Project Instructions for Reinforcement Learning (INF8250AE) (2025 Fall)

Amir-massoud Farahmand
Department of Software and Computer Engineering, Polytechnique Montréal
Mila – Quebec's AI Institute, Canada

1 Introduction

The goal of the project is to give you an opportunity to gain experience in doing research in reinforcement learning (RL).[1] The range of acceptable research topics is wide. Your work can be theoretical, empirical, algorithmic, applied, environment design, educational, or a combination of all. This is discussed more in Section 2.

The project contributes 30% to your final mark. It has three components. The percentage of each component and the deadline for each of them are (subject to minor change)

- Proposal (5%): November 14th, 2025 (Section 3)
- Presentation (5%): December 1st, 2025 (Section 4)
- Report (and Source Code) (20%): December 15th, 2025 (Section 5)

Collaboration. The project should be conducted collaboratively as a team of 3–4 human members all enrolled in the class (so not pets or AI friends, sorry!). Everyone should contribute to the project in a meaningful way, and they should be clear about their contributions. You can use the help of a machine in writing the code for your project. You can also use it to revise your writing. But you should not delegate writing the majority of the report to a machine.

Some frequently asked questions are answered in Section 6.

2 Types of Projects

You may choose any topic of your choice, as long as it is related to RL. Below I describe several general directions you may want to pursue, and how you can choose a topic within them. Your project might combine several of these directions.

You do not need to invent a new algorithm, reach a state of the art performance, or completely solve a new application domain to be successful in this project. I realize that there is not much time in a semester to learn about a new research area well enough to come up with an innovative idea and execute it completely (though sometimes your fresh perspective might lead to ideas that others have not thought before. In that case, you have to cherish that opportunity and pursue it). The goal is to give you a taste of what research in RL looks like. If you feel delighted enough after the end of this course, you have the option to continue working on the project with your team. This being said, you have to spend a good amount of time on this project and you must follow rigorous scientific methodology in pursuing your research.

[1] This document may be slightly changed in the coming weeks in order to clarify any questions that you have. This version: 2025 October 30.

Before providing specific suggestions for each type of project, I have a general suggestion on how you can start if you do not have any idea already: You can consult the articles published at AI/ML venues such as NeurIPS, ICML, RLC, ICLR, AAAI, IJCAI, AISTATS, COLT, JMLR, TMLR, MLJ, and JAIR, where RL papers are often published, to find many interesting papers. One of them may catch your attention. Read it carefully. Afterwards, read some of the papers that are cited within that paper, and backtrack. You often find the same set of papers are referred to again and again. It is also very helpful to see what other more recent papers have cited your originally selected paper. This helps you figure out what advances has been done since that original paper.[2]

2.1 Application

If your main research topic is in an application area, broadly defined, you can investigate whether you can formulate it as an RL problem and solve it using RL algorithms.

For example, if you are working on robotics, computer networks, computer games, healthcare, autonomous vehicles, energy management, scheduling, you can use RL, as many have already done so. Of course, there are many more application domains that can potentially fit well within the RL framework, but have not been explored yet. Discovering this possibility is an exciting area.

You need to make sure that the application can indeed be formulated as an RL problem. If the problem is better formulated as a supervised learning problem, for instance, that will not be a good research project.

If you decide to go through this path, you should design or use appropriate simulator for your environment, compare several RL algorithms (at least 2-3) that we have covered during the course or some new ones that you find in research papers, and compare them with each other. You need to follow high standards of empirical evaluations.

2.2 Empirical Study

Empirically investigating an already existing RL algorithm is a reasonable project. Think of yourself as an experimentalist who wants to understand the behaviour of an algorithm through careful design and conduct of experiments.

You want to know when the algorithm works and when it does not. The original paper that introduced the algorithm might have not explored all relevant questions. It is likely that their authors only reported successful results. Your goal is to empirically investigate the range of problems and conditions that results in success and failure of an algorithm.

