



# Review report of a final thesis

**Reviewer:** Mgr. Alexander Kovalenko, Ph.D.  
**Student:** Patrik Laurinc  
**Thesis title:** Text to query using large language models on local infrastructure  
**Branch / specialization:** Knowledge Engineering  
**Created on:** 21 May 2023

## Evaluation criteria

### 1. Fulfillment of the assignment

- [1] assignment fulfilled
- [2] assignment fulfilled with minor objections
- ▶ **[3] assignment fulfilled with major objections**
- [4] assignment not fulfilled

Upon completion of the final thesis assignment, the student successfully accomplished the task of translating the natural language into SQL queries utilizing large language models. However, it remains uncertain from the thesis content if a functioning prototype for the specified use case has been fully developed. The thesis provides minimal insight into aspects of performance and resource utilization.

### 2. Main written part

60/100 (D)

1. The inconsistent use of tenses within the thesis, particularly in Chapter 3, hampers the readability and flow of the text. It is crucial in academic writing to maintain tense consistency for clarity and precision - generally past tense for actions or findings that have occurred, and present tense for established facts and analyses.
2. The incorporation of figures and images is less effective due to their lack of informational value. For example, Figure 4.2 (a) and (c) show a training loss decline, which should decline anyways. Then the author describes the inconsistency between evaluation loss and evaluation "exact match", which may be just the wrong choice of the loss function, however, this is not further elaborated. Figure 4.4 (a) has no valuable information as well.
3. The labeling of images as "image was taken from [source]" is an unorthodox practice in academic writing and could affect the thesis's credibility. The common practice is either redrawing the figures or getting permission from the publishers.
4. Some concepts are not well described, for example, the PICARD algorithm which seems to be crucial for the results obtained

### **3. Non-written part, attachments**

80 /100 (B)

The diploma thesis includes the attached code, yet it lacks an operational prototype. Instruction with a download link to test the model would be useful.

### **4. Evaluation of results, publication outputs and awards**

70 /100 (C)

This work could serve as a foundation for future discoveries that may be deployable following comprehensive reassessment and reconstitution.

## **The overall evaluation**

70 /100 (C)

The final thesis presents plausible results, which, however, don't quite measure up to those achieved by state-of-the-art models. Assessing the novelty of the work is challenging, as there is a lack of detailed descriptions and evaluations of resource usage and performance. Additionally, tools that could enable more resource-efficient models, such as model quantization, have been overlooked.

## **Questions for the defense**

1. Explain the inconsistency between the evaluation loss (Figure 4.2 b) and exact match accuracy (Figure 4.3 a).
2. Explain the difference in the self-attention weight distribution shown in Figure 4.5 c and Figure 4.5 d.

## **Instructions**

### **Fulfillment of the assignment**

Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.

### **Main written part**

Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies?

Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 52/2021, Art. 3.

Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.

### **Non-written part, attachments**

Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.

### **Evaluation of results, publication outputs and awards**

Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.

### **The overall evaluation**

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.