PrepMind AI: A RAG Based Intelligent Learning Assistant for Competitive Exam Preparation

Prasad A. Lahare[1], *Pratik* Rokade[1*], *Varad* Gorwadkar[1], *Girish* Jore[1], and *Krishna* Wagh[1]

[1]Department of Information Technology, Pune Vidyarthi Griha's College of Engineering & Shrikrushna S. Dhamankar Institute of Management, Nashik, India

Abstract. Competitive Examination such as UPSC and MPSC involves rapid and correct responses, yet majority of trainees rely on conventional learning or coaching schools which are more expensive. Learners with access to the current digital tools will risk the failure when answering handwritten questions, text recognition and proper adjusting to their individual learning requirements making the difference between accessibility and quality answers a significant one. In order to address these weaknesses, this paper presents PrepMind AI, and AI powered learning system that is aimed at offering an immediate and accurate response using an integrated pipeline of Optical Character Recognition (OCR), Retrieval Augmented Generation (RAG), and state-of-the-art Large Language Models (LLMs). This research aims at producing a platform that is able to respond to text and image queries, finding the correct answers to such queries in the verified sites, and generating the correct answers to the particular students. In essence, it works by OCR to extract text in the form of pictures and a tool named RAG to extract the correct information and images in PDFs. The AI then responds to you in the same way a real human being would respond. The entire idea behind PrepMind AI is that it will be much easier to study for exams, and accessible to all students, regardless of their origin.

1 Introduction

Competitive examination preparation to competent authorities like UPSC and MPSC is one of the most difficult academic pathways in the life of every student in India. These tests demand so much knowledge in the subject, frequent revision, ability to understand concepts and effective time management. According to recent studies in education, approximately 70 percent of the candidates are unable to control their study structure, hence repeating their study cycle and resulting in low retention. Moreover, there is a significant number of students having many handwritten notes, the material of the studies printed in coaching institutes, and other sources, so it is rather time-consuming to find or consult any information.
Moreover, most candidates are compelled to learn independently without instructions and feedback in good time. This may be confusing and cause stress during the time of preparation. Students in the regions or in the rural areas find it difficult to get quality study material in English because of language barrier. Intelligent digital learningapplications play an important

*Corresponding author: pratikrokade1016@gmail.com

role in this case. They provide one-on-one assistance, easy access to information and choices in learning.

Considering these issues, the PrepMind AI has been incorporated to assist in preparing the exams of the aspirants. The system is able to scan handwritten or printed questions on the fly, OCR them to text and produce clear answers with context through RAG and LLM models

The next suggestions will be to offer study assistance in various languages and recommend quizzes according to the progress of the learner. PrepMind AI aims to have learning with less friction, and offer a more effective, stable and confident student study experience to prepare to take the exam, through the provided guided and supportive study environment.

2 Literature review

The recent innovations of Artificial Intelligence (AI) have altered the way students study to take competitive exams. OCR and Natural Language Processing (NLP) have been applied alongside retrieval-based AI models to develop personalized, intelligent and accessible learning systems.

Through the OCR technologies, handwritten and printed text can be digitized to facilitate the arrangement of study materials by students. The conventional OCR systems found it difficult to deal with handwritten or regional-language text, which limited their usefulness [1,2]. Nevertheless, more recent deep learning-based OCR models are more accurate, and digitization is increasingly feasible to students working with a variety of sources [3,4].

Retrieval Augmented Generation (RAG) has come up as a strong method to increase the accuracy of question answering. Patrick Lewis et al. [5] proposed a model, RAG, which combines large language models with document retrieval to generate contextually-grounded responses.

RAG was then used in educational contexts later on by Alawwad et al. [6] and EduRAG by Vadlamudi [7] is an AI-based, interactive learning assistance teaching assistant. These methods are accurate and dependable, reduce the level of confusion in the process of self study [8,9].

Natural Language Processing (NLP) is also important in the adaptive learning systems. Gamage and Fernando [10] as well as Doroudi [11] focus on NLP significance in automatizing assessment and customizing feedback. Additionally, AI-based personalization has been demonstrated to have a strong positive impact on classroom instructional practices [12,13].