For example, you may want to know the effect of stochasticity of the environment, learning rate, or the exploration technique on the performance of an online algorithm. Or you may choose to see how an algorithm works on a wide range of environments. You may choose a state of the art algorithm and implement it on domains that were not included in the original paper.

When you perform an empirical study, you need to follow good statistical practices. For example, if there is any randomness in the algorithm (e.g., random initial weights of neural network) or the environment, you need to run the algorithm multiple times in order to compute the mean performance as well as its confidence interval. You may want to consult Patterson et al. [2024] to learn more about how to conduct empirical studies properly.

2.3 Algorithm Design

You may decide to design a new algorithm by varying certain components of an already existing algorithm, and effectively search in the space of "adjacent possible". For example, what happens if you use a DNN in a Fitted Q-Iteration framework? (Voila! You have re-invented DQN!). Or what happens if you use various variance reduction techniques in a policy gradient method? Or what

[2] You may also consult the following good set of suggestions by Csaba Szepesvári, especially if you want to work on a theoretical project. In writing this instruction, in addition to Csaba's page, I consulted and borrowed ideas from Animesh Garg's course on 3D and Geometric Deep Learning, Roger Grosse's project instructions, and Sarath Chandar's project instruction. You may find good advice there too. Note that their evaluation criteria and what is acceptable or not is not the same as this course, so do not rely on that.

happens if instead of using scalar gains in an online algorithm such as TD, you use a matrix gain? And perhaps design an adaptive mechanism to change the gain?

The space of all possible algorithms is combinatorially large. Maybe we can use a computer to automatically explore it. For this course, however, I want you to explore it based on the insight gained in the course as well as your other courses and research experience.

You do not want to randomly wander in the space of all algorithms. Any new algorithm should be justified. You do not need to rigorously prove that the algorithm works before trying it empirically, but it is good to have a theoretical insight before your start implementing.

2.4 Theoretical Analysis

You may decide to work on a theoretical understanding of an RL algorithm or problem. Some example research directions are (by no means comprehensive):

- Understanding the convergence proofs of different RL algorithms (TD with function approximation, policy gradient, actor-critic) under various assumptions (e.g., linear function approximator, nonlinear function approximator, on-policy vs. off-policy sampling distribution, etc.).
- Understanding the sample complexity guarantee for RL problems and algorithms, investigate what the upper and lower bounds are, and whether they match.
- Investigate how the discounted MDP framework can be extended to other settings such as semi-Markov Decision Processes, Partial Observable MDPs, risk minimization instead of maximizing the mean return, etc., and study what this entails on standard algorithms such as VI, PI, LP, and their corresponding RL algorithms.
- Theoretically analyze an algorithm with a good empirical performance and try to analyze its theoretical properties, perhaps in a simplified setting (linear FA or even finite state-action MDP).

You should understand an aspect of theory literature very well and figure out

- What are interesting questions to ask? For example, is the stability of the algorithm the main concern? Or the sample complexity is?
- What are known and unknown about the topic? For example, do we have any upper bound for the sample complexity? What about a lower bound? Do they match?
- What assumptions are required to make the analysis work? Is there any assumption that can be relaxed? For example, do we need an i.i.d. assumption in the proofs? What changes if we relax that assumption? Or is boundedness of some quantities assumed? Is that necessary?
- Is there any part of the theoretical analysis that can be improved? Is there any tool that the authors of the paper use that is known to be improvable? For example, if they use Hoeffding's inequality, can the use of Bernstein's inequality lead to any improvement?

2.5 Educational Notebook

You can create an educational Jupyter/Colab Notebook-based tutorial on a topic related to this course. The goal is to create a Notebook that can be used in the future offering of this course as a tutorial in order to help students understand an RL topic better. The topic can be anything mentioned in the class or in the Foundations of Reinforcement Learning textbook [Farahmand, 2025], or any topic that we did not cover, but is still relevant to RL. If you have several ideas in mind, I can help you decide among them and ensure that you are not working on the same topic as another team.

Your notebook will be evaluated based on the pedagogical nature of the notebook, clarity in explaining algorithms, insights obtained on how the algorithm works based on the set of experiments you run, and the quality and insightfulness of visualization of complex concepts through well-designed experiments. Your tutorial should be different from the existing online tutorials. Some example of what I have in mind are Understanding RL Vision, The Paths Perspective on Value Learning (though your deliverable should be in the form of a Jupyter Notebook and not a blog post), but you do not need to follow their specific styles.

2.6 Environment Design

Creating a novel and interesting RL environment can be a basis for a good project too. The goal of this type of project is to create a new RL environment that either (a) captures a novel real-world application domain, or (b) highlights a specific algorithmic challenge not adequately addressed by existing benchmarks. Your contribution increases the set of possible and interesting environments available to RL researchers.

A good new environment allows RL researchers to study some aspects of their algorithms that have not been carefully studied before. For example, if the environment allows a controlled level of stochasticity, it helps researchers to understand the behaviour of their algorithm under stochasticity of the environment. Moreover, if the environment is particularly challenging for existing algorithms and most of them fail to perform well, it encourages the researchers to come up with better algorithms. A good and challenging environment accelerates the RL research. The introduction of the Atari Suite or MuJoCo are such examples in the last 10–15 years.

Some example of existing environments are Acrobat, Cart Pole, Mountain Car, Bipedal Walker, Car Racing, Taxi, Frozen Lake, Ant, Half Cheetah, and Atari Suite. You have worked with some of them in your homework assignments. Most, if not all, of these environments are "solved" in the sense that there are many RL algorithms obtaining very good performance on them. Your environment should be meaningfully different from them or other existing ones. For example, a slight variation of Taxi environment or Frozen Lake is not a good choice for your project. Take a look at Gymnasium External Environments, and the list of Projects maintained by the Farama Foundation to get an idea of the scope of existing environments. I want you to go beyond them! Note that what you see on these pages are not comprehensive, so if you come up with an environment idea, you need to do your research to make sure it does not already exist.

Coming up with a new good environment is not straightforward. One approach to come up with a new environment is similar to how one may come up with an Application type project (Section 2.1): If you have an expertise in an application domain that is not yet formulated within the RL framework, your domain might be a basis for coming up with a good new environment. Another approach is that you design an environment that tests a specific aspect of RL algorithms. An environment with a controlled level of stochasticity is an example, as already mentioned. Or you may specifically design an environment with a very large action space. Or an action space with a particular structure, for example, a combinatorial one. Or an environment with observations that have a controlled level of irrelevant dimensions or controlled amount of partial observability. This approach may not be very easy, as it requires you pinpointing a specific challenge in RL and coming up with an environment focusing on that challenge, but if your endeavour is successful, it can help the field.

For this type of project, you should design an environment with a Python package that others can easily use to test their RL algorithms. I recommend that the interface be compatible with Gymnasium API (or PettingZoo for Multi-agent RL, or similar interfaces of the Farama Foundation).

In addition to designing the environment, you need to apply several standard RL algorithms to the environment and report the result. As opposed to the Algorithm Design type of project (Section 2.3), you do not need to come up with a new algorithm – though if you do, that is welcomed. As opposed to the Empirical Study type of project (Section 2.2), your goal is not to empirically study a single algorithm through extensive empirical studies – though if you do, that is welcomed too. The goal is to show that the environment is solvable by existing algorithms, but perhaps not very efficiently, and can be an interesting benchmark for others.

Some examples of benchmarks and papers doing so are: Atari Suite [Bellemare et al., 2013], Industrial Benchmark [Hein et al., 2017], Procgen [Cobbe et al., 2020], MiniHack [Samvelyan et al., 2021], CropGym [Kallenberg et al., 2023], Optical Control Environment [Abuduweili and Liu, 2023], and Gym4ReaL [Salaorni et al., 2025].

3 Proposal

Your proposal should be a maximum of two (2) page summary of your intended research direction (references excluded in the page limit). You need to

- clearly define the problem,

- provide a brief summary of prior work,
- what you intend to achieve.

