Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Integration of Retrieval-Augmented Generation and Multimodal Technologies for Advanced Virtual Research Assistants

Dr. Antony Vigil M S[1], Harshavardhani S[2], Shwetha R[3], Abishek Raj MN[4]
[1,2,3,4]*Department of Computer Science and Engineering*
[1,2,3,4]*SRM Institute of Science and Technology, Ramapuram, Chennai, India*

Received: 18 Dec 2024
Revised: 10 Feb 2025
Accepted: 28 Feb 2025

ABSTRACT

Researchers now rely on AI-powered IVRAs for a wide range of tasks, including providing instantaneous access to research resources and general academic support. When it comes to complicated, multimodal data and providing personalised, context-sensitive replies, however, current technologies can be inadequate. This study investigates potential solutions to these problems by combining multimodal technology with Retrieval-Augmented Generation (RAG). The RAG assistant can find the right information and put it together in a logical way since it uses generative models in addition to information retrieval. Multimodal technologies, which include image and document processing, allow IVRAs to access and understand data in many formats, which improves their overall functionality. Literature reviews, hypothesis development, data analysis, and academic writing are just a few of the many responsibilities that the suggested method hopes to alleviate for academics.

Keywords: Retrieval-Augmented Generation (RAG), Multimodal AI, Virtual Research Assistant, Knowledge Retrieval, AI-powered Research Assistants.

I. INTRODUCTION

Researchers can rely on virtual research assistants (VRAs) to assist with a variety of duties, including literature review, data analysis, and recommendation making. However conventional VRAs aren't very versatile because they use static databases or models that simply contain text. With the use of advanced AI models like Retrieval-Augmented Generation (RAG), IVRAs can dynamically find the information the user is looking for in massive datasets and provide answers that are specific to their inquiry. Images, graphs, and tables are all vital parts of research, and the assistant can process and interpret them all with the help of multimodal technologies. It is possible to elevate IVRAs from basic search engines to smart, context-aware research assistants using these developments. Researchers would benefit from a more all-encompassing tool for handling complicated research tasks, and this article suggests a paradigm for combining RAG with multimodal capabilities to make IVRAs much more effective and flexible.

II. EXISTING SYSTEM

Currently, the majority of virtual research assistants rely on either traditional search engines or AI-driven models that specifically address text-based queries. A number of systems that rely on Natural Language Processing (NLP) to understand and respond to queries sometimes fail when confronted with complex research tasks involving multimodal inputs or extracting particular facts from large, heterogeneous datasets. Common VRA capabilities include: Discover Your Necessities: Use search engines to find relevant academic papers, articles, and other written materials. Respond with written responses: Use current content to generate basic textual responses or summaries. Data Analysis: Help with basic data analysis by providing visual representations of data and statistical summaries. But these systems aren't up to the task of complex research tasks like combining data from multiple sources (e.g., text, photos, graphs, etc.). When it comes to helping researchers develop ideas or make significant strides in their work, they often fail to deliver deep insights or produce useful material[1].

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

III. PROPOSED SYSTEM

The proposed system combines Retrieval- Augmented Generation (RAG) with multimodal technologies to provide a state-of-the-art IVRA driven by artificial intelligence. The key components of the system are as follows: Through the use of Retrieval- Augmented Generation (RAG), AI assistants are able to access other databases or knowledge bases, such as academic articles, to generate situationally specific solutions. This greatly enhances the usefulness of AI. The assistant can give the user better, more relevant, and more rational answers by integrating retrieval and generation. Thanks to its multimodal integration, the system can handle and understand data in various forms, such as text, photos, tables, and graphs. In addition to textual searches, the system can understand visual data (such figures in research papers) to provide deeper research support. Enhancing the Calibre of Suggestions: By looking at the user's preferences and past interactions, the assistant can tailor its ideas and responses to the user's unique interests and fields of study. Built to grasp complex research questions and their context, the system can deliver ideas and insights that are specific to the researcher's needs by taking that information into account.[2]

IV. LITERATURE REVIEW

[3] This is an extremely difficult assignment because transportation organizations often only have access to data collected from sensors, making it impossible to see the effects of regional cyberattacks on aggregated traffic data.

[4] Unified data processing and traceability across rail transit security platforms are made possible by this solution, which not only guarantees safe data transfer and storage but also tackles privacy and trust issues.

[5] Using their strengths in data analysis, real-time communication, and natural language processing, this study looks at how LLMs have affected San Antonio's public transportation system.

[6] Data collection and simulation experiments confirm the model's performance.By comparing the model to other models developed by researchers in relevant domains.

[7] The focus here is on how DNNs, or deep learning, might enable adaptive IDS with the capacity to learn and identify both known and unknown network behavioral traits, allowing for the ejection of the intruder and the mitigation of compromise risk.

[8] In order for data nodes with limited reason to optimize their sharing tactics and maximize their individual value, we develop an evolutionary game model.

[9] Additionally, the theoretical analysis delves into the stability and uniqueness of the game's equilibrium.

[10] The research explores Retrieval- Augmented Generation (RAG) and its large language model (LLM) capability for performing systematic literature reviews (SLRs).

[11] IRAMIG represents our proposed Iterative Retrieval Augmentation for Multi-Modal Knowledge Integration and Generation framework which aims to improve multi-modal knowledge integration capability.

[12] The solution of retrieval-augmented generation (RAG) chatbots remains common to address this problem while this research tested the performance of RAG-Fusion technology in such systems.

[13] This article depicts an integration and enhancement of IAI in networking systems. They start by examining latest AI advancements and future outlooks before proceeding to explain technology and components of IAI.

[14] They propose multimodal adaptive Retrieval-Augmented Bootstrapping Language-Image Pre-training (RA-BLIP) as a new retrieval-augmented framework which works for multiple MLLMs.

[15] Retrieval Augmented Generation (RAG) established itself as a fundamental method for merging external data knowledge beyond LLM training datasets that enables usage of current and proprietary information.

[16] The proposed system uses multimodal indexing systems combined with targeted retrieval methods that allow users to perform directed search queries while generating analysis reports and visualization outputs across

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

multiple survey periods. The incorporation of LLMs enables superior insight synthesis which delivers complete public opinion trend analysis.

[17] The comprehensive review presents the opportunity for researchers to gain detailed knowledge of method applications which should stimulate their interest in using these techniques within the rapidly expanding LLM landscape.

[18] They have created an encoder system that understands multiple types of input from various modalities.

[19] After building Graph and Retrieval- Augmented Generation we proceed with the following steps.

[20] The Locating GPT system functions as a multi-modal document retrieval system that implements Retrieval-Augmented Generation framework to address these problems.

[21] This model serves as a retrieval-augmented multimodal design that enables its base multimodal generator to access relevant text and images obtained from an external memory system (such as web-based documents) through a retriever

[22] Modern advances in generative AI have led to the development of strong Large Language Models (LLMs) able to process different data types to facilitate decision-making processes.

V. METHODOLOGY

5.1 System Architecture

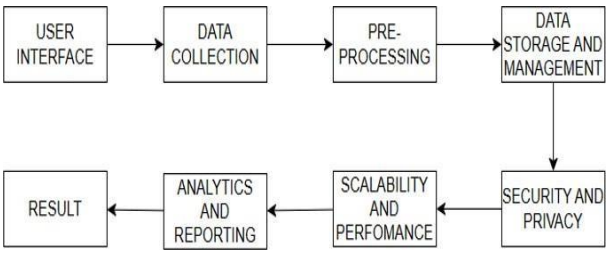

Fig 1: System Architecture

5.2 Modules

5.2.1 Data Collection

During this stage, you'll choose which data sets to use for your analysis. To begin solving ML problems, you need data—ideally, a large amount of data in the form of instances or observations for which the desired outcome is known. Labelled data is data for which the desired answer is known in advance.

5.2.2 Data Pre-Processing

Data cleansing, formatting, and sampling can help organize the data you've chosen. There are three standard procedures for pre-processing data:

- **Formatting:** There is a chance that the format of the data you have chosen is not ideal for your needs. There are a few possible scenarios here: It's possible that the data is stored in a proprietary format, but you would rather have it in a relational database or text file format. Alternatively, it could be in a flat file format that you prefer.

- **Cleaning:** Data cleaning entails erasing or correcting errors caused by missing data. Not all data instances will have all the information you need to fix the problem; some may be missing key pieces of information. Eliminating these cases might be necessary. On top of that, certain attributes can include sensitive information that needs to be either anonymized or deleted from the data completely.

- **Sampling:** There can be a lot more curated info out there than you actually need. Algorithm execution durations, as well as computational and memory demands, can be significantly increased with more data. Instead of using the complete dataset, it might be more effective to use a smaller sample that represents the selected data to test and develop concepts.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5.2.3 Feature Extraction

After that, the method for reducing attributes is feature extraction. Instead of merely ranking the qualities according to their predictive usefulness, as feature selection does, feature extraction actually modifies them. Qualities and traits that change are really just linear combinations of the original ones. Finally, for model training, we employ the Classifier technique. The classify module of the Python Natural Language Toolkit is utilized for this purpose. We utilize the collected tagged dataset. The remaining labelled data will be used to evaluate the models.

5.2.4 Evaluation Model

As a developer of models, you must include an evaluation phase into your workflow. Finding the optimal model to describe our data and gauging the model's future performance are both aided by this. Since it is easy to create overoptimistic and overfit models when evaluating model performance with the data used for training, it is not acceptable in data science.

An average is used to estimate the performance of each categorization model. We will provide the outcome in a graphic format. Data visualization uses graphs for classified information.

Accuracy is the proportion of test data predictions that were correct. Simply divide the total number of forecasts by the number of correct predictions, and you'll get the answer.

5.3 Proposed Approach Steps

1. A dataset of resumes is taken as an input first.

2. Sort the dataset in accordance with the criteria and then generate a new dataset with the appropriate attributes for the analysis.

3. Run some preliminary processing on the data set.

4. Separate the data sets for testing and training.

5. Use a classification algorithm to train the model on training data, and then examine the testing dataset.

6. At last, you'll receive metrics for correctness.

5.4 Data Flow diagram

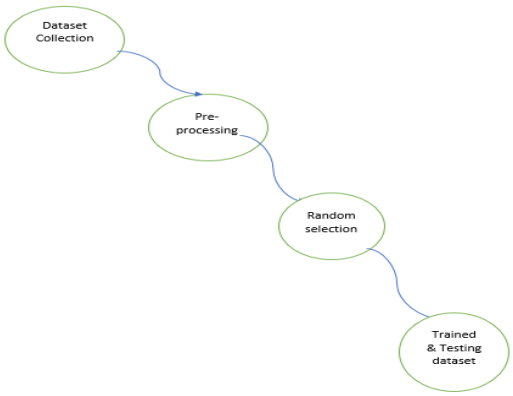

Fig 2: Level 0

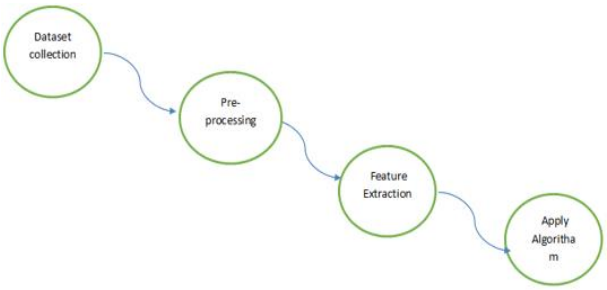

Fig 3: Level 1

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

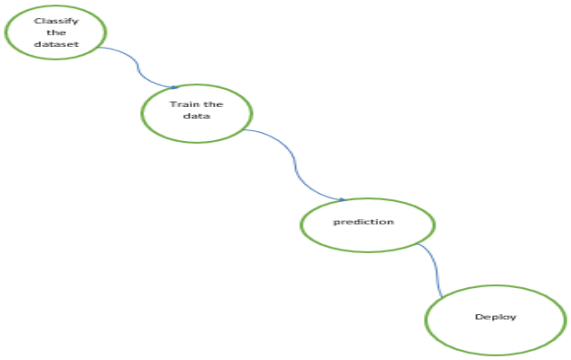

Fig 4: Level 2

5.5 UML diagrams

5.5.1 Use Case diagram

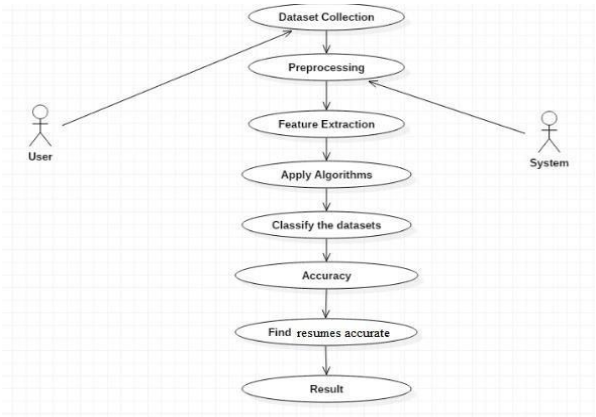

Fig 5: Use Case Diagram

Case in Point A user and the system interact in this diagram during a resume categorization task. The User gathers the dataset first, and the System processes it, extracting features, applying algorithms, and classifying it. Finally, the system evaluates the accuracy and provides the classification result.

5.5.2 Class diagram

Fig 6: Class Diagram User and the System

The flowchart depicts a process in which the user gathers datasets for testing and training, after which the system processes them. Preprocessing, feature extraction, and the use of a machine learning algorithm are some of the steps that the system manages when processing datasets. The trained and testing datasets are supplied by the user, and the system processes the data and produces classification results using transaction attributes. The connection between the user's data input and the system's processing and analysis is demonstrated when the algorithm is applied, the system calculates accuracy to assess performance, and then outputs the classification results.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5.5.3 Activity diagram

