	2025 – 2026	Enterprise QA	Confidence Scoring + Uncertainty Estimation	Confidence Calibration	Reduced overconfident errors	Limited explainability	Separate from attribution	Confidence Module

Chapter 4 Proposed Approach (Methodology)

Chapter 5 :Experimental Results and Discussion

Chapter 6 Conclusion and Future Work

References

[Abiodun, Oludare Isaac, et al., 2018]

Abiodun, Oludare Isaac, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad. “State-of-the-art in artificial neural network applications: A survey”. *Heliyon* 4. 2018.

[Abu Saa, Amjed, et al., 2019]

Abu Saa, Amjed, Mostafa Al-Emran, and Khaled Shaalan. “Factors Affecting Students’ Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques”. *Tech Know Learn* 24, 567–598., 2019.

[Adewale, Amoo M., et al., 2018]

Adewale, Amoo M., Alaba O. Bamidele, and Usman O. Lateef. “Predictive modelling and analysis of academic performance of secondary school students:

Artificial Neural Network approach”. International Journal of Science and Technology Educational Research 9(1), 1-8, 2018.

[Ali S, Rasheed Mansoor, and S Perumal, 2022]

Ali S, Rasheed Mansoor, and S. Perumal. “Multi-class LDA classifier and CNN feature extraction for student performance analysis during Covid-19 pandemic”. International Journal of Nonlinear Analysis and Applications, 13(1), pp. 13291339., 2022.

[Alpaydin, Ethem, 2010]

Alpaydin, Ethem (2010). Introduction to Machine Learning. London: The MIT Press. ISBN 978-0-262-01243-0