International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 10 | Oct 2026 www.irjet.net p-ISSN: 2395-0072

Toward Next-Generation AI Research Systems:

A Comparative Evaluation of Multi-Model Intelligence, Computational Efficiency, Portability, and Data Confidentiality

Om Bajirao Renuse[1], Nmeesh Dalvi[2], Roopali Patil[3]

[12]Masters of Computer Applications, School of Humanities and Sciences, D. Y Patil Deemed to be University, India.
[3]Associate Professor, School of Humanities and Sciences, D. Y Patil Deemed to be University, India.

Abstract: *Artificial intelligence (AI) is increasingly used to support research activities, but challenges related to response quality, computational requirements, portability, and data confidentiality remain. This study focuses on a portable, multi-model AI research assistant designed to provide a common interface for interacting with compatible AI models. The research aims to compare models based on response quality, processing time, resource consumption, and usefulness for research tasks. It also examines hardware compatibility and privacy considerations associated with local and remote processing. An experimental and comparative methodology is proposed, combining performance measurements with user feedback to evaluate the system's practical value. The study seeks to identify the trade-offs between model flexibility, efficiency, and confidentiality, providing insights into the design and evaluation of AI research assistants. Final conclusions will be based on experimental results and collected user responses.*

Key Words: Artificial intelligence, large language models, multi-model systems, research assistant, portability, data confidentiality.

I. INTRODUCTION

Artificial intelligence (AI) is increasingly used to support research activities such as summarization, information retrieval, and question answering. However, the effectiveness of an AI research assistant depends on response quality, processing time, hardware requirements, and data confidentiality. Applications that rely on a single model may also limit flexibility for users with different computing resources.

This study explores a portable, multi-model AI research assistant that allows compatible models to be evaluated through a common interface. It compares their performance, resource consumption, portability, and usefulness for research tasks while examining the privacy considerations of local and remote processing.

The main objectives are to compare AI models, evaluate hardware compatibility, assess data confidentiality, and identify trade-offs between response quality, speed, and computational requirements.

II. RELATED WORK

1) A. Retrieval-Augmented Generation

Lewis et al. [1] introduced Retrieval-Augmented Generation (RAG), which combines information retrieval with language models to generate answers using external information. Gao et al. [2] reviewed RAG techniques and highlighted challenges such as hallucinations and retrieval quality. These studies provide a foundation for developing AI assistants that work with research information.

2) B. Evaluation of AI Systems

Yu et al. [3] examined evaluation methods for RAG systems, emphasizing the importance of relevance, accuracy, and reliability. Their work supports comparing AI models using consistent evaluation criteria rather than relying only on the fluency of generated answers.

3) C. Privacy and Confidentiality

Zeng et al. [4] investigated privacy risks in RAG systems and highlighted the possibility of exposing information from private document collections. This shows why data storage, access controls, and the use of external AI services must be considered when developing research assistants.

4) D. Research Gap

Existing studies provide useful foundations for information retrieval, AI evaluation, and privacy. Building on this research, the present study focuses on evaluating a portable, multi-model AI research assistant by comparing response quality, processing time, hardware requirements, and data confidentiality. The aim is to understand the practical trade-offs between model flexibility, performance, and privacy through experimental evaluation.

III. RESEARCH PROBLEM, RESEARCH QUESTIONS, OBJECTIVES, AND HYPOTHESES

5) A. Research Problem

AI research assistants can simplify research tasks, but their effectiveness depends on answer quality, processing time, hardware requirements, portability, and data confidentiality. This study investigates whether a portable, multi-model AI research assistant can provide useful research support across compatible models and computing environments.

6) B. Research Questions

1. How does model selection affect response quality?
2. How do models differ in response time and resource consumption?
3. How compatible is the application with different computing environments?
4. What privacy risks arise from local and remote processing?
5. How useful is the assistant for research-related tasks?

7) C. Research Objectives

1. Compare multiple AI models on research tasks.
2. Evaluate response quality, speed, and resource requirements.
3. Examine portability and hardware compatibility.
4. Assess data confidentiality and privacy considerations.
5. Identify the assistant's usefulness and limitations.

8) D. Research Hypotheses

Hypothesis 1: Model Performance

- **H_{01}:** There is no statistically significant difference in response quality among the selected AI models.
- **H_{11}:** There is a statistically significant difference in response quality among at least some selected AI models.

Hypothesis 2: Response Time

- **H_{02}:** There is no statistically significant difference in response time among the selected AI models.
- **H_{12}:** There is a statistically significant difference in response time among at least some selected AI models.

These hypotheses will be tested using experimental data and appropriate statistical methods. The significance level and p-values will be reported after the experiments are completed.

IV. PROPOSED SYSTEM

9) A. System Overview

The proposed system is a portable, multi-model AI research assistant developed to support research activities through a common web interface. It uses Django as the application framework and aims to provide flexibility in model selection, research assistance, and data handling.

10) B. System Architecture

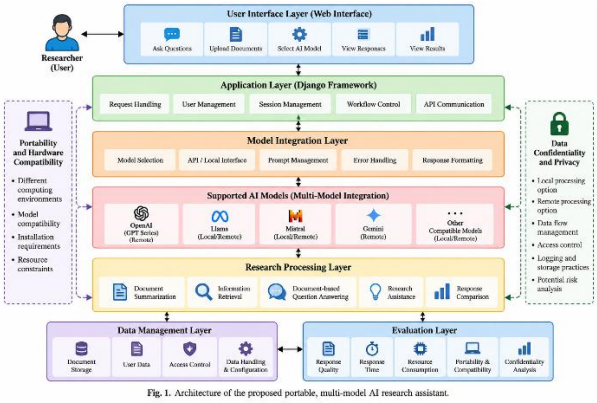


Fig. 1. Architecture of the proposed portable, multi-model AI research assistant.

The architecture consists of six main layers:

1. **User Interface Layer:** Allows users to submit questions and view responses.
2. **Application Layer:** Manages requests and application workflows using Django.
3. **Model Integration Layer:** Connects the application with compatible AI models.
4. **Research Processing Layer:** Supports summarization, question answering, and information retrieval where implemented.
5. **Data Management Layer:** Handles application data and document storage according to the system configuration.
6. **Evaluation Layer:** Assesses response quality, processing time, resource consumption, and compatibility.

11) C. Multi-Model Integration

The system aims to support multiple compatible AI models through a common interface. Models can be compared using consistent tasks and criteria, including response quality, processing time, and computational requirements. Only successfully integrated and tested models will be considered supported.

12) D. Research Support Functionality

The assistant is intended to help users answer research questions, summarize information, and explore relevant content. Document-based question answering and other advanced functions will be included only where supported by the implementation.

13) E. Portability and Hardware Compatibility

The system's portability will be assessed across supported computing environments. Testing will consider installation requirements, model compatibility, memory usage, and execution stability. Hardware limitations may affect which models can run successfully.

14) F. Data Confidentiality and Privacy

The study examines how research data is handled during local and remote processing. Factors such as data transmission, storage, logging, access controls, and external service usage will be considered. Local processing may reduce external data transmission but does not automatically guarantee confidentiality.

15) G. Expected System Contribution

The proposed system provides a framework for comparing AI models within a common research-assistant workflow. Its effectiveness will be evaluated through implementation testing and experimental results, focusing on performance, portability, research usefulness, and data confidentiality.

V. RESEARCH METHODOLOGY

A. Research Design: This study compares AI models based on response quality, speed, resource usage, portability, and research usefulness.

B. Sample Design: Selected research questions and tasks will be tested on each model under similar conditions.

C. Data Collection: Data will include model responses, processing time, resource consumption, compatibility, and privacy considerations.

D. Descriptive Analysis: Mean, median, standard deviation, tables, and charts will be used to summarize and compare results.

E. Inferential Analysis: Statistical tests will identify significant differences between models using a significance level of 0.05.

F. Hypothesis Testing: Hypotheses related to response quality and speed will be tested using actual experimental results and p-values.

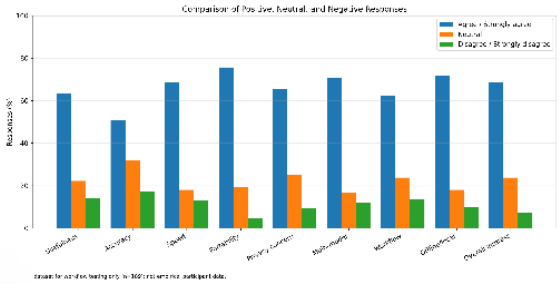
G. Interpretation of Data: Results will be analysed to identify the best-performing models while considering hardware compatibility, privacy, and study limitations.

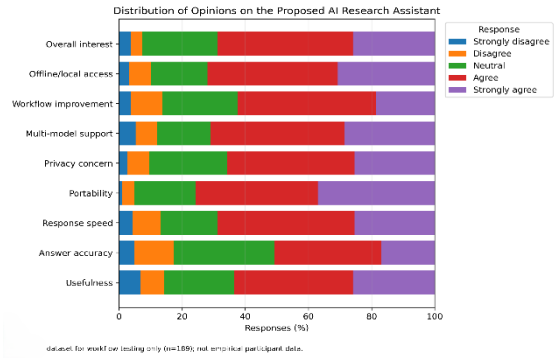
VI. RESULTS AND COMPARATIVE ANALYSIS

A. Evaluation Overview

The AI research assistant will be evaluated based on response quality, speed, resource usage, portability, research usefulness, and data confidentiality.

B. Response Quality





Model responses will be checked for accuracy, relevance, clarity, completeness, and consistency using common evaluation criteria.

C. Response Time

The time taken by each model to generate responses will be recorded and compared through average and median response times.

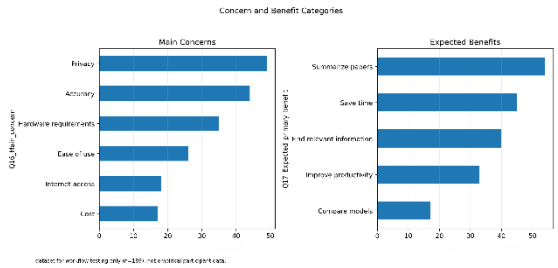
D. Resource Consumption

CPU usage, GPU usage, and memory consumption will be measured to understand each model's hardware requirements.

