Reproducibility Checklist

[Based on Joelle Pineau's ML Reproducibility checklist]

Descriptions (in reports, theses & papers)

- [ ] A clear description of the mathematical setting, algorithm, and/or model.
- [ ] A clear explanation of any assumptions.
- [ ] (If appropriate) An analysis of the complexity (time, space, sample size) of any algorithm.

Code Quality

- [x] Make training code available
- [x] Make evaluation code available
- [x] Well documented (e.g., DocStrings) and readable code
- [x] Unit test your code
- [x] README.md with precise instructions (commands) for installation and running the code
- [x] Dependencies and requirements (requirements.txt)

Experimental Reproducibility and Generalization

- [x] Several training repetitions with different random seeds [block seeds across settings be compared to one another]
- [x] [RL] Seeding of environments to control non-determinism of environments (at least 1 fixed seed in training and several for evaluation)
- [x] [RL] Several evaluation runs on the same environment
- [ ] [RL] Evaluation on several environments or variations of the same environment
- [ ] Run and report ablation studies to check the impact of different design decisions
- [x] Optimally, use the same amount (and technique) of hyperparameter optimization for all competitors

Reporting

- [x] All details regarding the experimental setting, incl. software versions and [RL] which env incl. reward function, gamma etc pp.
- [-] A description of results with central tendency (e.g. mean) & variation (e.g. error bars).
- [ ] (If possible and appropriate) Statistical hypothesis tests to show a significant difference in performance (beware of significant vs. substantial) – report alpha, type of test and test assumptions
- [x] The exact number of training and evaluation runs
- [x] Used random seeds (training, agents, environment)
- [-] All hyperparameters decisions (settings, ranges, optimization technique and resources)
- [x] Compute infrastructure (CPUs, GPUs, TPUs, RAM, OS)
- [-] The average runtime for each result, and (if possible) estimated energy cost.