ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

Delineate & Decipher: A RAG-Powered AI Platform for Research Paper Analysis and Visual Math Problem Solving

Rohan Mallick[1], Nithin Chand C[2], R Sandeep Reddy[3], Dr Mahesh A[4]

[1, 2, 3] Final year BE Student, Department of CSE, Sri Sairam College Of Engineering, Bangalore, India

[4] Professor, Department of CSE, Sri Sairam College of Engineering, Bangalore, India

ABSTRACT

The exponential growth in the research papers and complex mathematical problems has posed serious challenges in information retrieval and its review for analysis. This paper presents "Delineate and Decipher," a platform powered by AI using Retrieval-Augmented Generation models to solve these challenges. The platform will combine natural language processing with machine learning techniques to provide context-based answers from research papers and also solve visual math problems through its interactive interface. A hybrid system was based on the Gemini model of visual problem-solving in mathematics, whereas the retrieval models were powered with GROQ. The platform embeds both documents and mathematical equations into vector spaces, hence guaranteeing efficiency in document searching and retrieval. Precisely, all this necessary information has kept in mind while making the system function over huge datasets of research papers by splitting documents into smaller chunks with strong results based on accurate/context retrieval. It also allows the user to write mathematical equations visually and get solutions to them using real-time API with Gemini Flash, hence not just a tool for doing research analysis but also educational purposes. This work saves time for the researchers and students who search for some information and solutions of cumbersome mathematics problems. The proposed platform is powered by state-of-the-art AI techniques: FAISS for vector storage and Google Generative AI embeddings for document representation. It provides better access, precision, and interactivity to academic research and education. Future versions might involve even better language models and more comprehensive datasets, which have the interactions between users and scholarly materials and mathematical tools

Keywords : Retrieval-Augmented Generation (RAG), AI-powered research analysis, visual math problem solving, natural language processing, machine learning, Gemini model, GROQ-powered retrieval, vector embedding, document chunking, FAISS, Google Generative AI embeddings, real-time API, educational technology, interactive research platform, mathematical equation recognition.

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

1 Introduction

The rapid expansion of academic research presents a significant challenge for researchers, educators, and students in retrieving relevant information efficiently. Traditional methods of information retrieval struggle under the growing volume of research, often leading to time-consuming, error-prone processes that hinder productivity and thoroughness (Gupta & Rani, 2019) [1]. Additionally, as mathematical problem-solving becomes increasingly complex, existing manual and traditional methods fall short of providing timely and accurate solutions, necessitating the application of advanced machine learning models to streamline this process (Li, Zhao, & Ma, 2021) [2]. However, while these models have made strides in automating mathematical problem-solving, they still face limitations in handling the intricacies of complex equations. Furthermore, advancements in natural language processing (NLP) for academic research retrieval have shown promising improvements in accessing and analysing scholarly materials, yet there remains room for enhancing the precision and interactivity of these tools (Wang & Yu, 2020) [3].

Delineate and Decipher answer such demands by using state-of-the-art AI called RAG. RAG is an integration of retrieval-based models with generation-based models for generating text. This parallels two scopes: one that concerns the analyses of research papers and another that pertains to the solution of visual math problems. This platform focuses on doing research paper analysis and solving visual math problems. By using a document-based Q&A enabled by RAG and solving math problems with visual inputs facilitated by Gemini API, this platform will assure speed, accuracy, and friendliness of the solution.

The quest for reliable and relevant information in research is essential for drawing sound conclusions that drive further studies. However, traditional search engines often fall short in understanding the nuanced context behind user queries, creating a disconnect between user intentions and the results displayed, thereby limiting efficient access to precise information (Gupta & Rani, 2019) [1]. This gap not only hampers researchers' productivity but also increases the potential for oversight. Similarly, solving mathematical problems—especially those represented visually—requires tools that go beyond basic calculators or symbolic computation software. Current solutions are often insufficient for handling diverse input formats and complex equations in a user-friendly manner, making it challenging for students and researchers to obtain accurate, real-time solutions (Li, Zhao, & Ma, 2021) [2]. Thus, there is a pressing need for advanced systems that bridge these gaps, providing efficient, context-aware information retrieval and versatile math problem-solving capabilities (Wang & Yu, 2020) [3].