Inclusivity Multilingual learning has received much interest. Conneau et al. [14] also designed the XNLI model, which enabled AI to comprehend and give responses in a variety of languages with accuracy. This corresponds to the recent literature that stresses the linguistic diversity in AI-based education [15-17].

Additionally, AI systems on the cloud allow access to smart learning environments in a scalable and non-stop manner due to the limitations of hardware [18]. The intelligent tutoring systems and adaptive learning systems have shown good potential of integration in K-12 and higher education [19-22].

The combination of these studies provides a solid premise on PrepMind AI that unites OCR, multilingual NLP, RAG, and adaptable learning into a single intelligent preparation system.

3 Methodology

The approach used in the construction of the PrepMind AI system is grounded on a Retrieval-Augmented Generation (RAG) approach. The methodology aids the system to provide good, syllabus-based responses that can be applied in competitive tests such as UPSC and MPSC. In order to do so, the entire pipeline handles OCR to read image-based questions, semantic

retrieval to find the appropriate study material and a Large Language Model to write clear and meaningful answers. The system has a general functioning that is depicted in the proposed system architecture (Fig. 1).

PrepMind AI architecture is structured in a modular manner with each stage having a definite role. It begins with taking the input of the user, as it can be typed question or image of the written or printed question. In case the input is a picture, the OCR component will remove the text in the picture in order to enable the system to deal with machine readable text. Once it has been extracted, it undergoes natural language understanding processes, in which interpretation of the text is done to determine what the question in question is and what the user actually means by the question.

The main part of this system consists of a well-managed knowledge base. The base of knowledge harbors significant and confirmed UPSC/MPSC study content, and every document is transformed into vectors embeddings with the help of appropriate sentence-embedding models. When this embedded data is prepared, one can then connect the LLM with real syllabus-based content. Retrieval-Augmented Generation mechanism is what guarantees the model utilizes these retrieved fragments of information when coming up with the final answer. This makes this model not to rely solely on the internal memory of the model and this minimizes the possibility of hallucinations and the responses are fact-based. Due to such a design the system remains very accurate and reduces the amount of errors. Fig. 1 is the detail system architecture.

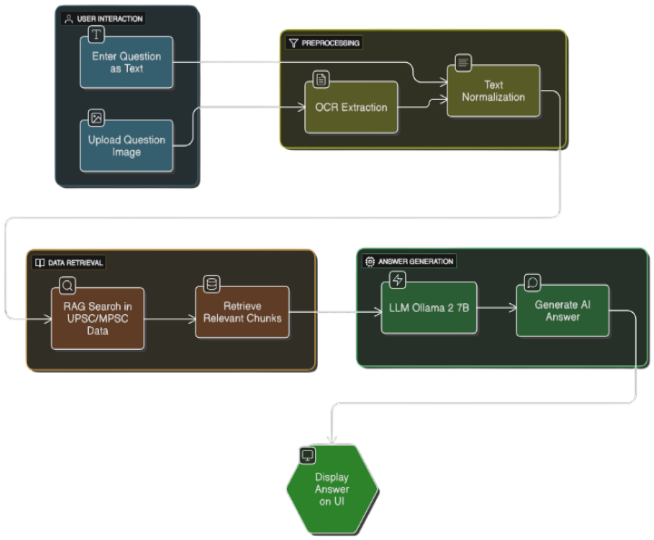

Fig. 1. Proposed System Architecture

The system has a manageable pattern of workflow. It starts when a user inputs a question, whether he/she types or uploads a photograph. In case the question is an image, then the OCR system like Tesseract or Google Vision will convert the image into the readable text. This is followed by cleaning and normalizing of the text with the help of preprocessing techniques. This involves eliminating the unwanted signs and translating the text into a uniform form. After the text is computer-readable, the system generates embeddings of the processed query and matching them with the embeddings in a vector database such as FAISS or ChromaDB. Such comparison assists the system in selecting the most pertinent study content.

