

I. IDENTIFICATION DATA

Thesis title:	Reinforcement learning for multi-robot navigation
Author's name:	Kairat Bekbolinov
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Karel Zimmermann
Reviewer's department:	Department of Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment	ordinarily challenging
<i>How demanding was the assigned project?</i>	
Please insert your comments here.	

Fulfilment of assignment	fulfilled
<i>How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.</i>	
Probably yes, but I am not sure if all scenarios from [3] has been used in the evaluation as required in the assignment.	
[3] C. Jestel, H. Surmann, J. Stenzel, O. Urbann and M. Brehler, "Obtaining Robust Control and Navigation Policies for Multi-robot Navigation via Deep Reinforcement Learning," 2021 7th International Conference on Automation, Robotics and Applications (ICARA), Prague, Czech Republic, 2021, pp. 48-54, doi: 10.1109/ICARA51699.2021.9376457.	

Methodology	correct
<i>Comment on the correctness of the approach and/or the solution methods.</i>	
Please insert your comments here.	

Technical level	C - good.
<i>Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?</i>	
<ol style="list-style-type: none"> 1. I am kind of missing explicit definition of the task. I mean, the introduction makes impression that the policy should be able to determine its high-level goals and understand the unknown environment. In the end, the task is reduced to "the navigation to a given goal with local obstacle avoidance". The simpler task is ok, but it should be explicitly defined in the beginning of Chapter 3. 2. Definition of the laser reward is unclear to me. Is it some kind of safety margin? If so, why don't you say it. 3. Direction reward motivates the greedy motion towards to goal, however you claim that it motivates the robot to follow the shortest path, which are two different things. 	

Formal and language level, scope of thesis	A - excellent.
<i>Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?</i>	
Please insert your comments here.	

Selection of sources, citation correctness	B - very good.
<i>Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?</i>	
Please insert your comments here.	

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

Please insert your comments here.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.

The student implemented 2D PyGame environment, defined reward function and used RL method (PPO) to learn a policy, which navigates robot towards a given goal in the environment consisting of static and dynamic obstacles. It is not clear, if PPO has been reimplemented, but I assume that it has just been downloaded from stable baselines (<https://stable-baselines3.readthedocs.io/en/master/modules/ppo.html>). Consequently, I assume that main effort has been focused in implementing the PyGame environment and tuning the reward function. The overall results are not fully satisfying and there is still a space for an improvement (78% success rate for Circle robot and 34% success rate for Car-like robot). Nevertheless, student correctly argues that the issues can come from multiple sources (training testing distribution, architecture, hyperparameters) and it is complicated to debug it. It would not be fair to judge the bachelor work based on the achieved results since efficient usage of RL algorithms requires an experienced user. Hence, it must have been complicated for the student to do some systematic research without having some prior experience. It is hard for me to reveal the actual time spent with experimentation, but I suspect that he spent more time than visible from the text.

The grade that I award for the thesis is **C - good**.

Question for the student:

1. You correctly mention that the source of problems could stem from „the implementation of the environment, the architecture of the neural network, the reward system, the learning algorithm, or the selected parameters“. How would you design experiments that could allow to isolate the source of problems?
2. Would it be possible to use a linear policy?
3. Would it be possible to solve the same task without the RL (e.g. by a fast online replanning)? If so, where do you see the main advantages and drawbacks in using the RL and other methods.
4. Would it be possible to use planned trajectories (or human control trajectories) in an imitation learning setup?
5. What is the portion of time you spend with particular subtasks (e.g. implementation PyGame environment, studying papers, designing rewards,...) ?

Date: **6.6.2024**

Signature: