

## Lab 5.4: Ethical Foundations— Responsible AI Coding Practices

2303A52206

CHENCHU RAJU

### AI Assisted Coding

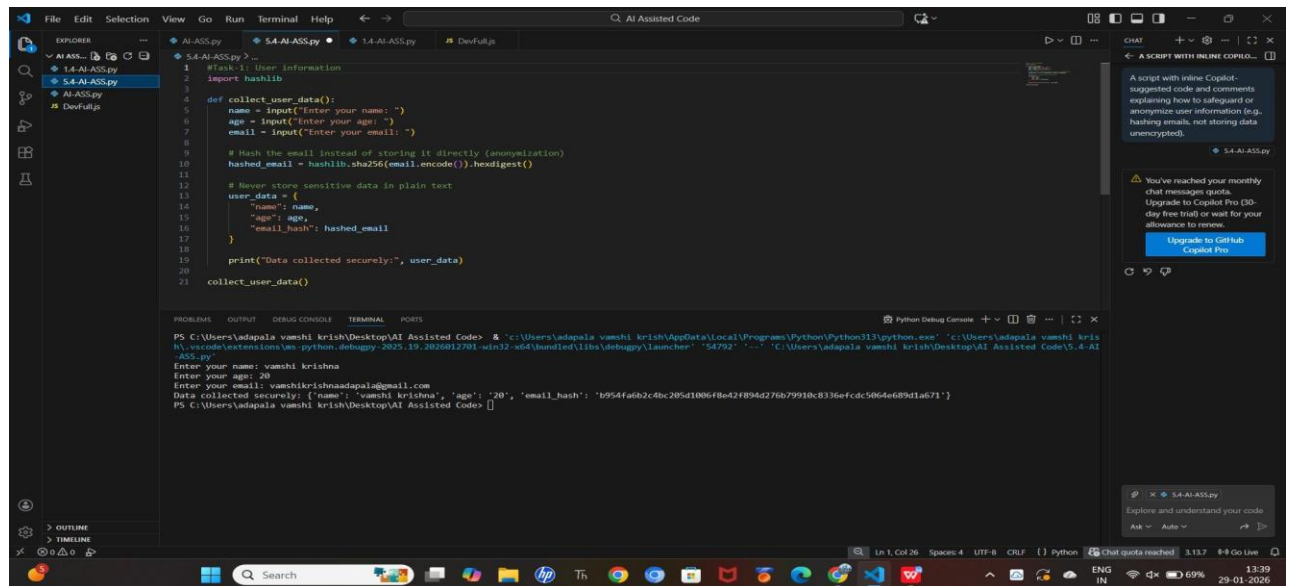
#### Task 1: Secure User Data Collection

**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:



```
1 #Task-1: User Information
2 import hashlib
3
4 def collect_user_data():
5     name = input("Enter your name: ")
6     age = input("Enter your age: ")
7     email = input("Enter your email: ")
8
9     # Hash the email instead of storing it directly (anonymization)
10    hashed_email = hashlib.sha256(email.encode()).hexdigest()
11
12    # Never store sensitive data in plain text
13    user_data = {
14        "name": name,
15        "age": age,
16        "email_hash": hashed_email
17    }
18
19    print("Data collected securely:", user_data)
20
21    collect_user_data()
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

```
PS C:\Users\adapala vamsi krish\Desktop\AI Assisted Code> & "c:\Users\adapala vamsi krish\AppData\Local\Programs\Python\Python313\python.exe" "c:\Users\adapala vamsi krish\AppData\Local\Programs\Python\Python313\python.exe" "c:\Users\adapala vamsi krish\Desktop\AI Assisted Code\5.4 AI Assisted Code.py"
Enter your name: vamsi krishna
Enter your age: 20
Enter your email: vamsikrishnaadapala@gmail.com
Data collected securely: {'name': 'vamsi krishna', 'age': '20', 'email_hash': 'b954fab2c4bc205d1006f8e42f894d276b79910c81336efcd5064e608d1a671'}
PS C:\Users\adapala vamsi krish\Desktop\AI Assisted 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:

The image shows a Windows desktop with a VS Code editor open. The editor has a dark theme and displays a Python script named `analyze_sentiment.py` in the `5.4-AI-ASS.py` file. The script defines a function `analyze_sentiment(text)` that uses a simple keyword-based approach to classify sentiment. It includes lists for `positive_words` and `negative_words`, a `score` variable, and a `print` statement for each input. The terminal window at the bottom shows the command `python analyze_sentiment.py` being executed, and the output displays the sentiment classification for each input string. A chat window on the right side of the editor shows a message from the AI assistant, suggesting the user upgrade to Copilot Pro. The Windows taskbar at the bottom shows various application icons and the system clock.

```
#Task-2: Sentiment Analysis
def analyze_sentiment(text):
    # Simple keyword-based approach (may be biased)
    positive_words = ["good", "happy", "excellent"]
    negative_words = ["bad", "sad", "terrible"]

    score = 0
    for word in positive_words:
        if word in text.lower():
            score += 1
    for word in negative_words:
        if word in text.lower():
            score -= 1

    # Bias mitigation: avoid offensive terms and ensure balanced dataset
    if score > 0:
        return "Positive"
    elif score < 0:
        return "Negative"
    else:
        return "Neutral"

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"))
```

PS C:\Users\adapala\ Desktop\AI Assisted Code> python analyze\_sentiment.py

Positive  
Negative  
Neutral  
Neutral

PS C:\Users\adapala\ Desktop\AI Assisted Code>

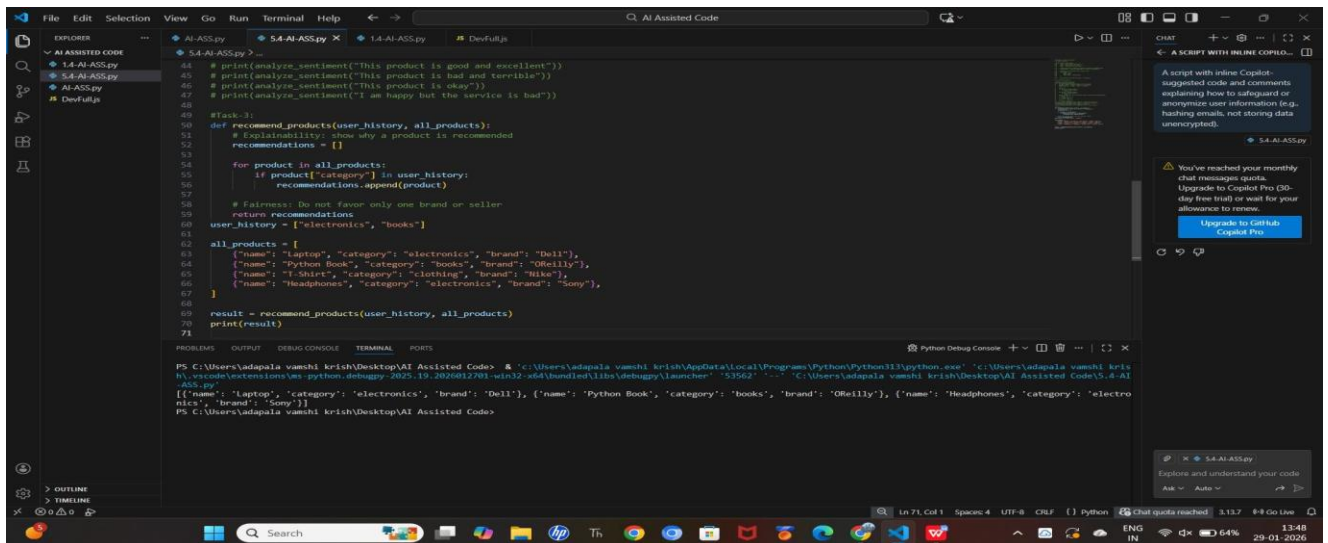
**Explanation:** The comments mention balancing datasets and avoiding offensive or biased words.

## Task 3: Ethical Product Recommendation System

**Prompt:** "Write a Python program that recommends products ethically."

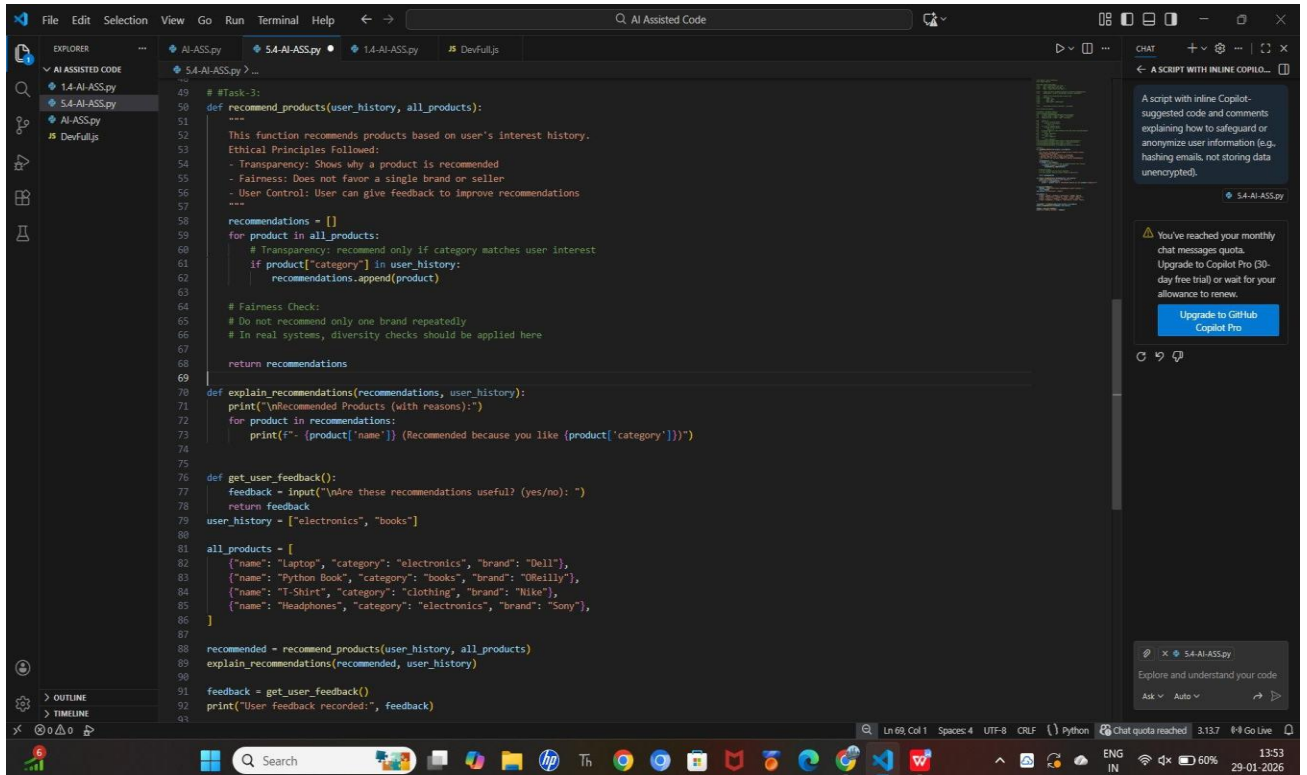
write a Python program that recommends products based on user history. Ask it to follow ethical guidelines like transparency and fairness.

**Code:**



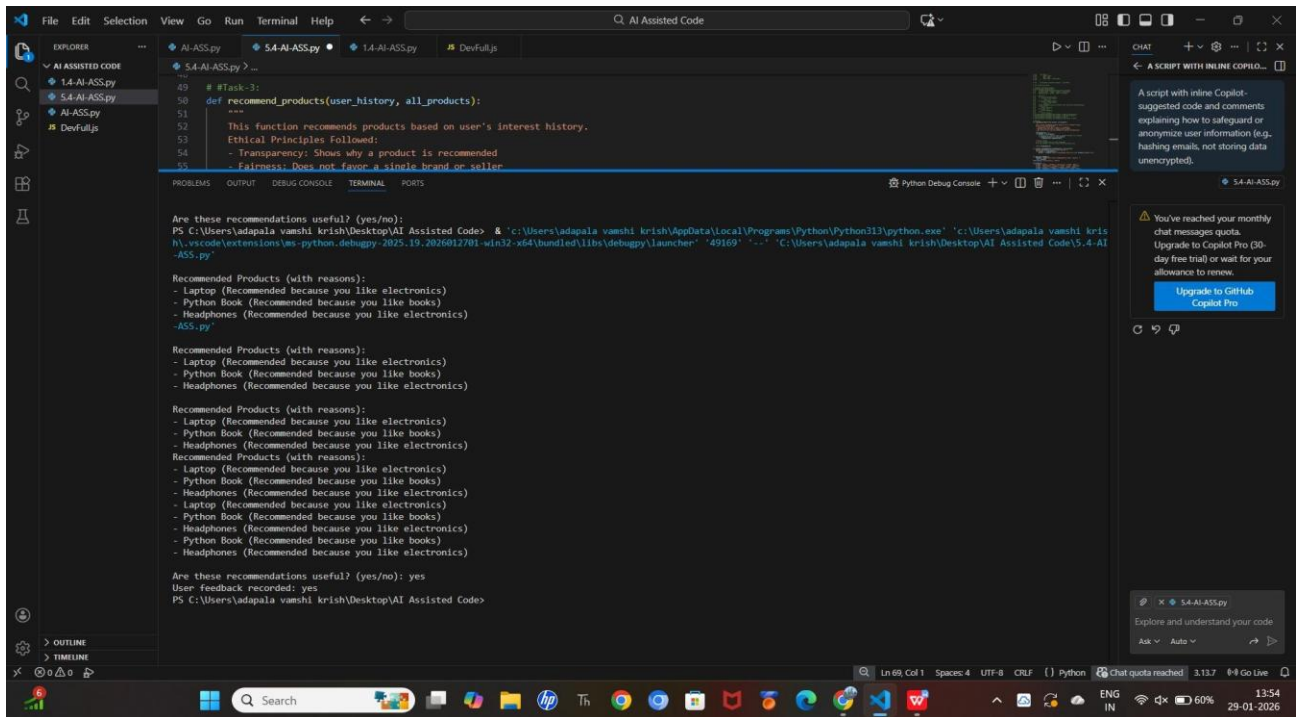
The screenshot shows a VS Code editor with a Python file named `5.4-AI-ASSpy`. The code implements a function `recommend_products` that takes `user_history` and `all_products` as input. It filters products based on the user's history and returns a list of recommendations. The `all_products` list includes items like 'Laptop', 'Python Book', 'T-Shirt', and 'Headphones'. The terminal output shows the execution of the script, displaying the list of recommended products: 'Laptop', 'Python Book', and 'Headphones'.

```
44 # print(analyze_sentiment("This product is good and excellent"))
45 # print(analyze_sentiment("This product is bad and terrible"))
46 # print(analyze_sentiment("This product is okay"))
47 # print(analyze_sentiment("I am happy but the service is bad"))
48
49 #Task-3:
50 def recommend_products(user_history, all_products):
51     # Explainability: show why a product is recommended
52     recommendations = []
53
54     for product in all_products:
55         if product["category"] in user_history:
56             recommendations.append(product)
57
58     # Fairness: Do not favor only one brand or seller
59     return recommendations
60
61 user_history = ["electronics", "books"]
62
63 all_products = [
64     {"name": "Laptop", "category": "electronics", "brand": "Dell"},
65     {"name": "Python Book", "category": "books", "brand": "O'Reilly"},
66     {"name": "T-Shirt", "category": "clothing", "brand": "Nike"},
67     {"name": "Headphones", "category": "electronics", "brand": "Sony"},
68 ]
69
70 result = recommend_products(user_history, all_products)
71 print(result)
```



The screenshot shows a VS Code editor with a Python file named `5.4-AI-ASSpy`. The code implements a more complex recommendation system. It includes a function `recommend_products` that filters products based on user history and ethical principles (Transparency, Fairness, User Control). It also includes a function `explain_recommendations` that prints the recommended products with reasons. The `all_products` list is the same as in the first screenshot. The terminal output shows the execution of the script, displaying the list of recommended products and the feedback recorded.

```
49 #Task-3:
50 def recommend_products(user_history, all_products):
51     """
52     This function recommends products based on user's interest history.
53     Ethical Principles Followed:
54     - Transparency: Shows why a product is recommended
55     - Fairness: Does not favor a single brand or seller
56     - User Control: User can give feedback to improve recommendations
57     """
58     recommendations = []
59     for product in all_products:
60         # Transparency: recommend only if category matches user interest
61         if product["category"] in user_history:
62             recommendations.append(product)
63
64     # Fairness Check:
65     # Do not recommend only one brand repeatedly
66     # In real systems, diversity checks should be applied here
67
68     return recommendations
69
70 def explain_recommendations(recommendations, user_history):
71     print("\nRecommended Products (with reasons):")
72     for product in recommendations:
73         print(f"- {product['name']} (Recommended because you like {product['category']})")
74
75
76 def get_user_feedback():
77     feedback = input("\nAre these recommendations useful? (yes/no): ")
78     return feedback
79
80 user_history = ["electronics", "books"]
81
82 all_products = [
83     {"name": "Laptop", "category": "electronics", "brand": "Dell"},
84     {"name": "Python Book", "category": "books", "brand": "O'Reilly"},
85     {"name": "T-Shirt", "category": "clothing", "brand": "Nike"},
86     {"name": "Headphones", "category": "electronics", "brand": "Sony"},
87 ]
88
89 recommended = recommend_products(user_history, all_products)
90 explain_recommendations(recommended, user_history)
91
92 feedback = get_user_feedback()
93 print("User feedback recorded:", feedback)
```



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

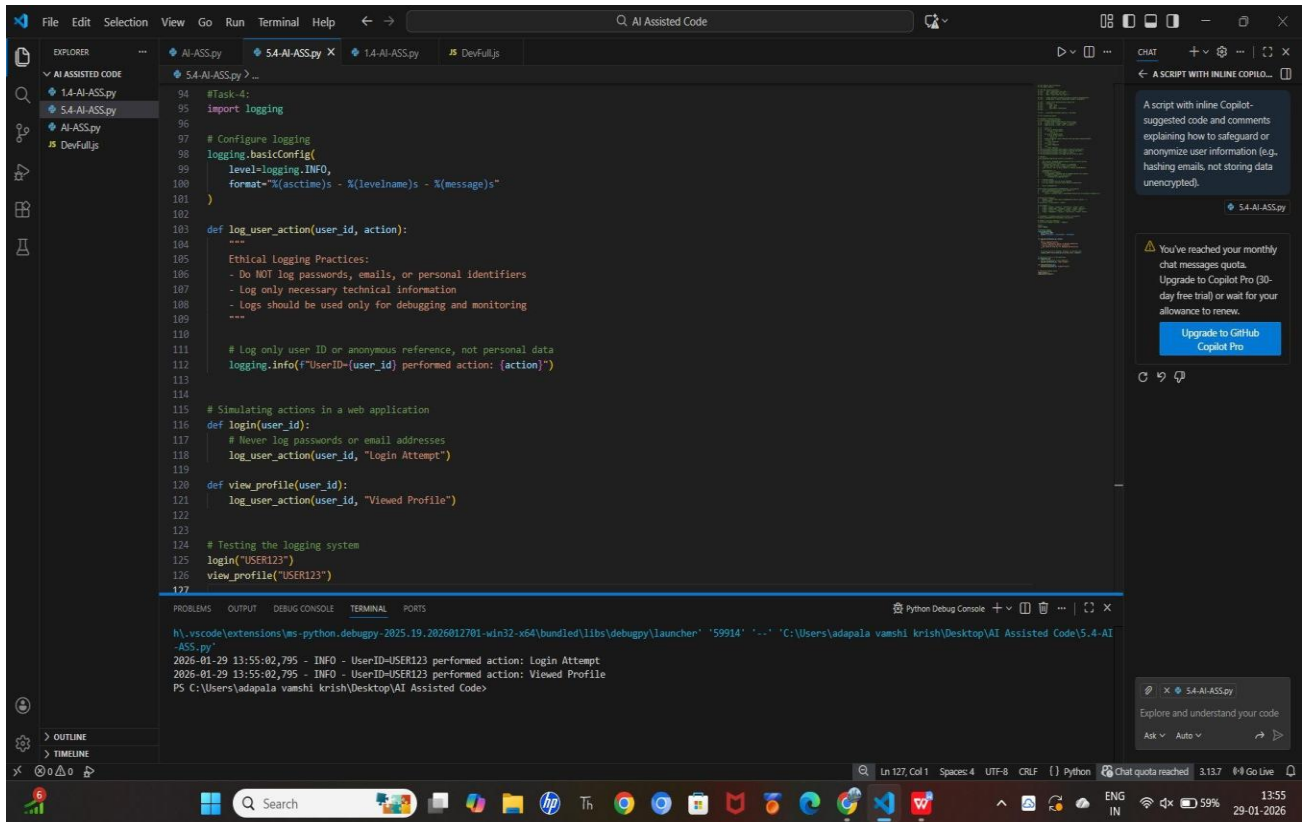
## Task 4: Ethical Logging

**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:**



```
File Edit Selection View Go Run Terminal Help
S4-AI-Assay.py
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 = [
    (9, 80, "Pass"),
    (12, 85, "Pass"),
    (8, 75, "Pass"),
    (11, 90, "Pass"),
    (7, 70, "Fail"),
    (10, 88, "Pass"),
    (6, 65, "Fail"),
    (13, 92, "Pass"),
    (5, 60, "Fail"),
    (14, 95, "Pass"),
]

# Model Training (Simplified Logic)
def train_model(data):
    """
    This function simulates training by learning simple rules.
    In real life, training would involve complex algorithms.
    """
    model = {
        "min_study_hours": 8,
        "min_attendance": 80
    }
    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
if __name__ == "__main__":
    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(f"Prediction Result: {result}")
    # Explanation (Transparency)
    print(f"Explanation:")
    print(f"The system predicts PASS if:")
    print(f"- Study hours >= 8 AND")
    print(f"- Attendance >= 80%")
    print(f"Otherwise, it predicts FAIL.")
    # Ethical Warning
    print(f"⚠️ Ethical Warning! ⚠️")
    print(f"This is only a demonstration AI system.")
    print(f"Do NOT use this for real academic, hiring, or life decisions.")
    print(f"Always involve human judgment.")
```

```
File Edit Selection View Go Run Terminal Help
S4-AI-Assay.py
def train_model(data):
    """
    This function simulates training by learning simple rules.
    In real life, training would involve complex algorithms.
    """
    model = {
        "min_study_hours": 8,
        "min_attendance": 80
    }
    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
if __name__ == "__main__":
    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(f"Prediction Result: {result}")
    # Explanation (Transparency)
    print(f"Explanation:")
    print(f"The system predicts PASS if:")
    print(f"- Study hours >= 8 AND")
    print(f"- Attendance >= 80%")
    print(f"Otherwise, it predicts FAIL.")
    # Ethical Warning
    print(f"⚠️ Ethical Warning! ⚠️")
    print(f"This is only a demonstration AI system.")
    print(f"Do NOT use this for real academic, hiring, or life decisions.")
    print(f"Always involve human judgment.")
```

```
File Edit Selection View Go Run Terminal Help
S4-AI-Assay.py
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
if __name__ == "__main__":
    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(f"Prediction Result: {result}")
    # Explanation (Transparency)
    print(f"Explanation:")
    print(f"The system predicts PASS if:")
    print(f"- Study hours >= 8 AND")
    print(f"- Attendance >= 80%")
    print(f"Otherwise, it predicts FAIL.")
    # Ethical Warning
    print(f"⚠️ Ethical Warning! ⚠️")
    print(f"This is only a demonstration AI system.")
    print(f"Do NOT use this for real academic, hiring, or life decisions.")
    print(f"Always involve human judgment.")
```

**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.