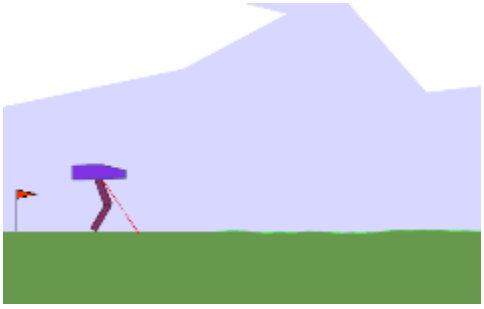
AISF Application

Due: Dec 26th

The Project

The primary goal of this project will be to train a multi-layer perceptron (MLP) with reinforcement learning (RL) to beat a bipedal walker environment. The premise is fairly simple, we have a 4-joint walker robot and the goal is to get it to walk a few meters to the right without the hull touching the ground. We are requiring everyone to use the gymnasium ai environment, which also contains more information about the specifics of the action and observation space. Initially we'll have everyone start with the normal environment, however, should you beat that we encourage you to try the hardcore environment. Below is a picture of the environment we'll be working with.



The main goal of this exercise is to see your research skills. As such we not only allow the use of AI tools, we encourage their use for initial setup as gymnasium is a bit annoying to work with. Similarly we encourage the use of libraries such as stable_baselines3. However you will quickly find that a standard PPO algorithm (that you can one-shot from gemini 3) won't be able to solve this. Instead we want you to look into the literature and find strategies for improving your walker. Note that we expect someone with some experience in coding and RL to take around 5-6 hours on this.

New To RL

If you're new to RL you may be having a hard time understanding all of the jargon above. That is ok. We are not expecting a very in depth theoretical knowledge of RL, this was just the best way to standardize our testing. As such to get started I would recommend simply asking an LLM what all the terms mean at a high level, then, if you have further questions, you can read more from open ai's RL starter blog. Do note that we will expect you to understand the concepts you used well enough to answer our questions on them, but we will take into account how new you are to the concept. We are not looking for your RL knowledge, but care more about your research abilities.

Where to Start

The way I'd approach this is:

- Get gemini 3 to create a basic stablebaselines 3 PPO solution and an environment renderer (you need to see what your walker is doing so you should always be saving gif's of it)
- Do some basic reward crafting
- Tune some parameters by hand (maybe 5-10 experiments to get a feel for it)
- Search the literature for improvements

In general there are a few main areas of improvement you can look into:

- Exploration vs exploitation
- Observation normalization
- Advantage vs reward
- Action clipping

Finally we want to see:

- Ablation studies to understand how much better the rewards/ other metrics is based on your improvements over your baseline
- Hyperparameter tuning
- Graphs from training

Requirements

Time Line

Please submit your code and write up through the following link by December 26th. We expect a minimum of 4 hours of work, while someone with some experience should be able to complete the requirements in 5-6 hours. If you are unable to complete the task by the due date (if you are on vacation, don't have access to a computer, Internet, etc.), please let us know ASAP.

Restrictions

- We also ask that you do this without the help of any other humans.
- If you've already completed quad walker hardcore mode before, email us.

Writeup

Should be around 3 pages, including references, not a strict max or min though.

Here is what we want listed:

- Your name and email
- Any prior experience you have in RL, Deep learning, and coding
- The number of hours spent and a break down of how you spent them
- Compute resources (can mostly be done on CPU so mention the number of logical threads and the model)
    - If you want GPUs, use Colab (though note this will likely be slower than cpu)
    - If compute restrained, you can list detailed experiments you'd want to run in greater detail (though you should try to use google colab first).
- A video of your walker's best performance (this should list the total number of time steps to train)
- Techniques used
    - And the justification of why the techniques work here
- Ablation studies
- Discussion of issues encountered (we also want to see what ideas you thought were good, didn't work out, and how you pivoted from them)
- Conclusion of what worked well and what else could be done with more time
- Citations for papers used

Code

A complete GitHub repository with the complete implementation and all experiments ran, with a readme file detailing all of the contents of the repository

Presentation

For those selected to give a final presentation (we will email you), prepare a 5 minute presentation about your implementation, and all experiments ran/references used to reach your final implementation.

Resources

Free gemini pro for students:
https://gemini.google/students/
Free github copilot pro for students:
https://docs.github.com/en/copilot/how-tos/manage-your-account/get-free-access-to-copilot-pro
Free colab pro for students:
https://blog.google/outreach-initiatives/education/colab-higher-education/
Walker environment:
https://gymnasium.farama.org/environments/box2d/bipedal_walker/
Open ai spinning up (RL starter guide):
https://spinningup.openai.com/en/latest/
Submission form:
https://forms.gle/cXDewf62q4Ljig4VA

FAQ

Q: What do we mean by Research Skills?

A: Primarily when we write about research skills we mean the following abilities:

- Locating relevant information (e.g. papers)
- Evaluating if a technique is likely to work
- Generating numerical analysis of various methods
- Organization/ methodicalness
- Communication

Q: How does MOR and RL help the HRM with ARC-AGI?

A: First of all, for Mixture of Recursions, we'd likely be using this as a router to determine how many iterations the low-level module will run for. To train for this we can just add some penalty for taking longer than needed. Thus we don't really need to do any balancing. It should also be noted this mostly just helps with saving compute as the low-level module will converge regardless so having a large number of recursions gives the same result as one that is shorter with the MOR setup.

As for RL, well that's a bit of a spoiler, however I will note that there are a few ways RL can help. In my opinion the best will be a recent paper that allows RL to be used to teach a model modular steps that can be composed for a complex task.

Q: Is it alright to participate without taking a lot of maths classes already.

A: Yes! We won't be doing any hardcore math. All that we are looking for is research skills, coding ability, and dedication.

Q: What day(s) would I have to commit?

A: For meetings the mentor will organize a time that works for everyone in the group. Most of the work will be async.

Q: What would the timeline/structure of this project look like?

A: The structure will be that you work with a mentor on their project with a few other participants. Depending on the interests of the participants and ideas for new directions, participants may work in groups or solo. Since we are aiming to publish this as a workshop in a major conference we will be working from Jan - April (since this is when most deadlines are). The basic cycle will be to come up with a research idea, try it, figure out what worked and what didn't, come up with new ideas, and retry them.

Q: Which skills would be most helpful for the projects?

A: For all of the projects the main things we are looking for is a willingness to learn and work hard. We of course can only measure these things via proxy, which is why this project is so sparse on details. We want to see who is willing to put in the time to learn about research skills and reinforcement learning. Other skills like PyTorch can be learned quickly.

Q: How much freedom and responsibilities are we tasked with?

A: Participants will be given as much freedom as possible. Any ideas will be heard out, and viable ones can be pursued so long as they are related to the topic chosen by the mentor. As for responsibilities, most of the coding work will be done by the participants, that is you will be primarily responsible for doing the research.

Q: How many people will be accepted to each project?

A: No project will take more than six participants due to logistical constraints and wanting to give everyone the best shot at publishing. That said this is an upper bound and the number may be subject to change.