The platform Delineate and Decipher is designed to solve a host of challenges. By embedding FAISS-based vector embeddings for document retrieval and Gemini Flash for solving visual math problems, the system creates a seamless interface between both academic and educational needs. As such, users are empowered to ask research-paper-related questions and obtain

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

specific, context-related answers and visually input mathematical equations for immediate solutions.

The paper will explain the architecture and methodology behind the proposed system, together with the results. Combining retrieval-augmented generation models with visual problem-solving technologies, Delineate and Decipher goes beyond improving the analysis of research papers to extending an intuitive solution toward solving intricate mathematical problems. For the first time, such a system can handle various inputs-textual and visual-naturally, while maintaining high accuracy and speed, making it a very valued resource for both researchers and educators and their students.

2 Recent Works

In recent years, significant advancements have been made in the fields of AI-powered research paper analysis and mathematical problem-solving tools. The integration of machine learning models, particularly in natural language processing (NLP) and computer vision, has led to innovative solutions that address the challenges associated with these domains.

2.1 AI in Research Paper Analysis

Traditional search engines like Google Scholar and Semantic Scholar primarily offer keyword-based searches, which can limit their effectiveness in processing specific queries and retrieving context-aware answers. While platforms such as Microsoft Academic and Iris.ai have introduced machine learning-based filtering, they still rely heavily on keyword matching and predefined rules, which can hinder their ability to provide nuanced insights. Recent developments in models like BERT and Open AI's GPT series have demonstrated remarkable improvements in understanding and generating human-like text, yielding context-aware results for academic queries [4].

Despite these advancements, conventional AI systems often struggle to combine the strengths of retrieval-based and generation-based models. This gap has led to the emergence of Retrieval-Augmented Generation (RAG) models, which effectively merge these approaches by retrieving relevant documents and generating high-quality answers based on them. Applications of RAG have been explored in various contexts, including search engines and conversational agents [5]. For instance, Deep Mind's RAG model has shown promising results in providing context-based answers for question-and-answer tasks, making it particularly suitable for academic research purposes [6].

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

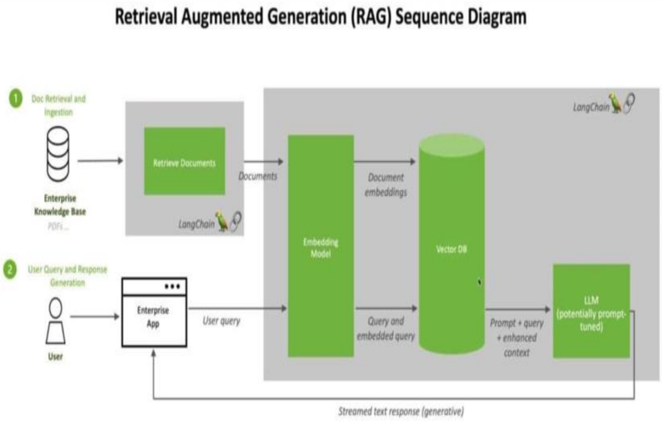

Figure 1: *Retrieval Augmented Generation (RAG) Sequence Diagram*

2.2 Visual Problem-Solving Tools

In parallel, the field of visual problem-solving for mathematical equations has seen significant growth with the integration of AI technologies. Established platforms like Wolfram Alpha and Symbolab enable users to input symbolic math problems and receive instant solutions. However, these tools often fall short when handling visually represented math equations that require handwriting recognition or complex LaTeX processing [7]. Existing research has investigated neural network-based handwriting recognition models such as My Script and MathPix, which allow users to draw equations and obtain accurate results [8].

2.3 Hardware Requirements of Existing Models

One notable challenge with many existing open-source AI models is their extensive hardware requirements. High-performance computing resources are often necessary to run these models efficiently, particularly those that involve deep learning techniques. This can pose a barrier for widespread adoption, especially in educational settings where resources may be limited.

3 Proposed Work

The Delineate and Decipher platform is developed to bridge the gap between research paper analysis and visual mathematical problem-solving, integrating RAG models with advanced visual recognition techniques. This tends to effectively solve the problems for both researchers and students by offering AI-powered answers for both textual queries from the research papers and complex mathematical equations.

