Project Report: CASP - AI-Powered Career, Plagiarism, and Study Assistant

Problem Statement 4

AI-Powered Interactive Learning Assistant for Classrooms: Build a Multimodal AI assistant for classrooms to dynamically answer queries using text, voice, and visuals while improving student engagement with personalized responses.

---

1. Objective

To develop an AI-powered web-based assistant that supports students and educators through:

- Career analysis and personalized guidance
- Vision-based plagiarism detection
- Smart notes generation and document-based chatbot with voice integration(speech-to-text)

---

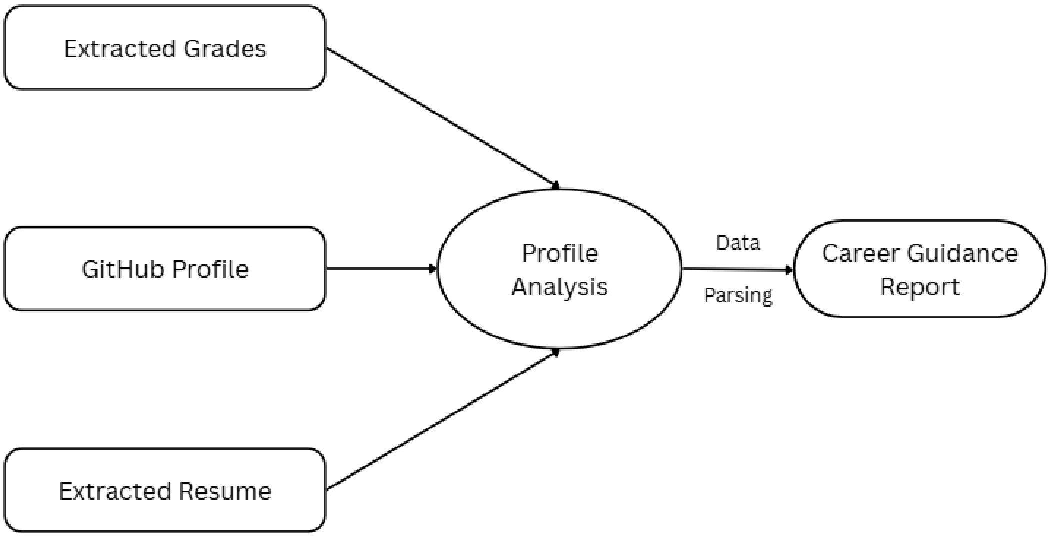
2. Team Details

Role	Name	Contributions
Team Member	Penugurthi Manoj	Career DHI (`views/page1.py`)
Team Member	Payal Choudhary	Plagiarism Checker (`views/page2.py`)
Team Member	P V Abhijitha	Study Assistant (`views/page3.py`)
Faculty Guide	Dr. Lade S Chakravarthi	Provided guidance and ensured timely progress by helping the team stay aligned with project deadlines.

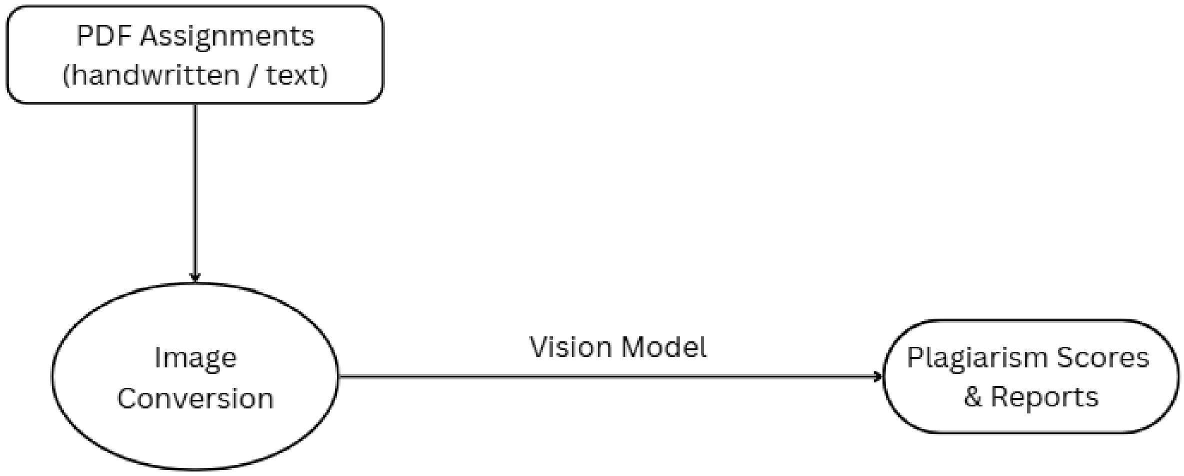
---

3. Data Flow Diagram

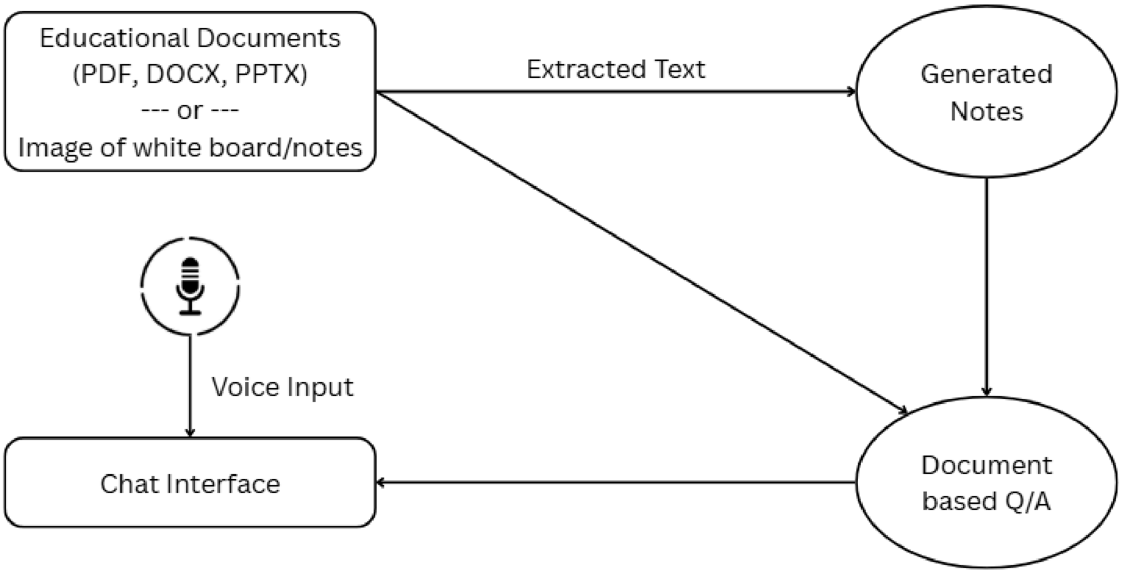
Career DHI (Data, Highlights & Insights)



Plagiarism Checker


PDF Assignments (handwritten / text)
Image Conversion
Vision Model
Plagiarism Scores & Reports

AI Buddy (Smart Study Assistant)


Educational Documents (PDF, DOCX, PPTX) --- or --- Image of white board/notes
Extracted Text
Generated Notes
Voice Input
Chat Interface
Document based Q/A

4. Methodology

The **CASP system** was developed using a **modular architecture** combined with an **agile-based development approach**, ensuring rapid, iterative progress and smooth collaboration among team members. The primary goal was to create a scalable, secure, and responsive AI-powered assistant for educational environments. The development methodology involved the following key stages:

4.1 Modular Development

Each core functionality—**Career DHI**, **Plagiarism Checker**, and **AI Buddy**—was treated as an independent module with clearly defined inputs, processes, and outputs. This modular design enabled:

- Parallel development by different team members without dependencies.
- Ease of testing and debugging module-wise functionality.
- Smooth integration of third-party tools and APIs as needed per module.

4.2 Role-Based Agile Workflow

- The team followed **agile principles**, dividing work into weekly sprints.
- Tasks were assigned based on individual module responsibilities:
    - Career DHI: Handled grade extraction, GitHub parsing, and resume summarization.
    - Plagiarism Checker: Focused on PDF image conversion, OCR, and similarity detection.
    - AI Buddy: Handled document summarization, voice input, and chatbot development.

Weekly stand-up discussions helped identify blockers, sync progress, and adjust workloads accordingly.

4.3 Rapid Prototyping using Streamlit

- **Streamlit** was used as the primary front-end framework due to its simplicity and flexibility for rapid prototyping.
- The UI was designed with tabs, containers, and chat components for seamless user experience.
- Minimal setup allowed quick testing of LLM integrations and real-time feature demonstration.

4.4 Model Optimization

- For vision tasks, **Groq's API** was used due to its high accuracy and speed with large document images.