E. Portability and Compatibility

Models will be tested across available computing environments to check installation, compatibility, stability, and execution requirements.

F. Data Confidentiality



Data handling, network requests, storage, and logging will be examined to identify potential privacy risks.

G. Comparative Results

The models will be compared using the following criteria:

Criterion	Measurement
Response quality	Accuracy and relevance scores
Response time	Seconds per response
Resource usage	CPU, GPU, and RAM
Portability	Compatibility test results
Confidentiality	Data-handling assessment
Research usefulness	Task-based evaluation

H. Statistical Analysis

Appropriate statistical tests will be used to identify significant differences between models. P-values and effect sizes will be reported where applicable.

I. Summary

The final analysis will highlight the strengths, limitations, and practical suitability of each model for research tasks.

VII. DISCUSSION

A. Model Performance and Response Quality

The study compares AI models based on accuracy, relevance, clarity, and completeness to identify their suitability for research tasks.

B. Response Time and Computational Requirements

Response speed, CPU/GPU usage, and memory requirements will be compared to understand each model's efficiency, especially on low-resource systems.

C. Portability and Hardware Compatibility

The study examines whether models can run across different computing environments and identifies compatibility issues and hardware limitations.

D. Data Confidentiality and Privacy

Data storage, network communication, and privacy risks will be examined. Local processing may reduce external data sharing but does not guarantee complete security.

E. Practical Usefulness for Research

The assistant aims to support literature summarization, information retrieval, and research-related queries. Its usefulness depends on answer quality, time savings, and the need to verify generated information.

F. Limitations of the Study

The study may be limited by model availability, hardware differences, subjective evaluation, task coverage, and the scope of privacy testing. Therefore, results may not apply to all AI systems.

G. Overall Discussion

The study evaluates the trade-offs between performance, speed, resource usage, portability, and confidentiality. Final conclusions will be based on verified experimental results rather than assumptions.

VIII. CONCLUSION AND FUTURE WORK

A. Conclusion

This research proposes a portable, multi-model AI research assistant to support academic tasks. It evaluates models based on response quality, speed, resource usage, portability, research usefulness, and data confidentiality.

The framework enables compatible AI models to be compared under consistent testing conditions. It also highlights the importance of balancing performance with hardware requirements and privacy considerations.

Final conclusions will be based on verified experimental results, helping identify suitable models for different research needs and computing environments.

B. Future Work

Future improvements may include:

1. **More AI Models:** Evaluate additional open-source and cloud-based models.
2. **Document Support:** Improve research-paper processing and summarization.
3. **Better Evaluation:** Use larger datasets, expert reviews, and statistical tests.
4. **Improved Portability:** Test across more operating systems and hardware configurations.

5. **Enhanced Security:** Strengthen data protection, access control, and privacy testing.
6. **Advanced Research Tools:** Add citation verification and literature-review assistance.
7. **Adaptive Model Selection:** Recommend models based on task requirements and available resources.
8. **User Evaluation:** Test usability and usefulness with students and researchers.

Future work will focus on implementing these improvements and validating their effectiveness through systematic testing.

REFERENCES

[1] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems, vol. 33, pp. 9459–9474, 2020. [Online]. Available: https://papers.nips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf

[2] Y. Gao et al., "Retrieval-Augmented Generation for Large Language Models: A Survey," arXiv preprint arXiv:2312.10997, 2023. [Online]. Available: https://arxiv.org/abs/2312.10997

[3] H. Yu, A. Gan, K. Zhang, S. Tong, Q. Liu, and Z. Liu, "Evaluation of Retrieval-Augmented Generation: A Survey," arXiv preprint arXiv:2405.07437, 2024. [Online]. Available: https://arxiv.org/abs/2405.07437

[4] S. Zeng et al., "The Good and The Bad: Exploring Privacy Issues in Retrieval-Augmented Generation (RAG)," in Findings of the Association for Computational Linguistics: ACL 2024, pp. 4505–4524, 2024. doi: 10.18653/v1/2024.findings-acl.267. [Online]. Available: https://aclanthology.org/2024.findings-acl.267/

[5] P. Liang et al., "Holistic Evaluation of Language Models," Transactions on Machine Learning Research, 2023. [Online]. Available: https://arxiv.org/abs/2211.09110

[6] M. T. R. Laskar et al., "A Systematic Survey and Critical Review on Evaluating Large Language Models: Challenges, Limitations, and Recommendations," in Proc. EMNLP, 2024. [Online]. Available: https://aclanthology.org/2024.emnlp-main.764/

[7] K. Chen et al., "A Survey on Privacy Risks and Protection in Large Language Models," Journal of King Saud University Computer and Information Sciences, vol. 37, article 163, 2025. [Online]. Available: https://doi.org/10.1007/s44443-025-00177-1