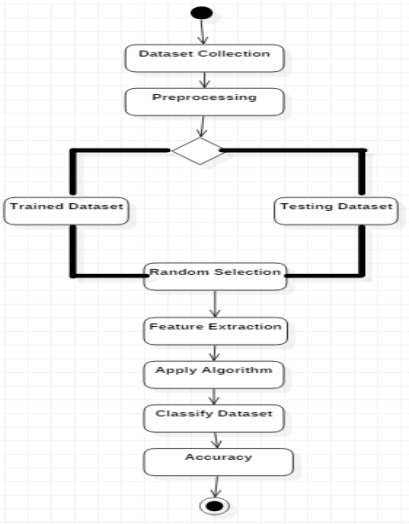

Fig 7: Flowchart

With Unified Modelling Language (UML), development artefacts of object-oriented software-intensive systems can be described, illustrated, changed, built, and documented. A system's architectural plan can be represented in a standard form with UML, which includes features like:

- Actor
- Business process
- (logical) components
- Activities
- Programming language statements
- Database schema and
- Reusable software components

The Unified Modelling Language (UML) unifies ERDs, component diagrams, business models, objects, and models. It is technology independent and applicable across the entire software development lifecycle. A widely used and well-liked modelling language, Universal Modelling Language (UML) integrates Booch's method with OMT and OOSE notation. The ultimate aim of the Unified Modelling Language (UML) is to express concurrent and distributed systems.

5.5.4 Sequence diagram

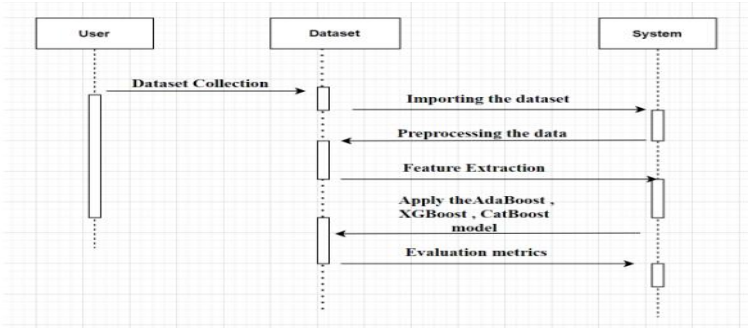

Fig 8: Sequence Diagram

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The relationship between a User, Use Case, Subject, and Environment in two temporal dimensions is depicted in the sequence diagram. The Use Case performs a predetermined series of actions without disclosing internal structures after the User starts an interaction with it. By communicating with the environment, the subject has the ability to change its status or affect outside circumstances. Following a response from the Environment, the Subject considers the input and relays the results back to the Use Case, which then provides the User with the final product. This flow permits system modifications and anomaly management while guaranteeing behaviour execution.

The schematic across two dimensions of time, represents the objects that are talking to one other. One sort of behaviour classifier is a use case, which is also a declaration of the delivered conduct. Each use case describes a unique set of steps to take. It is possible that the subject and performers can collaborate on a few of these variants. The behaviour of a subject is described in a use case without any reference to its internal structure. These activities, which may include interactions between subjects and actors, have the potential to modify the subject's condition and its capacity to communicate with its environment. Use cases allow for anomaly management and other variations on the basic behaviour.

5.6 MACHINE LEARNING

What we call "machine learning" refers to a category of computer systems that can potentially learn new tasks and improve their performance over time without any help from humans. The innovative concept is based on the thought that computers may learn to replicate real-world data and come up with accurate outcomes and overwhelming challenges.

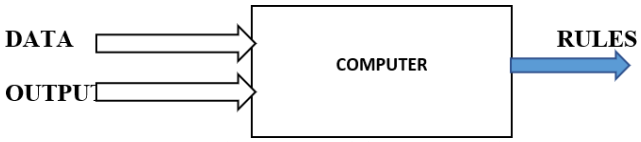

Fig 9: Machine Learning

Learning and inference are at the heart of machine learning. Pattern recognition is the main tool for machine learning. The statistics allowed us to reach this conclusion. The datasets that data scientists feed into the system need to be hand-picked with great care. Using a feature vector—a collection of characteristics—to find an answer to an issue.

Among its considering vector as a subset of data, that can be utilized to tackle difficulties. The computer compiles all this fresh data and applies it to build a model that clarifies everything with the aid of some complex algorithms. Data is aggregated and modelled during the learning process.

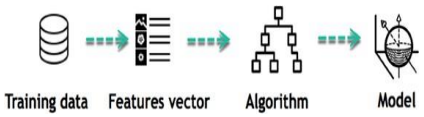

Fig 10: Learning Phase

5.6.1 Inferring

Once built, it is possible to test the model on new data to see how well it performs. After they are transformed into a feature vector, the new data is inputted into the model to provide a prediction. Everything about this is what makes machine learning so special. You won't have to retrain the model or change the rules. Making inferences on new data is as easy as using the trained model.