4.5 Secure API Integration

- All API keys and credentials (e.g., for Groq LLM and Vision APIs) were stored securely in a **`.env` file**, which was excluded from version control using `.gitignore`.
- This practice helped in:
    - Preventing accidental exposure of keys

    - Maintaining different environments for development, testing, and deployment

4.6 Testing and Integration

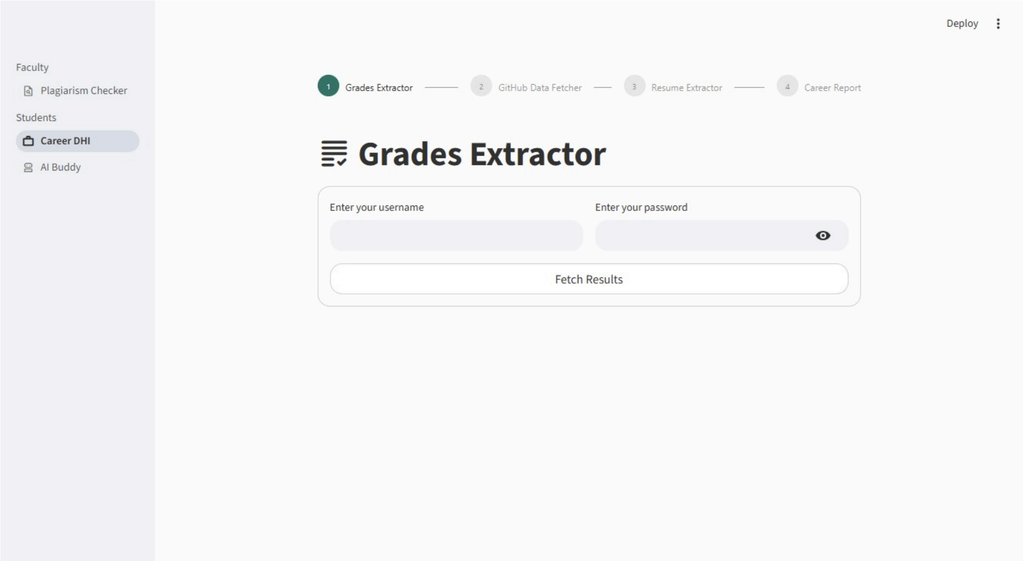
- After individual module development, integration was performed to ensure smooth data flow between modules.
- Extensive manual testing was done:
    - To verify multi-format document uploads (PDF, DOCX, PPTX)
    - To check response correctness of AI-generated summaries and career suggestions
    - To evaluate the accuracy of sentence-level plagiarism detection

---

5. Modules and Features

5.1 Career DHI (Data, Highlights & Insights)

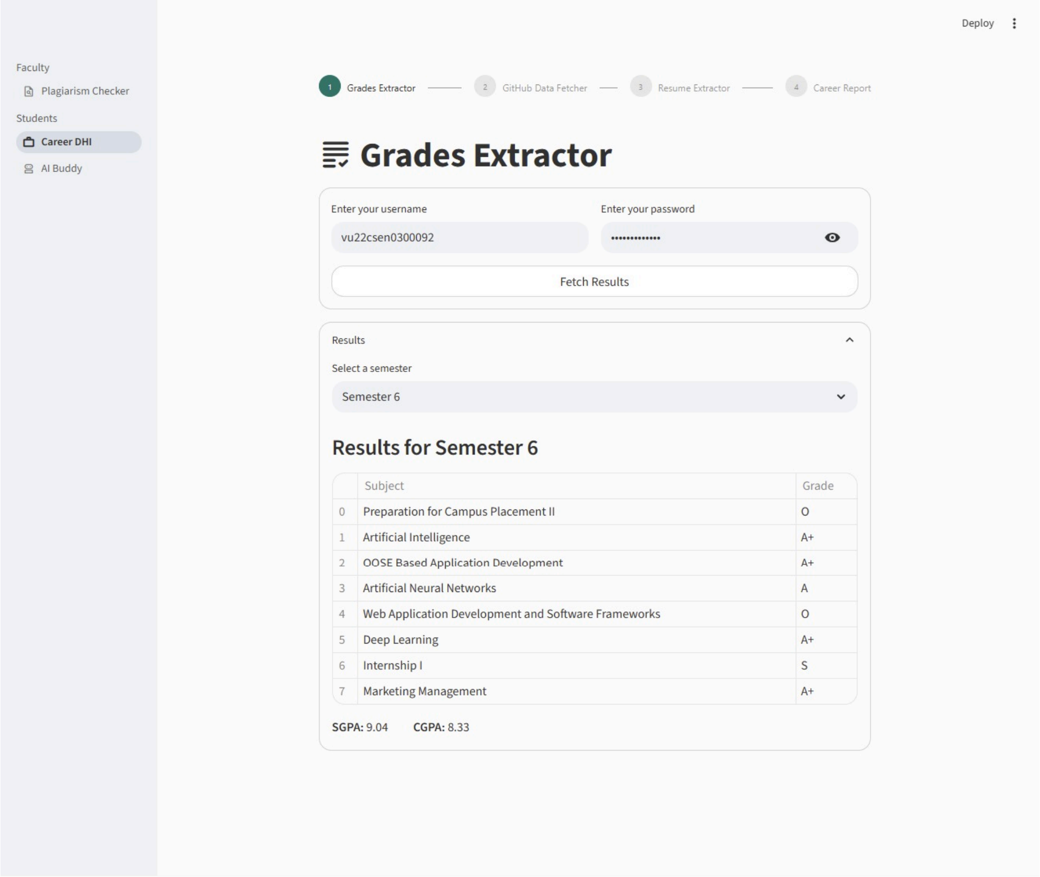
- **Grades Extractor:** Automates login and extraction of semester-wise grades from university portals using Selenium



Grades Extractor

Enter your username

Enter your password



Grades Extractor

Enter your username: vu22csen0300092

Enter your password: •••••••••••••

Results

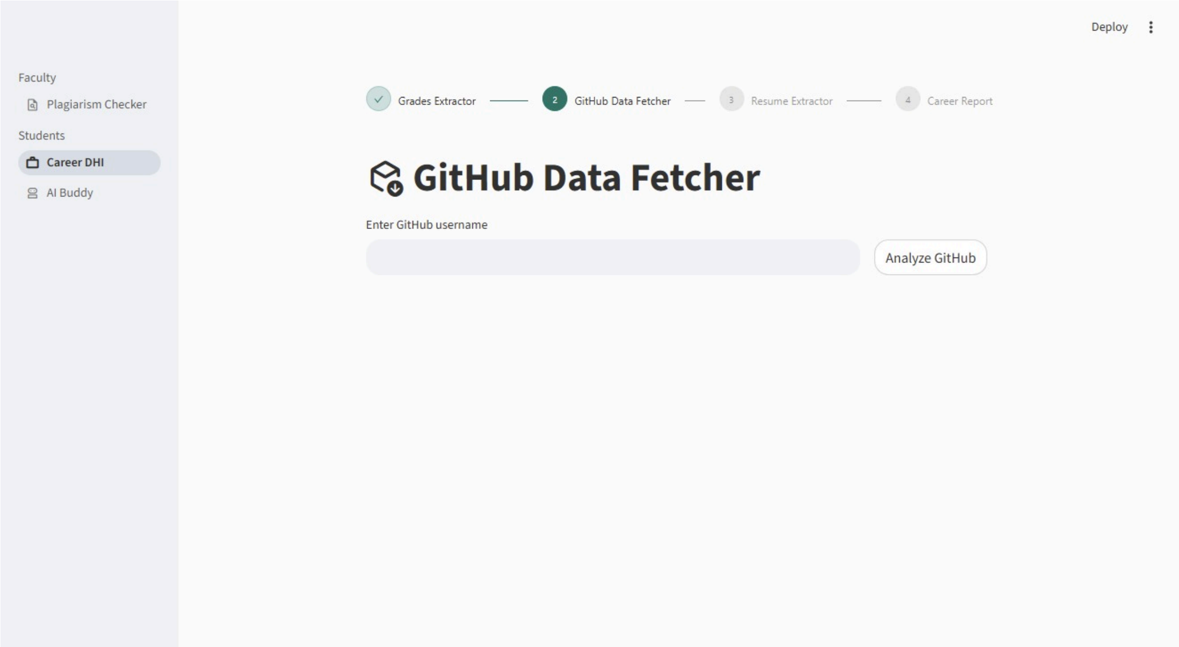
Select a semester: Semester 6

Results for Semester 6

	Subject	Grade
0	Preparation for Campus Placement II	O
1	Artificial Intelligence	A+
2	OOSE Based Application Development	A+
3	Artificial Neural Networks	A
4	Web Application Development and Software Frameworks	O
5	Deep Learning	A+
6	Internship I	S
7	Marketing Management	A+

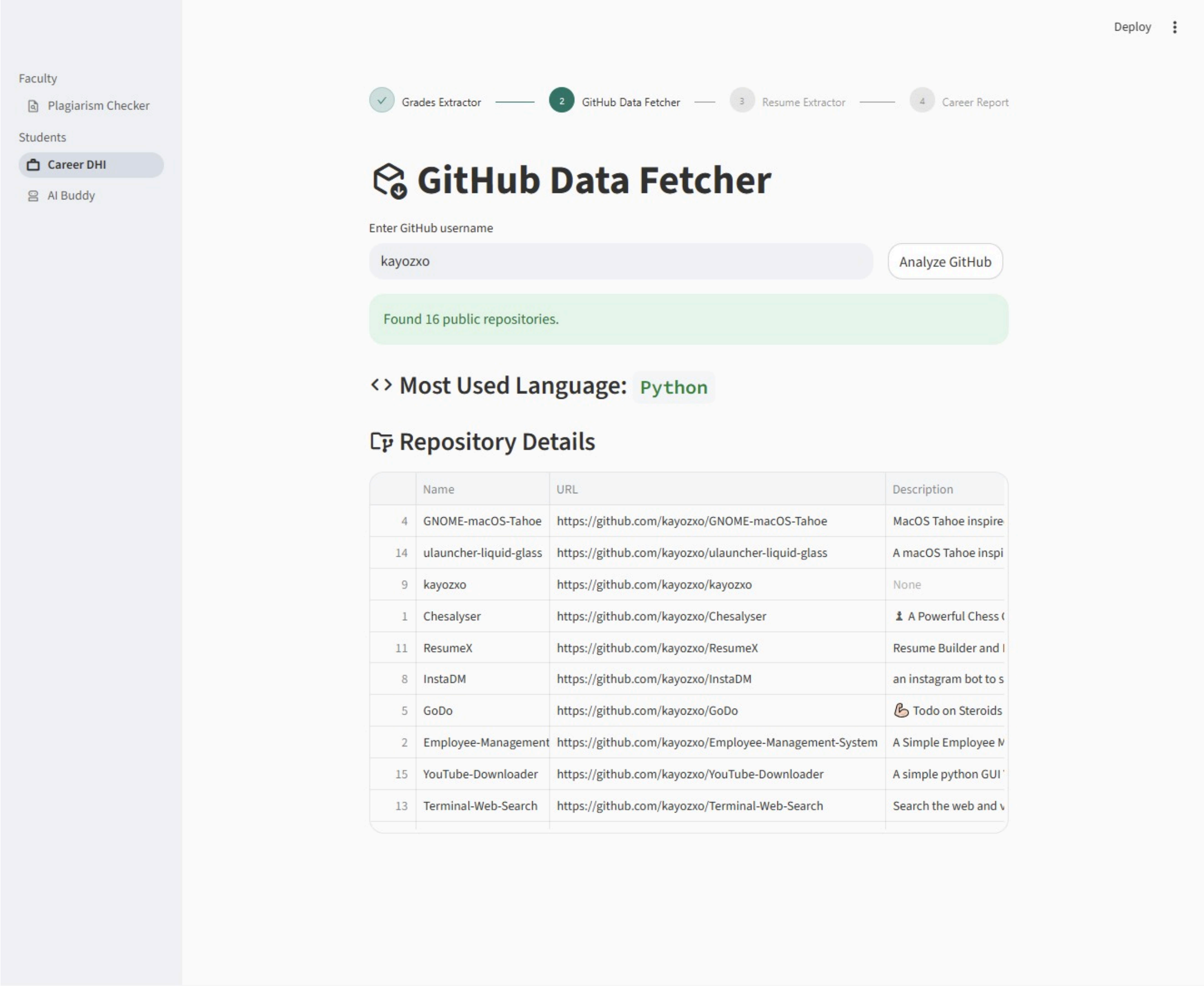
SGPA: 9.04 **CGPA:** 8.33

- **GitHub Data Fetcher:** Analyzes public GitHub profiles to determine the user's primary tech stack.



GitHub Data Fetcher

Enter GitHub username



GitHub Data Fetcher

Enter GitHub username

kayozxo

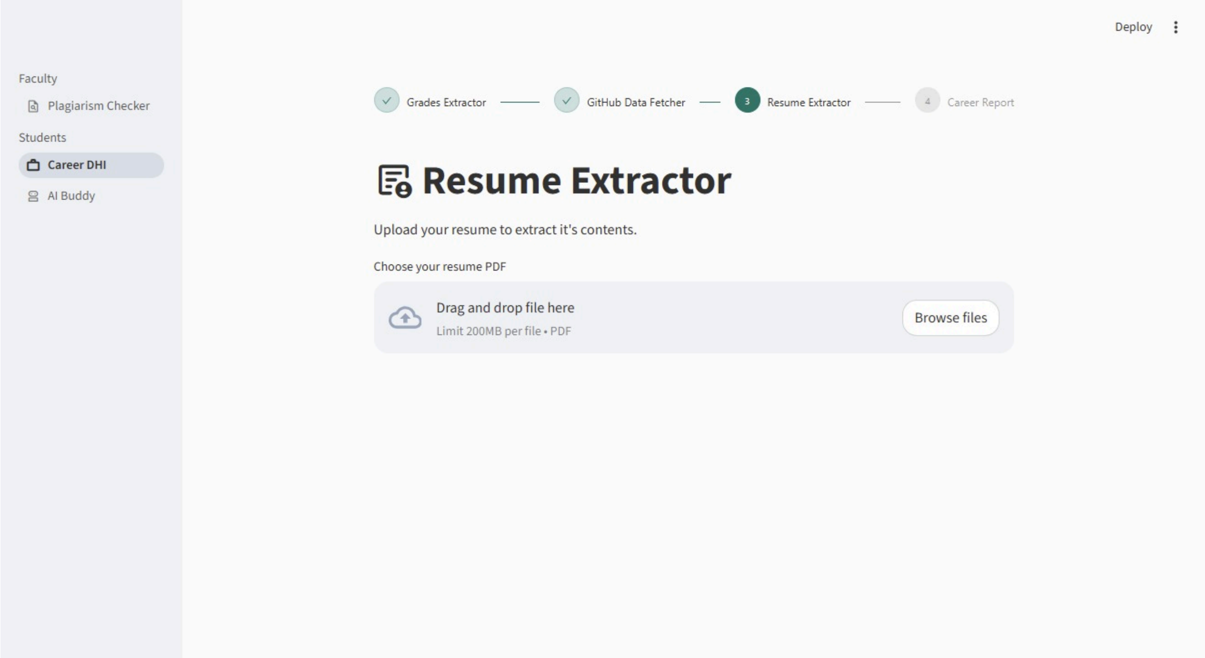
Found 16 public repositories.

<> Most Used Language: `Python`

Repository Details

	Name	URL	Description
4	GNOME-macOS-Tahoe	https://github.com/kayozxo/GNOME-macOS-Tahoe	MacOS Tahoe inspire
14	ulauncher-liquid-glass	https://github.com/kayozxo/ulauncher-liquid-glass	A macOS Tahoe inspi
9	kayozxo	https://github.com/kayozxo/kayozxo	None
1	Chesalyser	https://github.com/kayozxo/Chesalyser	♟ A Powerful Chess (
11	ResumeX	https://github.com/kayozxo/ResumeX	Resume Builder and I
8	InstaDM	https://github.com/kayozxo/InstaDM	an instagram bot to s
5	GoDo	https://github.com/kayozxo/GoDo	💪 Todo on Steroids
2	Employee-Management	https://github.com/kayozxo/Employee-Management-System	A Simple Employee M
15	YouTube-Downloader	https://github.com/kayozxo/YouTube-Downloader	A simple python GUI
13	Terminal-Web-Search	https://github.com/kayozxo/Terminal-Web-Search	Search the web and v

- **Resume Extractor:** Parses uploaded resumes (PDF) using PyMuPDF4LLM for content extraction.



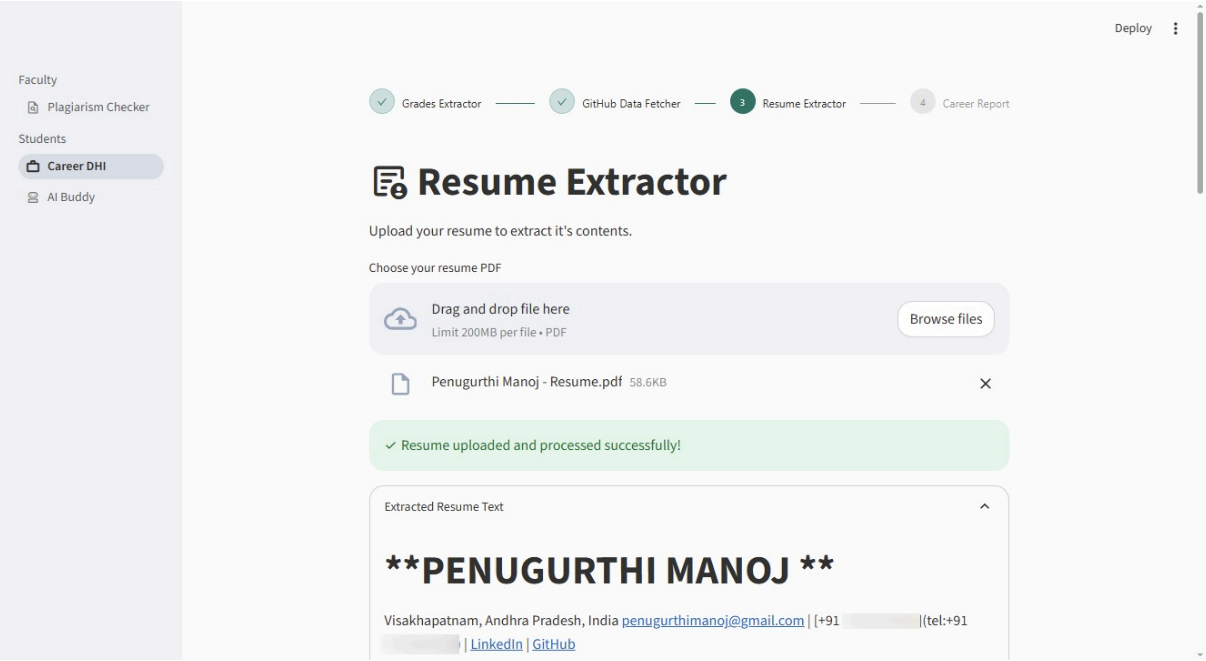
Resume Extractor

Upload your resume to extract it's contents.

Choose your resume PDF

Drag and drop file here

Limit 200MB per file • PDF



Resume Extractor

Upload your resume to extract it's contents.

Choose your resume PDF

Drag and drop file here
Limit 200MB per file • PDF

Penugurthi Manoj - Resume.pdf 58.6KB

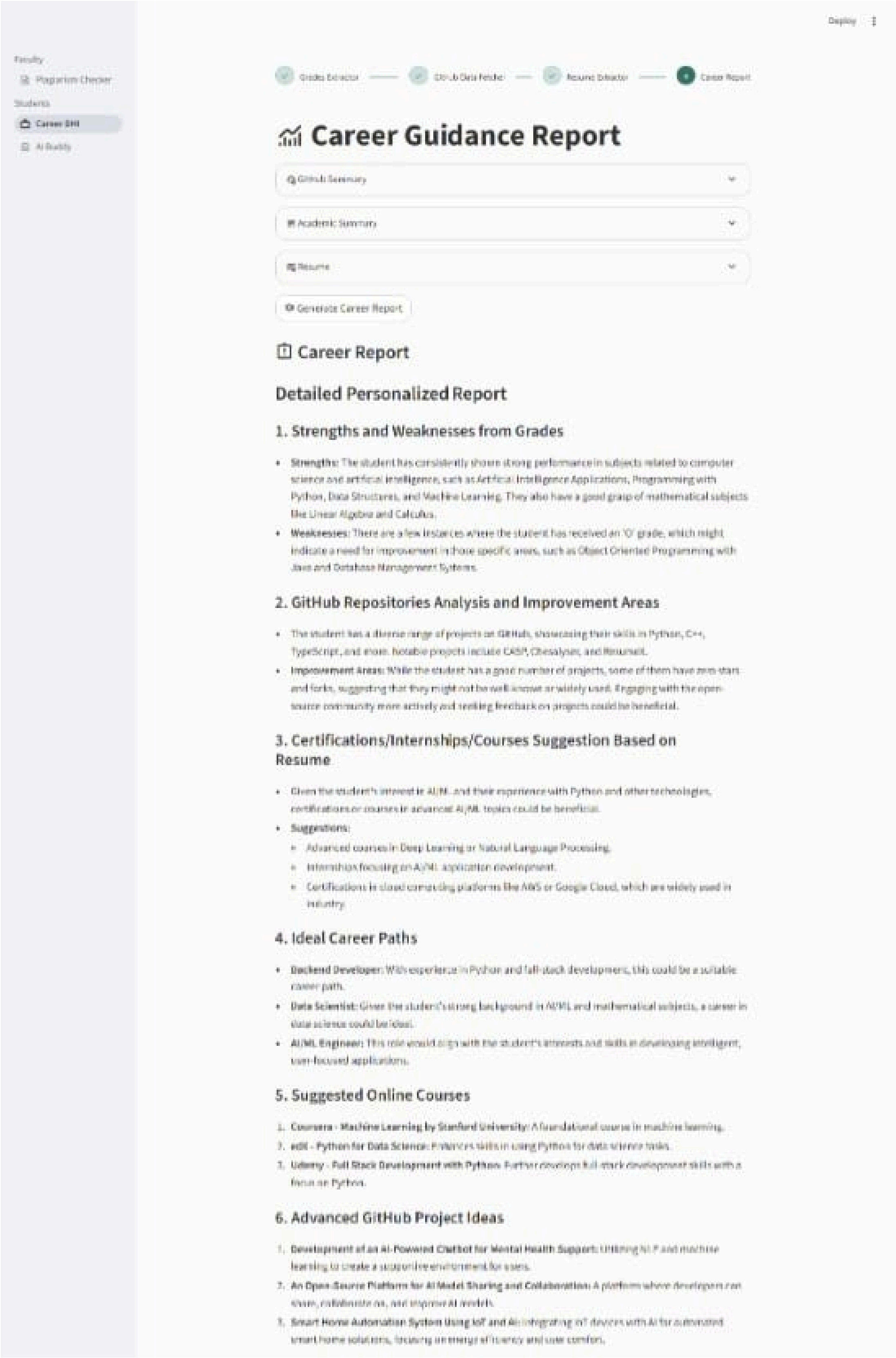
✓ Resume uploaded and processed successfully!

Extracted Resume Text

**PENUGURTHI MANOJ **

Visakhapatnam, Andhra Pradesh, India penugurthimanoj@gmail.com | [+91 [redacted] |(tel:+91 [redacted]) | LinkedIn | GitHub

- **Career Report Generator:** Combines academic and coding profiles to generate:
    - Strengths & weaknesses
    - Suggestions for improvement
    - Recommended certifications/courses
    - Ideal career paths and real project ideas



Career Guidance Report

Career Report

Detailed Personalized Report

1. Strengths and Weaknesses from Grades

- **Strengths:** The student has consistently shown strong performance in subjects related to computer science and artificial intelligence, such as Artificial Intelligence Applications, Programming with Python, Data Structures, and Machine Learning. They also have a good grasp of mathematical subjects like Linear Algebra and Calculus.
- **Weaknesses:** There are a few instances where the student has received an 'O' grade, which might indicate a need for improvement in those specific areas, such as Object Oriented Programming with Java and Database Management Systems.

2. GitHub Repositories Analysis and Improvement Areas

- The student has a diverse range of projects on GitHub, showcasing their skills in Python, C++, TypeScript, and more. Notable projects include CASP, Chessalyser, and Resumeit.
- **Improvement Areas:** While the student has a good number of projects, some of them have zero stars and forks, suggesting that they might not be well-known or widely used. Engaging with the open-source community more actively and seeking feedback on projects could be beneficial.

3. Certifications/Internships/Courses Suggestion Based on Resume

- Given the student's interest in AI/ML and their experience with Python and other technologies, certifications or courses in advanced AI/ML topics could be beneficial.
- **Suggestions:**
  - Advanced courses in Deep Learning or Natural Language Processing.
  - Internships focusing on AI/ML application development.
  - Certifications in cloud computing platforms like AWS or Google Cloud, which are widely used in industry.

4. Ideal Career Paths

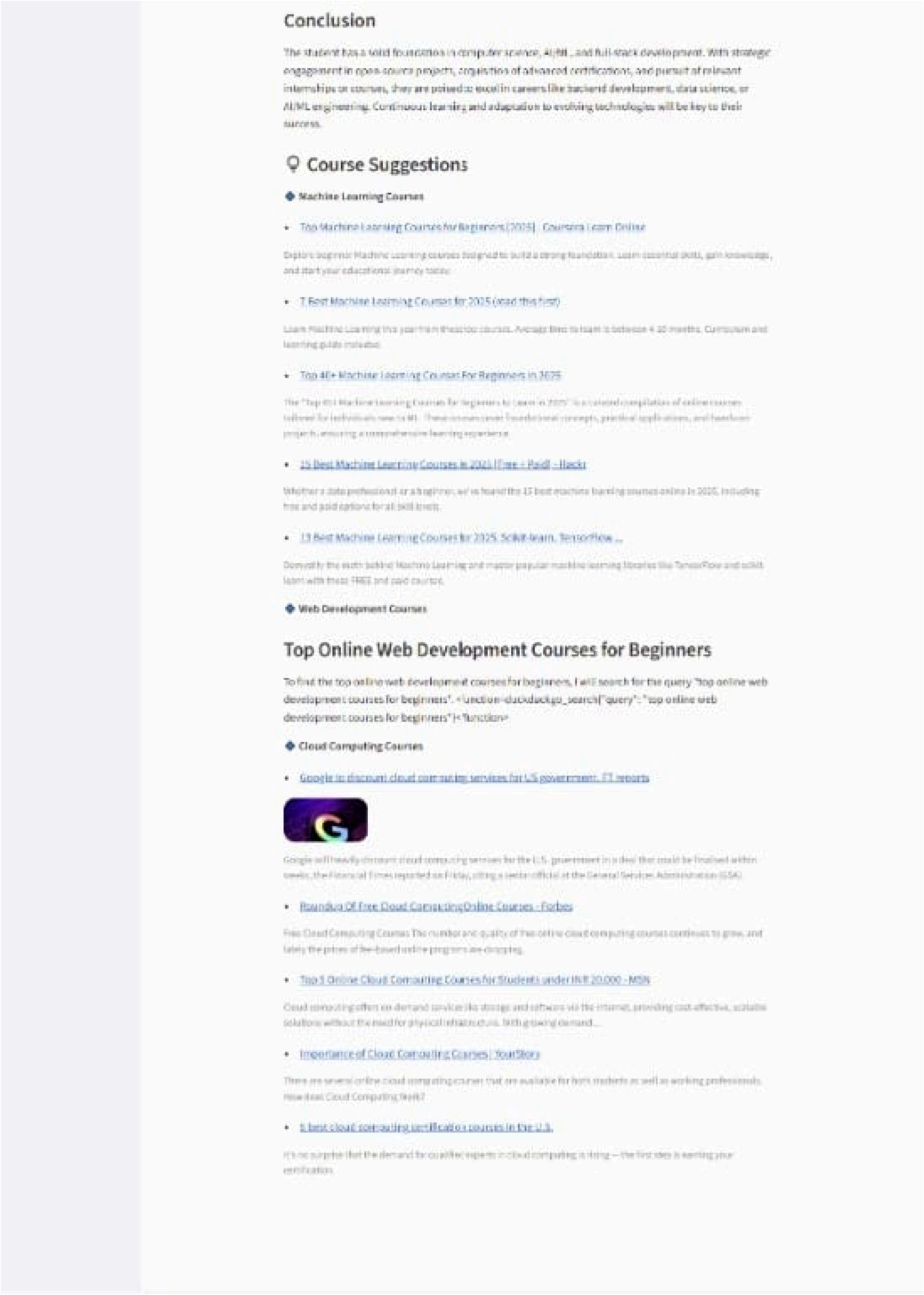
- **Backend Developer:** With experience in Python and full-stack development, this could be a suitable career path.
- **Data Scientist:** Given the student's strong background in AI/ML and mathematical subjects, a career in data science could be ideal.
- **AI/ML Engineer:** This role would align with the student's interests and skills in developing intelligent, user-focused applications.

5. Suggested Online Courses

1. **Coursera - Machine Learning by Stanford University:** A foundational course in machine learning.
2. **edX - Python for Data Science:** Enhances skills in using Python for data science tasks.
3. **Udemy - Full Stack Development with Python:** Further develops full-stack development skills with a focus on Python.

6. Advanced GitHub Project Ideas

1. **Development of an AI-Powered Chatbot for Mental Health Support:** Utilizing NLP and machine learning to create a supportive environment for users.
2. **An Open-Source Platform for AI Model Sharing and Collaboration:** A platform where developers can share, collaborate on, and improve AI models.
3. **Smart Home Automation System Using IoT and AI:** Integrating IoT devices with AI for automated smart home solutions, focusing on energy efficiency and user comfort.



Conclusion

The student has a solid foundation in computer science, AI/ML, and full-stack development. With strategic engagement in open-source projects, acquisition of advanced certifications, and pursuit of relevant internships or courses, they are poised to excel in careers like backend development, data science, or AI/ML engineering. Continuous learning and adaptation to evolving technologies will be key to their success.

💡 Course Suggestions

🔷 **Machine Learning Courses**

- Top Machine Learning Courses for Beginners [2025] | Coursera Learn Online

Explore beginner Machine Learning courses designed to build a strong foundation. Learn essential skills, gain knowledge, and start your educational journey today.

- 7 Best Machine Learning Courses for 2025 (read this first)

Learn Machine Learning this year from these top courses. Average time to learn is between 4-10 months. Curriculum and learning guide included.

- Top 40+ Machine Learning Courses For Beginners in 2025

The "Top 40+ Machine Learning Courses for Beginners to Learn in 2025" is a curated compilation of online courses tailored for individuals new to ML. These courses cover foundational concepts, practical applications, and hands-on projects, ensuring a comprehensive learning experience.

- 15 Best Machine Learning Courses in 2025 [Free + Paid] - Hackr

Whether a data professional or a beginner, we've found the 15 best machine learning courses online in 2025, including free and paid options for all skill levels.

- 13 Best Machine Learning Courses for 2025: Scikit-learn, TensorFlow, ...

Demystify the math behind Machine Learning and master popular machine learning libraries like TensorFlow and scikit-learn with these FREE and paid courses.

🔷 **Web Development Courses**

Top Online Web Development Courses for Beginners

To find the top online web development courses for beginners, I will search for the query "top online web development courses for beginners". <function=duckduckgo_search{"query": "top online web development courses for beginners"}</function>

🔷 **Cloud Computing Courses**

- Google to discount cloud computing services for US government, FT reports

Google will heavily discount cloud computing services for the U.S. government in a deal that could be finalised within weeks, the Financial Times reported on Friday, citing a senior official at the General Services Administration (GSA).

- Roundup Of Free Cloud ComputingOnline Courses - Forbes

Free Cloud Computing Courses The number and quality of free online cloud computing courses continues to grow, and lately the prices of fee-based online programs are dropping.

- Top 5 Online Cloud Computing Courses for Students under INR 20,000 - MSN

Cloud computing offers on-demand services like storage and software via the internet, providing cost-effective, scalable solutions without the need for physical infrastructure. With growing demand...

- Importance of Cloud Computing Courses | YourStory

There are several online cloud computing courses that are available for both students as well as working professionals. How does Cloud Computing Work?

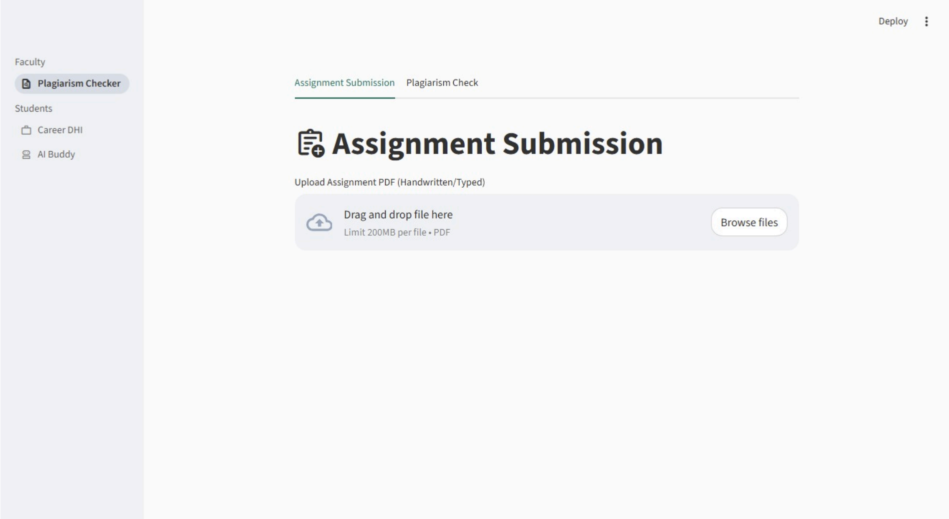
- 5 best cloud computing certification courses in the U.S.

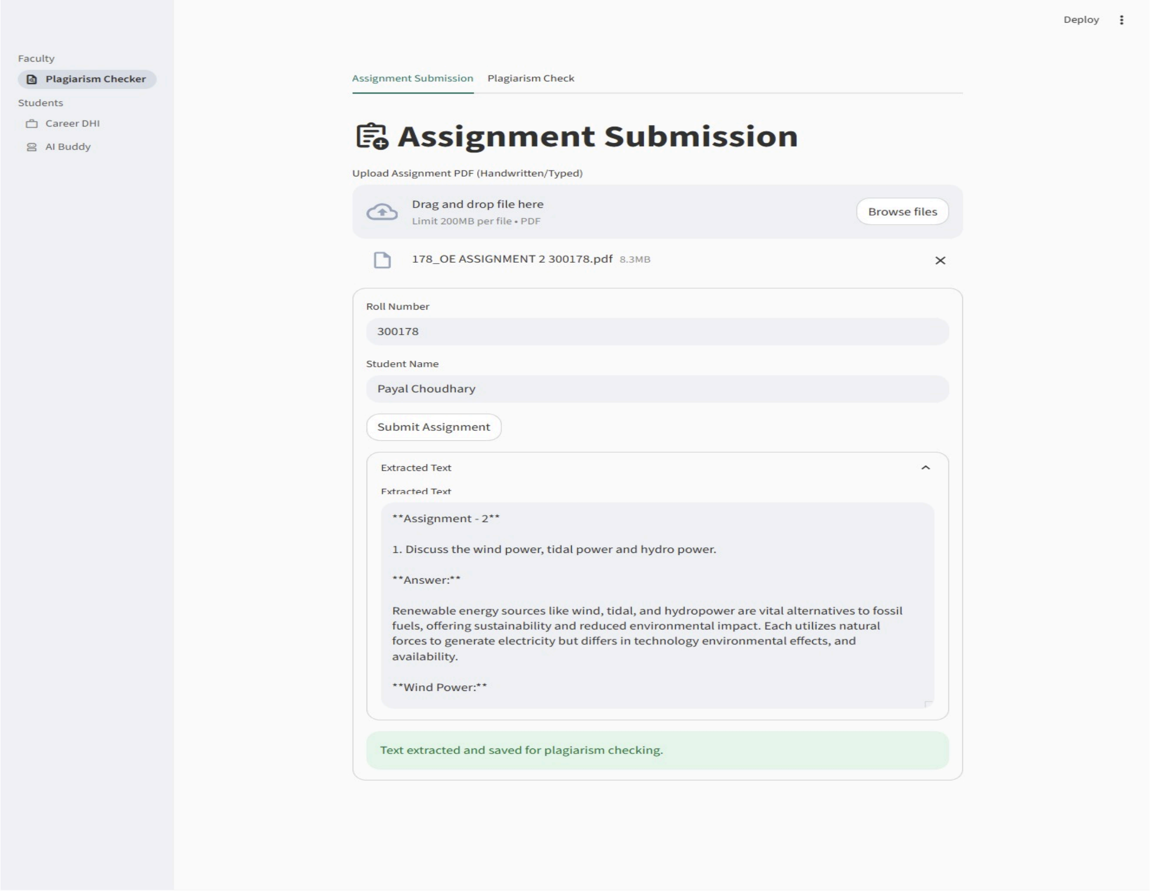
It's no surprise that the demand for qualified experts in cloud computing is rising — the first step is earning your certification.

5.2 Plagiarism Checker

- **PDF to Image Conversion:** Converts uploaded PDFs (typed or handwritten) into stitched images.
- **Vision LLM OCR:** Uses Groq vision API to extract text accurately from images.

- **Chunked OCR Processing:** Ensures stable OCR even for large image sizes by chunking inputs.
- **Sentence-Level Comparison:** Uses Jaccard similarity to detect sentence-level overlaps between submissions.
- **Database Storage:** Stores submissions, extracted text, and similarity results for future review.





178_OE ASSIGNMENT 2 300178.pdf 8.3MB

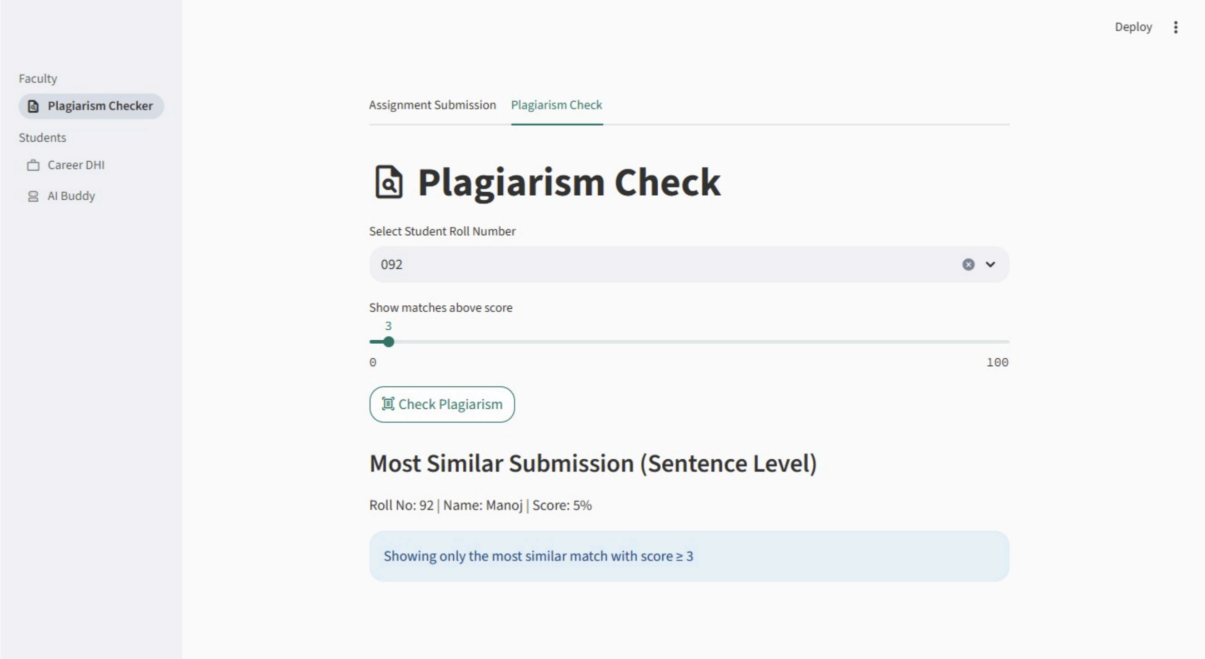
Roll Number: 300178

Student Name: Payal Choudhary

Extracted Text

```
**Assignment - 2**

1. Discuss the wind power, tidal power and hydro power.

**Answer:**

Renewable energy sources like wind, tidal, and hydropower are vital alternatives to fossil
fuels, offering sustainability and reduced environmental impact. Each utilizes natural
forces to generate electricity but differs in technology environmental effects, and
availability.

**Wind Power:**
```

Text extracted and saved for plagiarism checking.



Plagiarism Check

Select Student Roll Number

092

Show matches above score

3

0 100

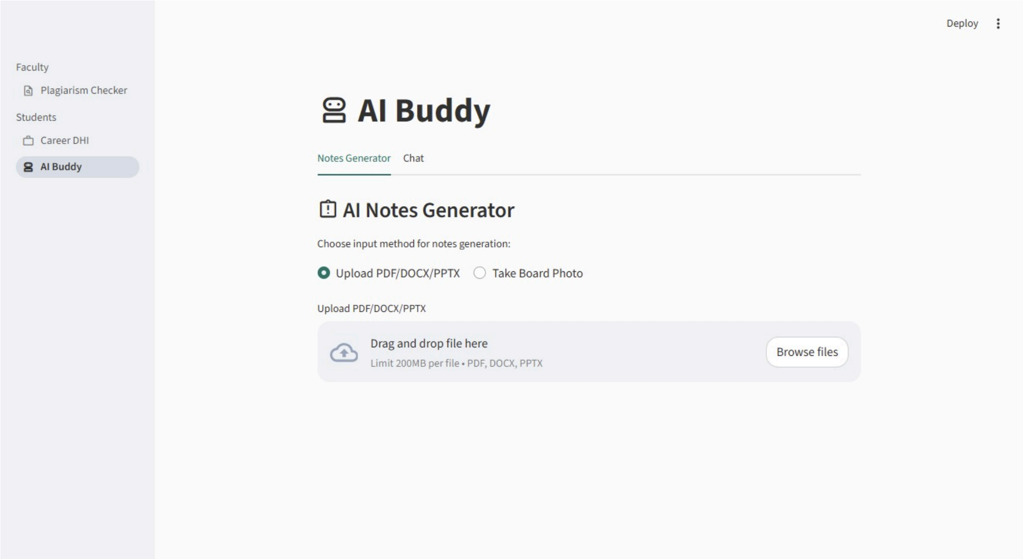
Most Similar Submission (Sentence Level)

Roll No: 92 | Name: Manoj | Score: 5%

Showing only the most similar match with score ≥ 3

5.3 AI Buddy (Smart Study Assistant)

- **Notes Generator:** Accepts PDFs, DOCX, and PPTX files and creates concise, high-quality notes.
- **Document Chatbot:** Users can ask questions from uploaded documents, answered contextually using LLMs.
- **Voice Integration:** Supports speech-to-text (input)
- **Recent Chat Feature:** Displays the latest queries at the top for continuous interaction.



AI Buddy

AI Notes Generator

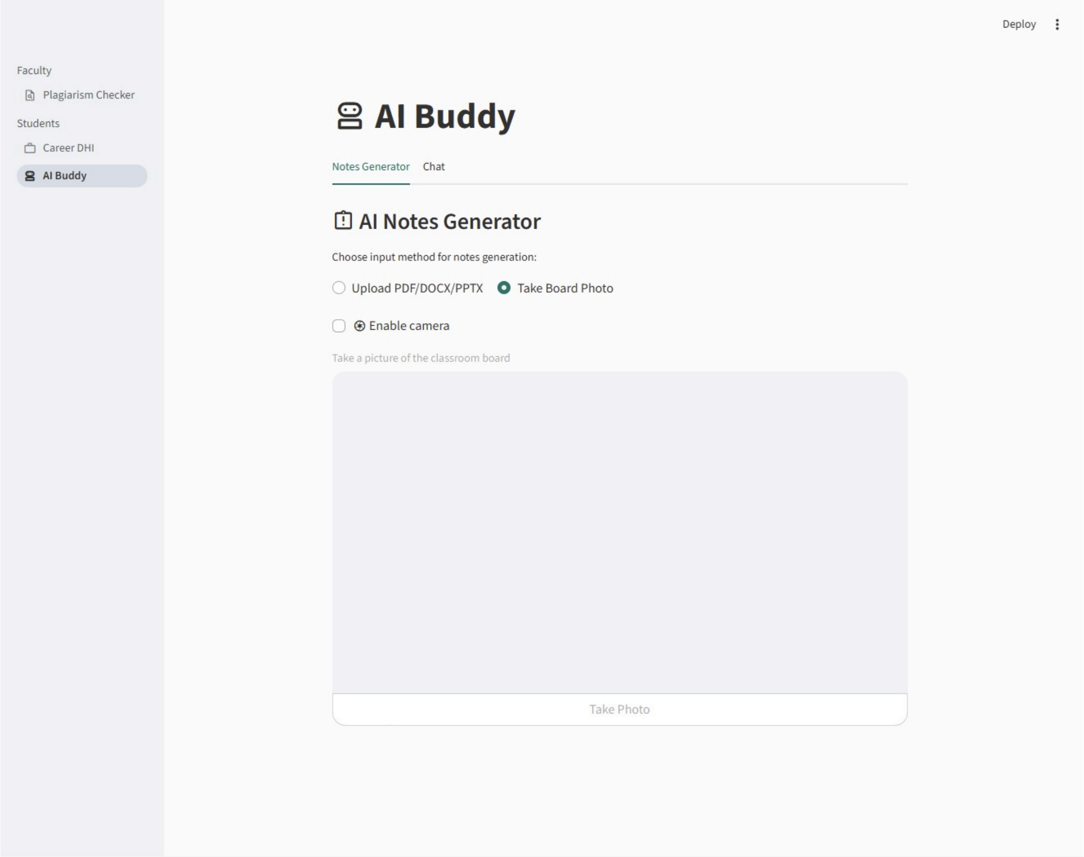
Choose input method for notes generation:

● Upload PDF/DOCX/PPTX ○ Take Board Photo

Upload PDF/DOCX/PPTX

Drag and drop file here

Limit 200MB per file • PDF, DOCX, PPTX



AI Buddy

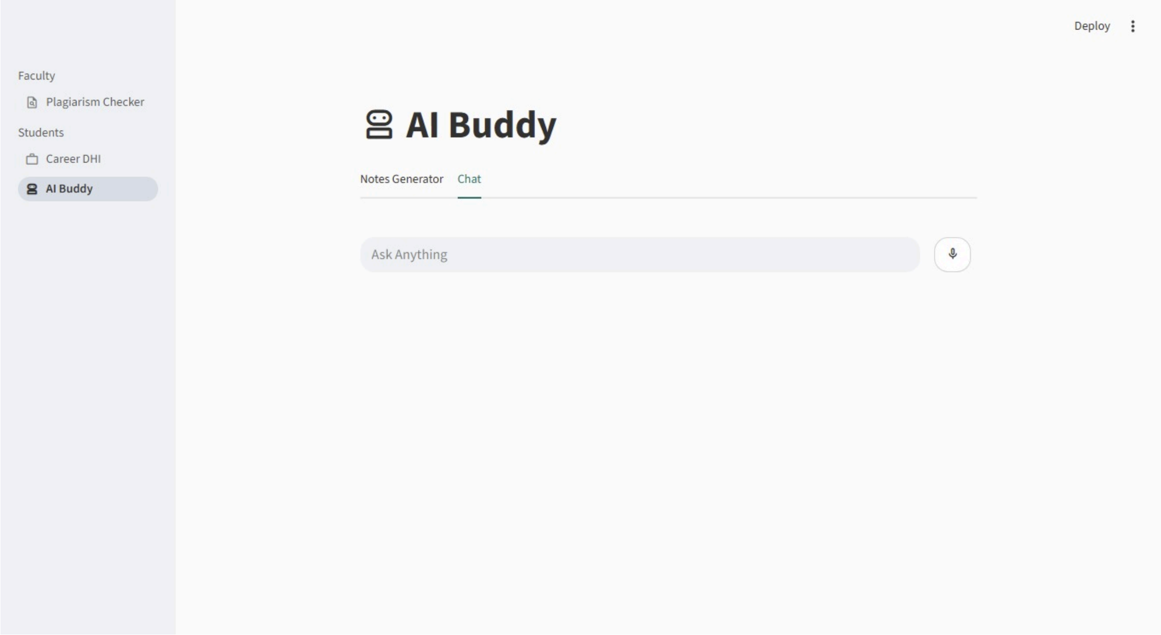
AI Notes Generator

Choose input method for notes generation:

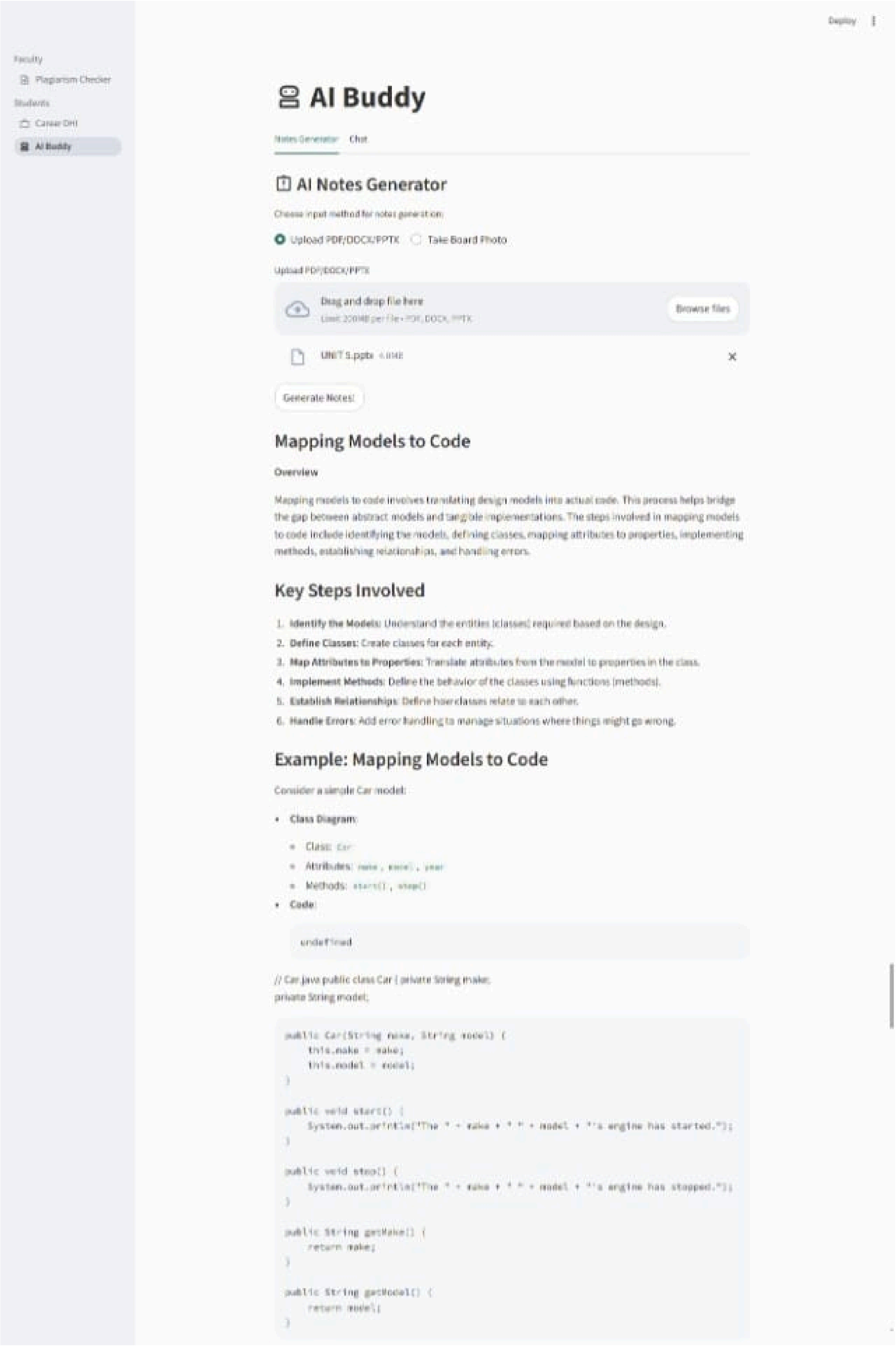
○ Upload PDF/DOCX/PPTX ● Take Board Photo

☐ Enable camera

Take a picture of the classroom board



AI Buddy



AI Buddy

AI Notes Generator

Choose input method for notes generation:

● Upload PDF/DOCX/PPTX ○ Take Board Photo

Upload PDF/DOCX/PPTX

Mapping Models to Code

Overview

Mapping models to code involves translating design models into actual code. This process helps bridge the gap between abstract models and tangible implementations. The steps involved in mapping models to code include identifying the models, defining classes, mapping attributes to properties, implementing methods, establishing relationships, and handling errors.

Key Steps Involved

1. **Identify the Models:** Understand the entities (classes) required based on the design.
2. **Define Classes:** Create classes for each entity.
3. **Map Attributes to Properties:** Translate attributes from the model to properties in the class.
4. **Implement Methods:** Define the behavior of the classes using functions (methods).
5. **Establish Relationships:** Define how classes relate to each other.
6. **Handle Errors:** Add error handling to manage situations where things might go wrong.

Example: Mapping Models to Code

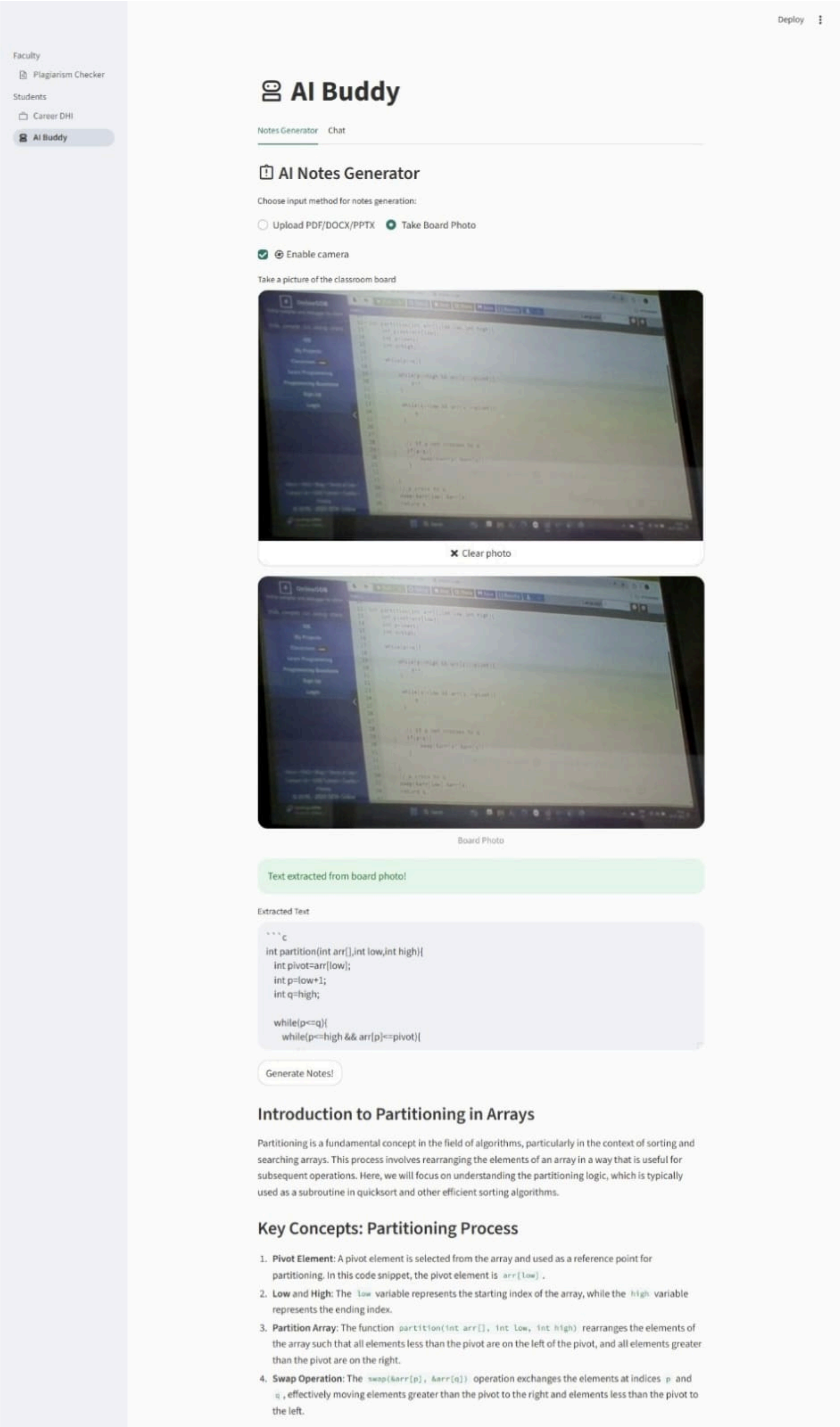
Consider a simple Car model:

- **Class Diagram:**
  - Class: `Car`
  - Attributes: `make` , `model` , `year`
  - Methods: `start()` , `stop()`
- **Code:**

```
undefined
```

// Car.java public class Car { private String make;
private String model;

```
public Car(String make, String model) {
    this.make = make;
    this.model = model;
}

public void start() {
    System.out.println("The " + make + " " + model + "'s engine has started.");
}

public void stop() {
    System.out.println("The " + make + " " + model + "'s engine has stopped.");
}

public String getMake() {
    return make;
}

public String getModel() {
    return model;
}
```



AI Buddy

AI Notes Generator

Choose input method for notes generation:

○ Upload PDF/DOCX/PPTX ● Take Board Photo

☑ Enable camera

Take a picture of the classroom board

Board Photo

Text extracted from board photo!

Extracted Text

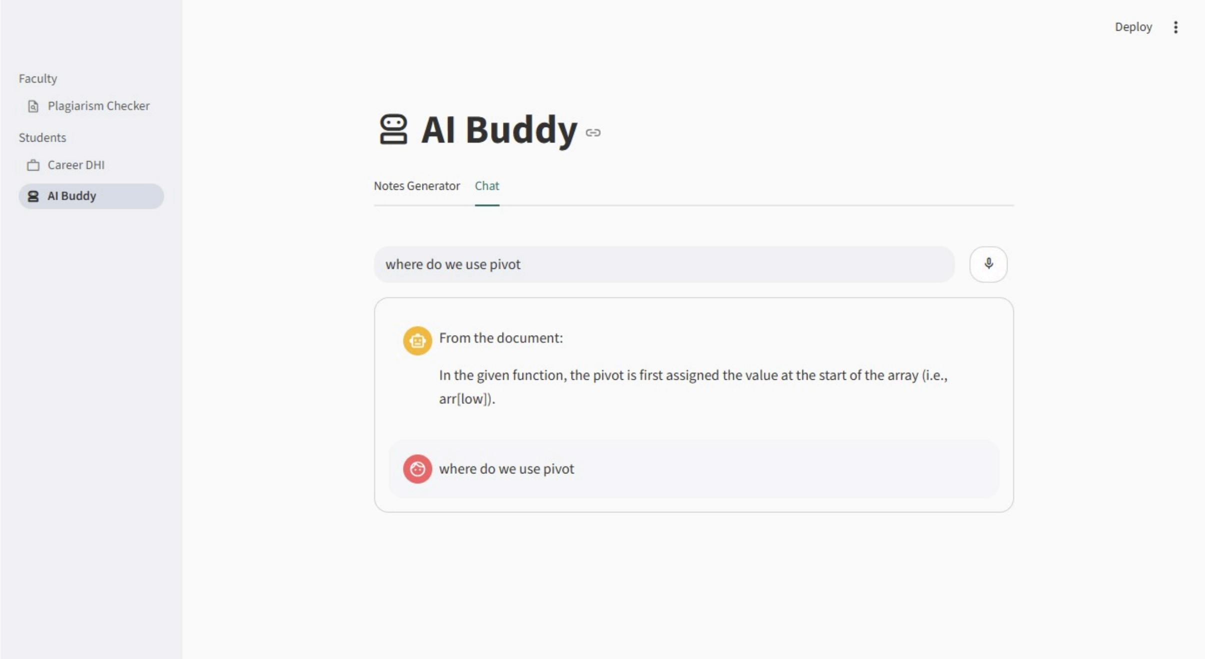
~~~
```c
int partition(int arr[],int low,int high){
  int pivot=arr[low];
  int p=low+1;
  int q=high;

  while(p<=q){
    while(p<=high && arr[p]<=pivot){
~~~

Introduction to Partitioning in Arrays

Partitioning is a fundamental concept in the field of algorithms, particularly in the context of sorting and searching arrays. This process involves rearranging the elements of an array in a way that is useful for subsequent operations. Here, we will focus on understanding the partitioning logic, which is typically used as a subroutine in quicksort and other efficient sorting algorithms.

Key Concepts: Partitioning Process

1. **Pivot Element**: A pivot element is selected from the array and used as a reference point for partitioning. In this code snippet, the pivot element is `arr[low]`.
2. **Low and High**: The `low` variable represents the starting index of the array, while the `high` variable represents the ending index.
3. **Partition Array**: The function `partition(int arr[], int low, int high)` rearranges the elements of the array such that all elements less than the pivot are on the left of the pivot, and all elements greater than the pivot are on the right.
4. **Swap Operation**: The `swap(&arr[p], &arr[q])` operation exchanges the elements at indices `p` and `q`, effectively moving elements greater than the pivot to the right and elements less than the pivot to the left.



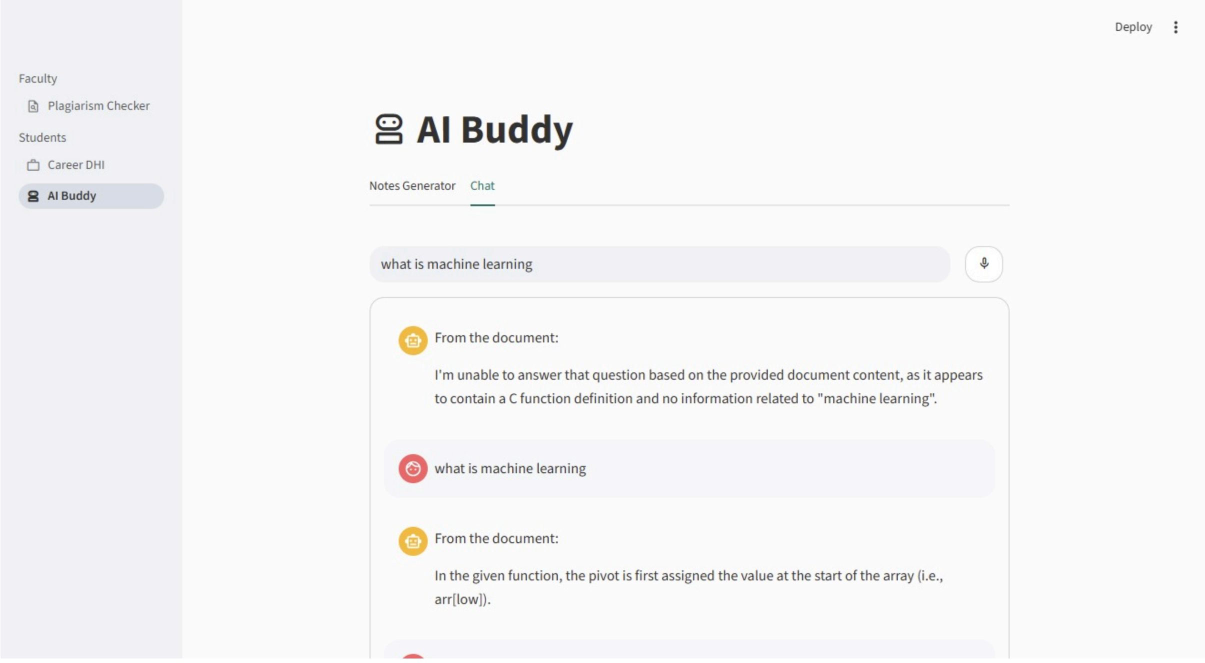
AI Buddy

where do we use pivot

From the document:

In the given function, the pivot is first assigned the value at the start of the array (i.e., arr[low]).

where do we use pivot



AI Buddy

what is machine learning

From the document:

I'm unable to answer that question based on the provided document content, as it appears to contain a C function definition and no information related to "machine learning".

what is machine learning

From the document:

In the given function, the pivot is first assigned the value at the start of the array (i.e., arr[low]).

6. Technology Stack

- **Frontend/UI:** Streamlit (with custom layout, tabs, and chat components)
- **AI Models:** Groq LLM (via Agno for text, Groq API for vision)
- **OCR & PDF Parsing:** PyMuPDF4LLM, pdf2image, PIL
- **Automation:** Selenium (grades extraction)
- **Similarity Detection:** Jaccard index for text overlap
- **Speech:** SpeechRecognition library
- **Data Storage:** Local JSON for submissions and results

- **Programming Language:** Python 3.13
- **Environment Security:** .env for storing API keys securely

---

5. Folder Structure

None

```
├── main.py                     # Main Streamlit entrypoint
├── views/
│   ├── page1.py                # Career DHI Module
│   ├── page2.py                # Plagiarism Checker
│   └── page3.py                # AI Buddy
├── utils/
│   ├── cdhi/                   # Career module helpers
│   └── plag/                   # Plagiarism processing logic
├── uploads/                    # Uploaded assignments
├── stitched/                     # Converted images
├── vision_text_db.json         # JSON database for extracted text storage
├── .env                        # Secure API keys storage
├── .gitignore                  # Ignored files
└── requirements.txt            # Python package dependencies
```

---

6. Security & Privacy

- All API keys are hidden via .env and never hardcoded.
- All data (submissions, reports) are stored locally and not shared with third-party services.
- Only authorized users (faculty/students) can access sensitive features.

---

7. How to Run the Application

1. Clone the repository:

Shell

```
git clone https://github.com/kayozxo/CASP.git
cd CASP
```

2. Install dependencies:

Shell

```shell
pip install -r requirements.txt
```

3. Set up .env file:

Shell

```shell
GROQ_API="your-groq-api-key"
GROQ_PLAG_API="your-groq-plag-key"
```

4. Run the application:

Shell

```shell
streamlit run main.py
```

---

8. Conclusion

CASP serves as a unified and secure AI platform for students and educators. It brings together intelligent career counselling, advanced plagiarism detection, and personalised study assistance, all under one lightweight and interactive Streamlit application.