

QueryMintAI: Multipurpose Multimodal Large Language Model for Personal Data

Sathe Siddhi Subhash¹ and Prof. D. S. Kshirsagar²

^{1,2}Department of Computer Engineering

^{1,2}Adsul's Technical Campus, Chas, India

sathesiddhi196@gmail.com¹, dipask2003@gmail.com²

Abstract: *Personal data is generated through everyday activities such as messages, documents, photographs, voice recordings, calendars, and digital transactions. Finding useful information across these different formats can be difficult when data is stored in separate applications. This paper presents QueryMintAI, a proposed multipurpose multimodal large language model system for understanding and searching personal data. The system aims to accept natural-language questions and retrieve relevant information from text and other supported media, while keeping privacy and user control as central design goals. The proposed architecture includes data ingestion, modality-specific processing, a unified retrieval layer, a language-model reasoning component, and a response-generation module. This paper describes the motivation, related concepts, architecture, possible applications, privacy considerations, evaluation plan, limitations, and future work. QueryMintAI is presented as a proposed design; no experimental results are claimed.*

Keywords: Large language models, multimodal AI, personal data, information retrieval, privacy, retrieval-augmented generation

I. INTRODUCTION

People store personal information in many formats, including emails, PDFs, notes, images, audio recordings, and calendar entries. Although this information can be useful, locating one specific detail often requires searching multiple applications and remembering where the information was saved.

Large language models (LLMs) can interpret natural-language questions and produce readable answers. Multimodal systems extend this ability to different kinds of input, such as text and images, and in some systems audio or video. When combined with retrieval methods, these capabilities can support a personal assistant that answers questions using information selected by the user.

QueryMintAI is proposed as a multipurpose multimodal system that helps users search and understand their personal data through a single conversational interface. Its goal is not simply to generate fluent text, but to connect answers to relevant retrieved information and clearly communicate uncertainty. Because personal data can be sensitive, privacy, permission, and transparent data handling are essential parts of the design.

The main objectives are: (1) to support natural-language queries over multiple personal-data formats; (2) to retrieve relevant information before generating an answer; (3) to provide understandable, source-grounded responses; and (4) to protect user data through access control, data minimization, and secure processing.

II. RELATED WORK AND BACKGROUND

A. Large Language Models. LLMs are trained on large collections of text and can perform tasks such as summarization, question answering, and language understanding. However, a model may produce unsupported statements when it lacks reliable evidence. For personal-data applications, answers should therefore be grounded in retrieved user-authorized sources.



B. Multimodal Models. Multimodal AI systems process or connect information from more than one modality, such as text and images. Optical character recognition (OCR) can extract text from scanned documents and photographs, while speech recognition can convert audio into searchable text. These components allow different media to be indexed in a common retrieval system.

C. Retrieval-Augmented Generation. Retrieval-augmented generation (RAG) first retrieves relevant documents or data fragments and then provides them as context to a language model. This can make answers more specific to a user's information and allows the system to show which sources support a response.

D. Privacy-Aware Personal AI. A personal assistant may encounter private conversations, financial records, identification documents, or location information. Privacy-aware design should limit collection, use explicit permissions, encrypt stored data, and provide ways to review or delete indexed information.

III. PROPOSED QUERMINTAI ARCHITECTURE

QueryMintAI is designed as a modular system. The proposed workflow begins when the user connects or uploads selected data sources. The system processes each format, creates searchable representations, retrieves relevant items for a question, and uses a language model to formulate a response.

A. Data Ingestion Layer. This layer accepts user-authorized sources such as text files, PDFs, images, notes, and audio recordings. The system should record source metadata and apply permission checks before processing. Unsupported or corrupted files should be reported rather than silently ignored.

B. Modality Processing Layer. Text documents can be parsed directly. Images and scanned pages can be processed using OCR or visual encoders. Audio can be transcribed when the user permits it. Extracted content should retain references to the original file, page, timestamp, or other useful location.

C. Indexing and Retrieval Layer. Content is divided into manageable segments and indexed using keyword search, embeddings, or a hybrid approach. A query is used to retrieve relevant segments. Access-control rules must be applied during retrieval so that the model only receives data the user is allowed to access.

D. Language Model and Response Layer. The retrieved evidence is supplied to the LLM with instructions to answer from the available sources, distinguish facts from uncertainty, and avoid inventing missing details. Responses may include source names or links so users can verify the answer.

E. User Interface. A chat interface allows questions such as "Find the date mentioned in my meeting notes" or "Summarize this document." Users should be able to inspect sources, correct extracted information, manage permissions, and remove data from the index.