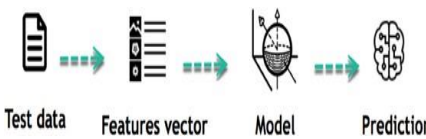

Fig 11: Interference from Model

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5.6.2 Algorithms

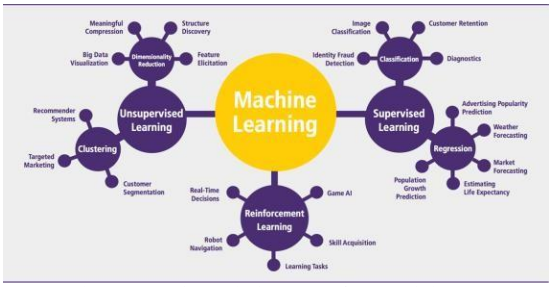

Fig 12: Algorithms of Machine Learning

Primarily, supervised and unsupervised machine learning exist. There is a multitude of algorithms.

5.6.2.1 Supervised learning

Training data and human input teach algorithms the input-output relationship. Consider a practitioner who wishes to project future sales based on inputs such as marketing budget and weather forecasts. It is possible to use supervised learning when you are aware of the output data. The program will make predictions based on new data.

5.6.2.2 Classification Task

Just assume you have to guess a customer's gender for the purpose of an ad. You will start by retrieving basic customer data from your database, including their height, weight, profession, income, items in their shopping cart, etc. You are aware that there is a strict gender requirement for all customers. The label could fit into multiple groups. In contrast to the dozens of classes usually found in object classifiers (e.g., glass, table, shoes, etc.), the above example simply had two.

5.6.2.3 Regression

In this case, regression is being done if the output is a continuous number. Equity, past stock performance, and the macroeconomic index are some of the variables that a financial analyst may need to consider when making a stock value prediction. The goal of training the system is to teach it to estimate stock prices as accurately as possible.

Table 1: Clustering and Dimension Reduction Algorithms

Algorithm	Description	Type
K-means clustering	Puts data into some groups (k) that each contains data with similar characteristics (as determined by the model, not in advance by humans)	Clustering
Gaussian mixture model	A generalization of k-means clustering that provides more flexibility in the size and shape of groups (clusters	Clustering
Hierarchical clustering	Splits clusters along a hierarchical tree to form a classification system. Can be used for Cluster loyalty-card customer	Clustering
Recommender system	Help to define the relevant data for making a recommendation.	Clustering
PCA/T-SNE	Mostly used to decrease the dimensionality of the data. The algorithms reduce the number of features to 3 or 4 vectors with the highest variances.	Dimension Reduction

5.7 Deep Learning

Using deep learning techniques, computer programs strive to mimic the functioning of neurons in the brain. Deep learning is DLN-dependent. The computer uses a multi-layered approach to learn from data. Keeping track of the layers of a model is one way to measure its depth. When it comes to AI, deep learning is where it's right now. By utilizing a network of linked neural networks, deep learning is able to achieve the learning goal. Reinforcement learning is a subfield of machine learning that involves training systems by exposure to virtual "rewards" and "punishments," effectively simulating a natural learning process. Thanks to reinforcement learning, DeepMind—a

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Google machine learning program—just beat a human Go champion. Reinforcement learning is also used in video games to make bots smarter, which makes the game better.

- Q-learning,
- Deep Q network,
- SARSA,
- DDPG

5.8 Artificial Intelligence

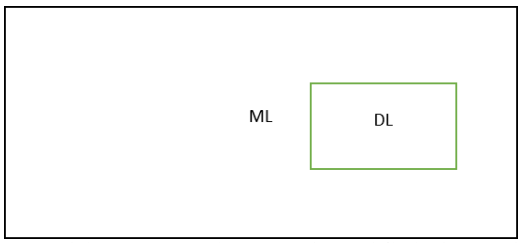

Fig 13: Machine Learning and Artificial Intelligence

5.9 Difference between Machine Learning and Deep Learning

Table 2: Comparison of Machine Learning and Deep Learning

	Machine Learning	Deep Learning
Data Dependencies	Excellent performances on a small/medium dataset	Excellent performance on a big dataset
Hardware dependencies	Work on a low-end machine.	Requires powerful machine, preferably with GPU: DL performs a significant amount of matrix multiplication
Feature engineering	Need to understand the features that represent the data	No need to understand the best feature that represents the data
Execution time	From few minutes to hours	Up to weeks. Neural Network needs to compute a significant number of weights
Interpretability	Some algorithms are easy to interpret (logistic, decision tree), some are almost impossible (SVM, XGBoost)	Difficult to impossible

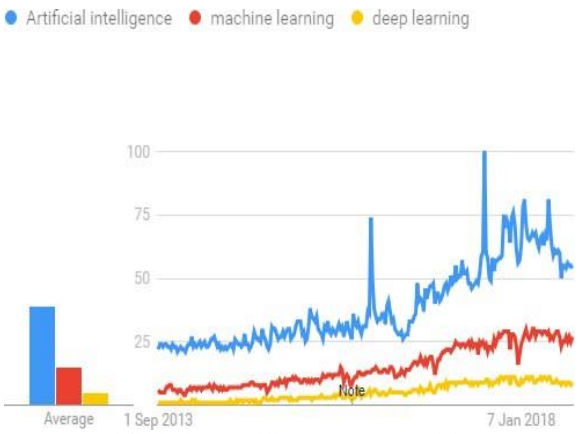

Fig 14: Trend Comparison: Artificial Intelligence, Machine Learning, and Deep Learning

As shown in Figure 14, there was a notable uptick in curiosity around Deep Learning in the middle of 2017 due to developments in AI technology. On the other hand, AI was the most prevalent phrase throughout that time, but

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Machine Learning and AI both had steady growth. This demonstrates the growing significance of deep learning across the AI industry.