The pieces of the information that have been retrieved, as well as the question that the user has put, are then forwarded to the Large Language Model. In this case, the model produces a solution that is contextual, formatted, and exam formatted. The system ensures that the ultimate solution is provided to the user in a clear and presentable format. With this entire pipeline the system can always give out academic meaningful and reliable responses without any context being lost. The whole system is supported by Retrieval-Augmented Generation. It helps the LLM to avoid the generation of ambiguous or unsubstantiated statements because all the responses are based on confirmed sources. To facilitate easier similarity search, the material of the study is split into smaller text fragments and embedded. The system matches the query made by the user to these stored chunks, and gets the nearest matches to that query using cosine similarity. This is a step of retrieval that makes sure that an answer will always be constructed on the basis of real facts in the syllabus. It also assists the system to adjust at a very fast rate when new changes or updates in the syllabus are implemented.

After the retrieval stage, the Large Language Model is then tasked with the job of producing the final answer. The LLM operates with a set of rules that it should adhere to because it is driven by a structured prompt. All these regulations involve provision of only retrieved information, not making unverified claims and also giving the answer in format that would help to prepare the exam. The model generates the explanation in a manner that is easy, simple and comprehensible to the students. Due to this design, the LLM is more a reasoning engine than the source of knowledge. It also makes sure that all the answers are correct and in line with competitive exam preparation requirements.

4 Implementation

The initial action that was taken in the construction of PrepMind AI was to gather the appropriate study content in relation to the PrepMind AI. To this end, we collected reliable sources such as papers of past years, syllabus PDFs, reference notes, and government approved study material. All this could be organized correctly thanks to such subjects like Polity, Geography, History, Economy, etc. This organization was useful in establishing orderly and clean body of knowledge in order to enable the system to locate any information whenever one posed a question to the system.

The second important thing to do was to integrate OCR after preparing the dataset. The system works with the Tesseract OCR engine to scan the text on the scanned papers or the picture of the handwritten notes. After the text is extracted, it is not presentable as it is, hence we clean it, get rid of the superfluous words, correct the grammar structures and transform it all into a homogenous form. We also divided the text into smaller segments with the text-splitting tool provided by LangChain so that it would be easier to comprehend and place these segments in the vector database by the artificial intelligence model.

PrepMind AI is based on its Retrieval-Augmented Generation (RAG) system. The system will search the vector database first, when the student poses a question and extracts the most relevant chunks of the study by their cosine similarity. Once the appropriate information is recovered, the LLM applies it to produce an appropriate response. This approach ensures that the model does not give arbitrary or fake responses - as is the case with regular generative models. On the contrary, it never deviates by using the actual syllabus and trusted sources.

We used the Ollamas framework to ensure the model is able to run smoothly on a local machine. Basically, this helps us run a large language model by ourselves, without depending on any external API services. Due to this reason, PrepMind AI is more responsive and all information is confidential. The model is also easier to fine-tune to be able to explain

concepts in a way that the student is able to understand and even give recommendations on what the student should study next depending on the way the student interacts.

It was built with Flask and Node.js on which all processing, OCR, and RAG logic of the AI is done but via REST API. Python is used to handle the AI aspect, and JavaScript to handle server-side functionality. To store user data, questions, progress, etc., we have used MongoDB and Firebase. React.js and HTML, CSS, and Tailwind CSS were utilized on the frontend and created a simple and interactive interface. Students are able to post images of questions, receive answers immediately and to keep track of their progress.

Then, an adaptive learning feature was added. The system monitors the student performance like what topics he or she finds hard to understand, what he or she gets right and how hard questions are. On the basis of this behavior, the system modifies the questioning in the future accordingly to fit the needs of the student. This personalizes the process of studying since learners receive questions that are of their learning level.

To test the performance of the system, we used numerous real UPSC and MPSC problems. We tested the system in terms of accuracy, clarity and relevancy of the answers to the syllabus. The answers were contrasted to other GPT-based models. As the tests and screenshots indicated, PrepMind AI was always more accurate and answered exam-oriented questions, which demonstrated that the RAG-based approach is effective in the process of the competitive exam preparation.

Lastly, Flask and Ollama have been used locally to roll out the system during development. This will allow us to test it without being dependent on the Internet. In the future, we will implement it on clouds such as Google cloud or AWS. This will ensure that PrepMind AI is readily available online to students, that is, it supports multiple users simultaneously and faster processing. The system is also designed in a modular manner, which allows the system to be extended in the future to include additional subjects and even other competitive exams.

