

---

# **CAPSTONE PROJECT**

## **SEMANITCAAI – AI-POWERED PLAGIARISM DETECTION AGENT**

**Presented By**

**Student name :Tulip Lenka**

**College Name & Department : Odisha University of Technology  
and Research**

# OUTLINE

- Problem Statement
- Technology used
- Wow factor
- End users
- Result
- Conclusion
- Git-hub Link
- Future scope
- IBM Certifications

---

# PROBLEM STATEMENT

Academic institutions face increasing difficulty in detecting nuanced forms of plagiarism, especially when assignments are paraphrased or generated by AI tools. Current plagiarism detectors lack contextual sensitivity to instructor-specific styles and grading patterns. The challenge lies in creating an adaptive AI system that learns from historical assignment submissions and instructor feedback to identify inconsistencies and potential misconduct dynamically. This would enhance academic integrity by flagging suspicious entries with improved accuracy and contextual awareness.

## ***Proposed Solution:***

**SemanticaAI** is a smart plagiarism detection assistant developed using IBM Watsonx and Granite LLM. Unlike typical tools, it catches semantic plagiarism by analyzing meaning, not just exact word matches. Users can interact naturally and get both plagiarism scores and rewrite suggestions — promoting ethical academic writing.

---

# TECHNOLOGY USED

- IBM Cloud Lite Services
- Natural Language Processing (NLP)
- Retrieval Augmented Generation (RAG)
- IBM Watsonx AI Studio
- IBM Granite Foundation Model
- LangGraph & ReAct Architecture
- JSON-based prompt configuration

---

# IBM CLOUD SERVICES USED

- IBM Cloud Watsonx AI Studio
- IBM Cloud Watson AI Runtime
- IBM Cloud Functions
- IBM Granite via Foundation Models
- IBM Cloud Object Storage (for text prompts/data)
- IBM Cloud Agent Lab

# WOW FACTORS

- **Semantic Detection:** Goes beyond surface-level matching to detect plagiarism based on meaning, even if the content is rephrased or paraphrased cleverly.
- **Smart Flagging:** Highlights suspicious sections and explains *why* they might be plagiarized, offering users deeper understanding — not just red marks.
- **Rewriting Support:** Doesn't just point out plagiarism — helps users rephrase ideas in their own words to improve originality and writing skills.
- **Conversational Interface:** Powered by IBM Granite, users interact naturally with the agent, like talking to a smart AI tutor.
- **Education-Oriented:** Promotes ethical writing and originality — ideal for students, educators, and academic institutions aiming to reduce unintentional plagiarism.
- **Easy Cloud Integration:** Built entirely on IBM Cloud — scalable, secure, and easy to extend for real-world academic use cases.

---

# END USERS

- College Students
- Research Scholars
- Content Writers
- Teachers and Academic Institutions
- Reviewers & Editors

# RESULTS

The screenshot displays the IBM watsonx AI Agent Lab interface. The top navigation bar includes the IBM watsonx logo, an 'Upgrade' button, a notification bell, a user profile for 'Tulip Lanka's Account', a location dropdown for 'Dallas', and a user icon. Below the navigation bar, the breadcrumb trail reads 'Projects / SemanticaAI / Agent Lab'. The main interface is divided into two panels: 'Build' on the left and 'Agent preview' on the right.

**Build Panel:**

- Model:** llama-3-3-70b-instruct
- Setup:** A dropdown menu is open, showing 'Configuration' as the selected option.
- Configuration:**
  - Framework:** LangGraph
  - Architecture:** ReAct
- Instructions:**
  - SemanticaAI - Your Academic Integrity Assistant**
  - Tasks SemanticaAI Performs:**
    - Analyze submitted text for semantic and n-gram based similarity...
- Knowledge:**
- Tools:**
  - Add a tool** and **Create custom tool** buttons.
  - Added tools (1):**
    - Google search:** Retrieve information from the internet with the Google search engine.

**Agent preview Panel:**

- SemanticaAI 07:17 PM:**
  - Welcome to SemanticaAI**
  - Hi! I'm SemanticaAI - your AI buddy to check for plagiarism and improve originality.**
- Sample questions:**
  - "Can you check this assignment for plagiarism?"
  - "How similar is my text to previous submissions?"
  - "Can you explain why this section is flagged?"
  - "How can I improve the originality of my work?"
- Activate Windows:** Go to Settings to activate Windows.



# RESULTS

The screenshot displays the IBM Watsonx Agent Lab interface. The top navigation bar includes the IBM Watsonx logo, an 'Upgrade' button, a notification bell, a user profile for 'Tulip Lenka's Account', and a 'Dallas' location selector. The main header shows the project path 'Projects / SemanticaAI / Agent Lab' and action buttons for 'Share feedback', 'Unsaved', 'New agent', and 'Deploy'.

The interface is divided into two main sections: 'Build' on the left and 'Agent preview' on the right.

**Build Section:**

- Model:** Set to 'granite-3-3-8b-instruct'.
- Configuration:**
  - Framework:** 'LangGraph'.
  - Architecture:** 'ReAct'.
- Instructions:** A text box containing the agent's role: 'SemanticaAI - Your Academic Integrity Assistant. SemanticaAI is a context-aware AI agent designed to uphold academic integrity by identifying semantic plagiarism...'. A link for 'Advanced configuration' is available.
- Knowledge:** A section for adding vector stores, with a 'Clear chat' button.
- Tools:** Includes buttons for 'Add a tool' and 'Create custom tool'.
- Added tools (4):** A list showing 'Google search' with the description 'Retrieve information from the internet with the Google search engine.'.

**Agent preview Section:**

The preview shows the agent's response to a query. The response text is: 'Both texts convey the same core message about the transformative impact of Artificial Intelligence (AI) across various industries. They highlight how AI automates tasks, boosts efficiency, and facilitates data-driven decision-making. Despite their similarities, there are subtle differences in wording and phrasing, which prevent them from being considered instances of plagiarism.'

The response is followed by three numbered analysis points:

- 1. Variation in Phrasing:**
  - Text 1: "Artificial Intelligence is transforming industries..."
  - Text 2: "AI is revolutionizing different fields..."
- 2. Synonym Usage:**
  - Text 1: "transforming," "improving efficiency," "enabling data-driven decision-making"
  - Text 2: "revolutionizing," "streamlining work," "enhancing productivity," "helping organizations make better decisions using data"
- 3. Sentence Structure:**
  - Text 1 presents a straightforward subject-verb-object structure.
  - Text 2 uses a more complex sentence structure with a relative clause ("which helps organizations make better decisions using data").

A concluding paragraph states: 'While these texts share a common theme and convey similar ideas, the differences in phrasing, synonym usage, and sentence structure prevent them from being flagged as plagiarized. SemanticaAI would recognize these variations and provide a plagiarism likelihood score with contextual explanation, along with actionable suggestions for improving originality. In this case, the score would be low, indicating minimal risk of plagiarism, and the suggestions would focus on enhancing the variety of language used to describe AI's impact.'

Below the analysis, there is a section titled 'How did I get this answer?' with a refresh icon.

The bottom of the preview shows a text input field with the placeholder 'Type something...' and an 'Activate Windows' watermark.

# RESULTS

The screenshot displays the IBM Watsonx Agent Lab interface, divided into two main panels: 'Build' and 'Agent preview'.

**Build Panel:**

- Setup:** Includes 'Framework' (LangGraph) and 'Architecture' (ReAct).
- Configuration:** Includes 'Instructions' and 'Knowledge'.
- Instructions:** A text box containing the following text: "SemanticaAI – Your Academic Integrity Assistant. SemanticaAI is a context-aware AI agent designed to uphold academic integrity by identifying semantic plagiarism – where ideas or structure are copied but words are changed. Unlike traditional detectors that rely on exact matches, SemanticaAI goes deeper: it understands the meaning behind the words. Built using IBM Granite models and RAG techniques, it offers real-time analysis, actionable insights, and promotes ethical learning."
- Knowledge:** A section for adding vector stores, with a 'Clear chat' button.
- Tools:** Includes 'Add a tool' and 'Create custom tool' buttons.
- Added tools (4):** A list of tools, including 'Google search' with the description 'Retrieve information from the internet with the Google search engine.'

**Agent preview Panel:**

- Agent preview:** A section for previewing the agent's output.
- Text 1:** "Artificial Intelligence is reshaping sectors by mechanizing processes, amplifying effectiveness, and empowering informed decision-making through data analysis."
- Text 2:** "revolutionizing," "streamlining work," "enhancing productivity," "helping organizations make better decisions using data"
- 3. Sentence Structure:**
  - Text 1 presents a straightforward subject-verb-object structure.
  - Text 2 uses a more complex sentence structure with a relative clause ("which helps organizations make better decisions using data").
- While these texts share a common theme and convey similar ideas, the differences in phrasing, synonym usage, and sentence structure prevent them from being flagged as plagiarized. SemanticaAI would recognize these variations and provide a plagiarism likelihood score with contextual explanation, along with actionable suggestions for improving originality. In this case, the score would be low, indicating minimal risk of plagiarism, and the suggestions would focus on enhancing the variety of language used to describe AI's impact.**
- How did I get this answer?**
- User:** "Can you rewrite this phrase , plagiarism free"
- SemanticaAI:** "Of course, I can help you rephrase that. Here's a plagiarism-free version:  
"Artificial Intelligence is reshaping sectors by mechanizing processes, amplifying effectiveness, and empowering informed decision-making through data analysis."  
Please note that while I strive to create unique phrasing, it's essential to ensure that the content remains faithful to the original intent and meaning. Always double-check for any unintentional plagiarism using appropriate tools or services."
- How did I get this answer?**
- Input field:** "Type something..."

# RESULTS

IBM watsonx

Upgrade ⓘ 🔔

Tulip Lenka's Account ▾ Dallas ▾ TL

Deployment spaces / SemanticAI / SemanticAI /

SemanticaAI Deployed Online

API reference Test Preview

Endpoints for inferencing ⓘ

Private endpoint

https://us-south.ml.cloud.ibm.com/ml/v4/deployments/ccb2b91b-5b3c-4c84-bf5a-1db465a2d16c/ai\_service?version=2021-05-01

https://us-south.ml.cloud.ibm.com/ml/v4/deployments/ccb2b91b-5b3c-4c84-bf5a-1db465a2d16c/ai\_service\_stream?version=2021-05-01

Public endpoint

https://us-south.ml.cloud.ibm.com/ml/v4/deployments/ccb2b91b-5b3c-4c84-bf5a-1db465a2d16c/ai\_service?version=2021-05-01

https://us-south.ml.cloud.ibm.com/ml/v4/deployments/ccb2b91b-5b3c-4c84-bf5a-1db465a2d16c/ai\_service\_stream?version=2021-05-01

[Learn more](#) about the 2021-05-01 version query parameter

Code snippets

cURL Java JavaScript Python Scala

```
# NOTE: you must set $API_KEY below using information retrieved from your IBM Cloud account (https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml-authentication.html?context=)

export API_KEY=your API key

export IAM_TOKEN=$(curl --insecure -X POST --location "https://iam.cloud.ibm.com/identity/token" \
--header "Content-Type: application/x-www-form-urlencoded" \
--header "Accept: application/json" \
--data-urlencode "grant_type=urn:ibm:params:oauth:grant-type:apikey" \
--data-urlencode "apikey=$API_KEY" | jq -r '.access_token')

# TODO: manually define and pass values to be scored below

curl --location "https://us-south.ml.cloud.ibm.com/ml/v4/deployments/ccb2b91b-5b3c-4c84-bf5a-1db465a2d16c/ai_service_stream?version=2021-05-01" \
```

About this deployment

Name

SemanticaAI

Description

Hi! I'm SemanticaAI -- your AI buddy to check for plagiarism and improve originality.

Deployment Details

Deployment ID: ccb2b91b-5b3c-4c...

Serving name: No serving name.

Software specification: runtime-24.1-py3.11

Hardware specification: Extra extra small: 1 CPU and 2 GB RAM

Copies: 1

Tags

wx-agent

Associated asset

SemanticaAI

f9b9f4e6-8983-4291-b2cc-7d372bfe8376

Last modified

1 hour ago

Created on

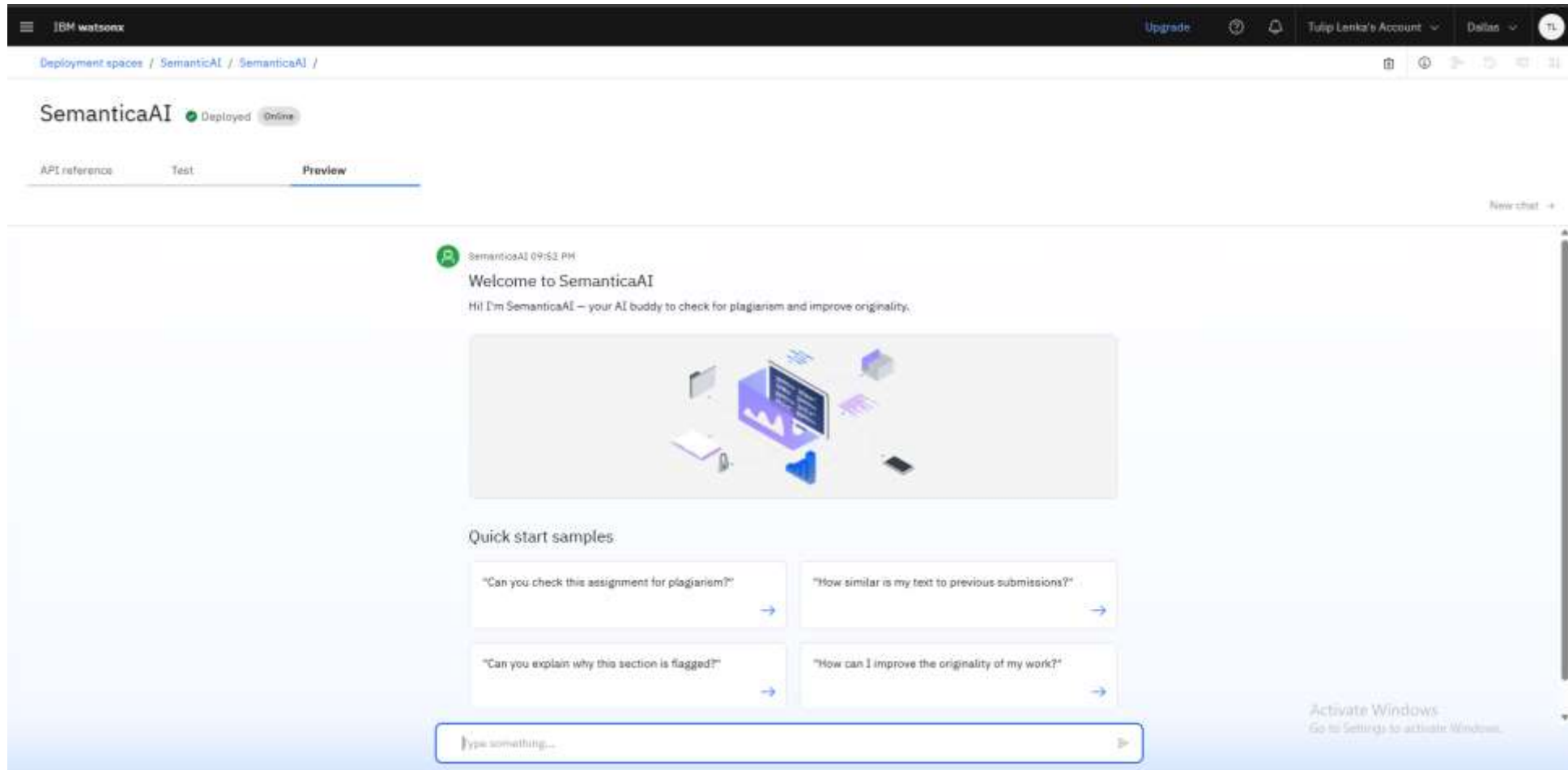
Jul 30, 2025

Activate Windows

Go to Settings to activate Windows.

# RESULTS

## *Deployed AI Agent :*



---

# CONCLUSION

- **SemanticaAI** is more than just a plagiarism checker — it's an AI-powered academic companion that understands the *meaning* behind words. By detecting hidden and paraphrased plagiarism, it empowers users to write with integrity and confidence.
- Built entirely using IBM's cutting-edge Granite foundation models and deployed on the watsonx platform, SemanticaAI demonstrates how AI can transform education by promoting ethical learning, originality, and smart writing habits.
- With its potential for real-time use in classrooms and integration with learning platforms, SemanticaAI sets the stage for a new era of AI-guided academic support.

---

# FUTURE SCOPE

- Add AI-generated text detection (like ChatGPT-style content)
- Language translation + multilingual plagiarism detection
- Support document uploads (PDF, DOCX)
- Integration with learning platforms (e.g., Moodle)
- Real-time classroom deployment

# IBM CERTIFICATIONS



IBM **SkillsBuild**

Completion Certificate



This certificate is presented to

Tulip Lenka

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



In recognition of the commitment to achieve  
professional excellence



# Tulip Lenka

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/a628c412-753e-45f0-a8c7-5a1da9c8e35a>



---

# GITHUB LINK

- <https://github.com/tuliip1807/SemanticaAI-Plagiarism-Checker>



# THANK YOU