
CAPSTONE PROJECT

RESEARCH AGENT

Presented By:

Abinash Sharma

**College Name- Maharaja Sriram Chandra Bhanja Deo University, Baripada, Mayurbhanj,
Odisha**

Department Of Computer Application

OUTLINE

- **Problem Statement**
- **Proposed System/Solution**
- **System Development Approach (Technology Used)**
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

Research Agent The Challenge- 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. Technology - Use of IBM cloud lite services /IBM Granite is mandatory

PROPOSED SOLUTION

The proposed system aims to create an AI-powered Research Agent capable of automating literature search, summarization, and reference management for academic and industrial researchers. This solution leverages **IBM Granite models** on **IBM Cloud Lite services** to ensure scalability, accuracy, and enterprise-grade reliability. The system will consist of the following components:

Data Collection

Collect academic papers and research articles from sources like IEEE, Springer, ArXiv, and PubMed.
Support multiple file formats (PDF, DOCX, URLs) for flexible research input.
Utilize real-time watsonx AI for updated research data.

Data Preprocessing

Extract and clean text from uploaded research documents.
Handle metadata (titles, abstracts, keywords) and remove irrelevant content (ads, formatting).
Convert data into embeddings for semantic search using IBM Granite embeddings.

PROPOSED SOLUTION

Natural Language Processing & AI Model

Implement **Retrieval-Augmented Generation (RAG)** for precise literature search and contextual summarization.
Use IBM Granite models (large language models) for question answering and hypothesis suggestion.
Support multi-turn conversational research queries for better user interaction.

Deployment

Build a user-friendly web interface integrated with IBM Cloud Lite services.
Deploy the AI agent on **Watsonx AI** with Granite for real-time performance.

Evaluation

Measure summarization quality using and accuracy of retrieved information using **precision-recall** metrics.
Gather researcher feedback to fine-tune the system continuously.
Conduct stress testing for performance and scalability on cloud infrastructure.

SYSTEM APPROACH

Strategy & Methodology:

The Research Agent is developed using IBM Cloud Lite and IBM Granite to automate literature retrieval, summarization, and reference management. The approach ensures scalability, real-time performance, and academic-grade accuracy.

System Requirements:

Hardware: Cloud deployment via IBM Cloud Lite (lite tier for R&D).

Software: Python 3.x, IBM Watsonx services, Granite models.

Storage: Lightweight vector index (~1–5 GB).

Libraries & Tools:

NLP & AI: IBM Granite LLM.

Deployment: IBM Cloud Functions (Watsonx).

Workflow:

Summarizing & preprocessing (papers, abstracts).
semantic indexing.

Query handling with RAG (Granite model).

Cloud deployment & continuous evaluation.

ALGORITHM & DEPLOYMENT

Algorithm Selection

The Research Agent employs a **Retrieval-Augmented Generation (RAG)** architecture integrated with **IBM Granite large language models**. Watsonx ai is chosen because it combines semantic search (retrieving relevant documents) with generative AI (summarizing and answering queries), ensuring accurate and context-aware responses suitable for academic research.

Data Input

Research papers and academic articles

Metadata: title, abstract, keywords

User queries (natural language research questions)

Training Process

Preprocess documents and convert them into **vector embeddings** using IBM Granite embeddings..

Fine-tune the prompt templates for summarization and hypothesis generation to align with academic tone and accuracy.

Prediction / Query Process

User submits a research question through the web interface.

Semantic search retrieves the most relevant documents from the vector index.

IBM Granite LLM generates a **concise, context-aware summary or answer**.

Results are displayed in structured format (summary, references, and possible hypotheses).

RESULT

The screenshot displays the IBM watsonx Agent Lab web interface. The browser's address bar shows the URL: `dataplatfom.cloud.ibm.com/wx/agents?context=wx&project_id=5b3f4a76-9618-4442-aa55-84f117a35834`. The interface is divided into two main panels: 'Build' on the left and 'Agent preview' on the right.

Build Panel:

- Model:** Set to 'granite-3-3-8b-instruct'.
- Framework:** Set to 'LangGraph'.
- Architecture:** Set to 'ReAct'.
- Instructions:** A text box contains the following text:
You are a Research Agent designed to assist users with academic and scientific research tasks. Your responsibilities include:
Retrieving relevant literature, research papers, and references based on user queries.
Understanding research questions using natural language processing.
- Knowledge:** A section for vector index, with a 'Clear chat' button.

Agent preview Panel:

- Header:** 'Agent preview'.
- Text:** 'Large scale embedded systems. An embedded system supported by 32-64 bit processors/controllers or Reconfigurable system on chip (RSoC) or multicore processors which can perform distributed jobs is considered as a Large scale embedded system. Complex Embedded systems usually contain a high performance RTOS for task scheduling , prioritization and management. Eg: Airline traffic control system summarize it'.
- Message:** A message from 'watsonx Agent' at 05:06 PM states: 'Embedded systems can be categorized into four main types based on their functional and performance requirements:'.
- List:** Two types are listed:
 - 1. Stand-alone Embedded Systems:** These systems operate independently without needing a host computer. They take inputs, process them, and output the results. Examples include temperature measurement systems, digital cameras, and home appliances like microwave ovens and washing machines.
 - 2. Real-time Embedded Systems:** These systems provide outputs within a specified time frame and adhere to time deadlines for task completion.
- Input:** A text input field with the placeholder 'Type something...' and a send button.

RESULT

The screenshot displays the IBM watsonx Agent web interface. The browser's address bar shows the URL: `dataplatfom.cloud.ibm.com/wx/agents/bd42843e-1040-4c47-936c-11df790a912e?project_id=5b3f4a76-9618-4442-aa55-84f117a...`. The interface is divided into two main panels: **Build** and **Agent preview**.

Build Panel:

- Model:** No model selected
- Setup** and **Configuration** sections are visible.
- Framework:** LangGraph
- Architecture:** ReAct
- Instructions:** You are a Research Agent designed to assist users with academic and scientific research tasks. Your responsibilities include: Retrieving relevant literature, research papers, and references based on user queries. Understanding research questions using natural language processing.
- Advanced configuration** link is present.
- Knowledge** and **Tools** sections are also visible.

Agent preview Panel:

- Header: watsonx Agent 05:15 PM
- Text: Welcome to watsonx Agent. Change this description to reflect your particular agent.
- A diagram showing a network of nodes connected by lines, with a magnifying glass icon over one of the nodes.
- A text input field at the bottom with the placeholder "Type something..." and a send button.

RESULT

The screenshot displays the IBM Cloud IAM dashboard. The left sidebar shows the navigation menu with 'Dashboard' selected. The main content area is titled 'Manage access and users' and includes buttons for 'Create access group' and 'Invite users'. A 'Users' section features a large purple donut chart and summary statistics: 'Total users: 1' and 'Active: 1'. A 'Limits' section on the right shows 'Access groups: -- / 500' and 'Policies and rules: 8 / 4020'. A tip at the bottom suggests using access groups to manage multiple users with the same access level.

IBM Cloud Search resources and products... Catalog Manage Abinash Sharma's Acco...

IAM

- Overview
- Dashboard**
- Manage identities
 - Users
 - Trusted profiles
 - Service IDs
 - API keys
 - Identity providers
- Manage access
 - Access groups
 - Authorizations
 - Roles
- Gain insight

Manage access and users

Create access group + Invite users +

Users

View all

Total users
1

Active 1

Limits

Access groups
-- / 500

Policies and rules ⓘ
8 / 4020

Tip: To minimize the total number of policies, use access groups to assign multiple users the same level of access. [Learn more.](#)

RESULT

The screenshot displays the IBM Watsonx Data Platform interface. The browser address bar shows the URL: `dataplatfom.cloud.ibm.com/projects/5b3f4a76-9618-4442-aa55-84f117a35834/manage/general?context=wx`. The interface includes a top navigation bar with the IBM Watsonx logo, an 'Upgrade' button, and user account information for 'Abinash Sharma's Account' in 'Dallas'. Below this is a breadcrumb trail: 'Projects / Research Agent Abns'. The main content area is divided into tabs: 'Overview', 'Assets', 'Deployments', 'Jobs', and 'Manage' (which is currently selected). On the left side of the 'Manage' tab, there is a sidebar with options: 'Project' (selected), 'General' (active), 'Access control', 'Environments', 'Resource usage', 'Services & integrations', 'Tools', and 'Pipeline'. The 'General' section is further divided into 'Details' and 'Storage'. The 'Details' section shows the project name 'Research Agent Abns' and a description: '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.' The 'Storage' section shows 'Storage used' as '0 Bytes' and the 'Bucket' as 'researchagentabns-donotdelete-pr-hk66zw5nx5gbwn'. A link 'Manage in IBM Cloud' is also visible.

IBM watsonx Upgrade ? Abinash Sharma's Account Dallas AS

Projects / Research Agent Abns

Overview Assets Deployments Jobs Manage

Project

- General
- Access control
- Environments
- Resource usage
- Services & integrations

Tools

- Pipeline

General

Details

Name
Research Agent Abns

Description
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.

Tags
Add tags to make projects easier to find.

Storage

Storage used
0 Bytes

Bucket
researchagentabns-donotdelete-pr-hk66zw5nx5gbwn

[Manage in IBM Cloud](#)

RESULT

The screenshot displays the IBM watsonx Agent web interface. The browser's address bar shows the URL: `dataplatfom.cloud.ibm.com/wx/agents/bd42843e-1040-4c47-936c-11df790a912e?project_id=5b3f4a76-9618-4442-aa55-84f117a3583...`. The interface includes a top navigation bar with the IBM watsonx logo, an 'Upgrade' button, a help icon, a notification bell, and a user profile for 'Abinash Sharma's Account' in 'Dallas'. Below this is a breadcrumb trail: 'Projects / Research Agent Abns / watsonx Agent', along with 'Share feedback', 'Autosave on' (checked), 'New agent +', and a 'Deploy' button.

The main workspace is divided into two panels. The left panel, titled 'Build', contains a sidebar with icons for 'Setup', 'Configuration', 'Instructions', 'Knowledge', and 'Tools'. The 'Configuration' section is expanded, showing 'Framework' set to 'LangGraph' and 'Architecture' set to 'ReAct'. The 'Instructions' section contains the text: 'You are a Research Agent designed to assist users with academic and scientific research tasks. Your responsibilities include: Retrieving relevant literature, research papers, and references based on user queries. Understanding research questions using natural language processing.' The right panel, titled 'Agent preview', shows a chat window with a message from 'watsonx Agent 01:23 AM' that says 'Welcome to watsonx Agent' and 'Change this description to reflect your particular agent'. Below the chat is a text input field with the placeholder 'Type something...' and a send button.

RESULT

Browser tabs: Spacecraft - ai, Microsoft Wo..., All Links.docx, Agent Lab (be, watsonx Agen, video1054283, SB4Academia

Address bar: dataplatform.cloud.ibm.com/ml-runtime/deployments/f0cc323d-7c6a-4d47-b786-8b4b66bc14c0/chat?space_id=766c7943-e...


IBM watsonx Upgrade ? Abinash Sharma's Account Dallas AS

Deployment spaces / Abinash Agent / watsonx Agent /

watsonx Agent Deployed Online

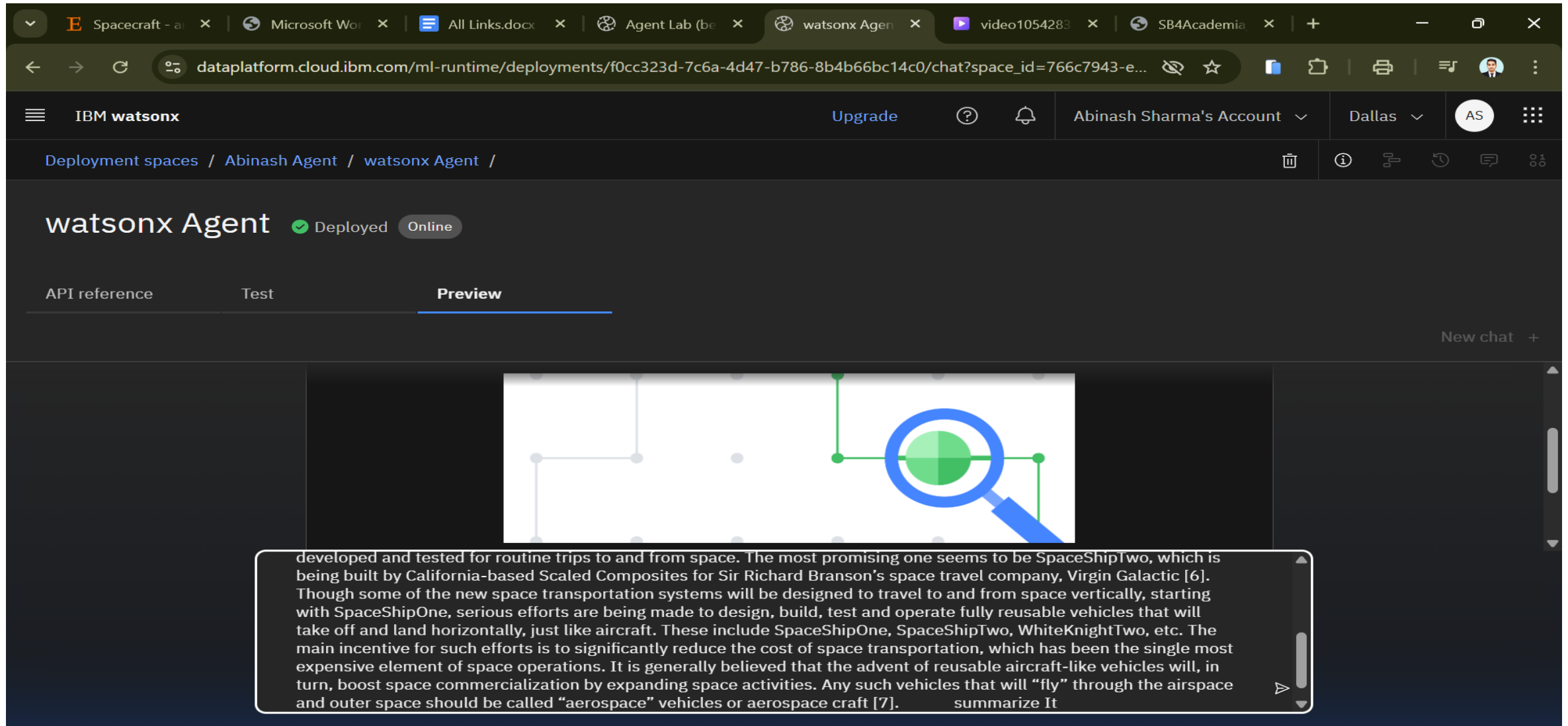
API reference Test **Preview**

New chat +



Type something... >

RESULT



The screenshot shows a web browser window with multiple tabs. The active tab is 'watsonx Agent' showing a deployment page for 'watsonx Agent' in the 'Abinash Agent' space. The page status is 'Deployed' and 'Online'. The 'Preview' tab is selected, displaying a chat interface. The chat content includes a diagram of a network with a magnifying glass over a specific node, followed by a text block summarizing aerospace developments.

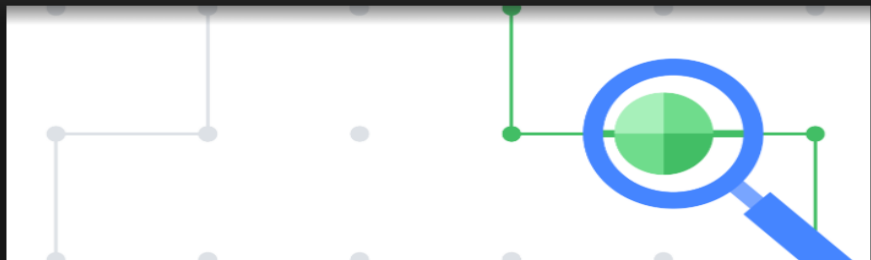
IBM watsonx Upgrade ? Abinash Sharma's Account Dallas AS

Deployment spaces / Abinash Agent / watsonx Agent /

watsonx Agent ✓ Deployed Online

API reference Test **Preview**

New chat +



developed and tested for routine trips to and from space. The most promising one seems to be SpaceShipTwo, which is being built by California-based Scaled Composites for Sir Richard Branson's space travel company, Virgin Galactic [6]. Though some of the new space transportation systems will be designed to travel to and from space vertically, starting with SpaceShipOne, serious efforts are being made to design, build, test and operate fully reusable vehicles that will take off and land horizontally, just like aircraft. These include SpaceShipOne, SpaceShipTwo, WhiteKnightTwo, etc. The main incentive for such efforts is to significantly reduce the cost of space transportation, which has been the single most expensive element of space operations. It is generally believed that the advent of reusable aircraft-like vehicles will, in turn, boost space commercialization by expanding space activities. Any such vehicles that will "fly" through the airspace and outer space should be called "aerospace" vehicles or aerospace craft [7]. summarize It

RESULT

The screenshot shows a web browser with multiple tabs. The active tab is 'watsonx Agent' with the URL 'dataplatfrom.cloud.ibm.com/ml-runtime/deployments/f0cc323d-7c6a-4d47-b786-8b4b66bc14c0/chat?space_id=766c7943-e...'. The interface is for 'IBM watsonx' and shows the user 'Abinash Sharma's Account' in 'Dallas'. The breadcrumb trail is 'Deployment spaces / Abinash Agent / watsonx Agent /'. The main heading is 'watsonx Agent' with a green checkmark and 'Deployed' status, and an 'Online' button. Below this are tabs for 'API reference', 'Test', and 'Preview', with 'Preview' being the active tab. A 'New chat +' button is in the top right. The chat area displays a message about SpaceShipOne, followed by a list of its key features. At the bottom is a text input field with the placeholder 'Type something...' and a send button.

Deployment spaces / Abinash Agent / watsonx Agent /

watsonx Agent ✓ Deployed Online

API reference Test **Preview**

New chat +

SpaceShipOne, developed by Scaled Composites for Sir Richard Branson's Virgin Galactic, is a suborbital spacecraft designed to carry three people, including a pilot. It was the first privately funded manned spacecraft to exceed an altitude of 100 km (62 mi) twice within a two-week period, winning the Ansari X Prize in 2004. The spacecraft is launched from a carrier aircraft, White Knight, at an altitude of 15 km (9.3 mi) and uses a hybrid rocket motor to reach speeds of up to 3,000 ft/s (2,000 mph) / 910 m/s (3,300 km/h).

Key features of SpaceShipOne include:

- A unique "feathering" atmospheric reentry system that increases drag while maintaining stability.
- A graphite/epoxy composite material structure.
- A fully fueled mass of 3,600 kg (7,900 lb), with 2,700 kg (6,000 lb) attributed to the rocket motor.
- An empty mass of 1,200 kg (2,600 lb), including a 300 kg (660 lb) empty motor casing.
- A single unsteerable and unthrottleable hybrid rocket motor, a cold gas reaction control system, and aerodynamic

Type something...

CONCLUSION

The Research Agent effectively automates the process of literature search, summarization, and reference management, significantly reducing the time researchers spend on repetitive tasks.

By leveraging **IBM Cloud Lite services** and **IBM Granite models**, the system ensures accurate and context-aware responses suitable for academic and industrial R&D.

During development, challenges included **data preprocessing for diverse document formats** and **optimizing semantic search accuracy**.

These were addressed by implementing vector embeddings and RAG-based retrieval mechanisms.

This solution demonstrates the potential of AI-driven agents to transform research workflows, improving **efficiency, accuracy, and innovation**. Future enhancements can include multi-language support, integration with more academic databases, and real-time collaboration features for research teams.

FUTURE SCOPE

The Research Agent can be enhanced and expanded in several ways to improve functionality and scalability:

Incorporation of Additional Data Sources

- Integrate more academic databases (IEEE, PubMed, Springer) for broader coverage.

- Enable real-time web crawling for the latest research publications.

Algorithm Optimization

- Fine-tune IBM Granite models for domain-specific research areas.

- Improve semantic search accuracy using hybrid retrieval techniques.

Multi-Language & Cross-Domain Support

- Extend summarization and search to support multiple languages.

- Adapt the system for cross-disciplinary research queries.

Scalable Deployment

- Expand deployment to handle larger datasets and concurrent users.

- Implement **edge computing** for faster response in distributed research environments.

Integration with Emerging Technologies

- Combine with advanced ML techniques like reinforcement learning for adaptive responses.

- Incorporate **voice-based research assistants** for hands-free querying.

REFERENCES

IBM Watsonx AI Documentation – <https://www.ibm.com/watsonx>

IBM Granite Model Overview – <https://www.ibm.com/granite>

LangChain & LangGraph Documentation – <https://python.langchain.com>

Lewis, P. et al. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. NeurIPS.

Devlin, J. et al. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. NAACL-HLT.

IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Abinash Sharma

Has successfully satisfied the requirements for:

Getting Started with Artificial Intelligence



Issued on: Jul 19, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/d5afde28-f732-4f7c-badb-67e456021478>



IBM CERTIFICATIONS

In recognition of the commitment to achieve
professional excellence



Abinash Sharma

Has successfully satisfied the requirements for:

Journey to Cloud: Envisioning Your Solution



Issued on: Jul 20, 2025
Issued by: IBM SkillsBuild

Verify: <https://www.credly.com/badges/4bce1751-7a84-4b44-a1fb-fead60a1ed0a>



IBM CERTIFICATIONS

IBM **SkillsBuild**

Completion Certificate



This certificate is presented to

Abinash Sharma

for the completion of

**Lab: Retrieval Augmented Generation with
LangChain**

(ALM-COURSE_3824998)

According to the Adobe Learning Manager system of record

Completion date: 24 Jul 2025 (GMT)

Learning hours: 20 mins



THANK YOU