

# Bachelor Thesis Review

Faculty of Mathematics and Physics, Charles University

<b>Thesis author</b>	Teodora Stojcheska
<b>Thesis title</b>	Fine-tuning Code Generation Models with Compiler Feedback
<b>Year submitted</b>	2024
<b>Study program</b>	Computer Science
<b>Specialization</b>	Artificial Intelligence
<b>Review author</b>	doc. Mgr. Martin Pilát, Ph.D. Supervisor
<b>Department</b>	Department of Theoretical Computer Science and Mathematical Logic

## Overall

good OK poor insufficient

	good	OK	poor	insufficient
Assignment difficulty	X	X		
Assignment fulfilled	X	X		
Total size <i>... text and code, overall workload</i>		X		
<p>The thesis deals with the complex problem of improving the quality of code-generation LLMs with compiler feedback, instead of the more traditional human feedback. The student was able to implement, train, and evaluate such models on the task of generating code in Python and Java. The experiments show promising results. The student demonstrated that she is capable of using these complicated techniques for a complex task and the results of the thesis are a good contribution to the area.</p>				

## Thesis Text

good OK poor insufficient

	good	OK	poor	insufficient
Form <i>... language, typography, references</i>		X	X	
Structure <i>... context, goals, analysis, design, evaluation, level of detail</i>		X	X	
Problem analysis		X		
Developer documentation		X		
User Documentation		X		
<p>The text of the thesis is the weakest part. The student started working on the text quite late in the process and the overall quality of the text is rather low. The techniques are described very briefly, with many details missing. Unfortunately, this is also true for the techniques developed by the student. The experiments are well-executed, but their description could also be more detailed. On the other hand, the experiments are much more challenging to run than experiments in other bachelor thesis and require large amounts of GPU RAM for fine-tuning the LLMs.</p> <p>The documentation is in a README file and contains all the necessary information.</p>				

**Thesis Code**

good    OK    poor    insufficient

Design	<i>... architecture, algorithms, data structures, used technologies</i>		X		
Implementation	<i>... naming conventions, formatting, comments, testing</i>		X		
Stability			X		
<p>The quality of the implementation is good. It consists mostly of Python scripts for training and evaluation of the models and uses standard Python deep learning libraries. At the same time, the scripts use useful external resources for logging the results (Weights &amp; Biases). The trained models are stored at the Hugging Face repository.</p>					

**Overall grade**    Very Good  
**Award level thesis**    No

Date August 27, 2024

Signature