THESIS SUPERVISOR'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Benchmarking Techniques for Evaluation of Large Language Models**
Author's name:	**Adam Jirkovský**
Type of thesis :	master's
Faculty/Institute:	Faculty of Electrical Engineering
Department:	Department of Cybernetics
Thesis reviewer:	Ing. Jan Šedivý, CSc.
Reviewer's department:	CIIRC

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment	**challenging**

How demanding was the assigned project?

The main task of the thesis was developing a Czech Bench, a collection of Czech data sets and scripts to evaluate Large Language Models (LLM) in Czech. The thesis is focusing on testing foundational and instruction-aligned LLMs in several categories, question-answering systems, generating text, etc. It also tries to evaluate coherence and factuality of the generated text. The work was based on similar existing LLM tests for English. Provide estimation of the accuray of exidsting models in Czech.

Fulfilment of assignment	**fulfilled**

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.

The student has delivered a comprehensive study of the task at hand. He reviewed a significant amount of literature and summarized it in the initial sections of the thesis. He then learned evaluation of several English LLM platforms and selected the most promising ones and translated the scripts to Czech. This included not only the translation but also creation and modification of testing scripts to adjust for Czech. He also tested to trained several Czech LLMs. All training and experiments are documented in detail in the results section.

Activity and independence when creating final thesis	**A - excellent.**

Assess whether the student had a positive approach, whether the time limits were met, whether the conception was regularly consulted and whether the student was well prepared for the consultations. Assess the student's ability to work independently.

Adam had a very positive and enthusiastic approach. We met weekly to discuss further steps and potential problems. Our discussions were always fast-paced and on topic, as Adam consistently came well-prepared with numerous new suggestions. During the preparation of the thesis, he worked very independently.

Technical level	**A - excellent.**

Is the thesis technically sound? How well did the student employ expertise in his/her field of study? Does the student explain clearly what he/she has done?

The development of an evaluation platform for LLMs is one of the latest most important trends in AI in Czechia. Solving this problem required learning many new technologies that are not part of the regular curriculum. Adam proved to be very quick in grasping even the finest details from the literature. He also demonstrated significant skill in using cloud AWS infrastructure to crawl and run experiments.

Formal level and language level, scope of thesis	**A - excellent.**

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?

The thesis is written with attention to every important detail. It is aesthetically pleasing and uses the recommended LaTeX template. It is written in clear and understandable English, conforming to standard notations and other formal requirements.

Selection of sources, citation correctness	A - excellent.
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?	
Adam skillfully collected and studied a vast array of internet and literature resources. Through this ongoing process, he was steadily discovering the field's newest and most sophisticated papers. Notably, he maintained deep attention to detail, ensuring all statements were properly referenced in accordance with ethical citation practices. His citations adhere to the standard format	

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

Adam's thesis is the first of its kind on the Czech AI landscape and is a significant contribution to the field. It provides a comprehensive overview of the current state of the art of evaluating LLMs and represents an essential tool for development of Czech LLMs. Congratulations!!

The grade that I award for the thesis is **A - excellent.**

Date: : **7.6.2024**   Signature: