Lab 5.4: Ethical Foundations – Responsible AI Coding Practices

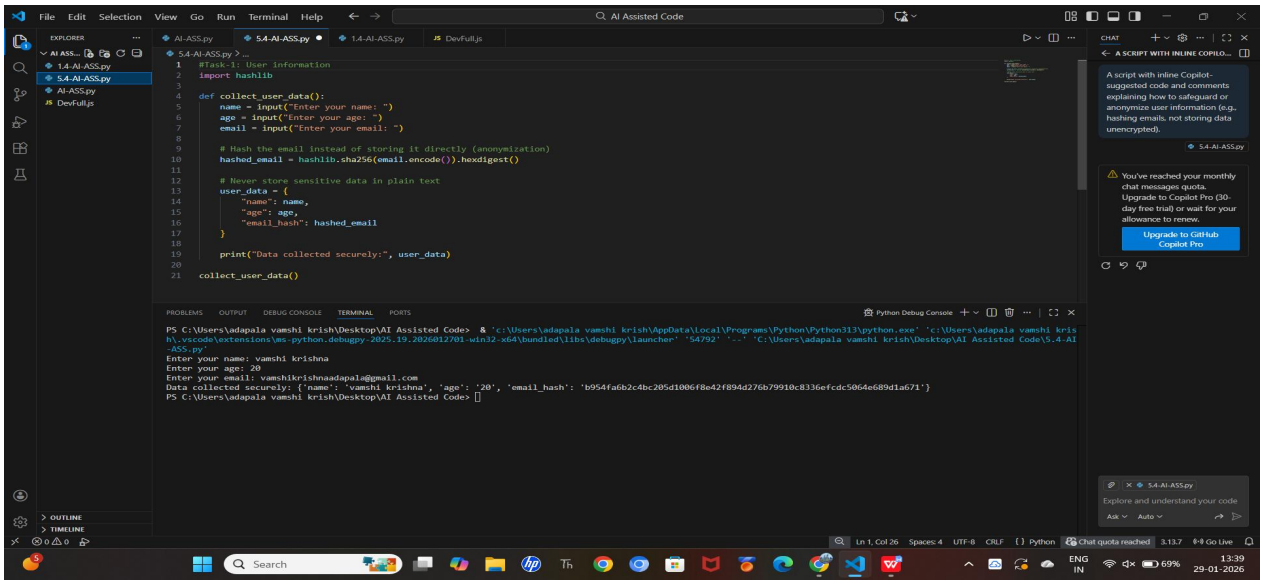
2303A52249
A.VAMSHI KRISHNA
AI-ASS

Task 1: Secure User Data Collection

Prompt: "Generate a Python script that collects user data (name, age, email) and add comments on how to anonymize or protect this data."

Generate a Python script that collects user data (e.g., name, age, email). Then, ask Copilot to add comments on how to anonymize or protect this data.
• A script with inline Copilot-suggested code and comments explaining how to safeguard or anonymize user information (e.g., hashing emails, not storing data unencrypted).

Code:



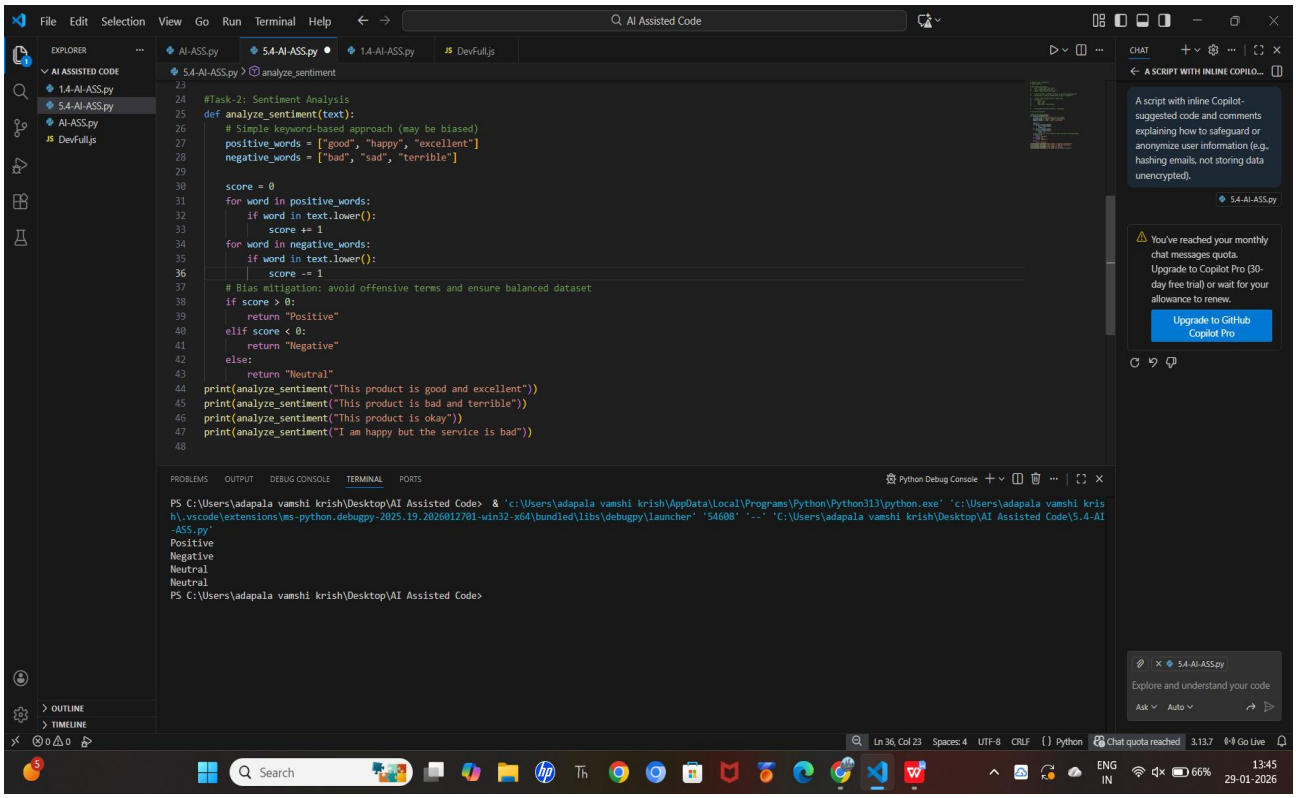
Explanation:

The email is hashed before storage. Sensitive data is not stored in plain text.

Task 2: Sentiment Analysis with Bias Handling

Prompt: "Generate a Python function for sentiment analysis and identify potential biases." Generate a Python function for sentiment analysis. Then prompt Copilot to identify and handle potential biases in the data.
• Copilot-generated code with additions or comments addressing bias mitigation strategies (e.g., balancing dataset, removing offensive terms).

Code:

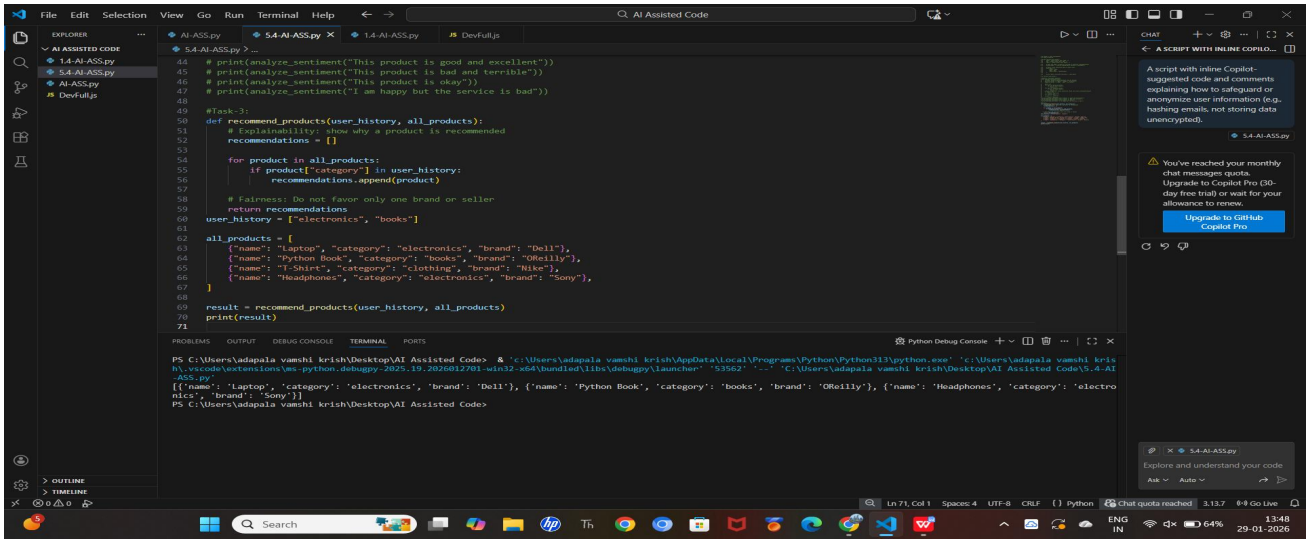


Explanation: The comments mention balancing datasets and avoiding offensive or biased words.

