Homework Assignment:

Comparative Study of REINFORCE and A2C Variants

The goal of this assignment is to develop a rigorous understanding of policy-gradient reinforcement learning methods by implementing and comparing several variants of the REINFORCE and Advantage Actor–Critic (A2C) algorithms.
You will evaluate their behavior across environments of increasing complexity and analyze how architectural choices, such as shared vs. separate networks and the use of Generalized Advantage Estimation (GAE), influence learning stability and performance.

1. **Select three environments:**
   a. One from Classic Control (e.g., Acrobot, MountainCar, Pendulum).
   b. One from Box2D (e.g., LunarLander-v2, Car Racing with discrete action).
   c. One from Atari (e.g, Adventure, Air Raid, Amidar, etc)

2. **You must implement the following algorithms:**
   a. REINFORCE (Monte-Carlo policy gradient).
   b. A2C with separate actor and critic networks.
   c. A2C with a shared network (shared encoder + two output heads).
   d. A2C with GAE(λ).

3. **Experimental Protocol for Each Environment**
   a. Train all algorithms for the same number of episodes.
   b. Use the same random seeds.
   c. Record metrics including learning curves and convergence speed.
   d. Generate plots comparing all algorithms.

4. **Your analysis should address the following points:**
   a. Which A2C variant performs the best.
   b. The impact of GAE, and network sharing.
   c. Strengths and weaknesses of each method.

5. **Deliverables**
   You must submit the following materials:
   a. Complete source code (***notebooks***) for all implemented algorithms
   b. Reward curves and comparison plots for each environment
   c. Performance comparison tables (e.g., average return, standard deviation, convergence time)

d. A 4–6 page report containing:
    i. Introduction, Methodology, Experimental setup, Results, Analysis and discussion