IV. WORKING METHODOLOGY

The proposed workflow consists of six steps. First, the user selects the personal data sources to include. Second, files are parsed or transformed into searchable text and representations appropriate to each modality. Third, the system stores indexed segments with source metadata and access permissions. Fourth, the user asks a question in natural language. Fifth, the retrieval module selects relevant evidence and passes it to the language model. Sixth, the system returns a concise answer with supporting source references where available.

To reduce unsupported answers, the system should return an uncertainty message when retrieval finds insufficient evidence. For questions involving dates, amounts, names, or other exact values, the response should be checked against the retrieved source. The design should also prevent content inside a retrieved document from overriding system security rules or user permissions.

V. APPLICATIONS

QueryMintAI could support personal document search, summarization of study notes, retrieval of information from meeting records, searching text inside scanned documents, and finding details across user-authorized notes and files.



With suitable processing components, it could also search transcripts of voice notes or identify relevant moments in media.

For students, it could help locate definitions and explanations across notes and PDFs. For professionals, it could retrieve information from meeting notes and project documents. For everyday users, it could help organize receipts, reminders, and personal records. These are potential applications; their usefulness would need to be validated through implementation and testing.

VI. PRIVACY, SECURITY, AND ETHICAL CONSIDERATIONS

Personal data requires stronger safeguards than a general public-information search. QueryMintAI should follow data minimization: process only sources and fields needed for the user's task. Consent should be clear, permissions should be revocable, and users should be able to remove files and associated index entries.

Data should be protected in transit and at rest. Where possible, sensitive processing may be performed locally or through privacy-preserving services. Logs should avoid storing unnecessary personal content. Access controls must be enforced before retrieval, not merely described in a prompt to the language model.

The system should also address prompt injection, accidental disclosure, and incorrect extraction. Retrieved documents must be treated as data rather than trusted instructions. High-impact answers should show evidence and invite verification. The system should not imply that it has searched sources that were not connected or successfully processed.

VII. EVALUATION PLAN

Because QueryMintAI is presented here as a proposed system, no measured performance results are reported. A future prototype can be evaluated using a test collection of user-approved sample documents and questions with known answers.

Useful measures include retrieval precision and recall, answer correctness, source attribution accuracy, response time, and the rate of unsupported answers. Multimodal tests should separately measure OCR or transcription quality. Privacy testing should examine permission enforcement, deletion of indexed content, and resistance to malicious instructions embedded in files.

Baselines could include ordinary keyword search and text-only retrieval. Tests should cover direct fact lookup, multi-document questions, summarization, ambiguous queries, missing evidence, and requests for data that the user has not authorized.

VIII. LIMITATIONS AND FUTURE WORK

The system's quality will depend on the accuracy of document parsing, OCR, transcription, indexing, and retrieval. Poor-quality scans, incomplete files, ambiguous questions, or outdated records may lead to missing or incorrect answers. Processing many files can also require substantial storage and computing resources.

Future work includes building a working prototype, comparing retrieval strategies, testing supported modalities, adding source-level explanations, and measuring privacy and security performance. Further development may explore on-device processing, user-controlled retention periods, multilingual support, and integrations with common personal-data formats.

IX. CONCLUSION

QueryMintAI is a proposed multipurpose multimodal LLM system for searching and understanding personal data through natural-language questions. Its architecture combines data ingestion, modality-specific processing, indexing, retrieval, and evidence-grounded response generation. The design emphasizes that useful personal AI must provide not only convenient answers but also source transparency, access control, privacy, and user control.



The next step is to implement the proposed architecture and evaluate it on a carefully prepared dataset. Until such testing is completed, the benefits described in this paper should be understood as design goals rather than experimentally demonstrated results.

REFERENCES

- [1] A. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems, vol. 33, 2020.
- [2] OpenAI, "GPT-4 Technical Report," arXiv:2303.08774, 2023.
- [3] A. Radford et al., "Learning Transferable Visual Models From Natural Language Supervision," Proceedings of ICML, 2021.
- [4] K. Lee et al., "Language Models as Tool Makers," arXiv preprint, 2023.
- [5] National Institute of Standards and Technology, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," NIST AI 100-1, 2023.
- [6] European Union, "Regulation (EU) 2016/679 (General Data Protection Regulation)," Official Journal of the European Union, 2016.
- [7] D. S. Sahoo et al., "A Survey of Large Language Models," arXiv preprint, 2023.
- [8] T. B. Brown et al., "Language Models are Few-Shot Learners," Advances in Neural Information Processing Systems, vol. 33, 2020.