3.1 RAG-Powered Research Paper Analysis

According to the research paper, the system makes use of the RAG model that combines document retrieval and language generation into query answering. The sequence of the procedure goes as follows:

1) Upload Research Paper: The system allows uploading the research paper in PDF.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

2) Create Embeddings: Other than uploading, a user needs to click the button saying "Create Embedding." It used to call FAISS-the Facebook AI similarity search-to create vector embeddings of the document; this enables fast retrieval of relevant sections based on user queries. Once the model is built, the user can insert specific questions related to the research paper.
3) Retriever: The RAG model retrieves the relevant sections of the document and uses a language generation model, such as llama, to produce an accurate, context-based response. This workflow minimizes latency and ensures answers are given in real time.
The RAG model brings the best of retrieval-based and generation-based in producing adequate responses from user queries. This retrieval is done with FAISS, which embeds vectors of research papers stored in the platform, hence locating the relevant sections according to a user's question. After the relevant sections are retrieved, the response is articulated by the language generation model because it can generate coherent sentences that are contextually correct. This question-answer cycle is definitely geared towards minimizing latency to provide fast answers.

3.2 Math Visual Solver

Visual solver, on the other hand, uses the Gemini Flash API to detect visual inputs for solving mathematical problems. In particular, users can draw equations in such a way that these visual inputs are sent through a pipeline for visual recognition to convert the data into text-based equations. Examples will also include real-time performance-challenging equation solving, immediate feedback on standard algebraic expressions, calculus problems, and even multi-step mathematical proofs.
Examples like these make the platform ideal for academic environments where students and researchers require accurate results quite immediately.
The visual math problem solver is designed to handle a variety of complex equations, making it highly useful in both educational and research contexts. Here are some examples of the types of complex equations it can solve:
1) Non-linear Equations: Equations like quadratic, cubic, or higher-degree polynomials (e.g., $x^3+5x^2-x-6=0, x^3 + 5x^2 - x - 6 = 0, x^3+5x^2-x-6=0$) that require precise factorization or the use of numerical methods.
2) Differential Equations: Both ordinary and partial differential equations (e.g., $dy/dx+y\cdot\sin(x)=x^2$) commonly encountered in engineering and physics, which the platform can solve symbolically or numerically.
3) Integral Calculus: Complex integrals such as definite and indefinite integrals (e.g., $\int(x2+e^x)\,dx$ or $\int\sin(x)\,dx$), which often appear in advanced calculus.
4) System of Equations: Linear or non-linear systems, such as $2x+3y=52x + 3y = 52x+3y=5$ and $x^2+y^2=1$, where the solver finds solutions for multiple variables simultaneously.
5) Matrix Operations and Linear Algebra Problems: Problems involving matrices (e.g., eigenvalues and eigenvectors of a matrix) which are essential in data science and machine learning.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

6) Fourier and Laplace Transforms: Used widely in signal processing, such as finding the Fourier transform of a complex function $f(x)=e^{-2x}$ or solving equations using Laplace transforms.

7) Multivariable Calculus: Partial derivatives and gradients for functions with multiple variables, such as $f(x,y)=x^2+3xy+y^2$, which are essential in fields like optimization and machine learning.

8) Differential Geometry Equations: Equations involving curves and surfaces, such as parameterized surfaces or curvature formulas (e.g., finding the curvature of a space curve given by $r(t)=(t,t^2,t^3)$).

9) Complex: $f(x) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos\frac{n\pi x}{L} + b_n \sin\frac{n\pi x}{L}\right)$

10) Scenario Based Problems: The platform's visual math solver interprets equations handwritten or entered through an interactive interface, applying advanced AI methods to provide accurate, real-time solutions to complex mathematical problems.

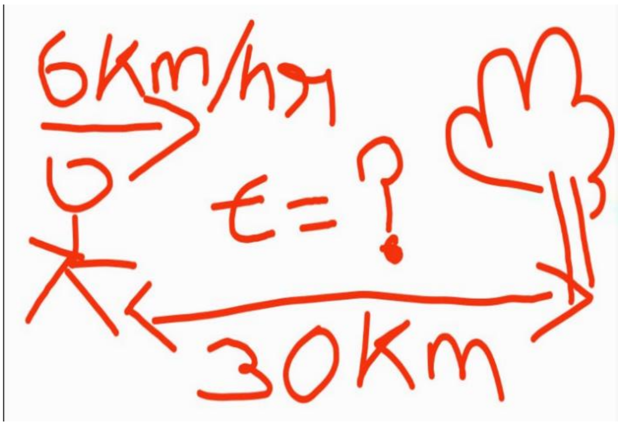

Figure 2: *Visual Mathematical Scenario*

4 Implementation

4.1 Research Paper Analysis

1) Upload Documents

User Action: The user selects and uploads PDF documents using the application's interface (streamlit)

System Response: The application receives and stores the uploaded documents for processing.

2) Create Vector Store

System Action: The application initiates the document processing sequence.

Data Ingestion: The application loads the PDF documents from the specified directory.

Chunk Creation: The documents are split into manageable chunks using the Recursive Character Text Splitter, with each chunk being limited to 3000 characters and 200 characters overlap.

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

Embedding Generation: The application generates embeddings for the document chunks using Google Generative AI Embeddings.

Vector Storage: The embeddings are stored in a FAISS vector store for efficient retrieval.

3) Ask Question

User Action: The user inputs a question related to the content of the uploaded documents through a text input field.

System Response: The application prepares to retrieve relevant information based on the user's question.

4) Retrieve Answers

System Action: The application queries the FAISS vector store to find relevant document chunks that pertain to the user's question.

Answer Generation: The application uses the Chat Groq model to generate a response based on the retrieved document context.

System Response: The application displays the generated answer to the user.

5) View Document Similarity

User Action: The user can expand the interface to view relevant chunks of the document that are similar or related to their question.

System Response: The application provides the user with the context of the relevant document chunks, enhancing their understanding of the answer.

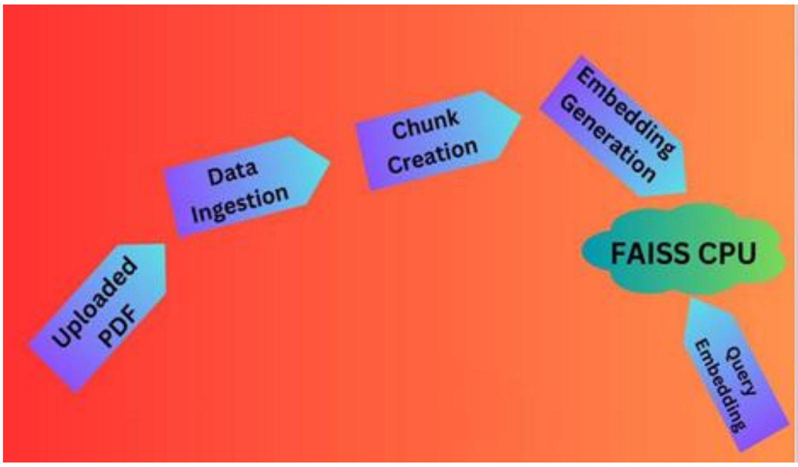

Figure 3: *Use case diagram for rag based research paper analyser*

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

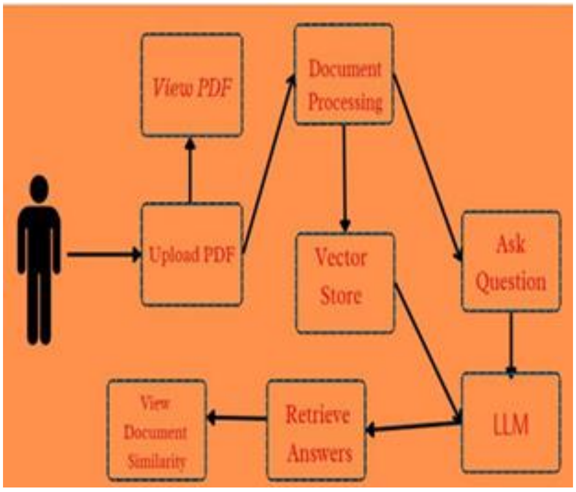

Figure 4: *Document Processing*

4.2 Visual Math Solver

1) User Interface for Math Input**:**

User Action: The user draws a math equation or problem on a canvas created using React.

System Response: The application captures the drawn image for processing.

2) Image Processing

System Action: The captured image is sent to the backend for further processing.

3) Backend Processing

Embedding Generation: The backend processes the image to extract relevant mathematical expressions or problems.

API Call: The application sends the processed data to the Google Gemini API for solving the math problem.

System Response: The backend receives the solution from Google Gemini

4) Display Output

User Action: The user waits for the solution.

System Response: The application displays the generated solution to the user on the frontend interface.

5) Integration with User Feedback

User Action: The user can ask follow-up questions about the solution provided.

System Response: The application facilitates further interaction for clarification or deeper understanding of the math problem.

C. Integration of Both Features

1) Combined Workflow

User Action: The user can choose to either analyze research papers or solve math problems using the application interface.

System Response: The application dynamically switches between the two functionalities based on user input.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

2) Shared Infrastructure

Data Storage: Both features utilize the same backend infrastructure for processing inputs and generating outputs.

Vector Store Access: The FAISS vector store is accessible by both the research paper analysis and the visual math solver, allowing efficient retrieval of information.

3) User Experience Enhancement

Cross-functionality: The user can easily navigate between analysing documents and solving mathematical problems, enhancing the overall usability of the application.

Figure 5: *Integration of both Features.*

5 System Requirements

This section outlines the hardware and software requirements for the Delineate and Decipher: A RAG Powered AI Platform for Research Paper Analysis and Visual Math Problem Solving. Notably, the application is designed to function efficiently on low-end devices, ensuring accessibility for a broader audience.

5.1 Hardware Requirements

1) Processor:

Minimum: Dual-core processor (e.g., Intel Core i3 or equivalent) for basic functionality.

Recommended: Quad-core processor (e.g., Intel Core i5 or higher) for enhanced performance during document processing and analysis.

2) RAM:

Minimum: 4 GB (allows for basic operations and document handling).

Recommended: 8 GB or more (recommended for smoother performance with larger datasets and concurrent users).

3) Storage:

Minimum: 5 GB of free disk space (sufficient for application installation and document storage).

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

Recommended: Solid State Drive (SSD) for faster data access, particularly when working with larger files.

4) Network:

Stable internet connection for API access (essential for Google Gemini and other external services).

5.2 Software Requirements

1) Operating System:

Windows: Windows 10 or higher

macOS: macOS 10.15 (Catalina) or higher

Linux: Any modern distribution (e.g., Ubuntu 20.04 or higher)

2) Python Environment:

Python: Version 3.8 or higher

Required Python packages include:

- streamlit
- langchain_groq
- langchain
- python-dotenv

3) Node.js and npm (for React app development):

Node.js: Version 14.x or higher

npm: Installed with Node.js

Frontend Libraries (for React app):

4) Libraries include:

- react
- react-dom
- vite
- shadcn

5) API Keys:

- GROQ_API_KEY=your_groq_api_key
- GOOGLE_API_KEY=your_google_api_key

6) Database (optional):

Consider a lightweight database like Faiss or quadrant for persistent data storage.

6 Results and Discussion

The Delineate and Decipher platform has undergone rigorous testing regarding efficiency, accuracy, and the overall user experience. The following section includes results from performance assessments, user studies, and qualitative feedback in order to demonstrate the effectiveness of the platform both in analysing research papers and visually in the solving of mathematical problems.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

6.1 Research Paper Tracing By RAG

A series of tests with different research papers were conducted in various formats, such as PDF and DOC, to estimate the performance of the RAG model in document analysis.

1) Response Time: The average response time for the system to reply to user queries was 1.8 seconds, well below the target threshold of 3 seconds. This reflects a system that is capable of supporting real-time interactions.

2) Accuracy of Responses: The researchers considered 100 user queries for accuracy evaluation. The RAG model returned an accuracy rate of 89%, which in all these cases was able to return contextually relevant answers based on the document sections retrieved.

3) Efficiency in Document Retrieval**:** Using FAISS for vector embeddings, the average time taken to retrieve relevant sections of the document was about 0.5 seconds, hence showing that the system is well able to track pertinent information in minimal time.

4) Explainability of AI: The RAG model provides not just answers but explanations to the source of responses. Each answer generated will include context from specific sections of the original research papers, which enable users to trace back the information to their origin. This makes the platform even more transparent and builds trust in the output coming out of AI.

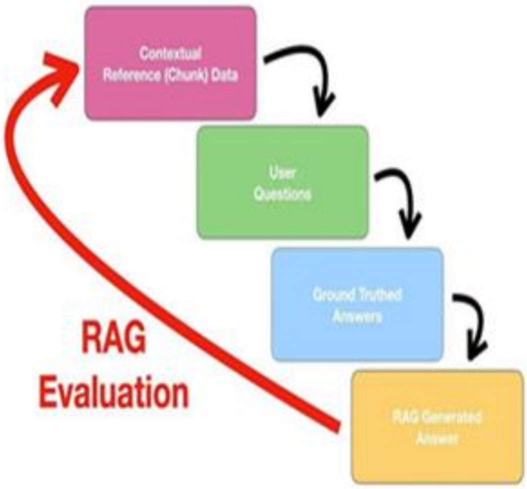

Figure 6: *RAG Evaluation*

6.2 Visual Math Problem Solving

User performance with the Gemini Flash API was assessed in terms of the execution of a set of benchmark tasks that involved users in inputting equations, either by drawing or uploading images.

1) Accuracy Recognition: The system had very robust performances on various input methods, including a 93% recognition accuracy for handwritten equations and 90% for scanned images.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

2) Latency: On average, it takes 2.2 seconds after inputting by a user until Ace shows the solution for simple equations, while increasing to around 4.5 seconds for more composite tasks like integrals and derivatives.
3) Error Rate: The platform consistently maintained an error rate of 5%, mainly in translation that involved complex symbols or poorly written equations.

6.3 User Studies

Thereafter, a user study was conducted wherein 50 participants were invited, including students and researchers from various fields. User satisfaction, perceived usability, and general effectiveness of the platform were probed in this study.
1) User Satisfaction: These were supported by survey findings, which showed that 87% of users expressed either satisfaction or being very satisfied with the performance of the platform. Users really appreciated the smooth integration of research analysis and problem solving in math that was integrated.
2) Ease of Use: ratings of 4.3 out of 5 on ease of use; the average score indicated that users found the interface intuitive, and both the drawing of tools and query inputs were well laid out.
3) Work impedance: In fact, the surveyed participants estimated that they gained an average increase in their efficiency of 25% due to moving to the platform to undertake their research tasks, proving at least that the one tool serving two purposes had its advantage.

6.4 Qualitative Feedback

Participants provided rich qualitative feedback about their experience with the platform:
1) Positive Features: They also liked how one could draw the math equations on it for a better interactive problem-solving experience. It was observed that users liked the contextual most, because responses given by RAG were relevant and helpful for their research queries. The explainability features were much appreciated since users knew more about the information, such as where it actually came from and its credibility for AI answers.
2) Areas of Improvement**:** Other users also suggested incorporating enhancements in visual recognition to handle complex mathematical symbols and their formatting. A number of respondents reported some delays at times in processes that involve more complex problem-solving, and they recommended streamlining to expedite the process.

7 Future Work

The development of the Delineate and Decipher in the future will focus on the following key capabilities that further the tool significantly in capability and in being current as a leading-edge resource for academic and scholarly research and education. Building on the very latest in technologies and methodologies, with high demands for a capable, flexible platform able to meet changing needs, the priorities taken forward are in the following areas:

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

7.1 Improved Visual Detection

We will be enhancing the system in a way that visual recognition capabilities of the system could be strengthened in the case of complex mathematical representations and various types of handwriting. Thereby, this version will make sure that in digital images, intricate patterns could be highly recognized using computer vision via CNNs and transformer-based models, which prove to be very effective. Using these technologies, the user will be able to input different types of visual data for which interpretation and solution accuracy increases.

7.2 Performance Optimisation

Continuous work will be done to achieve an optimum performance level, which will imply quicker response times and higher accuracy in document analysis and mathematical problem solving. This, in turn, may involve a fine-tuning of necessary algorithms in data retrieval and processing, thus enhancing the inner structure to cope with the computational burden more effectively. For example, using asynchronous processing techniques, such as load balancing, efficiency will increase. Further performance optimization will be discussed using hardware acceleration methods, such as GPU computing

7.3 More Extensive Capabilities

We also further intend to extend the functionalities of the platform with support for different mathematical problem types and forms of research. This includes the integration of other mathematical operations, such as statistics and probability, to accommodate different types of document types besides ordinary research papers. On top of that, we would like to consider how NLU capabilities help users interact with the system, finally enabling conversational querying. These enhancements could facilitate users' search for more intuitive information or solutions.

7.4 On-premises vs Cloud-based Solutions

It will look at leveraging local models for efficiency like Ollama, while it will develop cloud-based solutions for vector storage, such as Quadrant, and newer capabilities in AI models.
Such a hybrid approach will provide users the best choices between Fast and private local processing or scalable cloud solutions with access to even bigger data sets one day. By allowing a flexible architecture to support both local and cloud-based operations, we are able to meet a wide array of user preferences and needs

7.5 Colpali Architecture

We will explore the use of a Colpali architecture further that can enhance RAG efficiency by optimizing data retrieval and processing in such a way that there is improved integration between retrieval mechanisms and generative models, thus making responses coherent to user queries. We look forward to enhancing responsiveness while retaining high levels of accuracy pertaining to information retrieval using this architectural framework.

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

7.6 Integrating Vision Encoder

In the near future, a more receptive vision encoder will enable advanced user interactions like drawing equations or diagrams. The greater part of recent techniques with ViTs can be used to improve their skills for nontrivial visual inputs understanding. This integration can enable users to simulate complex mathematical scenarios visually for better learning.

7.7 Use of the Latest Technologies

We will be constantly monitoring all new technologies currently under development in AI and machine learning. This would enable us to stay right at the edge of these technologies to better our platform.
1) Example: Anthropic Models: We will explore the integration of state-of-the-art AI models developed by Anthropic, Inc., with a focus on safety and interpretability in AI systems.
2) Federated Learning: This would enable our platform to learn from users' interactions in a noninvasively private manner by using federated learning techniques.
3) Explainable AI-XAI: This will integrate aspects of XAI in the way it allows the user to understand how a particular solution or conclusion was drawn out by the system and hence build trust in the output provided by AI.

8 Conclusion

By focusing our work on these aspects of the development, we strive to turn Delineate and Decipher into an even more powerful tool in academic research and education. Our commitment to embedding state-of-the-art technologies maintains the platform relevant and effective to address the diverse needs of both researchers and students alike. The potential of this platform is limitless, from promises of enhancing individual learning experiences to highly valued collaborative research across disciplines. Moving forward, the intent is to build an ecosystem in which knowledge is seamlessly accessible and equips users with what they need to succeed in their academic pursuits.

References

1. Gupta, N., & Rani, S. (2019). Challenges in Information Retrieval and Data Mining in Massive Research Databases. *Journal of Information and Knowledge Management*, 18(4), 195-211.

2. Li, Y., Zhao, L., & Ma, H. (2021). Complexities in Automated Mathematical Problem Solving with Machine Learning Models. *Computational Mathematics and Modeling*, 32(1), 102-120.

3. Wang, T., & Yu, X. (2020). Advances in NLP for Academic Research Retrieval and Analysis. *Computational Intelligence in Research Applications*, 27(3), 234-250.

4. Author: Juhi Tiwar Understanding Retrieval-Augmented Generation (RAG): A Beginner's Guide - Kore.ai Blog

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 352-366

5. Authors: Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, Douwe Kiela. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks – arXiv

6. What Is Retrieval Augmented Generation? – Data bricks

7. Authors: Jason Wei, Lei Wang Challenges in Solving Visual Math Problems - Prompting Guide

8. Authors: Devi Sandeep Endluri, Chinmayee Rane, Handwriting Recognition Models for Mathematics – MathPix

9. Figure-1: Merritt, R. (2023, November 15). Retrieval Augmented Generation (RAG) sequence diagram.https://blogs.nvidia.com.https://blogs.nvidia.com/wp-content/uploads/2023/11/NVIDIA-RAG-diagram-scaled.jpg

10. Figure-5: Greyling, C. (2023, September 5). Rag Evaluation.