

I. IDENTIFICATION DATA

Thesis title:	Integrating Text and Image Models for Question Answering
Author's name:	Jan Čuhel
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Cybernetics
Thesis reviewer:	Giorgos Talias
Reviewer's department:	Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment	ordinarily challenging
<i>How demanding was the assigned project?</i>	
The assignment mainly consists of reviewing and putting together different models and tools in a more or less straight-forward way. Nevertheless, there are challenges regarding the range of different models and tools used and the underlined technicalities when working with and training very large models.	

Fulfilment of assignment	fulfilled
<i>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.</i>	
The goals are achieved. There is a thorough literature review, and extensive evaluation of a variety of models. A complete pipeline for handling the task a hand is suggested.	

Methodology	correct
<i>Comment on the correctness of the approach and/or the solution methods.</i>	
The thesis includes a thorough literature review, a dataset construction process with appropriate ground-truth and extensive evaluation of a set of different models using quantitative metrics. All steps are carefully designed, and the results are convincing.	

Technical level	A - excellent.
<i>Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?</i>	
The technical details of the approach are clearly described. The same holds for the experimental details. It appears that the student obtained a good level of understanding regarding the capabilities of VLMs and LLMs and how to use and train them.	

Formal and language level, scope of thesis	A - excellent.
<i>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?</i>	
The manuscript is properly organized and easy to follow. The use of English is satisfactory and the use of pseudo-code and flow charts was helpful. There are no mathematical formulations in the thesis.	

Selection of sources, citation correctness	A - excellent.
<i>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?</i>	
The related work section is clearly written, extensive and up to date.	

Additional commentary and evaluation (optional)
<i>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.</i>

Please insert your comments here.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.

The objectives of the thesis project are meaningful and the underlined challenges were ordinary. The student has managed to properly satisfy all objectives and to deliver a good quality manuscript and a practical system for a real-world application.

Questions:

- 1. According to fig. 3.3. and the description in the manuscript, the retrieved image does not depend on the generated answer. Should an improve system jointly generate the answer and retrieve the image so that each one is conditioned on the other? How can such a design be pursued?*
- 2. In the thesis the threshold on cosine similarity is found and then precision and recall are evaluated. In some cases, a comparison is not conclusive, i.e. between two methods one has a bit higher recall and the other a bit higher precision. What about performing evaluation for a varying value of the threshold generating precision-recall curves and choosing Average Precision (equal to the area under the curve) as the metric? Why was such a metric not considered?*

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

Date: **9.6.2024**

Signature: