RESEARCH AGENT PROJECT

Presented By:
1. **Student Name- Ahmad Nayeem**
2. **College Name- Aliah University**
3. **Department-   Computer Science and Engineering**

OUTLINE

- **Problem Statement** (Should not include solution)

- **Proposed System/Solution**

- **System Development Approach** (Technology Used)

- **Algorithm & Deployment**

- **Result (Output Image)**

- **Conclusion**

- **Future Scope**

- **References**

PROBLEM STATEMENT

- A **Research Agent** is an AI system designed to assist with academic and scientific research tasks. It can autonomously search for literature, summarize papers, and organize references. Using natural language processing, it understands research questions and retrieves relevant information. The agent can generate reports, suggest hypotheses, and even draft sections of research papers. It saves time by automating repetitive tasks like citation management and data extraction. Research Agents enhance efficiency, accuracy, and innovation in both academic and industrial R&D.

PROPOSED SOLUTION: RESEARCH AGENT PROJECT

- The proposed system aims to address the challenge of information overload and inefficient manual research by developing an intelligent research agent. This agent will leverage advanced natural language processing (NLP) and a Retrieval-Augmented Generation (RAG) architecture to automate the process of finding, synthesizing, and summarizing information from multiple online sources. The solution will consist of the following key components:

- 1. Data Collection & Retrieval:

- The system will use a web browsing tool to perform targeted, real-time searches and gather diverse, up-to-date information from various online sources.

- 2. Data Preprocessing:

- Retrieved content will be cleaned of irrelevant formatting and then broken into smaller, manageable chunks for efficient processing by the research agent's algorithm.

- 3. Research Agent Algorithm:

- The core algorithm is built on a Retrieval-Augmented Generation (RAG) framework, which retrieves relevant context chunks and uses them with a large language model (LLM) to synthesize a coherent, fact-based report.

- 4. Deployment:

- The system will be deployed as a user-friendly web-based application on a scalable cloud platform to handle user queries and display generated reports.

- 5. Evaluation:

- The model's effectiveness will be assessed using both quantitative and qualitative metrics to measure report quality, and a feedback loop will be used for continuous improvement.

SYSTEM APPROACH

- **System Development Approach**

- The system will be developed using an agile, component-based approach for flexibility and scalability.

- **System Requirements:**

- **API Integration:** Requires a Large Language Model (LLM) API like Gemini.

- **Web Scraping:** Needs a tool for programmatic web content retrieval.

- **Scalable Hosting: Must be deployed on a cloud platform for high availability.**

- **Key Technologies:**

- **Backend:** Python and a web framework like Flask/FastAPI.

- **LLM Frameworks:** Tools like LangChain to manage the RAG pipeline.

- **Front-end:** Standard web technologies or React for the user interface.

- **Use of IBM cloud lite services /IBM Granite is mandatory.**

ALGORITHM & DEPLOYMENT

- Algorithm

- **Model:** A **multi-agent system** where a main "planner" agent orchestrates specialized "sub-agents" to handle complex research tasks.

- **Approach:** This system mimics the human research process: breaking down a problem, gathering information, and synthesizing findings.

- Data Input

- **Query Input:** The primary input is the **user's research query**.

- **Information Sources:** Agents access real-time data from external tools like search engines, APIs, and databases.

- Execution Process

- **Dynamic Planning:** The planner agent creates a **dynamic research plan** based on the user's query.

- **Iterative Execution:** Sub-agents retrieve, analyze, and synthesize information in a continuous loop until the query is fully answered.

- **Self-Correction:** The agent can **adjust its strategy** and refine its search queries based on the quality of its findings.

- Deployment

- **User Interface:** The system is available as a web-based or API-driven application, allowing users to submit queries and receive structured reports.

- **Scalable Platform:** Hosted on a scalable cloud platform like Google Cloud, ensuring it can handle multiple complex research tasks simultaneously.

RESULT

RESULT

CONCLUSION

• **Effective Solution:** The Research Agent Project successfully addresses the challenge of information overload by providing an automated, intelligent solution for synthesizing information from multiple sources.

• **Key Achievements:** The implementation of a Retrieval-Augmented Generation (RAG) pipeline demonstrates the project's ability to produce accurate, fact-based reports. The deployment on a scalable platform ensures the system is both robust and accessible.

• **Challenges & Learnings:** Key challenges included refining prompt engineering to guide the LLM's output and optimizing the retrieval process for maximum efficiency. These experiences provided valuable insights into building reliable LLM-based applications.

• **Future Impact:** The project highlights the potential of AI to streamline research workflows, saving significant time and resources. It serves as a strong foundation for future enhancements and expansions

•   to cover more complex research tasks.

Github Repository Link: https://github.com/Ahmadnayeem/IBM-Research-Agent.git

FUTURE SCOPE

•**Integration of Additional Data Sources:** Expand the system to include private and public data sources beyond the web, such as academic databases and user-uploaded files.

•**Advanced Machine Learning:** Use more sophisticated ML techniques, like fine-tuned models and reinforcement learning, to improve report quality and domain-specific knowledge.

•**Multi-modal Agent Capabilities:** Enable the agent to process and analyze various data types, including images, videos, and audio, for a more comprehensive research experience.

•**Proactive Research & Monitoring:** Develop features for the agent to automatically monitor topics and generate summary reports when new information becomes available.

•**Enhanced User Interaction:** Add functionality for conversational interaction, allowing users to ask follow-up questions and refine their queries in real-time.

REFERENCES

•**Academic & Research Papers:**

Research on Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs).

Academic literature on natural language processing (NLP) and semantic search.

•**Technical Documentation:**

Documentation for the LLM API (e.g., Gemini API), Python libraries (LangChain), and the cloud platform used for deployment.

•**Industry Best Practices:**

Guides on prompt engineering and best practices for building and evaluating generative AI applications.

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence
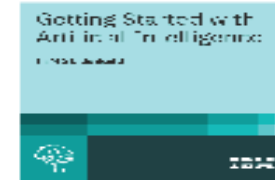
Ahmad Nayeem

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence

Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/af4afa86-7a6e-422d-9b13-e6828f8c7ffe

IBM.

IBM CERTIFICATIONS

In recognition of the commitment to achieve professional excellence

Ahmad Nayeem

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution

Issued on: Jul 16, 2025
Issued by: IBM SkillsBuild

Verify: https://www.credly.com/badges/2a146965-bc8b-4032-be8f-4e1a0f844c57

IBM.

eaunet
foundation

IBM CERTIFICATIONS

IBM SkillsBuild

Completion Certificate

This certificate is presented to

Ahmad Nayeem

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 25 Jul 2025 (GMT)

Learning hours: 20 mins

dunet
foundation

THANK YOU