Task 3: Ethical Product Recommendation System

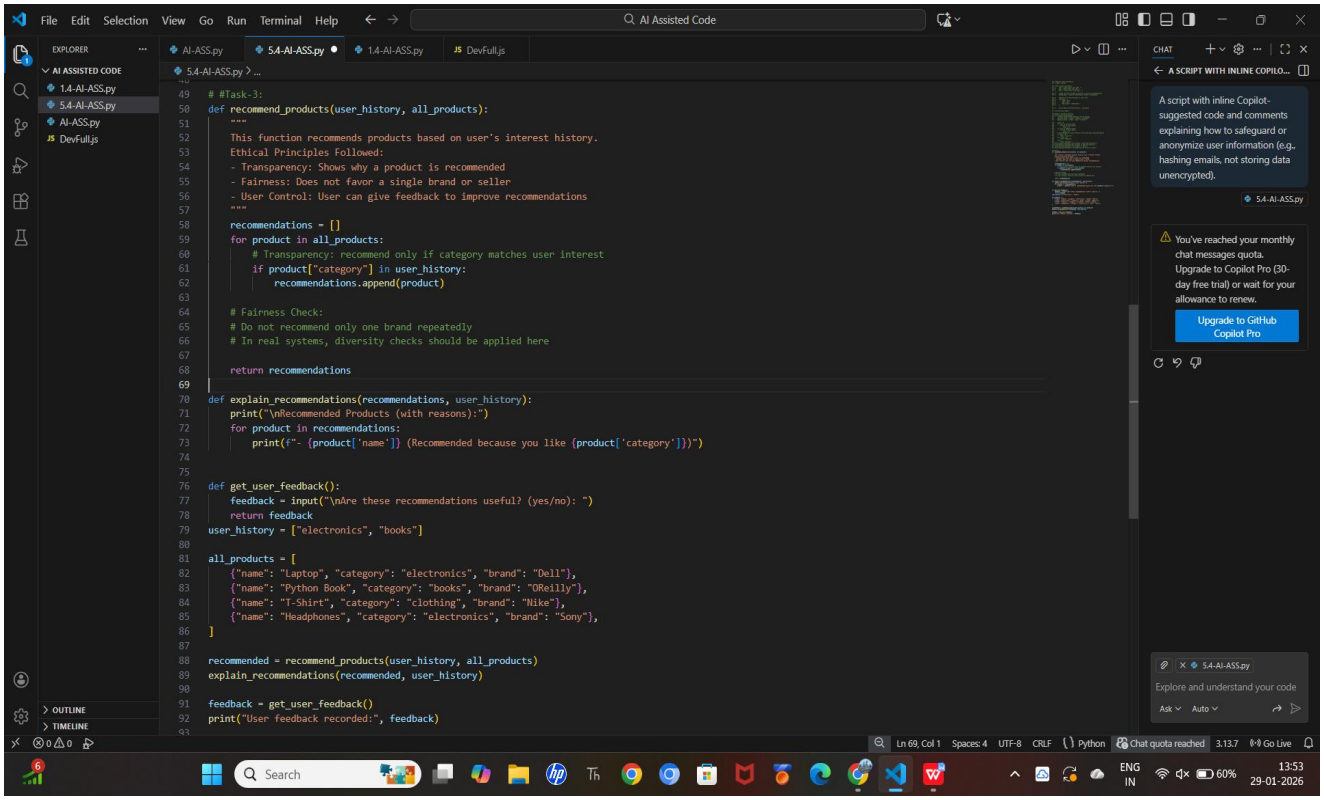
Prompt: "Write a Python program that recommends products ethically."
write a Python program that recommends products based on user history.
Ask it to follow ethical guidelines like transparency and fairness.

Code:

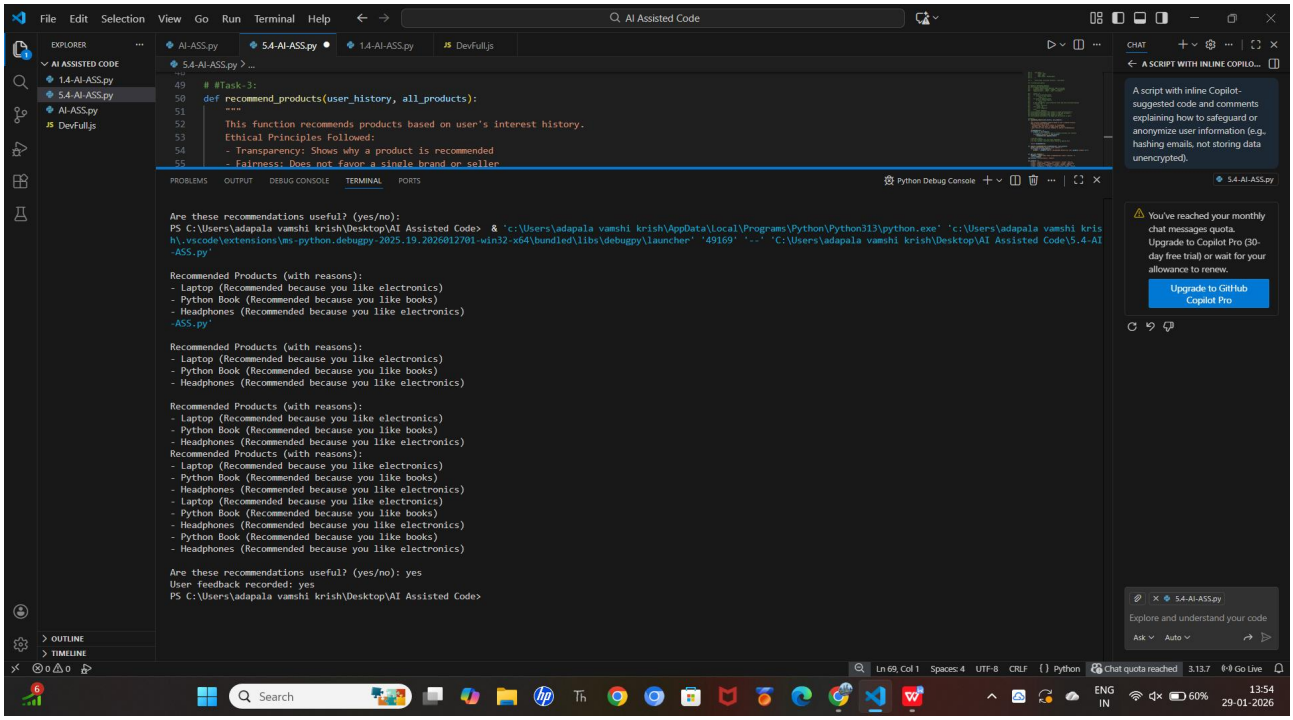


```
# print(analyze_sentiment("This product is good and excellent"))
# print(analyze_sentiment("This product is bad and terrible"))
# print(analyze_sentiment("This product is okay"))
# print(analyze_sentiment("I am happy but the service is bad"))

#Task-3:
def recommend_products(user_history, all_products):
    # Explainability: show why a product is recommended
    recommendations = []

    for product in all_products:
        if product["category"] in user_history:
            recommendations.append(product)

    # Fairness: Do not favor only one brand or seller
    return recommendations
user_history = ["electronics", "books"]

all_products = [
    {"name": "Laptop", "category": "electronics", "brand": "Dell"},
    {"name": "Python Book", "category": "books", "brand": "OReilly"},
    {"name": "T-Shirt", "category": "clothing", "brand": "Nike"},
    {"name": "Headphones", "category": "electronics", "brand": "Sony"},
]

result = recommend_products(user_history, all_products)
print(result)
```

```
PS C:\Users\adapala vamshi krish\Desktop\AI Assisted Code> & 'c:\Users\adapala vamshi krish\AppData\Local\Programs\Python\Python313\python.exe' 'c:\Users\adapala vamshi krish\.vscode\extensions\ms-python.debugpy-2025.19.2026012701-win32-x64\bundled\libs\debugpy\launcher' '53562' '--' 'C:\Users\adapala vamshi krish\Desktop\AI Assisted Code\5.4-AI-ASS.py'
[{'name': 'Laptop', 'category': 'electronics', 'brand': 'Dell'}, {'name': 'Python Book', 'category': 'books', 'brand': 'OReilly'}, {'name': 'Headphones', 'category': 'electronics', 'brand': 'Sony'}]
PS C:\Users\adapala vamshi krish\Desktop\AI Assisted Code>
```



```
# #Task-3:
def recommend_products(user_history, all_products):
    """
    This function recommends products based on user's interest history.
    Ethical Principles Followed:
    - Transparency: Shows why a product is recommended
    - Fairness: Does not favor a single brand or seller
    - User Control: User can give feedback to improve recommendations
    """
    recommendations = []
    for product in all_products:
        # Transparency: recommend only if category matches user interest
        if product["category"] in user_history:
            recommendations.append(product)

    # Fairness Check:
    # Do not recommend only one brand repeatedly
    # In real systems, diversity checks should be applied here

    return recommendations

def explain_recommendations(recommendations, user_history):
    print("\nRecommended Products (with reasons):")
    for product in recommendations:
        print(f"- {product['name']} (Recommended because you like {product['category']})")


def get_user_feedback():
    feedback = input("\nAre these recommendations useful? (yes/no): ")
    return feedback
user_history = ["electronics", "books"]

all_products = [
    {"name": "Laptop", "category": "electronics", "brand": "Dell"},
    {"name": "Python Book", "category": "books", "brand": "OReilly"},
    {"name": "T-Shirt", "category": "clothing", "brand": "Nike"},
    {"name": "Headphones", "category": "electronics", "brand": "Sony"},
]

recommended = recommend_products(user_history, all_products)
explain_recommendations(recommended, user_history)

feedback = get_user_feedback()
print("User feedback recorded:", feedback)
```



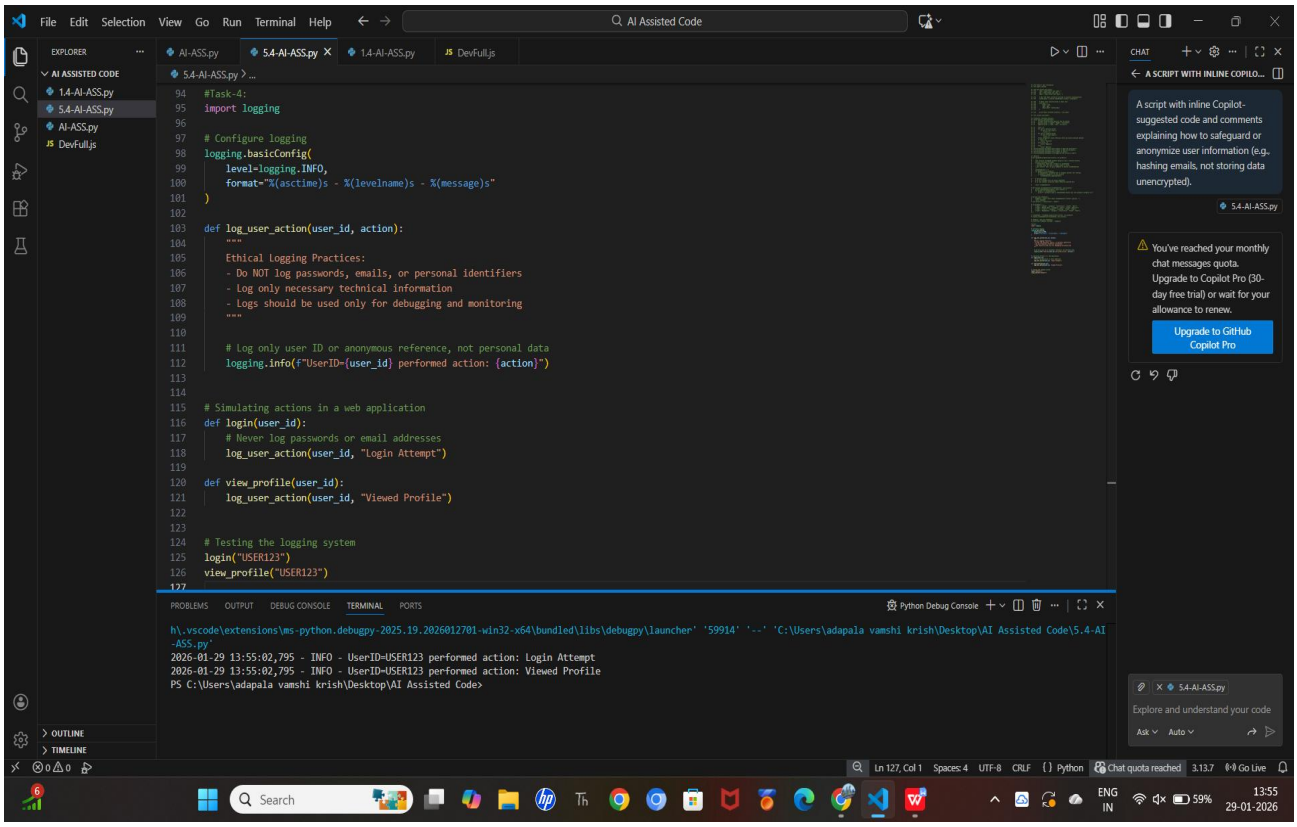
Explanation: The system avoids favoritism and is transparent about recommendations.

Task 4: Ethical Logging

Prompt: "Generate logging functionality that does not log sensitive data."
Generate logging functionality in a Python web application. Then, ask it to ensure the logs do not record sensitive information.

• Logging code that avoids saving personal identifiers (e.g., passwords, emails), and includes comments about ethical logging practices.

Code:



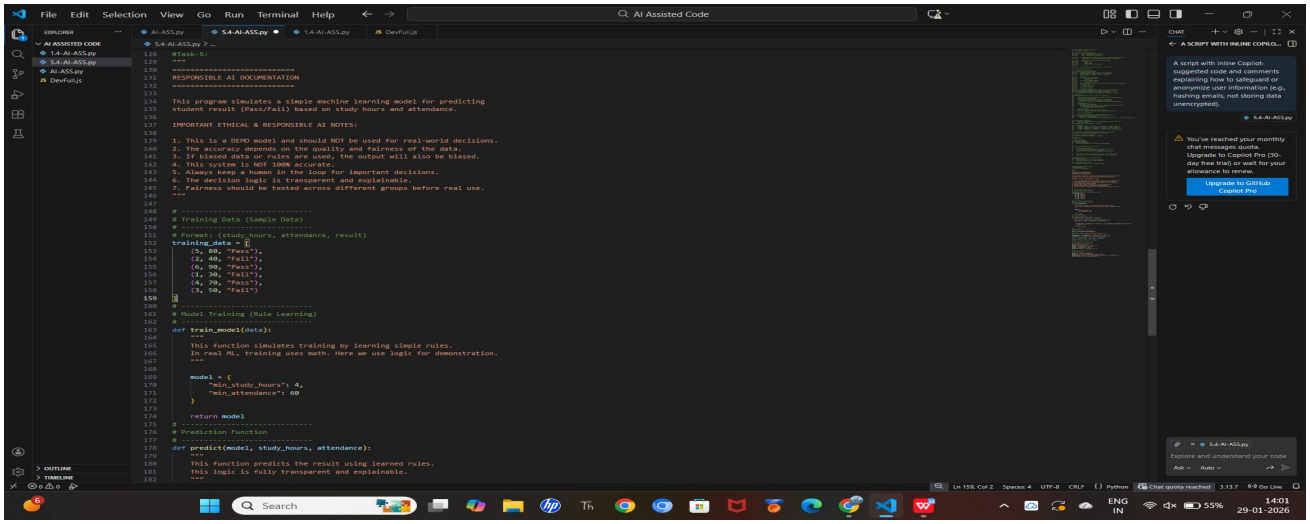
Explanation: Sensitive information is excluded from logs.

