## Introduction

Artificial intelligence (AI) models have become increasingly vulnerable to various types of attacks, including adversarial attacks, evasion attacks, and AI poisoning attacks. These attacks can compromise the integrity and reliability of AI systems, leading to severe consequences in critical applications such as healthcare, finance, and national security. Therefore, it is essential to develop robust defence mechanisms to protect AI models against these attacks.

## Adversarial Attacks

Adversarial attacks involve adding noise or perturbations to the input data to mislead the AI model into making incorrect predictions. These attacks can be categorized into two types:

- **White-box attacks**: The attacker has complete knowledge of the AI model's architecture, parameters, and training data.
- **Black-box attacks**: The attacker has limited or no knowledge of the AI model's architecture, parameters, and training data.

## Evasion Attacks

Evasion attacks involve manipulating the input data to avoid detection by the AI model. These attacks can be launched by adding noise or perturbations to the input data, making it difficult for the AI model to identify the attack.

## AI Poisoning Attacks

AI poisoning attacks involve manipulating the training data to compromise the AI model's performance. These attacks can be launched by injecting malicious data into the training dataset, causing the AI model to learn incorrect patterns and relationships.

## Defence Mechanisms

Several defence mechanisms can be employed to protect AI models against these attacks:

- **Adversarial Training**: Training the AI model on adversarial examples to improve its robustness against attacks.
- **Input Preprocessing**: Preprocessing the input data to detect and remove noise or perturbations.
- **Anomaly Detection**: Detecting and flagging unusual patterns or behaviour in the input data.
- **Ensemble Methods**: Combining the predictions of multiple AI models to improve robustness against attacks.
- **Game-Theoretic Approaches**: Modelling the interaction between the attacker and the AI model as a game, and using game-theoretic techniques to develop optimal defence strategies.

## Azure Prototype Requirements

To develop a robust defence mechanism against attacks on AI models, the following Azure prototype requirements are necessary:

- **Azure Machine Learning**: A cloud-based platform for building, training, and deploying AI models.

- **Azure Notebook**: it is help to run our code for models.

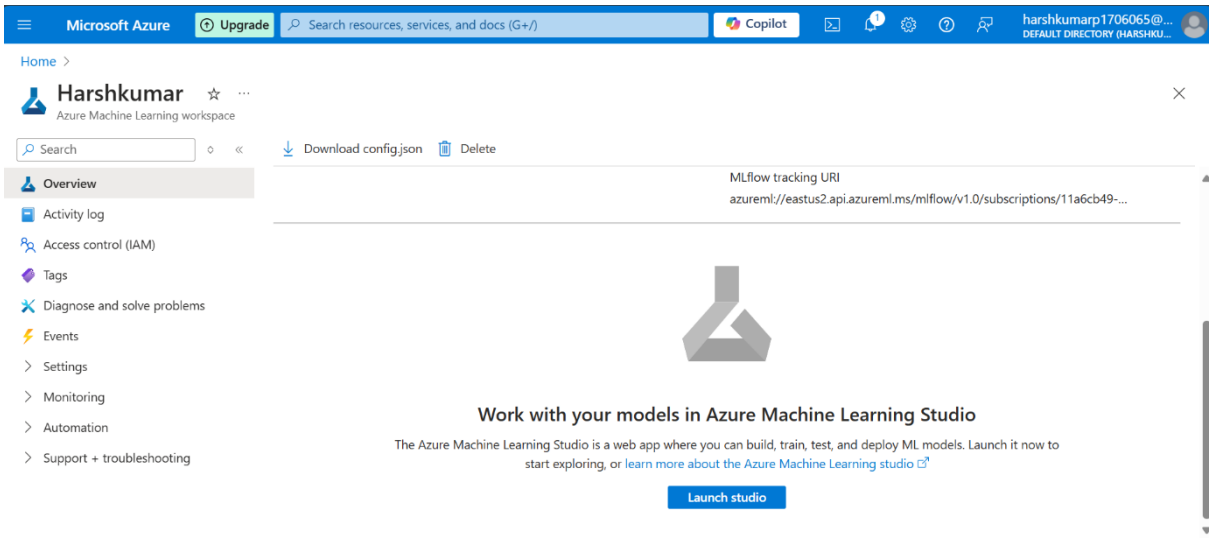
**Plan for the Capstone Project**

The plan for the capstone project involves the following steps:

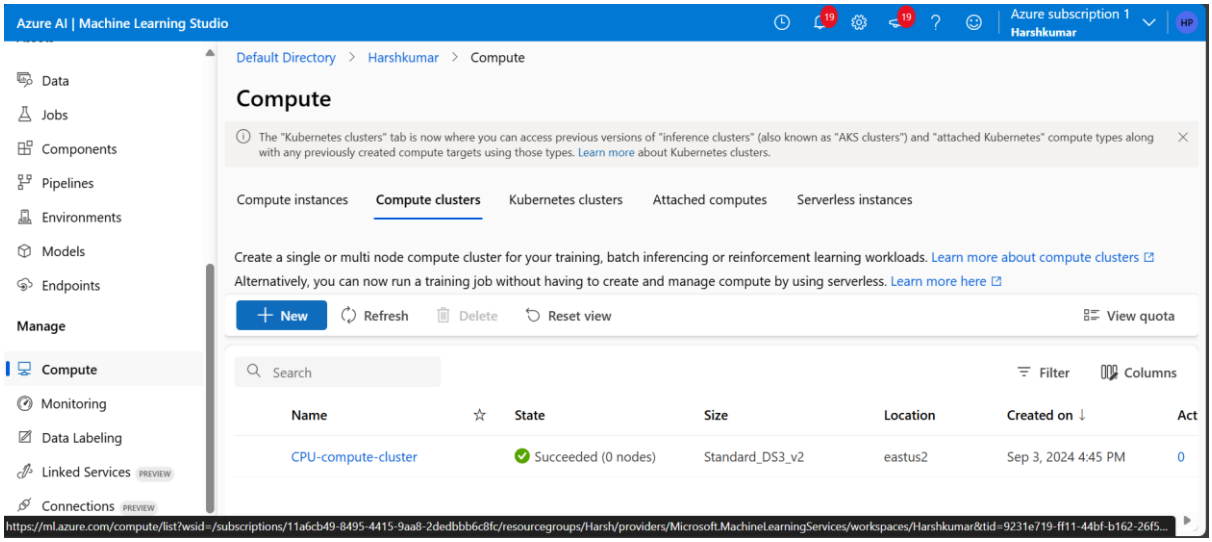
1. **Literature Review**: Conduct a comprehensive review of existing defence mechanisms against attacks on AI models.
2. **Data Collection**: Collect and preprocess a dataset for training and testing the AI model.
3. **AI Model Development**: Develop and train an AI model using Azure Machine Learning.
4. **Attack Simulation**: Simulate various types of attacks on the AI model, including adversarial attacks, evasion attacks, and AI poisoning attacks.
5. **Defence Mechanism Development**: Develop and implement defence mechanisms to protect the AI model against attacks.
6. **Evaluation and Testing**: Evaluate and test the effectiveness of the defence mechanisms against attacks.
7. **Deployment**: Deploy the AI model and defence mechanisms on Azure.

**Azure implementation:**

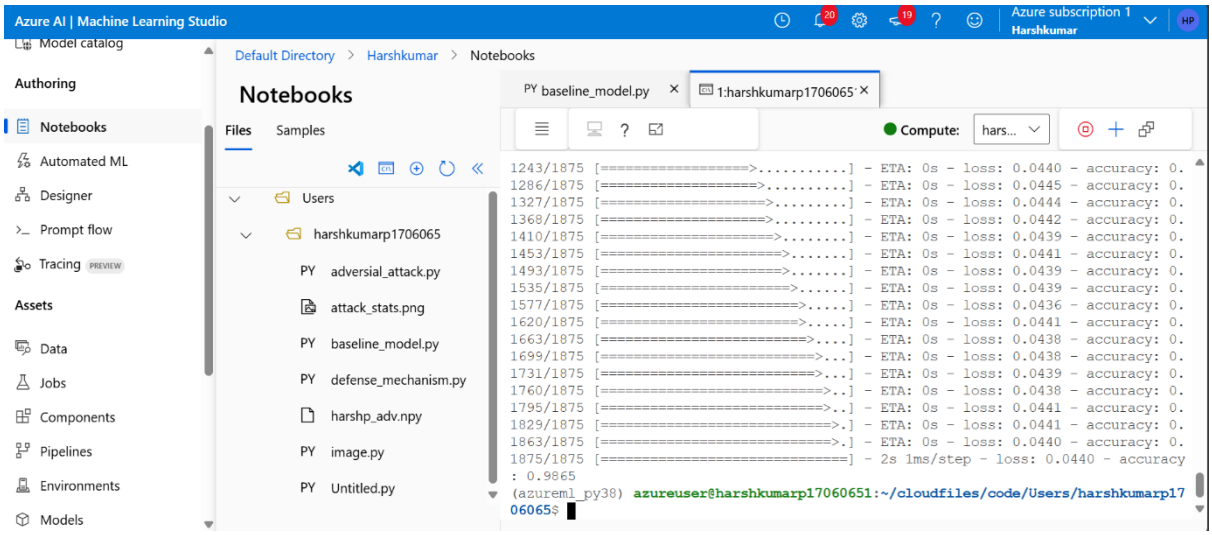
create azure machine learning studio



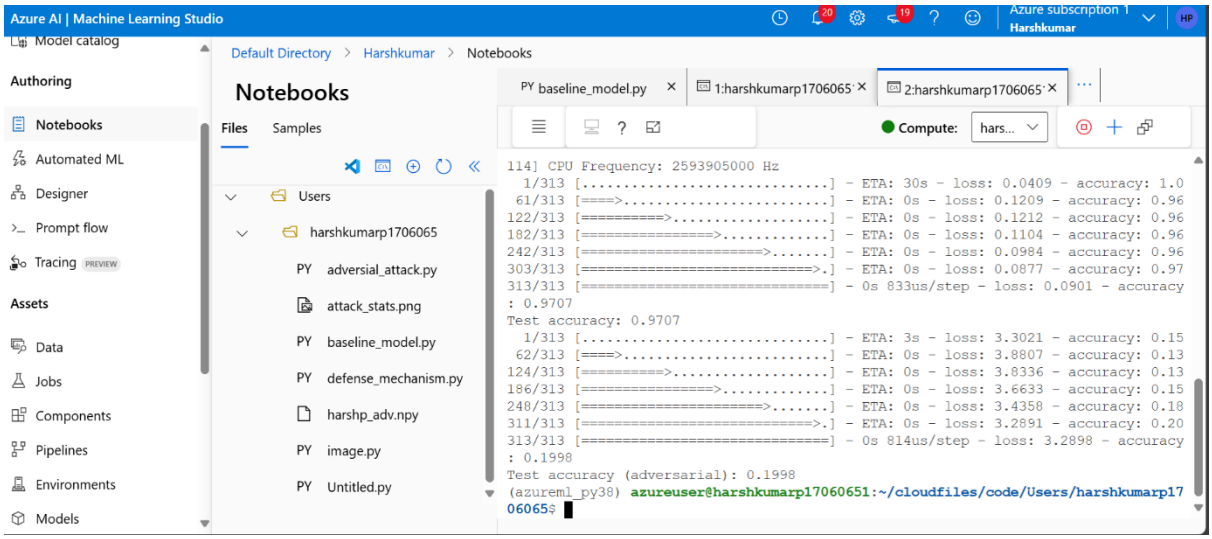

computer instance



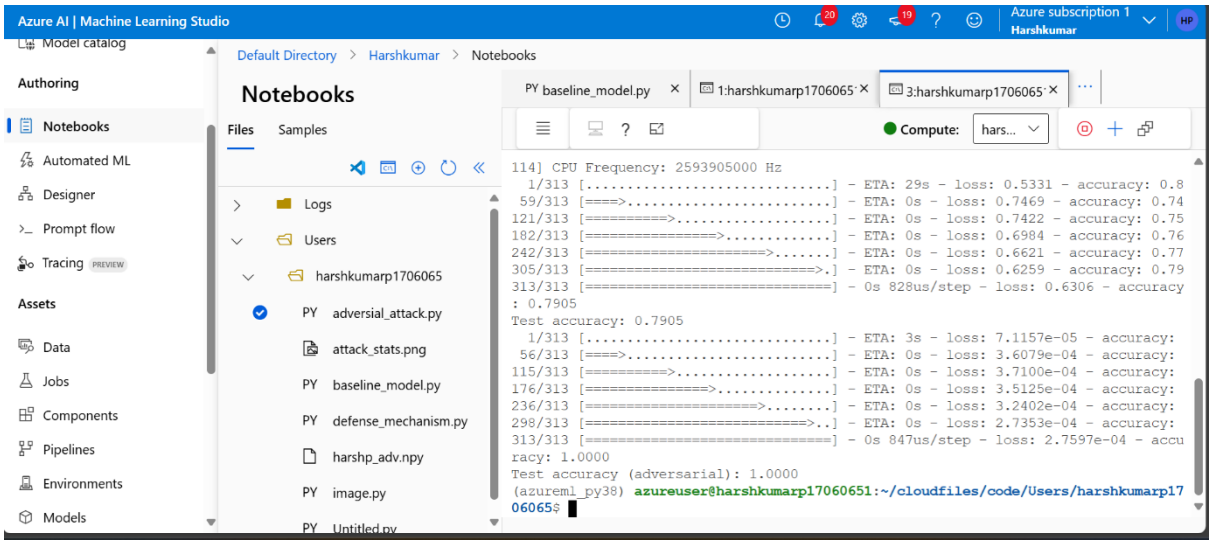

baseline model



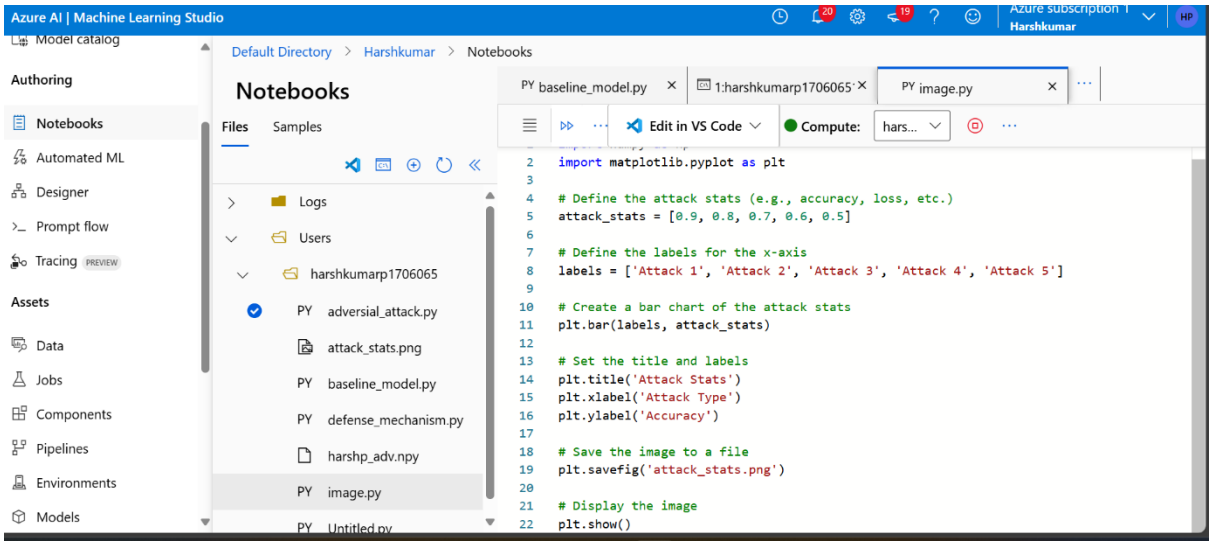

adversarial attack



defence mechanism




code for visual image




visual image of attacks