CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

SUPERVISOR'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	**Optimization of Text Tokenization for Efficient Language Models Training**
Author's name:	**Alina Haitota**
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Katedra kybernetiky
Thesis supervisor:	Ing. David Herel
Supervisor's department:	FAI: základní výzkum AI

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **2. challenging**

Evaluation of the difficulty of the thesis assignment.

The topic required a combination of theoretical understanding and practical implementation in an advanced and evolving field. The integration of tokenization strategies and architecture-level experimentation is notably challenging at the bachelor level.

Satisfaction of assignment **1. fulfilled**

Assess whether the handed thesis meets the assignment. Present the points of the assignment that fell short or were extended. Try to assess the importance, impact, or cause of each shortcoming.

The student addressed all the key points in the assignment: literature review, implementation of dual-stream tokenization pipelines, integration of subword and multiword units, and evaluation on language models. Minor limitations were acknowledged and discussed.

Activity and independence when creating the final thesis **B - very good**

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

The student worked independently and delivered a comprehensive thesis.

Technical level **B - very good**

Assess the level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.

The technical depth was appropriate and well executed for a bachelor's thesis. It demonstrated understanding of modern NLP workflows, including tokenization pipelines, language model training, preprocessing, and dual-stream architecture implementation. The use of both pretrained and custom tokenizers showed good command of the topic.

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

SUPERVISOR'S OPINION OF FINAL THESIS

Formal and language level, scope of thesis **B - very good**

Assess the correctness of the usage of formal notation. Assess the typographical and language arrangement of the thesis.

The thesis is well-written. Some minor grammatical issues or stylistic inconsistencies do not detract from overall clarity. The structure and formatting followed academic standards, and the scope of the thesis was well balanced.

Selection of sources, citation correctness **B - very good**

Present your opinion on student's activity when obtaining and using study materials for thesis creation. Characterize the selection of sources. Assess whether the student used all relevant sources. Verify that all used elements are correctly distinguished from own results and thoughts. Assess that citation ethics have not been breached and that all bibliographic citations are complete and in accordance with citation conventions and standards.

The student used relevant, high-quality sources. Citations appear accurate and complete.

Additional commentary and evaluation

Present your opinion to achieved primary goals of the thesis, e.g., level of theoretical results, level and functionality of technical or software conception, publication performance, experimental dexterity, etc.

Please insert your commentary (voluntary evaluation).

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Summarize thesis aspects that swayed your final evaluation.

This was a challenging topic, especially for a bachelor's thesis. The student fulfilled the assignment and showed good technical and theoretical insight. Some experiments could be more successful, but the implementation was solid. Given the effort, complexity, and quality, a grade of B is appropriate.

I evaluate handed thesis with a classification grade **B - very good**

Date: **9.6.2025** Signature: