Course Project

Code of Honor. All external resources used in the project, including research papers, open-source repositories, datasets, and any content or code generated using AI tools, e.g., ChatGPT, GitHub Copilot, Claude, Gemini, must be *clearly cited* in the final submission. The final report must also include *a clear breakdown of individual group member contributions.* Any lack of transparency in the use of external resources or in reporting group contributions will be considered academic dishonesty and will significantly impact the final evaluation.

Topic	Sequence Modeling for Reinforcement Learning with Decision Transformers
Category	Applications of Generative Models
Supervisor	Amirhosein Rostami

Objective Design and implement a Decision Transformer, a generative model that treats reinforcement learning (RL) as a sequence modeling task. The model should learn to predict the next action based on the historical trajectories and a desired return-to-go.

Motivation Decision Transformers offer a modern approach to RL by framing it as a *conditional generation problem*, eliminating the need for traditional value-based methods [2]. This project explores how Transformer-based models can be applied to policy learning.

Requirements The final submission should address the following requirements.

1. Choose a simple environment in OpenAI Gym[1] with discrete or low-dimensional continuous action space, e.g., CartPole, MountainCar, or a simple custom grid-world.
2. Generate or collect offline trajectories as state, action, reward, and return-to-go sequences. For this, traditional methods, e.g., use a PPO or DQN agent, from a pre-implemented code can be used with proper citation.
3. Implement a small Transformer, e.g., a GPT-style decoder, that takes the trajectory history and predicts the next action.
4. Evaluate the policy by deploying it in the environment and measuring cumulative return. Compare the result to
   - a standard deep RL agent,
   - **[Optional]** a behavioral cloning model, i.e., predict next action from state only [1].
5. **[Optional]** Explore how changing the desired return affects policy behavior or experiment with different tokenization strategies.

[1] `https://gymnasium.farama.org/`

Milestones

1. *Literature review and dataset creation.* Study the Decision Transformer paper [2]. Choose the environment and prepare an offline dataset using a reliable deep RL implementation from the literature.
2. *Model implementation.* Implement the Transformer-based policy model and train it on the trajectory data using autoregressive prediction.
3. *Evaluation.* Evaluate performance in the original environment and compare it to benchmark and optionally behavioral cloning.
4. [**Optional**] *Extensions.* Analyze the impact of conditioning on different return-to-go values and/or modify the model architecture to improve throughput.
5. *Final report.* Document all components of the project, including model design, training procedure, evaluation results, and challenges.

Submission Guidelines The main body of work is submitted through Git. In addition, each group submits a final paper and gives a presentation. In this respect, please follow these steps.

- Each group must maintain a Git repository, e.g., GitHub or GitLab, for the project. By the time of final submission, the repository should have
  - Well-documented codebase
  - Clear `README.md` with setup and usage instructions
  - A `requirements.txt` file listing all required packages or an `environment.yaml` file with a reproducible environment setup
  - Demo script or notebook showing sample input-output
  - *If applicable,* a `/doc` folder with extended documentation
- A final report (maximum *5 pages*) must be submitted in a PDF format. The report should be written in the provided formal style, including an abstract, introduction, method, experiments, results, and conclusion.
  Important: Submissions that do not use template are considered *incomplete.*
- A 5-minute presentation (maximum *5 slides including the title slide*) is given on the internal seminar on Week 14, i.e., *Aug 4 to Aug 8,* by the group. For presentation, any template can be used.

Final Notes While planning for the milestones please consider the following points.

1. You are encouraged to explore innovative approaches to conditioning or generation as long as the core objectives are met.
2. While computational resources are limited, carefully chosen datasets and training setups can make even diffusion models feasible. Trade-offs, e.g., resolution, training steps, are expected and should be justified.
3. Teams are expected to manage their computing needs and are advised to perform early tests to estimate runtime and training feasibility. As graduate students, team members can use facilities provided by the university, e.g., ECE Facility. Teams are expected to inform themselves about the limitations of the available computing resources and design the model accordingly.

References

[1] Michael Bain and Claude Sammut. A framework for behavioural cloning. *Machine Intelligence*, pages 103–129, 1995.

[2] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Misha Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision transformer: Reinforcement learning via sequence modeling. In *Proc. Advances in Neural Information Processing Systems (NuerIPS)*, volume 34, pages 15084–15097, 2021.