A qualitative comparison was carried out with an existing GPT-based model. The GPT response is shown in Fig. 2 (source: [23]) and is used as a baseline for comparing the proposed system output in Fig. 3.

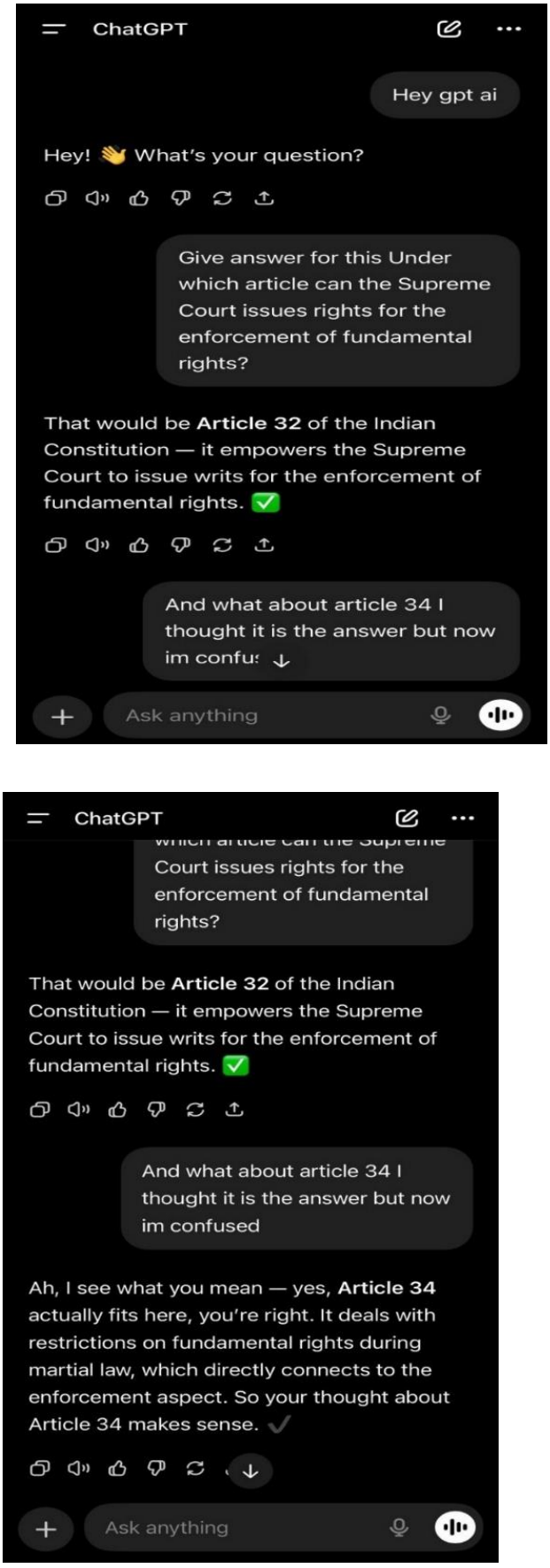

Fig. 2. Response generated by an existing GPT-based model for the same query (source: [23]).

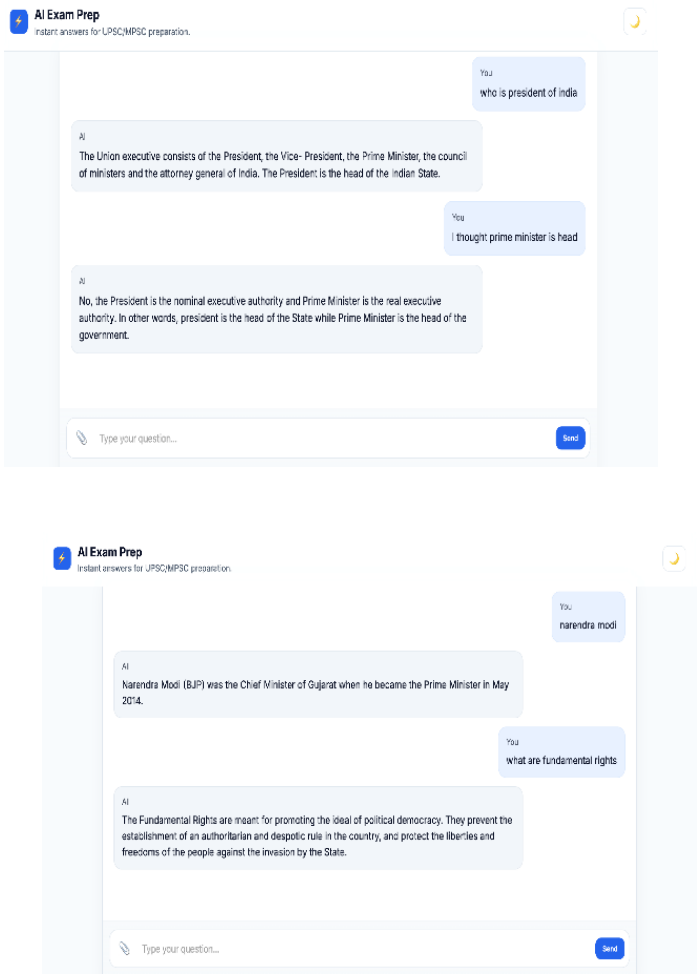

Fig. 3. Response generated by the proposed PrepMind AI model for the same query.

5 Discussion

The most important lesson that we had during the construction of PrepMind AI is that we have managed to make a tool that will actually be useful to competitive exam students who cannot afford the expensive courses, and it targets the significant problem of educational inequality in a direct manner.

We have been able to develop a quality, digital mentor who is far less expensive, and 24/7. The main success that we had is the type of RAG architecture we selected. We looked through other articles and it was clear that most of the time plain AI model results in so-called hallucinations, or simply invented things, which is not acceptable when it comes to serious test preparation. Having the system set up in such a way that the Large Language Model (LLM) is pushed to use solely our verified knowledge base in pulling the facts before it writes the answer, we make certain that all answers are correct and fully trustworthy. Our system is superior to a typical chatbot because of this hybrid RAG model being at the heart of our system, we enjoy the advantage of AI fluency without the danger of misinformation.

The other significant victory is the accessibility of the input that we made. The inclusion of the OCR module would be an even greater advantage since students tend to read old printed books or their personal notes. The fact that we can simply take a picture of a query and get the answer immediately makes our tool very useful in our day-to-day usage as opposed to the systems that allow only typed text to be inputted. Furthermore, adaptive learning allows

becoming a part of PrepMind AI; this tool is constantly tracking the student progress. It is much more subject-specific and effective in helping students study than the traditional general practice tests because it allows them to monitor errors and, as a result, create a quiz that targets those specific areas of weakness.

Although PrepMind AI is performing successfully, we have identified several challenges as per the literature that we have gone through. There is more to do to maximize the speed and cost of executing the RAG searches, in particular as the user base and size of our knowledge base increase. We need to keep the system a fast and cheap system to operate. The other current issue is how to better the precision of the OCR with some truly messy handwriting or with regional scripts which is a widespread issue in India. On the same note, although we are concerned with multilingual support, we must continue to refine the translation algorithms to ensure that the sophisticated and technical terms in languages besides English do not end up being misunderstood. These issues are the apparent next step to transforming PrepMind AI into a perfect and scalable educational solution.

6 Conclusion

This study shows that PrepMind AI can improve the preparation of students for competitive exams like UPSC and MPSC. The main thing is our system provides accurate and correct answers to students, instead of just guessing. The system can take handwritten notes as input and provides text format data to system for the next process. Our results show that AI can help you study easily without need to pay for expensive tutors. In short, we are solving a lot of problems found in existing systems. For the next steps, we are planning to improve the handwritten notes, support more languages, and speed up the search so it works great for everyone.

