AALTO UNIVERSITY SCHOOL OF SCIENCE

CCIS major Machine Learning, Data Science and Artificial Intelligence (Macadamia)

Course:	CS-E4875 - Research Project in Machine Learning, Data Science and Artificial Intelligence		
Student:	Nguyen Xuan Binh	**Student number:**	887799
Supervisor:	Yasmine Nahal, Computer Science Doctoral Candidate, yasmine.nahal@aalto.fi		
Title:	Human in the loop interaction in reinforcement learning for *de novo* molecular design		
Presentation:	Date: 09.08.2024, location: online (Zoom), audience: Yasmine Nahal and Prof. Samuel Kaski		

Brief description of the project

(*Describe the project, its goals, what was done and how. Explain how the project is related to machine learning and data mining, if not obvious.*)

The project focused on reinforcement learning with human feedback for molecular design. The main goal was to integrate three different user models into the source code of a molecular design tool, REINVENT, to enable scoring molecules according to user preferences provided via three different interaction types (liking or disliking the design, comparing a pair of designs, ranking from best to worst design). First, the user model architectures were implemented in REINVENT. Second, an active learning workflow was developed to query an oracle for feedback about the proposed REINVENT design and use it to train the user models. Third, an analysis pipeline was developed to compare the predictive performance of the different models with respect to the oracle and goodness of the generated designs.

Evaluation of the work, report and presentation

(*Evaluate the quality of the achieved results, the ability of the student to carry out research work independently, the quality of the report (structure, language, etc.) and the clarity of the presentation. The instructor should fill this part.*)

The student was able to carry out this research work in an independent way, with guidance from the supervisor when necessary. The student was able to achieve promising results and write a good quality report, describing the methodology and highlighting the main findings. The presentation was clear, conveying all the necessary information. Also, the student demonstrated a great ability to adapt to a new, multi-disciplinary research field at the interface of machine learning and generative chemistry, and contributed to advancing research in human in the loop machine learning for drug discovery.

Number of credit points (5-10): 5	**Proposed grade:** 5
Date: 12.08.2024	**Supervisor's signature:** Yasmine Nahal