Task 5: Machine Learning Model with Responsible Documentation

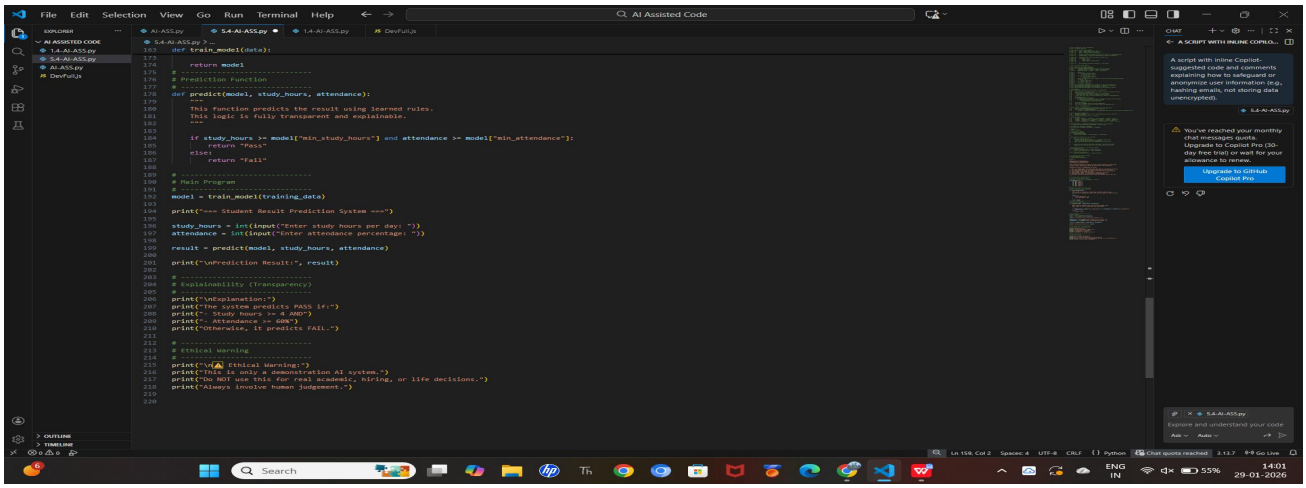
Prompt: "Generate a ML model and add documentation for responsible use."
Generate a machine learning model. Then, prompt it to add documentation on how to use the model responsibly (e.g., explainability, accuracy limits).

- Copilot-generated model code with a README or inline documentation suggesting responsible usage, limitations, and fairness considerations.

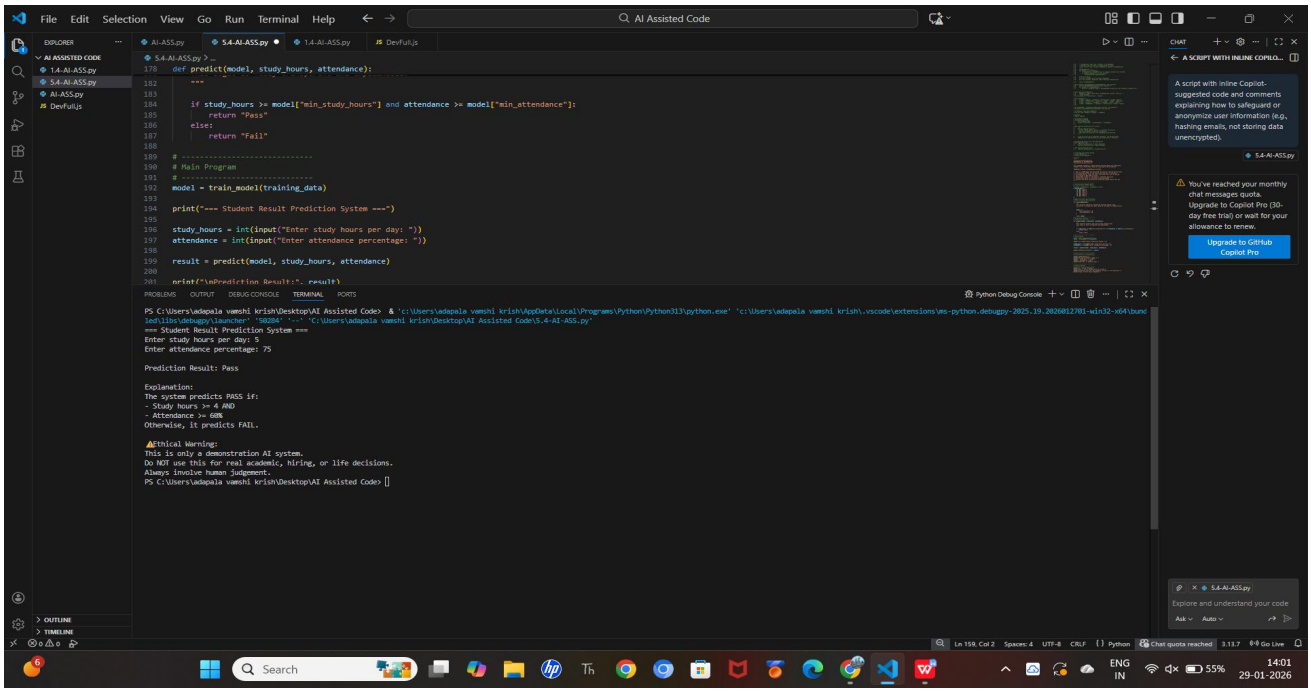
Code:



```
#Task-5:
"""
--------------------------
RESPONSIBLE AI DOCUMENTATION
--------------------------

This program simulates a simple machine learning model for predicting
student result (Pass/Fail) based on study hours and attendance.

IMPORTANT ETHICAL & RESPONSIBLE AI NOTES:

1. This is a DEMO model and should NOT be used for real-world decisions.
2. The accuracy depends on the quality and fairness of the data.
3. If biased data or rules are used, the output will also be biased.
4. This system is NOT 100% accurate.
5. Always keep a human in the loop for important decisions.
6. The decision logic is transparent and explainable.
7. Fairness should be tested across different groups before real use.
"""

# ----------------------------
# Training Data (Sample Data)
# ----------------------------
# Format: (study_hours, attendance, result)
training_data = [
    (5, 80, "Pass"),
    (2, 40, "Fail"),
    (6, 90, "Pass"),
    (1, 30, "Fail"),
    (4, 70, "Pass"),
    (3, 50, "Fail")
]

# ----------------------------
# Model Training (Rule Learning)
# ----------------------------
def train_model(data):
    """
    This function simulates training by learning simple rules.
    In real ML, training uses math. Here we use logic for demonstration.
    """

    model = {
        "min_study_hours": 4,
        "min_attendance": 60
    }

    return model
# ----------------------------
# Prediction Function
# ----------------------------
def predict(model, study_hours, attendance):
    """
    This function predicts the result using learned rules.
    This logic is fully transparent and explainable.
    """
```



```
def train_model(data):
    return model
# ----------------------------
# Prediction Function
# ----------------------------
def predict(model, study_hours, attendance):
    """
    This function predicts the result using learned rules.
    This logic is fully transparent and explainable.
    """

    if study_hours >= model["min_study_hours"] and attendance >= model["min_attendance"]:
        return "Pass"
    else:
        return "Fail"

# ----------------------------
# Main Program
# ----------------------------
model = train_model(training_data)

print("=== Student Result Prediction System ===")

study_hours = int(input("Enter study hours per day: "))
attendance = int(input("Enter attendance percentage: "))

result = predict(model, study_hours, attendance)

print("\nPrediction Result:", result)

# ----------------------------
# Explainability (Transparency)
# ----------------------------
print("\nExplanation:")
print("The system predicts PASS if:")
print("- Study hours >= 4 AND")
print("- Attendance >= 60%")
print("Otherwise, it predicts FAIL.")

# ----------------------------
# Ethical Warning
# ----------------------------
print("\n⚠️ Ethical Warning:")
print("This is only a demonstration AI system.")
print("Do NOT use this for real academic, hiring, or life decisions.")
print("Always involve human judgement.")
```



```
def predict(model, study_hours, attendance):
    """

    if study_hours >= model["min_study_hours"] and attendance >= model["min_attendance"]:
        return "Pass"
    else:
        return "Fail"

# ----------------------------
# Main Program
# ----------------------------
model = train_model(training_data)

print("=== Student Result Prediction System ===")

study_hours = int(input("Enter study hours per day: "))
attendance = int(input("Enter attendance percentage: "))

result = predict(model, study_hours, attendance)

print("\nPrediction Result:", result)
```

```
PS C:\Users\adapala vamshi krish\Desktop\AI Assisted Code> & 'c:\Users\adapala vamshi krish\AppData\Local\Programs\Python\Python313\python.exe' 'c:\Users\adapala vamshi krish\.vscode\extensions\ms-python.debugpy-2025.19.2026012701-win32-x64\bundled\libs\debugpy\launcher' '58204' '--' 'C:\Users\adapala vamshi krish\Desktop\AI Assisted Code\5.4-AI-ASS.py'
=== Student Result Prediction System ===
Enter study hours per day: 5
Enter attendance percentage: 75

Prediction Result: Pass

Explanation:
The system predicts PASS if:
- Study hours >= 4 AND
- Attendance >= 60%
Otherwise, it predicts FAIL.

⚠️Ethical Warning:
This is only a demonstration AI system.
Do NOT use this for real academic, hiring, or life decisions.
Always involve human judgement.
PS C:\Users\adapala vamshi krish\Desktop\AI Assisted Code>
```

Explanation: The documentation explains limitations, fairness, and responsible usage.

Conclusion:

This lab demonstrates how developers must remain responsible, transparent, and ethical while using AI-generated code.