
CAPSTONE PROJECT

RESEARCH AGENT

Presented By:

1. **Student Name: Chirag**
2. **College Name & Department: GURU GOBIND SINGH
INDRAPRASTHA UNIVERSITY with Btech in CSE AIML**

OUTLINE

- Problem Statement
- Proposed System/Solution
- System Development Approach
- Algorithm & Deployment
- Wow factors
- End users
- Result
- Conclusion
- Future Scope
- References

PROBLEM STATEMENT

The Challenge - Researchers, students, and professionals across academic and industrial domains face a significant challenge in managing and synthesizing the ever-increasing volume of academic publications, technical articles, datasets, and rapidly evolving research trends. The manual process of discovering relevant literature, filtering extraneous information, extracting key findings, and synthesizing insights from diverse sources is time-consuming, often inefficient, and can lead to missed connections or overlooked critical information. This information overload hinders the ability to stay abreast of the latest advancements, identify emerging trends, and accelerate the pace of innovation and discovery. There is a critical need for an intelligent system that can automate and enhance these research tasks, allowing individuals to focus on analysis, critical thinking, and generating new knowledge rather than being bogged down by information management.

PROPOSED SOLUTION

Based on the problem statement and the requirement to use IBM Cloud Lite services and IBM Granite, the proposed system solution is a cloud-based AI Research Agent with the following components:

- **User Interface:** A web-based interface (potentially built with Gradio as explored in the notebook) for users to input research questions and ask about relevant documents.
- **Natural Language Processing (NLP) Module:** To understand user queries, extract keywords, and identify the scope of research. This will leverage the capabilities of IBM Granite.
- **Information Retrieval Module:** To search for relevant literature and data from various sources (e.g., academic databases, web). The tools utilized in the provided notebook (GoogleSearch, DuckDuckGo, Wikipedia, TavilySearch, WebCrawler) can be integrated here.
- **Summarization and Analysis Module:** To summarize research papers, extract key findings, and perform basic data analysis. This will primarily utilize the IBM Granite model.
- **Report Generation Module:** To synthesize information and generate structured reports or draft sections of research papers.
- **Orchestration Layer:** To manage the workflow of the agent, coordinating the interaction between different modules (potentially using a framework like LangChain/LangGraph as seen in the notebook).
- **IBM Cloud Services:** Leveraging services available under the Lite plan, such as Watsonx.ai for accessing IBM Granite, and potentially other services for storage or deployment.

SYSTEM APPROACH

Phase 1: Core Functionality (MVP)

- Set up the IBM Cloud environment and access to IBM Granite.
- Develop the basic NLP and Information Retrieval modules.
- Implement document loading and text splitting.
- Integrate IBM Granite for basic summarization of retrieved text.
- Build a simple command-line or basic web interface (like the initial Gradio app) for testing.
- Focus on getting the core agent functionality working for simple queries and documents.

Phase 2: Enhanced Capabilities

- Improve the NLP module for more complex query understanding.
- Enhance information retrieval with more sophisticated search strategies and potentially more diverse sources.
- Refine the summarization and analysis capabilities.
- Develop the report generation module.
- Implement more robust error handling and feedback mechanisms.

Phase 3: Advanced Features (if Lite plan allows or for future iterations)

- Implement a basic knowledge base or simple persistent storage.
- Explore features like hypothesis generation or drafting research paper sections.
- Optimize performance and scalability.

ALGORITHM & DEPLOYMENT

1. Natural Language Processing (NLP):

- **Tokenization and Embedding:** Breaking down text into tokens and converting them into numerical representations (embeddings) using IBM Granite or related libraries.
- **Named Entity Recognition (NER):** Identifying key entities in the user query and retrieved text (e.g., names, organizations, keywords).
- **Keyword Extraction:** Identifying important terms for searching and summarization.
- **Semantic Search:** Using embeddings to find documents or passages semantically related to the user's query.

2. Information Retrieval:

- **Keyword-based Search:** Using extracted keywords to query search engines and databases.
- **Ranking Algorithms:** Ordering search results based on relevance.