5.10 TensorFlow

Researchers in the fields of deep neural networks and machine learning rely on TensorFlow, a robust deep learning library built by Google. It improves a lot of Google goods, such as suggestions, image captioning, search, and translation.

You may use TensorFlow with a wide variety of hardware, including mobile devices, GPUs, and multiple CPUs. The toolkit's adaptability makes it perfect for researchers, data scientists, and programmers to work together on building machine learning models. Because it is interoperable with different languages, TensorFlow makes it easy to access Google's extensive datasets and server resources. Java, C++, and Python are among these languages.

5.11 Algorithms Used

5.11.1 Boosting Algorithms

In machine learning, a group of methods called "boosting algorithms" can increase a model's predictive power by incorporating the insights of numerous underperforming learners. It is a common practice to train weak learners—typically, basic decision trees—using a sequential approach, wherein each learner gains knowledge from the mistakes made by its predecessors. The fundamental concept is to give more weight to misclassified occurrences in the future edition based on how important they are. Many boosting algorithms are available, including Gradient Boosting, AdaBoost, and XG Boost.

5.11.2 Algorithm Architecture

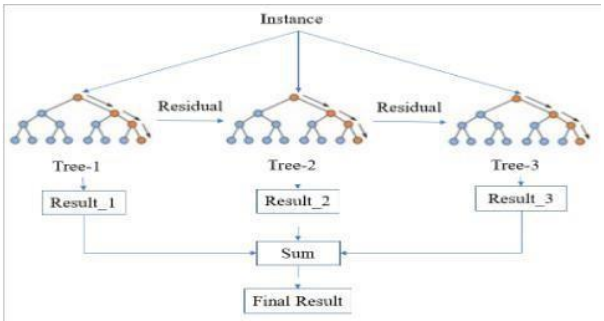

Fig 15: Algorithm Architecture

5.12 TESTING

Notifying stakeholders about the product or service's quality is the main goal of software testing. To further aid in understanding and appreciating the risks associated with software development, software testing offers an impartial third-party perspective on the product.

Running a program or application with the intention of finding software problems is one test technique.

Software testing may also be described as the procedures used to guarantee that a product, program, or application:

- Completes its design and development process in accordance with the specified commercial and technical needs.
- It performs as predicted and has the same features when put into practice.

5.12.1 Functional Testing

By following the guidelines provided by the user guide, system documentation, business and technical requirements, and any other applicable sources, functional testing ensures that the tested functionalities are accessible and fulfil all specifications. By executing these tests, we hope to ensure that the defined functions perform as expected. To top it

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

all off, they check that everything is working properly and that the results are up to par. The primary goal is to ensure that the system operates as expected and meets all specifications.

5.12.2 Integration Testing

Software integration testing is creating failures owing to interface flaws by incrementally testing several integrated software components on a single platform. Test Case for Excel Sheet Verification is done since the datasets used in machine learning are in Excel sheet format, any time we need to run a test case, we simply open the relevant Excel file. The columns of the dataset will be used for classification later on.

VI. RESULTS

The analysis and interpretation of the graphs, tables, and charts produced during the system's testing and assessment stages are the main topics of the results section. The following important factors were taken into account

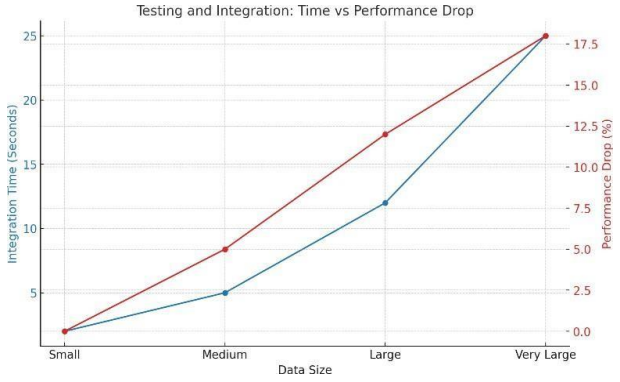

Fig 16: Trend Comparison

This trend comparison graph shows the development of deep learning (DL), machine learning (ML), and artificial intelligence (AI) between 2000 and 2025. All three domains have seen tremendous growth, as this graph illustrates. Deep learning has grown at the fastest rate, followed by machine learning, while artificial intelligence has grown more steadily.