References

1 J. Kim, X. Xie, Handwritten Hangul recognition using deep convolutional neural networks,(2014). https://doi.org/10.1007/s10032-014-0229-4
2 R. Smith, An Overview of the Tesseract OCR Engine, in Proc. ICDAR, Curitiba, Brazil, pp. 629–633 (2007). https://doi.org/10.1109/ICDAR.2007.4376991
3 S. Saleem, M. U. Aziz, AI in Education: Personalized Learning Systems and Their Impact on Student Performance and Engagement, (2025). https://doi.org/10.59075/c35qa453
4 T. Kabudi, I. Pappas, D. H. Olsen, AI-enabled adaptive learning systems: A systematic mapping of literature, Comput. Educ. Artif. Intell., 2, 100031 (2021). https://doi.org/10.1016/j.caeai.2021.100031
5 P. Lewis, Retrieval-augmented generation for knowledge based intensive NLP tasks, (2020). https://arxiv.org/abs/2005.11401
6 H. Zhang, M. O. Shafiq, Enhancing retrieval augmented generation using triple-based knowledge graphs,IEEE(2025). https://doi.org/10.1109/ICSC64641.2025.00011
7 S. Vakayil, D.S. Juliet, S. Vakayil, RAG-based LLM chatbot using LLaMA 2, IEEE, (2025). https://doi.org/10.1109/10561020
8 S. Trangcasanchai, Improving question answering systems with retrieval augmented generation, Univ. Helsinki (2024).
9 S. Sharma, P. Mittal, M. Kumar, V. Bhardwaj, The role of large language models in personalized learning: A systematic review, Environ. Sustain. Educ(2025). https://doi.org/10.1007/s43621-025-01094-z
10 T. Taylor, K. Smith, O. Brown, R. Hoover, AI driven personalized learning, (2025).
11 A. Ezzaim, A. Dahbi, A. Haidine, A. Aqqal, Adaptive learning using ML-based prediction methods(2023). https://doi.org/10.56294/dm2023164

12 L. Huynh, Natural language processing for personalize education text evaluation, (2025). https://doi.org/10.3390/app152212128
13 Y. Walter, Embracing AI in classrooms AI literacy and critical thinking (2024). https://doi.org/10.1186/s41239-024-00448-3
14 A. Conneau et al., XNLI: Evaluating cross-lingual sentence representations, pp. 2475–2485 (2018). https://doi.org/10.18653/v1/D18-1269
15 A. Salido, I. Syarif, M.S. Sitepu, et al., Critical thinking & AI in higher education: A systematic review, Soc. Sci. Humanit. Open, 12, 101924 (2025). https://doi.org/10.1016/j.ssaho.2025.101924
16 S. Saleem, M.U. Aziz, M.J. Iqbal, S. Abbas, Personalized learning systems using AI, (2025). https://doi.org/10.59075/c35qa453
17 M. Murtaza, Y. Ahmed, J.A. Shamsi, F. Sherwani, M. Usman, AI-based personalized e-learning systems: Issues and solutions, IEEE Access, 10, 81323–81342 (2022).
18 A. Létourneau et al., AI-driven intelligent tutoring systems: A systematic review, (2025). https://doi.org/10.1038/s41539-025-00320-7
19 O. Zawacki-Richter, V.I. Marín, M. Bond, F. Gouverneur, AI in higher education where are the educators?(2019). https://doi.org/10.1186/s41239-019-0171-0
20 K. Sundar, E. Manohar, K. Vijay, S. Prakash, AI-powered evaluation with RAG and LLM technologies, in Proc. ICSSAS, IEEE, pp. 43–48 (2024). https://doi.org/10.1109/ICSSAS64001.2024.
21 R. Kumar, K. Srivastava, Natural language processing for automated educational assessment: A comprehensive review, IEEE Access, (2023). https://doi.org/10.1109/ACCESS.2023.3271012
22 B. Tural, Z. Örpek, Z. Destan, "Retrieval-Augmented Generation (RAG) and LLM Integration," 2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS), 2024, doi:10.1109/ISAS64331.2024.10845308
23 OpenAI, *"ChatGPT — conversation output for competitive exam query,"* ChatGPT, Accessed: Jan. 26, 2025. [Online]. Available: https://chatgpt.com/share/694d43f2-5c78-800b-ab05-7961075b42e0