The instructor and the TAs will provide you with feedback on your proposal. We also have some office hours before the proposal deadline, in case you want to bounce ideas back and forth before submitting them in the written form.

4 Presentation

You need to prepare a presentation defining the problem, and very briefly summarizing what was known about it, your contributions and your results up to that point, and a brief conclusion. Depending on the type of project you have, the format might be slightly different (for instance, we do not expect any figure in a purely theoretical work). It is OK if your work is not complete by that the presentation date; you have about two more weeks before submitting your final report.

You have about 5min to present your work. The exact amount depends on how many teams we will have. As this is a short time for a presentation, you should be efficient in your communication.

You need to present the material in the class and all the team members should be present to answer questions (unless there is a valid excuse to miss the class, such as a medical issue). We also record the presentations and intend to upload them on YouTube to showcase your achievements. If you are against uploading your presentation on YouTube, please contact us well in advance of the presentation day.

I will evaluate your presentation based on its clarity, timing, and your responses.

5 Report and Source Code

You should write a 6–8 page report summarizing your work. We encourage you to use LaTeX.[3] You can use the NeurIPS style file, though you are not required to use this specific style.[4] Whatever style you use, make sure it has large margins, so I can leave you handwritten comments.

At a high-level, your report should include:

- **Problem Definition and Motivation:** Clearly state what problem you are tackling and why we should care about it.
- **Summary of Prior Work:** What other attempts have been made in order to address this problem.
- **Your Contributions:** Statement and proof of a new result or summary and critique of prior results (Theoretical); clear description of the algorithm and evidence (theoretical or empirical) supporting how it works (Algorithmic); the description of the algorithm, the experimental design to evaluate them, and the empirical results (Empirical); how you formulated your application, the description of algorithms you have tried, and the performance you achieved in comparison with other (non-RL) baselines (Application); what the introduced environment is (specify the MDP), how to use the codebase, and what the performances of baseline algorithms are (Environment Design).
- **Negative Results (optional, but encouraged):** If you face difficulties in obtaining good results in the research, you should report them and assess what might have caused them.
- **Conclusions:** What have you learned and what is remained to be done or figured out?
- **Individual Contributions:** What is the individual contribution of each of you?

It is common in research that each co-authors contribute to different aspects of the project. Some may come up with the high-level ideas, some design the algorithm, some study the idea theoretically, some design and conduct the experiments, and some others write the paper. I'd like to acknowledge

[3] If you do not know how to use LaTeX, this is a great opportunity to learn.

[4] The exception is if your project is in the form of Educational Notebook (Section 2.5, in which case, your report is effectively contained in the notebook.

that this is how modern science works and let you have different contributions. That being said, *you need to have a section describing the rule of each team member in the whole project in some detail*. The only requirement is that *all team members must be involved in writing the paper*. If you are not good in writing yet, the university is the right place to practice it.[5]

If your project has a source code, which most projects do, you should submit it too. All empirical results should be reproducible. Submit code and configuration files. If your environment or data includes ethical considerations (e.g., human data), describe them.

I will read your paper as if you want to submit it to an RL/ML venue and I am the PI supervising the project. I will most likely leave some handwritten comments on the PDF itself. Since this is a course and you need to be given a grade, I also evaluate your work based on its quality of writing and explanations (including the clarity of the problem definition and the precision of your statements), how well you cover the prior work, your contributions (which depends on the type of research you have conducted), and whether you have followed a good scientific methodology. Parts of your mark depends on the "excellence" of your work, which is reserved for the projects and reports that go beyond the expectation.

6 FAQ

Q: Is it acceptable to have a project that overlaps with my thesis research project?

A: Yes, and I encourage it. But you should be clear about what parts have been done before this project, and what contributions are new. The basic idea is that you should not reuse your prior work for this project; you have to spend a lot of time during this semester to work on this project, but you can use it for your thesis (of course, if your supervisor is OK with it).

Q: Is it acceptable to have a project that overlaps with another course project?

A: Try to avoid it! If there is a good reason to have a project that spans more than one course, that can be discussed. You need to get the permission of all instructors for this. Which means that you need to send an email to me and the other instructor(s) and get a joint permission. Since your research is done within a team and you may have different teams in different courses, this makes the credit assignment difficult, hence the discouragement.

Q: Can I extend the project from a previous course?

A: Yes! You should mention it in your proposal, include the report from the previous project in your submission, and be explicit about the new contributions specific to this course. In other words, be clear about the Δ.

Q: Can I have a team size of 1, 2, or 5+?

A: Teams should consist of three or four (preferred) members. A solo or two-member team is acceptable only if there is a very good reason. Any other arrangement requires a clear reason and my prior permission.

Q: What if I discover that someone else has done a very similar thing to what I am doing in this project?

A: That is OK. It is a part of science. Make sure to mention those paper(s) in your prior work, and provide a detailed comparison.

Q: How many times should I run the algorithm?

A: The exact number of runs depends on the problem and the hypothesis that you want to evaluate. As a very rough rule of thumb, aim to produce results based on at least 10+ runs. You may use standard error ($\frac{\sigma}{\sqrt{N}}$, with σ^2 being the variance of the quantity under study and N being the number of runs/seeds) as a reasonable measure of the size of your confidence interval – of course, rigorous statistical hypothesis test is always welcomed. If the confidence intervals between two methods overlap, you cannot claim one method is better than the other.

Q: Is it OK to copy a paragraph from another paper or a textbook?

[5] If there is any good reason that you cannot write the paper, for example a medical reason that limits your typing ability, you should discuss it with me beforehand.

A: Generally speaking, No! That would likely be a case of plagiarism, unless you explicitly quote it within quotation marks and cite the source right there. It is much better and safer practice if all that is written in your work is your own words. If you are not familiar with what plagiarism is and how to avoid it, please familiarize yourself.

References

Abulikemu Abuduweili and Changliu Liu. An optical control environment for benchmarking reinforcement learning algorithms. *Transactions on Machine Learning Research (TMLR)*, 2023. 4

Marc G. Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. The arcade learning environment:an evaluation platform for general agents. *Journal of Artificial Intelligence Research (JAIR)*, 47:253–279, 2013. 4

Karl Cobbe, Chris Hesse, Jacob Hilton, and John Schulman. Leveraging procedural generation to benchmark reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2020. 4

Amir-massoud Farahmand. *Foundations of Reinforcement Learning (draft)*. 2025. URL https://amfarahmand.github.io/IntroRL/lectures/FRL.pdf. 3

Daniel Hein, Stefan Depeweg, Michel Tokic, Steffen Udluft, Alexander Hentschel, Thomas A Runkler, and Volkmar Sterzing. A benchmark environment motivated by industrial control problems. In *IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017. 4

Michiel G.J. Kallenberg, Hiske Overweg, Ron van Bree, and Ioannis N. Athanasiadis. Nitrogen management with reinforcement learning and crop growth models. *Environmental Data Science*, 2, 2023. 4

Andrew Patterson, Samuel Neumann, Martha White, and Adam White. Empirical design in reinforcement learning. *Journal of Machine Learning Research (JMLR)*, 25(318):1–63, 2024. 2

Davide Salaorni, Vincenzo De Paola, Samuele Delpero, Giovanni Dispoto, Paolo Bonetti, Alessio Russo, Giuseppe Calcagno, Francesco Trovò, Matteo Papini, Alberto Maria Metelli, Marco Mussi, and Marcello Restelli. Gym4real: A suite for benchmarking real-world reinforcement learning, 2025. 4

Mikayel Samvelyan, Robert Kirk, Vitaly Kurin, Jack Parker-Holder, Minqi Jiang, Eric Hambro, Fabio Petroni, Heinrich Kuttler, Edward Grefenstette, and Tim Rocktäschel. Minihack the planet: A sandbox for open-ended reinforcement learning research. In *Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*, 2021. 4