The project must be carried out by a **team of two** students.
Deadline for preliminary analysis ("analyse préliminaire") : Monday March 30 at 11h59pm.
Deadline for final report ("rapport final") : Friday May 1st at 11:59pm.

1 Context

You work for an artificial intelligence consulting firm. A client is interested in using reinforcement learning (RL) in non-deterministic environments, where dangerous situations must be avoided. More specifically, your client is interested in applications where dangerous situations must be avoided as much as possible during learning, as they can be very costly. They therefore ask you to conduct a study showing the limitations of RL strategies under these constraints and to evaluate approaches proposed in the literature to address this problem.

2 Work to be done

The objective of the project is to study the impact of non-deterministic dynamics in an environment on the ability to avoid dangerous situations when learning a policy via RL. The project must be carried out in Python on the Frozen Laze (v1) environment from the Gymnasium [Towers et al., 2024]. In this environment, a player must cross a frozen lake with holes (dangerous situations) that must be avoided. The slippery surface of the lake limits the predictability of movements (non-deterministic environment). We are looking for a learning strategy that allows the player to avoid falling into the holes while learning their policy.

Your project must be hosted on GitHub. We will provide you with the usernames to add to the project for correction. Be sure to use the `commit` command, as the change history will be used to validate your contributions[1] to the project and control for plagiarism.

The work will be carried out in two stages: 1) preliminary analysis; 2) final report.

2.1 Preliminary analysis

First, your client asks you to produce a preliminary analysis demonstrating your understanding of the Frozen Lake environment using the Gymnasium library, describe the study you plan to conduct, and present the results of preliminary experiments. Your document should follow the structure below:

- Case studies
- Considered strategies
- Experimental methodology
- Results
- References

Your document can contain **up to 4 pages** (excluding references) and must be written in LaTeX using the following style: `https://github.com/kourgeorge/arxiv-style`. Do not include an abstract or keywords. You must submit the preliminary analysis in PDF format. It must contain the URL of the project GitHub containing the documented source code (Python) to reproduce your experiments.

Case studies The Frozen Lake environment is highly configurable. It is your responsibility to fully understand the environment in order to propose a relevant study that will assess the impact of non-deterministic environmental dynamics on the avoidance of dangerous situations. To this end, present the components of the environment, explaining how they can be manipulated to adjust the difficulty of the environment: degree of stochasticity in the dynamics; number of dangerous situations; positioning of dangerous situations. Fi-

[1] Possible penalty of up to 50% for lack of contribution.

nally, present **three case studies** that you wish to consider and explain in concrete terms how to configure your environment to obtain them. Describe the levels of difficulty/challenges captured by your case studies to justify their relevance.

Considered strategies As part of the project, you must compare different RL strategies. In the preliminary analysis stage, identify **two strategies** (e.g., DQN [Mnih et al., 2015], PPO [Schulman et al., 2017]) that will serve as *baselines* for the study. We are referring here to strategies that were not explicitly designed with mechanisms to handle the challenges found in this environment. Explain your choice of these strategies and describe how they work.

Experimental methodology You must conduct experiments with the strategies considered on the various case studies established. For each strategy considered, you must provide details of the hyperparameters used and how they were determined. You can use existing libraries that provide an implementation of the strategies, such as Stable-Baselines3 [Raffin et al., 2021]. Simply indicate which libraries you are using and how you are configuring the different strategies. Also describe all the experimental parameters necessary to reproduce your study. This includes, in particular, the duration of the episodes, the number of time steps for learning, and the number of repetitions (to obtain confidence intervals on performance). Finally, present the different performance indicators you use to quantify the achievement of the study objectives by the strategies considered. Please note that the objective is not limited to maximizing rewards on the final policy!

Results Present the results obtained using the methodology described above. Your results should include tables and/or figures that allow you to track the indicators of **performance during learning** for the various case studies considered and visualize the d**ifference between the policies** learned by the two strategies. In your presentation of the results, highlight the salient facts in each table/figure. This step is crucial as it will inform your selection of subsequent strategies to improve performance (for the final report).

References You must provide appropriate bibliographic references when referring to elements of the literature that do not belong to you. This includes, for example, environments, libraries, and algorithms/strategies. Ensure that the style of your entries in the bibliography is consistent. Also avoid including unnecessary information (e.g., DOI, ISSN, URL).

2.2 Final report

Following validation of the preliminary analysis by the client, you must complete your study to include an RL strategy with specific mechanisms to address the challenges of the problem at hand. Your final report must follow the following structure:

- Case studies (udpated)
- Considered strategies (updated)
- Experimental methodology (updated)
- Results (updated)
- Discussion
- References (updated)

Your document can contain **up to 6 pages** (excluding references) and must be written in LaTeX in the following style: `https://github.com/kourgeorge/arxiv-style`. Do not include the abstract or keywords. You must submit the final report in PDF format. Your report must contain the URL of the project GitHub containing the documented source code (Python) to reproduce your experiments.

Case studies No new material is expected at this stage. Simply apply the changes following the preliminary analysis correction, if applicable.

Considered strategies In preparation for the final report, search the scientific literature and identify **three published articles** (in scientific journals or conference proceedings), each presenting a method suited to the problem addressed by the study. Briefly describe (3-5 sentences) each strategy. Then select the one you consider most promising for improving the performance of the baseline strategies in the problem under consideration. Justify your choice of strategy and explain how it works. Add this new material to the baseline strategies updated following the correction of the preliminary analysis.

Experimental methodology You must conduct experiments with the strategies considered on the various case studies established. For each strategy considered, provide details of the hyperparameters used and how they were determined. You can use existing implementations of the strategies available via libraries or on GitHub. Simply indicate the sources of your implementations and how you use them. If methodological errors were identified in the preliminary report stage, be sure to correct them by updating the description of the methodology and repeating the experiments with the baseline strategies. This includes, in particular, performance indicators.

Results Present the results obtained using the methodology described above for all of the strategies considered. Your results should include tables and/or figures that allow the **performance indicators during learning** to be tracked for the various case studies considered and the **difference between the policies** learned by the different strategies to be visualized. In your presentation of the results, indicate the salient facts to be noted in each table/figure.

Discussion Discuss the results obtained in relation to what was expected based on the literature. Explain the results obtained (do not simply mention that approach A is better than approach B). Feel free to refer to the literature to support your hypotheses. Present your thoughts on the relevance of the different approaches, considering algorithmic complexity, implementation challenges, and computational cost. Present your general conclusions regarding the difficulties identified in your study: What would be your overall message to the client who commissioned this study?

References Update following the same instructions as for the preliminary analysis.

References

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

Antonin Raffin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268):1–8, 2021.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Mark Towers, Ariel Kwiatkowski, Jordan Terry, John U Balis, Gianluca De Cola, Tristan Deleu, Manuel Goulão, Andreas Kallinteris, Markus Krimmel, Arjun KG, et al. Gymnasium: A standard interface for reinforcement learning environments. *arXiv preprint arXiv:2407.17032*, 2024.