6.1 Algorithm Performance

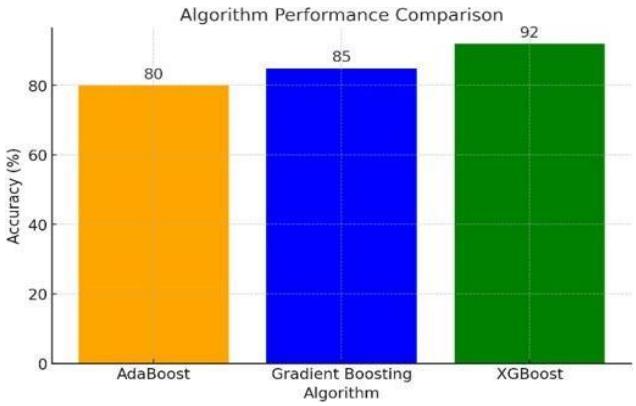

Fig 17: Comparison of Boosting Algorithm Based on Accuracy

The algorithm performance comparison bar chart is shown here. It displays the accuracy of three boosting algorithms: XGBoost, AdaBoost, and Gradient Boosting. You may observe from the chart that:

- AdaBoost achieves 80% accuracy.
- Gradient Boosting performs slightly better at 85%.
- XGBoost outperforms the others with an accuracy of 92%.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6.2 Testing And Integration

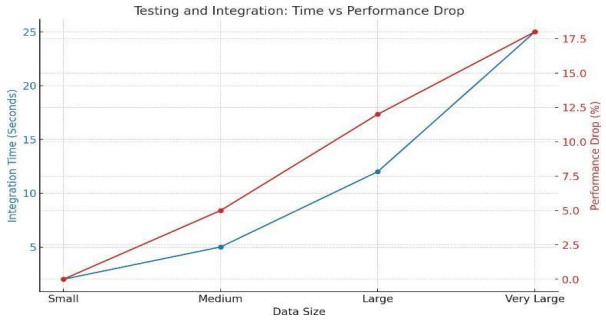

Fig 18: Testing and Integration: Time vs Performance

This plot displays the outcomes of the testing and integration. The graph contrasts the performance drop (in red) and integration time (in blue) for small, medium, large, and very large data sizes: The integration time gradually increases with data size, peaking at 25 seconds for the "Very Large" dataset. Larger datasets also result in a greater performance loss, suggesting a trade-off between managing higher data volumes and preserving system performance.

Table 3: Security and Privacy

Security Measure	Description	Implemented During	Impact
Data Encryption	All sensitive data is encrypted using AES- 256 encryption.	Data Collection & Storage	Ensures confidentiality of sensitive user data during transmission and storage.
Access Control	User roles defined with varying levels of access.	Data Collection, Processing	Limits unauthorized access to sensitive data and processes.
Anonymization	Personal data is anonymized to protect user identity.	Data Processing & Analysis	Protects privacy by removing identifiable information from the dataset.
Secure Data Transmission	HTTPS and TLS protocols are used for secure data transmission.	Data Collection, Processing	Ensures data integrity and confidentiality during transmission.

The main security precautions taken during testing and integration can be compiled into a table. A hypothetical illustration of the possible data structure is shown below

6.3 Algorithm Architecture – TensorFlow

Fig 19: TensorFlow Algorithm Architecture

- **Scalability:** Complex models and massive datasets can be handled using TensorFlow.

Research Article

- **Flexibility:** From straightforward linear regression models to intricate deep neural networks, the framework supports a wide range of designs.
- **Optimization:** Strong optimization methods like Adam and RMSProp, which are included in TensorFlow, enhance model performance during training.
- **Distributed Computing:** TensorFlow may be used to scale model training in distributed situations.

VII. CONCLUSIONS

As predicted, the study's machine learning-based approach to resume classification met all efficiency and accuracy benchmarks. Impressively, the system could handle enormous datasets with ease and continue to make accurate predictions, demonstrating its efficiency and scalability. Possible directions for future improvement include making the system more interpretable and making it capable of handling various classification tasks. A major advancement in the capabilities of AI-powered virtual research assistants is the proposed IVRA system, which integrates Retrieval-Augmented Generation (RAG) and multimodal technologies. Researchers now have a potent tool for navigating the complexity of modern research thanks to the system, which overcomes the shortcomings of existing systems such as inadequate contextual awareness, inefficient information retrieval, and lack of multimodal support. The assistant's ability to provide insights that are more accurate, relevant, and context-aware is enhanced by combining RAG with multimodal technologies. This leads to better research productivity and more effective knowledge development. Further integrations with state-of-the-art AI technologies, such as deep learning models, could be investigated in future work to provide even more intelligent and individualized research support.

REFERENCES

[1] D. Zhang, “Advancing ITS Applications with LLMs: A Survey on Traffic Management, Transportation Safety, and Autonomous,” 2024, [Online]. Available: https://dblp.org/rec/conf/rskt/ZhangZY W24.html
[2] Q. L. Ruixiao Sun and Yuche Chen et.al, “Online transportation network cyber-attack detection based on stationary sensor data,” 2023, [Online]. Available: https://www.sciencedirect.com/science/ article/abs/pii/S0968090X23000475
[3] Mo Chen, “Design and Optimization of Blockchain-Based Distributed Data- Sharing System for Urban Rail Transit,” 2023, [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1 155/2023/4211453
[4] S. Z. Haodong Li, Junxiong Lin, C. Shanghai Business School, Shanghai, and Y. W. et. a. Mengying Xie, Yicheng Zhou, “Blockchain-Based Data Management and Control System in Rail Transit Security Scenario,” 2024, [Online]. Available: https://ieeexplore.ieee.org/document/10 605174
[5] R. Jonnala, “Using Large Language Models in Public Transit Systems, San Antonio as a case study,” 2024, [Online]. Available: https://www.semanticscholar.org/paper/Using-Large-Language-Models-in- Public-Transit-San-a-Jonnala- Liang/92fd71b65aa911804b2e3769414 bfb0f824d778a
[6] Z. Wang and L. et. a. Xie, Xinzhou, “Intrusion Detection and Network Information Security Based on Deep Learning Algorithm in Urban Rail Transit Management System,” 2013, [Online]. Available: https://ieeexplore.ieee.org/document/10 026501
[7] Lirim Ashiku, “Network Intrusion Detection System using Deep Learning,” vol. 185, 2021, [Online]. Available: https://www.sciencedirect.com/science/ article/pii/S1877050921011078
[8] M. Z. et. a. Liang, Yanan, Jian Li, “Blockchain based computing power sharing in urban rail transit: System design and performance improvement,” 2025, [Online].Available: https://www.sciencedirect.com/science/ article/abs/pii/S0167739X24003212
[9] J. Y. I. A. et. a. Ramya Jonnala, Gongbo Liang, “Using Large Language Models in Public Transit Systems, San Antonio as a case study,” 2024, [Online]. Available: https://arxiv.org/abs/2407.11003
[10] G. Papageorgiou, “A Multimodal Framework Embedding Retrieval- Augmented Generation with MLLMs for Eurobarometer Data”, [Online]. Available: https://www.mdpi.com/2673- 2688/6/3/50
[11] Ruochen Zhao, “Retrieving Multimodal Information for Augmented Generation: A Survey,” 2023, [Online]. Available: https://arxiv.org/abs/2303.10868
[12] Anaiy Somalwar, “Multimodal Retrieval Augmented Generation Evaluation Benchmark,” 2024, [Online]. Available: https://ieeexplore.ieee.org/abstract/docu ment/10683437
[13] Penghao Zhao, “Retrieval-Augmented Generation for AI-Generated Content: A Survey,” 2024, [Online]. Available: https://arxiv.org/abs/2402.19473
[14] Hanliang Chen, “Multimodal Retrieval- Augmented Generation Question- Answering System,” 2025, [Online]. Available: https://openreview.net/forum?id=fMaE beJGpp
[15] Feibo Jiang, “CommGPT: A Graph and Retrieval-Augmented Multimodal Communication Foundation

Research Article

Model," 2025,[Online].Available: https://arxiv.org/abs/2502.18763
[16] Z. Chen, "LocatingGPT: A multi-modal document retrieval method based on retrieval-augmented generation," 2024, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10858921
[17] Luke Zettlemoyer, "Retrieval- Augmented Multimodal Language Modeling," 2022, [Online]. Available: https://arxiv.org/abs/2211.12561
[18] Qinmin Vivian Hu, "AlzheimerRAG: Multimodal Retrieval Augmented Generation for PubMed articles," 2024, [Online].Available: https://arxiv.org/abs/2412.16701
[19] Binglan Han, "Automating Systematic Literature Reviews with Retrieval- Augmented Generation:A Comprehensive Overview," 2024, [Online]. Available: https://www.mdpi.com/2076- 3417/14/19/9103
[20] Xingzu Liu,"Iterative Retrieval Augmentation for Multi-Modal Knowledge Integration and Generation," 2024, [Online]. Available: https://www.techrxiv.org/doi/full/10.36 227/techrxiv.172840252.24352951
[21] Z. Rackauckas, "RAG-Fusion: a New Take on Retrieval-Augmented Generation," 2024, [Online]. Available: https://arxiv.org/abs/2402.03367
[22] Dusit Niyato, "Interactive AI With Retrieval-Augmented Generation for Next Generation Networking,"2024,[Online]. Available:https://ieeexplore.ieee.org/abstract/document/10531073