3. Text Summarization:

- **Extractive Summarization:** Identifying and extracting the most important sentences or phrases from the text.
- **Abstractive Summarization:** Generating new sentences that capture the main points of the text (leveraging IBM Granite's generation capabilities).

4. Orchestration/Agent Logic:

- **Finite State Machines or Directed Acyclic Graphs (DAGs):** To define the workflow of the agent, transitioning between different stages (query analysis, search, summarization, reporting). Frameworks like LangGraph are well-suited for this.
- **Tool Use:** Enabling the language model (IBM Granite) to use external tools (search engines, document readers) to gather information.

ALGORITHM & DEPLOYMENT

The deployment strategy will focus on utilizing IBM Cloud Lite services:

- **Watsonx.ai:** The primary platform for accessing and running the IBM Granite model.
- **AI Service Deployment:** Deploying the core agent logic as an AI Service on Watsonx.ai, as demonstrated in the provided notebook. This allows the agent to be exposed as a REST API.
- **Gradio Interface Deployment:** Deploying the Gradio web interface separately. This could potentially be deployed on a small virtual server instance or a serverless platform if available within the Lite plan limitations. The interface will interact with the deployed AI Service API.
- **Storage:** Utilizing available IBM Cloud storage options (within Lite plan limits) for storing temporary files or potentially a small knowledge base.
- **Scalability (Limited in Lite):** The Lite plan will have limitations on resources and scalability. For a production environment, a paid plan would be necessary to handle increased load.
- **Monitoring and Logging:** Implementing basic monitoring and logging to track agent usage and identify errors.

WOW FACTORS

- **Accelerated Research:** Significantly reduce the time spent on manual literature search, review, and synthesis, allowing researchers to focus on analysis and critical thinking.
- **Enhanced Discovery:** Utilize AI to uncover hidden patterns, connections, and emerging trends within large volumes of research data that might be missed through traditional methods.
- **Streamlined Workflow:** Automate repetitive tasks such as citation management, data extraction, and initial report drafting, making the research process more efficient.
- **Improved Accessibility:** Democratize access to advanced information processing capabilities, enabling individuals with limited resources to conduct more comprehensive research.
- **Contextual Understanding:** Leverage advanced NLP (like IBM Granite) to understand complex research questions and provide more relevant and nuanced responses.

END USERS

- **Academic Researchers:** Scientists, scholars, and postgraduate students who need to conduct thorough literature reviews, stay updated in their fields, and accelerate their research process.
- **University Students:** Particularly those working on dissertations, theses, or extensive research papers who need help managing information and understanding complex topics.
- **R&D Professionals in Industry:** Individuals working in corporate research and development departments who need to track competitive landscapes, analyze market trends, and stay informed about technological advancements.
- **Analysts and Consultants:** Professionals who require quick synthesis of information from various sources to inform their reports and recommendations.
- **Librarians and Information Specialists:** Who can use the tool to assist patrons with complex research queries and curate relevant resources.

RESULT

GitHub link: https://github.com/ItsDeveloperChirag/CAPSTONE_PROJECT_Edunet.git

Some Code snapshots

```
import os
from ibm_watsonx_ai import APIClient, Credentials
import getpass
credentials = Credentials({
    url="https://us-south.ml.cloud.ibm.com",
    api_key=getpass.getpass("Please enter your api key (hit enter): ")
})
#IWjtMQpCrJT_SE2MJPF8h99w4v7b_q33Athk5FTi4oe3
```

```
from langchain_ibm import ChatWatsonx
from ibm_watsonx_ai import APIClient
from ibm_watsonx_ai.foundation_models.utils import Tool, Toolkit
from langchain_core.messages import AIMessage, HumanMessage
from langgraph.checkpoint.memory import MemorySaver
from langgraph.prebuilt import create_react_agent
import json
import requests

model = "ibm/granite-3-3-8b-instruct"

service_url = "https://us-south.ml.cloud.ibm.com"
# Get credentials token
credentials = {
    "url": service_url,
    "token": context.generate_token()
}
```

```
local_question = "Summarize the latest research on sustainable energy."
messages.append({ "role" : "user", "content": local_question })
context = RuntimeContext(api_client=client, request_payload_json={"messages": messages})
response = local_function(context)
result = ''
if (streaming):
    for chunk in response:
        print(chunk, end="\n\n", flush=True)
else:
    print(response)
```

```
iface = gr.Interface(
    fn=query_agent,
    inputs=gr.Textbox(lines=5, label="Your Question",
        placeholder="e.g., Summarize the latest research on sustainable energy."),
    outputs=gr.Textbox(label="Agent's Response"),
    title="Agent for Research UI",
    description="Interact with the deployed AI research agent using this interface."
)

iface.launch(share=True)
```

RESULT

Build

AI

Model: granite-3-3-8b-instruct



Setup

Name

Research Agent

Description

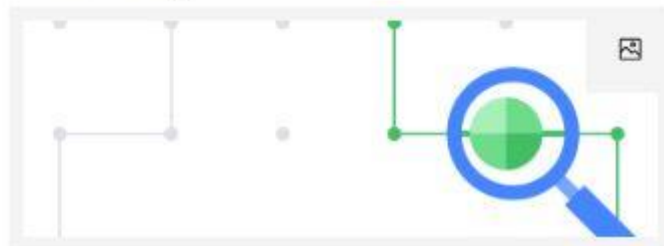
84/100

AI research assistant on Watsonx for quick insights, summaries, and trend discovery.

Icon



Placeholder image



Quick start questions

Add up to 4 quick start questions that your users can select.



Summarize the latest research on sustainable energy.



Recent paper on agriculture



Configuration

Framework

LangGraph

Architecture

ReAct

Instructions

[Advanced configuration](#)

You are a cognitive AI-powered assistant built on IBM's Watsonx platform, designed to streamline research and data analysis. It leverages advanced natural language processing and machine learning to understand complex queries, gather relevant information from diverse sources, and present insights in an easy-to-understand format. The agent can summarize large volumes of data, detect trends, suggest research directions, and integrate

RESULT

Knowledge

Tools



Add a tool

Create custom tool

Added tools (4)


Google search

Retrieve information from the internet with the Google search engine.





DuckDuckGo search

Retrieve information from the internet with the DuckDuckGo search engine.




Wikipedia search

Retrieve information from Wikipedia articles.




Webcrawler

Retrieve information from a website.

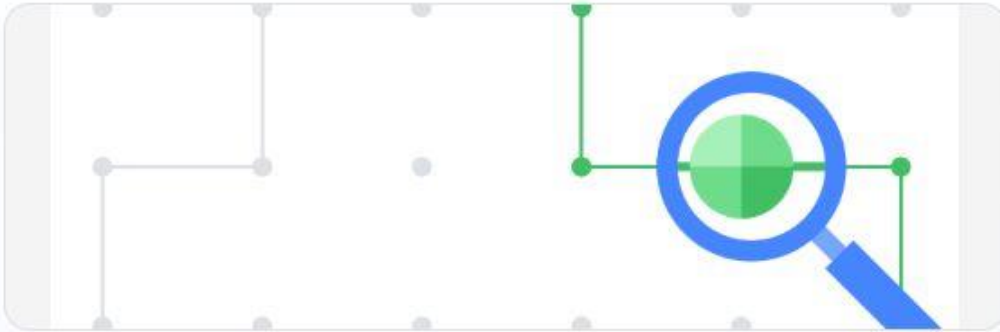


Agent preview

 Research Agent 06:50 PM

Welcome to Research Agent

AI research assistant on Watsonx for quick insights, summaries, and trend discovery.



Sample questions

Summarize the latest research on sustainable energy.

Recent paper on agriculture

Type something...

edunet
foundation

RESULT

Output:

Agent for Research UI

Interact with the deployed AI research agent using this interface.

Your Question

Summarize the latest research on sustainable energy.

Clear

Submit

Agent's Response

The latest research on sustainable energy, from 2020 to the present, has seen significant advancements in various areas:

1. **Smart Grids**: In 2022, a study simulated and analyzed "transactive energy mechanisms" to engage the large-scale deployment of flexible distributed energy resources (DERs), such as air conditioners, water heaters, batteries, and electric vehicles, in the operation of the electric power system.
2. **Super Grids**: Researchers proposed a novel strategy in 2022 to create a global sustainable interconnected energy system based on deep-ocean-compressed hydrogen transportation.
3. **Microgrids and Off-the-Grid Solutions**: A method for integrating multiple energy storage systems and distributed energy resources without dedicated communication or control systems was described in 2022. This approach could make microgrids more accessible and affordable, especially in areas where they are most needed, such as during power outages or after disasters.
4. **Solar Power**:
 - **Perovskite Solar Cells**: Efficiency has dramatically increased from 3.8% in 2009 to 25.2% in single-junction architectures and 28.1% in tandem cells by 2022, surpassing the

Flag

CONCLUSION

The proposed AI Research Agent, built upon IBM Cloud Lite services and leveraging the capabilities of IBM Granite, offers a promising solution for automating and enhancing academic and scientific research. By integrating natural language processing, information retrieval, document processing, and summarization techniques within an orchestrated framework, the agent can significantly reduce the time and effort spent on literature review, data extraction, and report generation. While operating within the constraints of the Lite plan, the phased development approach allows for building a functional core agent and progressively adding more advanced features.

FUTURE SCOPE

Based on the foundation established with the IBM Cloud Lite plan and IBM Granite, the future scope of the AI Research Agent project could include:

- **Integration with more data sources:** Connecting to a wider range of academic databases, repositories, and potentially enterprise internal knowledge bases.
- **Advanced Analysis:** Incorporating more sophisticated data analysis techniques, statistical modeling, and potentially machine learning for deeper insights.
- **Interactive Hypothesis Generation:** Developing features that allow the agent to suggest testable hypotheses based on synthesized research findings.
- **Collaborative Features:** Enabling multiple users to collaborate on research projects using the agent.
- **Improved User Interface:** Enhancing the web interface with more interactive visualizations and customization options.
- **Scalability and Performance Optimization:** Transitioning to a paid IBM Cloud plan to handle larger datasets, more complex queries, and a higher volume of users.
- **Specialized Agents:** Developing specialized versions of the agent tailored for specific research domains (e.g., medical research, material science).
- **Integration with Research Tools:** Connecting with citation management software, electronic lab notebooks, and other research workflow tools.

REFERENCES

While this proposal outlines a conceptual approach, the following resources and technologies are relevant to the development of an AI Research Agent:

- **IBM Watsonx.ai Documentation:** For details on accessing and using IBM Granite and other AI services.
- **IBM Cloud Documentation:** For information on available services and plan limitations.
- **LangChain and LangGraph Documentation:** For guidance on building language model applications and orchestrating agent workflows.
- **Academic Research Papers on AI in Research:** Relevant literature on the use of AI for literature review, summarization, and knowledge discovery.
- **Documentation for specific tools:** (e.g., Google Search API, DuckDuckGo API, Wikipedia API, Tavily Search API, Web Crawler libraries) - for details on integrating external information sources.
- **Gradio Documentation:** For building interactive web interfaces for the agent.

IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



CHIRAG .

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence



Issued on: Jul 18, 2025

Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/3481ebef-acbe-45a7-8a48-e9dbddc4dc4b>



IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



CHIRAG .

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution



Issued on: 11 FEB 2024

Issued by IBM

Verify: <https://www.credly.com/go/f0xu4Eu0>



IBM CERTIFICATIONS



This certificate is presented to

CHIRAG

for the completion of

Lab: Retrieval Augmented Generation with LangChain

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 19 Jul 2025 (GMT)

Learning hours: 20 mins



THANK YOU